

c2

STATISTICAL MECHANICS AND FIELD THEORY

RECEIVED
LAWRENCE
BERKELEY LABORATORY

JUL 2 1979

LIBRARY AND
DOCUMENTS SECTION

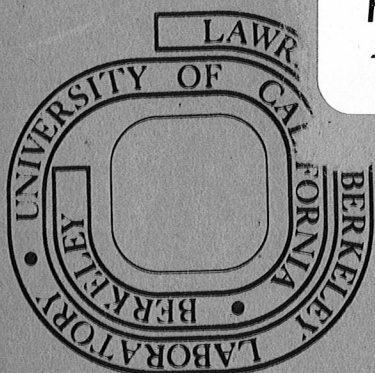
Stuart Alan Samuel
(Ph. D. thesis) ✓

May 1979

Prepared for the U. S. Department of Energy
under Contract W-7405-ENG-48

TWO-WEEK LOAN COPY

This is a Library Circulating Copy
which may be borrowed for two weeks.
For a personal retention copy, call
Tech. Info. Division, Ext. 6782



LBL-9168
c2

ABSTRACT

This thesis exploits the interplay of statistical mechanics and field theory. Generally speaking, field theories are abstract mathematical objects in which powerful mathematical methods have been developed. Field theories are useful for doing calculations but the essential physics is usually hidden and obscure. To unravel the important underlying ideas is difficult. Statistical mechanics, on the other hand, deals with physical systems, the physics of which is known via experiment and intuition. Statistical systems are concrete objects in which the essential underlying physical ideas are known. The roles of field theory and statistical mechanics are therefore complementary, one is powerful mathematically, the other provides physical insight. When a field theory is equivalent to a statistical mechanics system (or vice versa), one has the best of both possible worlds: one knows the physics and one has the mathematical tools to calculate. This thesis establishes and exploits the connection between field theory and statistical mechanics.

The thesis has two main parts. The first applies field theory methods to statistical mechanics. In particular, statistical systems are related to fermionic-

NOTICE

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Department of Energy, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

like field theories through a path integral representation. Such path integrals are over anticommuting variables. The basic definitions, ideas, and uses of anticommuting variables are presented in Chapters I and II. Previously solved models are resolved quickly and simply. Considered are the Ising model, the free-fermion model, and close-packed dimer problems on various lattices. Graphical calculational techniques are developed. They are powerful and yield a simple procedure to compute the vacuum expectation value of an arbitrary product of Ising spin variables. From a field theorist's point of view, these chapters are the simplest most logical derivation of the Ising model partition function and correlation functions. This work promises to open a new area of physics research when the methods are used to approximate unsolved problems.

Chapter III solves by the methods of Chapters I and II a new model named the \mathbb{Z}_2 pseudo-free vertex model.

The second part of the thesis applies statistical mechanics intuition to field theories. Chapter IV shows that certain relativistic field theories are equivalent to classical interacting gases. Using this analogy many results are obtained, particularly for the Sine-Gordon field theory. The main results are enumerated in the summary to which the reader is referred. Chapter V

addresses the most important problem in strong interaction physics: quark confinement. The difficult task of proving confinement has confronted theorists for a decade. Chapter V, although not a proof of confinement, presents a logical, esthetic, and simple picture of how confinement works. A key ingredient is the insight gained by using an analog statistical system consisting of a gas of macromolecules. This analogy allows the computation of Wilson loops in the presence of topological vortices and when symmetry breakdown occurs in the topological quantum number. Topological symmetry breakdown calculations are placed on approximately the same level of rigor as instanton calculations. The picture of confinement that emerges is similar to the dual Meissner type advocated by Mandelstam. Before topological symmetry breakdown, QCD has monopoles bound linearly together by three topological strings. Topological symmetry breakdown corresponds to a new phase where these monopoles are liberated. It is these liberated monopoles that confine quarks.

PRELIMINARY NOTES

A shortened version of Chapter IV has appeared in print in *Physical Review D* 18, 1916 (1978). The American Physical Society has granted permission for its inclusion in this thesis. Chapter V will be published in *Nuclear Physics B*. The other chapters have also been submitted for publication.

Each chapter has its own set of references, figures, and figure captions. These appear at the end of each chapter.

ACKNOWLEDGMENTS

I thank Harry Morrison for reading Chapters I, II, and III. I thank Korkut Bardakci for discussions concerning the material in Chapters IV and V. I also thank him for his excellent guidance. I am grateful to both Harry Morrison and Korkut Bardakci for encouragement.

TABLE OF CONTENTS

Abstract

Preliminary Notes (i)

Part I Applications of Field Theory Methodsto Statistical Mechanics 1

| | |
|--|----|
| Chapter I The Use of Anticommuting Variable Integrals in Statistical Mechanics I | 2 |
| I. Introduction | 3 |
| II. Integrals Over Anticommuting Variables | 4 |
| III. Sample Representations | 8 |
| IV. Quadratic Actions | 14 |
| V. The Operator Algebra | 18 |
| Appendix A | 21 |
| Appendix B | 22 |
| Footnotes and References for Chapter I | 24 |
| Figures for Chapter I | 25 |
| Chapter II The Use of Anticommuting Variable Integrals in Statistical Mechanics II | 44 |
| I. Introduction | 45 |
| II. The Partition Functions for the Dimer and Ising Models | 46 |

| | |
|---|-----|
| III. Graphical Evaluation of Partition Functions | 53 |
| IV. Solvable Two-Dimensional Dimer Problems | 56 |
| V. Anticommuting Variable Correlations | 63 |
| VI. The Ising Model Correlation Functions | 68 |
| VII. Summary | 73 |
| References for Chapter II | 75 |
| Figures for Chapter II | 76 |
| Chapter III The Pseudo-Free 128 Vertex Model | 105 |
| I. Introduction | 106 |
| II. The Model | 107 |
| III. The Solution | 115 |
| IV. The $128+8$ Pseudo-Free Vertex Model | 118 |
| V. Conclusion | 125 |
| Appendix A | 125 |
| References for Chapter III | 127 |
| Figures and Tables for Chapter III | 127 |

Part II Applications of Statistical Mechanics

| | |
|--|-----|
| <u>to Field Theory</u> | 155 |
| Chapter IV The Grand Partition Function in Field Theory With Applications to the Sine-Gordon | 156 |

| | |
|---|-----|
| I. Introduction | 157 |
| II. Gaussian Representation | 160 |
| III. Perturbative Verification | 172 |
| IV. The Two Dimensional Sine-Gordon | 183 |
| V. The Phases of the Sine-Gordon | 187 |
| VI. The Non-Linear σ -Model | 195 |
| VII. Renormalization | 197 |
| VIII. Tidbits | 210 |
| IX. Summary | 216 |
| References for Chapter IV | 219 |
| Tables and Figures for Chapter IV | 222 |
| Chapter V Topological Symmetry Breakdown and Quark Confinement | 228 |
| I. Introduction | 229 |
| II. Closed Loop Gas as a Field Theory | 234 |
| III. Wilson Loops in the Presence of Topological Vortices | 244 |
| IV. Z_N Vortices | 253 |
| V. Monopoles | 277 |
| VI. From 2+1 to 3+1 | 281 |
| VII. Relation to Mandelstam's Scheme | 294 |
| VIII. Open Questions | 296 |
| IX. Summary | 301 |
| References for Chapter V | 305 |
| Tables and Figures for Chapter V | 310 |

PART I

APPLICATIONS OF FIELD THEORY METHODS
TO STATISTICAL MECHANICS

CHAPTER I**THE USE OF ANTICOMMUTING VARIABLE
INTEGRALS IN STATISTICAL MECHANICS I**

I. INTRODUCTION

This paper introduces a new method of attacking certain problems in statistical mechanics. It uses integrals over anticommuting variables to express partition functions in terms of field theories.

The interplay of field theory and statistical mechanics is important. Many complicated field theories have simple underlying statistical mechanics analogues¹⁾. This supplies physical insight into these complicated field theoretic structures and allows one to extract the key concepts. On the other hand, when a statistical mechanics model is expressed as a field theory, various field theory techniques can be used such as perturbation theory, operator methods, variational methods, functional methods, etc. These are powerful avenues of attack, especially for extracting numbers. In short, the statistical mechanics point of view allows one physical insight whereas the field theory point of view supplies the powerful mathematical tools. It is therefore important to understand the connections between statistical mechanics and field theory. It is in this direction that this paper is written.

I shall use integrals over anticommuting variables. They were introduced to handle fermionic degrees of freedom in a path integral formulation²⁾. Until recently³⁾, they were usually used in formal ways, rarely being employed in actual calculations. In this paper and the following ones they will be used in a practical manner to obtain numbers. They are, without a doubt, powerful mathematical tools. They supply relations, relate unrelated models, organized unruly algebra, and evoke rapid

calculations often in a few steps.

I will try to follow a logical development with a pedagogical touch. First, this paper will introduce and review integrals over anticommuting variables (Sec. II). I have tried to summarize their key properties. Further details may be found in the references. Next (Sec. III) I will show how several problems may be expressed in terms of anticommuting variable integrals. This is a brute force method involving no elegance or ingenuity. Often a model has several different representations. It is important, therefore, to find the "best" and "efficient" ones. The fourth section will present a couple models in solvable form. Finally, I will discuss what these variables mean in the context of operator field theory.

This paper and the next deal only with solvable models. This is deliberate since it forms a testing ground on how these methods work. In the next paper, the actual solution of the two solvable models presented in Sec. IV will be carried out.

II. INTEGRALS OVER ANTICOMMUTING VARIABLES

This section will review¹⁾ some properties of integrals over Grassmann variables. More details may be found in reference four. A set of N Grassmann (or anticommuting) variables are objects, η_α ($\alpha = 1, 2, \dots, N$), satisfying

$$\eta_\alpha \eta_\beta + \eta_\beta \eta_\alpha = 0. \quad (2.1)$$

In particular, $\eta_\alpha^2 = 0$. Taking sums and products the most general

construct is

$$f = a_0 + \sum_{\alpha} a_{\alpha} \eta_{\alpha} + \sum_{\alpha < \beta} a_{\alpha\beta} \eta_{\alpha} \eta_{\beta} + \dots + a_{123\dots N} \eta_1 \eta_2 \dots \eta_N, \quad (2.2)$$

with the a 's real or complex numbers. Functions of these variables are defined via Taylor series, which because of eq.(2.1) terminate at the N th order. Equation (2.2) is the most general function, an N th order polynomial.

The anticommuting variable integral of a function, f , of the form of eq. (2.2) is defined by

$$\int d\eta f \equiv \int d\eta_1 d\eta_2 \dots d\eta_N f \equiv a_{123\dots N}. \quad (2.3)$$

The only term which contributes is the one where each η occurs precisely once, the sign being determined by the order (for example,

$\int d\eta_1 d\eta_2 \eta_2 \eta_1 = -1$). Often η 's are associated in pairs (or conjugates), one of which will have a dagger (i.e. η_{α} and η_{α}^{\dagger}).

This is convenient for determining the sign of an integral. For these the measure is defined as $\int d\eta d\eta^{\dagger} \equiv \int d\eta_1 d\eta_1^{\dagger} \dots d\eta_N d\eta_N^{\dagger}$.

Statistical mechanics problems will involve spins, atoms, bonds, etc. at sites, \vec{x} , to which anticommuting variables will be assigned. The variable, \vec{x} , will range over the region of interest; for a cubic crystal this might be a three dimensional lattice so that $\vec{x} = (\alpha, \beta, \gamma)$ has integer coordinates. Often several variables are needed at a site, in which case, an additional label, r , is required, and the η 's will appear as

$\eta_{\vec{x}}^r$, $\eta_{\vec{x}}^{r\dagger}$ ($r = 1, 2, \dots, T$) for T types. Graphically $\eta_{\vec{x}}^r$ and $\eta_{\vec{x}}^{r\dagger}$ may be represented by an "o" and an "x" at \vec{x} . Different types may be distinguished by using different colors. The important point to remember is that a contribution to an integral occurs only if each site is covered by one "o" and one "x" of each color (type).

Key properties of these integrals which are consequences of eq.(2.3) are the following:

1. Shift of variable. Given J_α which anticommute with themselves and with all the η 's,

$$\int d\eta f((\eta_\alpha)) = \int d\eta f((\eta_\alpha + J_\alpha)) . \quad (2.4)$$

2. Change of Variables. Let $\psi_\alpha = \sum_{\beta} A_{\alpha\beta} \eta_\beta$ (with A invertible) be linear combinations of η 's and hence an equivalent set of anticommuting variables. Then

$$\int d\eta f(\eta) = (\det A) \int d\psi f(A^{-1}\psi) . \quad (2.5)$$

Contrast this with normal (i.e. Riemann) integration where there is a factor $(\det A)^{-1}$ rather than $(\det A)$ in eq. (2.5).

3. Quadratic and Quadratic-like Actions.

$$\int d\eta d\eta^\dagger \exp\left(\sum_{\alpha\beta} \eta_\alpha A_{\alpha\beta} \eta_\beta^\dagger\right) = \det A . \quad (2.6)$$

$$\int d\eta \exp\left(\frac{1}{2} \sum_{\alpha\beta} \eta_{\alpha} A_{\alpha\beta} \eta_{\beta}\right) = \text{Pf } A. \quad (2.7)$$

$$\int d\eta d\eta^{\dagger} \int d\psi d\psi^{\dagger} \exp\left(\sum_{\alpha\beta} \eta_{\alpha} \eta_{\alpha}^{\dagger} A_{\alpha\beta} \psi_{\beta} \psi_{\beta}^{\dagger}\right) = \text{perm } A. \quad (2.8)$$

$$\int d\eta d\eta^{\dagger} \exp\left(\frac{1}{2} \sum_{\alpha\beta} \eta_{\alpha} \eta_{\alpha}^{\dagger} A_{\alpha\beta} \eta_{\beta} \eta_{\beta}^{\dagger}\right) = \text{hf } A. \quad (2.9)$$

These are respectively the determinant, Pfaffian⁵⁾, permanent, and haffian of A . Permanents and haffians are determinants and Pfaffians without the sign of permutation factor. In eqs. (2.7) and (2.9) A must be even dimensional. In eq. (2.7) A may be chosen to be antisymmetric. In eq. (2.9) it may be chosen to be symmetric, but must have zero's along the diagonal. These equations are easily proved by expanding the exponents: permutations of products of $A_{\alpha\beta}$ are obtained with the appropriate combinatorial and sign factors. Equation (2.6), however, is easier to prove by transforming $\eta^{\dagger} \rightarrow A^{-1} \eta^{\dagger}$ and using eq. (2.5).

Anticommuting variables are powerful objects. Let us demonstrate some of their power by proving the well known result⁶⁾ that $(\text{Pf } A)^2 = \det A$ for an antisymmetric even dimensional matrix. Usual proofs are quite cumbersome. Use eq. (2.6) and rewrite $\eta_{\alpha} = \sqrt{\frac{1}{2}} (\eta_{\alpha}^{(1)} + i\eta_{\alpha}^{(2)})$, $\eta_{\alpha}^{\dagger} = \sqrt{\frac{1}{2}} (\eta_{\alpha}^{(1)} - i\eta_{\alpha}^{(2)})$,

$d\eta_\alpha d\eta_\alpha^\dagger = id\eta_\alpha^{(1)} d\eta_\alpha^{(2)}$. Since A is antisymmetric
 $\eta_\alpha A_{\alpha\beta} \eta_\beta^\dagger = \frac{1}{2} \eta_\alpha^{(1)} A_{\alpha\beta} \eta_\beta^{(1)} + \frac{1}{2} \eta_\alpha^{(2)} A_{\alpha\beta} \eta_\beta^{(2)}$ (the cross terms cancel).
 The exponent factors into two exponents and the integral factorizes
 into two integrals, each of the form of eq. (2.7).

Finally, one may take derivatives of anticommuting variables.
 For example, $\frac{d}{d\eta_1} \eta_1 = 1$, $\frac{d}{d\eta_1} \eta_2 = 0$. All the usual rules of
 differentiation hold except for minus signs in the product rule due
 to anticommutation relations. Thus $\frac{d}{d\eta_1} (\eta_2 \eta_1)$
 $= (\frac{d}{d\eta_1} \eta_2) \eta_1 - \eta_2 \frac{d}{d\eta_1} \eta_1 = -\eta_2$. These derivatives act to the
 right. Derivatives acting to the left are defined analogously:
 $\eta_1 \frac{\overleftarrow{d}}{d\eta_1} = 1$. A powerful tool is the following:

4. Integration by parts. Given two functions, f and g ,

$$\int d\eta f \frac{\overrightarrow{d}}{d\eta} g = \int d\eta f \frac{\overleftarrow{d}}{d\eta} g. \quad (2.10)$$

In conclusion, anticommuting variables may be manipulated,
 integrated, and differentiated much like ordinary variables except that
 anticommutation must be taken into account.

III. SAMPLE REPRESENTATIONS

In a dimer problem^{6,7,8}) there are a set of sites and a
 set of bonds connecting certain pairs of sites. The bonds may absorb
 dimers. If E_b is the energy of a particular dimer,
 $z_b = \exp(-\beta E_b)$ is the Boltzmann factor associated with an
 absorption. A site may be used only once, so that no two dimers
 may overlap or even touch. Effectively any two dimers are infinitely
 repulsive. There are two kinds of problems: the close-packed

problem in which every site must be covered exactly once, and the usual dimer problem where some sites may be left uncovered.

The statistical mechanics of this system is determined by the partition function. This partition function may be represented as an anticommuting integral. As an example, let us consider the two dimensional close-packed dimer problem. The sites are the integer lattice points (α, β) in a two dimensional plane. Bonds occur between nearest neighbors in the vertical and horizontal directions; z_v is associated with vertical bonds and z_h with horizontal bonds. The partition function is

$$Z(z_h, z_v) = \int d\eta d\eta^\dagger \exp \left[\sum_{\alpha, \beta} \left(z_h \eta_{\alpha\beta} \eta_{\alpha\beta}^\dagger \eta_{\alpha+1\beta} \eta_{\alpha+1\beta}^\dagger + z_v \eta_{\alpha\beta} \eta_{\alpha\beta}^\dagger \eta_{\alpha\beta+1} \eta_{\alpha\beta+1}^\dagger \right) \right]. \quad (3.1)$$

There is an η and η^\dagger for each site, and the total measure is a product over all sites of the measure at each site. The operator $\exp(z_h \eta_{\alpha\beta} \eta_{\alpha\beta}^\dagger \eta_{\alpha+1\beta} \eta_{\alpha+1\beta}^\dagger) = 1 + z_h \eta_{\alpha\beta} \eta_{\alpha\beta}^\dagger \eta_{\alpha+1\beta} \eta_{\alpha+1\beta}^\dagger$ has the option of placing a dimer on the bond between (α, β) and $(\alpha + 1, \beta)$ (see fig.1). If the option is exercised, a weight z_h results and no more dimers may be placed on sites involving (α, β) and $(\alpha + 1, \beta)$. Since the integral is zero unless every site is covered exactly once, eq. (3.1) is the partition function for the two dimensional close-packed dimer problem. This model (and, in general, any close-packed dimer model) is by eq. (2.9) a haffian.

Modifying the measure of eq. (3.1) by

$$\int d\eta d\eta^\dagger \rightarrow \int d\eta d\eta^\dagger \exp \left(\sum_{\alpha\beta} \eta_{\alpha\beta} \eta_{\alpha\beta}^\dagger \right), \quad (3.2)$$

would produce the (usual) dimer partition function, since the $\eta_{\alpha\beta} \eta_{\alpha\beta}^\dagger$ piece of $\exp(\eta_{\alpha\beta} \eta_{\alpha\beta}^\dagger) \equiv 1 + \eta_{\alpha\beta} \eta_{\alpha\beta}^\dagger$ would cover any uncovered site (α, β) . On the other hand, for sites already covered by a dimer the 1 term would be used. The new action would be

$$A_{\text{dimer}}(z_h, z_v) = \sum_{\alpha\beta} \eta_{\alpha\beta} \eta_{\alpha\beta}^\dagger (1 + z_h \eta_{\alpha+1\beta} \eta_{\alpha+1\beta}^\dagger + z_v \eta_{\alpha\beta+1} \eta_{\alpha\beta+1}^\dagger). \quad (3.3)$$

Equation (3.3) may be interpreted as the partition function of monomers and dimers where the energy of a monomer is zero. If E_m is the energy of a monomer, then

$$A_{\text{dimer} + \text{monomer}}(z_m, z_h, z_v) = \sum_{\alpha\beta} \eta_{\alpha\beta} \eta_{\alpha\beta}^\dagger (z_m + z_h \eta_{\alpha+1\beta} \eta_{\alpha+1\beta}^\dagger + z_v \eta_{\alpha\beta+1} \eta_{\alpha\beta+1}^\dagger), \quad (3.4)$$

with $z_m = \exp(-\beta E_m)$, is the partition function for dimers and monomers. By rescaling $\eta_{\alpha\beta} \rightarrow \sqrt{\frac{I}{z_m}} \eta_{\alpha\beta}$ $\eta_{\alpha\beta}^\dagger \rightarrow \sqrt{\frac{I}{z_m}} \eta_{\alpha\beta}^\dagger$ one obtains

$$Z_{\text{dimer} + \text{monomer}}(z_m, z_h, z_v) = z_m^N Z_{\text{dimer}}(z_h/z_m^2, z_v/z_m^2), \quad (3.5)$$

where N is the number of sites. This result (that the partition function for dimers and monomers is simply related to the partition function for dimers alone) is easily derived using physical considerations. In general, there will be transformations on the Grassmann integral which yield results in a few steps that, unlike

this example, are difficult to obtain using physical arguments. This is one reason why anticommuting variables are powerful.

To deal with a general dimer problem, let α be a labelling of sites. The set of bonds, B , is a set of pairs (α, β) having Boltzmann factors $z_{\alpha\beta}$. Then

$$Z = \int d\eta d\eta^\dagger \exp \left(\sum_{\alpha} \eta_{\alpha} \eta_{\alpha}^\dagger + \sum_{(\alpha, \beta) \in B} z_{\alpha\beta} \eta_{\alpha} \eta_{\beta}^\dagger \right). \quad (3.6)$$

Dimer models are equivalent to $\eta \eta^\dagger \eta \eta^\dagger$ field theories with a kinetic energy term consisting only of a mass piece, $\eta \eta^\dagger$. The field theory methods that deal with $\eta \eta^\dagger \eta \eta^\dagger$ theories may be applied to dimer problems.

Almost all partition functions which have a graphical representation are expressible as anticommuting integrals. The d -dimensional Ising model⁹⁾ has such a graphical representation^{6,7,8)}, where one sums over closed non-overlapping but (possibly) intersecting polygonal curves; in two dimensions this is obtained by starting with configurations where all spins are down and drawing curves around regions of up spin. There is a Boltzmann factor for each unit of "Bloch" wall. Alternatively, one may use bond variables¹⁰⁾ (which works in any dimension) for which there is a similar representation with different Bloch wall Boltzmann factors.

Let us consider $d = 2$. Then

$$Z_{\text{Ising}}(J_h, J_v) = \int Z_{\text{closed polygons}}(z_h, z_v), \quad (3.7)$$

where $Z_{\text{Ising}}(J_h, J_v)$ is the Ising model partition function, with

horizontal and vertical spin couplings J_h and J_v , $Z_{\text{closed polygons}}(z_h, z_v)$ is the partition function for closed non-overlapping polygons with Boltzmann weights, z_h and z_v for horizontal and vertical Bloch walls, and f is a multiplicative factor. For the first representation

$$f = \exp \left[N(\beta J_v + \beta J_h) \right] ,$$

$$z_h = \exp(-2\beta J_v) ,$$

$$z_v = \exp(-2\beta J_h) ,$$
(3.8)

where N is the number of sites. For the bond variable representation

$$f = (2 \cosh \beta J_v \cosh \beta J_h)^N ,$$

$$z_h = \tanh \beta J_h ,$$

$$z_v = \tanh \beta J_v .$$
(3.9)

Duality is the well known fact that the Ising model has these two representations relating low and high temperatures, one using bond variables on the lattice and one using disorder variables on the dual lattice.

To express the Ising model as a field theory, use four sets of anticommuting variables at a site (α, β) , $\eta_{\alpha\beta}^r$ and $\eta_{\alpha\beta}^{r\dagger}$ with $r = R(\text{"right"}), L(\text{"left"}), U(\text{"up"}), \text{ or } D(\text{"down"})$ (see fig.2). To draw the sides of polygons use dimer operators

$\eta_{\alpha\beta}^R \eta_{\alpha\beta}^{R\dagger} \eta_{\alpha+1\beta}^L \eta_{\alpha+1\beta}^{L\dagger}$ and $\eta_{\alpha\beta}^U \eta_{\alpha\beta}^{U\dagger} \eta_{\alpha\beta+1}^D \eta_{\alpha\beta+1}^{D\dagger}$ (see fig. 3). They give rise to a wall action

$$A_{\text{wall}} = \sum_{\alpha\beta} \left[z_h (\eta_{\alpha\beta}^R \eta_{\alpha\beta}^{R\dagger} \eta_{\alpha+1\beta}^L \eta_{\alpha+1\beta}^{L\dagger}) + z_v (\eta_{\alpha\beta}^U \eta_{\alpha\beta}^{U\dagger} \eta_{\alpha\beta+1}^D \eta_{\alpha\beta+1}^{D\dagger}) \right]. \quad (3.10)$$

I next require "selection rules" at each (α, β) site. Suppose

$\eta_{\alpha\beta}^D \eta_{\alpha\beta}^{D\dagger} \eta_{\alpha\beta}^R \eta_{\alpha\beta}^{R\dagger}$ is inserted in the integral, then only the configuration of (fig. 4a) may occur. Figure 4 illustrates the eight possibilities which can happen. To limit the graphs to these possibilities insert $1 + g$ with

$$g = N_{NN}^{RD} + N_{NN}^{DL} + N_{NN}^{LU} + N_{NN}^{UR} \\ + N_{NN}^{UD} + N_{NN}^{RL} + N_{NN}^{UL} + N_{NN}^{DR}, \quad (3.11)$$

where $N^R = \eta^R \eta^{R\dagger}$. By using $1 + g = \exp \left[\ln(1 + g) \right] = \exp(g - \frac{1}{2}g^2)$ an action for these selection rules is obtained

$$A_{\text{s.r.}} = \sum_{\alpha\beta} (N_{\alpha\beta}^{RD} + N_{\alpha\beta}^{DL} + N_{\alpha\beta}^{LU} + N_{\alpha\beta}^{UR} \\ + N_{\alpha\beta}^{UD} + N_{\alpha\beta}^{RL} - 2 N_{\alpha\beta}^{RU} N_{\alpha\beta}^{LD}), \quad (3.12)$$

where again $N_{\alpha\beta}^R = \eta_{\alpha\beta}^R \eta_{\alpha\beta}^{R\dagger}$. The total action for the Ising model is $A = A_{\text{wall}} + A_{\text{s.r.}}$.

I call the above method of obtaining integral representations the "selection rule" method. By weighting the configurations of fig. 4, more general Ising-like models are obtained. Representations of more complicated models like the Baxter¹¹⁾ model can be derived in a similar manner.

The above representation of the Ising model is inefficient: It uses four sets of anticommuting variables per site; furthermore the action involves products of up to eight variables. Given a particular model, there will be many Grassmann integral representations. It is important to find efficient representations. Ingenuity in finding the "best" set of variables and the "best" actions will determine whether a model is exactly solvable and will determine how well approximation methods work. In the next section, efficient representations are found for these two ϵ -dimensional models.

IV QUADRATIC ACTIONS

Some models have quadratic action representations. I call these pseudo-free theories because they are exactly solvable by the techniques that solve free theories. In this section I will represent the two dimensional close-packed dimer and Ising models as pseudo-free theories. A later paper will calculate the partition functions and correlation functions.

The two dimensional dimer problem will be dealt with first. The method used to solve it closely follows the standard method¹²⁾ of attack. In fact, I will be essentially reproducing the known method in integral form, circumventing a few algebraic steps along the way.

Take the lattice plane, group sites into units of four, and use the labelling indicated in fig. 5. Sites 1 and 3 are called odd sites; sites 2 and 4 are called even sites. For each unit (α, β) , assign four sets of anticommuting variables, $\eta_{\alpha\beta}^r$, $\eta_{\alpha\beta}^{r\dagger}$ ($r = 1, 2, 3, 4$), one for each of the four original lattice points. It will be shown that

$$Z_{\text{dimer}}(z_h^A, z_v^A) Z_{\text{dimer}}(z_h^B, z_v^B) = \int d\eta d\eta^\dagger \exp(A), \quad (4.1)$$

where

$$\begin{aligned} A &= A^A + A^B, \\ A^A &= \sum_{\alpha\beta} \left[z_h^A (\eta_{\alpha\beta}^1 \eta_{\alpha\beta}^{2\dagger} + \eta_{\alpha\beta}^{4\dagger} \eta_{\alpha\beta}^3) \right. \\ &\quad + z_v^A (\eta_{\alpha\beta}^{2\dagger} \eta_{\alpha\beta}^3 + \eta_{\alpha\beta}^{4\dagger} \eta_{\alpha\beta}^1) \\ &\quad \left. + z_h^A (\eta_{\alpha\beta}^{2\dagger} \eta_{\alpha+1\beta}^1 + \eta_{\alpha\beta}^3 \eta_{\alpha+1\beta}^{4\dagger}) + z_v^A (\eta_{\alpha\beta+1}^1 \eta_{\alpha\beta}^{4\dagger} + \eta_{\alpha\beta}^3 \eta_{\alpha\beta+1}^{2\dagger}) \right], \end{aligned} \quad (4.2)$$

and A^B is obtained by replacing z_h^A by z_h^B , z_v^A by z_v^B , $\eta_{\alpha\beta}^r$ by $\eta_{\alpha\beta}^{r\dagger}$, and $\eta_{\alpha\beta}^{r\dagger}$ by $\eta_{\alpha\beta}^r$. Equation (4.2) may look complicated, but it has a simple graphical representation in fig. 6. Each of the eight dimer-like operators of fig. 6a corresponds to a term of eq.(4.2). The dimer object, $\eta_{\alpha\beta}^1 \eta_{\alpha\beta}^{2\dagger}$, produces an "o" at 1 and an "x" at 2 in the (α, β) unit. Arrows are used to indicate the order of the η 's as illustrated in fig. 7. The dimers weighted by z^A factors are the ones with "c"'s on odd sites and "x"'s on even sites

and are called A-dimers. The B-dimers have "x"'s on odd sites and "o"'s on even sites.

We can now make contact with the usual method of solution. By the "golden rule" of Grassman integrals, each site must have an "x" and an "o". This means each site is covered by exactly one A-dimer and one B-dimer. Therefore, we have a simultaneous A and B dimer problem: Expand the B-action exponent and choose one configuration, b , which covers all sites with B-dimers. Let w_b be its weight (that is, the product of the z_h^B and z_v^B factors; for example, if $z_h^B = z_v^B \equiv z^B$ then $w_b = (z^B)^{N/2}$ where N is the number of sites). Expanding the A-action exponent, each A-dimer covering results in diagrams of closed non-overlapping polygons and overlapping isolated dimer pairs (see fig. 21, p. 233 of reference 7) with the proper weight (up to possibly a minus sign). A minus sign could result because of reorderings of anticommuting variables in evaluating integrals. It is proven in Appendix A, however, that all terms are positive. The reader is invited to check some examples by using the rules of fig. 8. Each configuration, b , of B-dimers yields $w_b Z_{\text{dimer}}(z_h^A, z_w^A)$. Equation (4.1) results by summing over all B-coverings.

Every planar close-packed dimer problem, which is exactly solvable by the usual techniques, is expressible as an anticommuting integral over a quadratic action. At this stage, Grassmann integrals are used only as a bookkeeping device which organizes the algebra. No true progress has been made. The next example will obtain a quadratic action for the Ising model. Although similar to previous derivations, several simplifications are made.

I shall use eq. (3.7) which relates the Ising model to a sum over closed non-overlapping but (possibly) intersecting polygons. I shall then use the anticommuting variables to "draw" these configurations. Two sets of variables will be used at each (α, β) site: $\eta_{\alpha\beta}^h, \eta_{\alpha\beta}^{h\dagger}$, and $\eta_{\alpha\beta}^v, \eta_{\alpha\beta}^{v\dagger}$. The superscripts "h" and "v" stand for horizontal and vertical. Consider

$$Z_{\text{closed polygons}}(z_h, z_v) = (-1)^N \int d\eta d\eta^\dagger \exp(A), \quad (4.3)$$

where N is the number of sites and

$$A = A_{\text{Bloch wall}} + A_{\text{corner}} + A_{\text{monomer}},$$

$$A_{\text{Bloch wall}} = \sum_{\alpha\beta} \left(z_h \eta_{\alpha\beta}^{h\dagger} \eta_{\alpha+1\beta}^h + z_v \eta_{\alpha\beta}^{v\dagger} \eta_{\alpha\beta+1}^v \right), \quad (4.4)$$

$$A_{\text{corner}} = \sum_{\alpha\beta} \left(a_1 \eta_{\alpha\beta}^{h\dagger} \eta_{\alpha\beta}^v + a_3 \eta_{\alpha\beta}^{v\dagger} \eta_{\alpha\beta}^h + a_2 \eta_{\alpha\beta}^{v\dagger} \eta_{\alpha\beta}^{h\dagger} + a_4 \eta_{\alpha\beta}^v \eta_{\alpha\beta}^h \right),$$

$$A_{\text{monomer}} = \sum_{\alpha\beta} \left(b_h \eta_{\alpha\beta}^h \eta_{\alpha\beta}^{h\dagger} + b_v \eta_{\alpha\beta}^v \eta_{\alpha\beta}^{v\dagger} \right).$$

The Bloch wall action produces a unit of Bloch wall in either the horizontal or vertical direction [see fig. (9)] weighted by the appropriate Boltzmann factor. The term A_{corner} produces the four corners of fig. (10) necessary to construct a polygon. I have allowed for the most general quadratic form by weighting corners with the a_i . For the Ising model, set $a_i = 1$. Finally, A_{monomer} fills all unoccupied "h" and "v" sites with monomer.

Again, for the Ising model, set b_h and $b_v = 1$. The eight possible configurations which can occur at a site are shown in fig. (11) with their weights. There is an extra (-1) for each site because of the $(-1)^N$ in eq. (4.3). The minus signs in configurations (b) through (g) always cancel in pairs and may be dropped. The extra minus sign in fig. (11h) is explained in Appendix B. This Appendix deals with minus signs due to reorderings of Grassmann variables. Finally, the double corners of fig. (12a,b) do not occur because a single corner uses up both horizontal and vertical variables. Equations (3.7), (4.3), and (4.4) form the quadratic action representation of the two-dimensional Ising model.

V THE OPERATOR ALGEBRA

This section discusses the operator aspects of Grassmann variables and their probabilistic interpretation.

In the previous two sections, partition functions have been expressed as fermionic field theories. By taking expectation values of Grassmann variables (as well as functions of them) we may treat them as operators. They act like "local observables", measuring tools with probabilistic interpretations. Consider for example, the two dimensional dimer problem whose action is given by eq. (3.3).

$$Z(\eta_{\alpha\beta}, \eta_{\alpha\beta}^\dagger) = \int d\eta d\eta^\dagger \exp(A_{\text{dimer}})(\eta_{\alpha\beta}, \eta_{\alpha\beta}^\dagger), \quad (5.1)$$

is the sum over dimer configurations with the restriction that no dimer be placed on the (α, β) site. Therefore, $\langle \eta_{\alpha\beta}, \eta_{\alpha\beta}^\dagger \rangle$ is the probability that the (α, β) site is not covered by a dimer.

Likewise $\langle 1 - \eta_{\alpha\beta} \eta_{\alpha\beta}^\dagger \rangle$ is the probability that (α, β) is covered. In general, the expectation value of an operator will be the probability that a corresponding configuration will occur.

What do the equations of motion mean? The equation for $\eta_{\alpha\beta}$ is obtained by taking $\frac{d}{d\eta_{\alpha\beta}}$ A_{dimer} . Let O be an operator (i.e. some function of the η 's and η^\dagger 's) and use integration by parts [eq. (2.10)]:

$$\left\langle \frac{d}{d\eta_{\alpha\beta}} O \right\rangle = \left\langle O \frac{d}{d\eta_{\alpha\beta}} A \right\rangle. \quad (5.2)$$

Equation (5.2), which involves the equation of motion of $\eta_{\alpha\beta}$, will generate many probability relations and is quite useful. For example, let O be $\eta_{\alpha\beta}$ and let $P_{(\alpha,\beta)}^O$, $P_{(\alpha,\beta)}^u$, $P_{(\alpha,\beta)\text{and}(\alpha,\beta+1)}^u$, etc. be respectively the probabilities that (α, β) is occupied, that (α, β) is unoccupied, that (α, β) and $(\alpha, \beta+1)$ are unoccupied, etc., then

$$\begin{aligned} P_{(\alpha,\beta)}^O &= z_h \left[P_{(\alpha,\beta)\text{and}(\alpha+1,\beta)}^u + P_{(\alpha,\beta)\text{and}(\alpha-1,\beta)}^u \right] \\ &+ z_v \left[P_{(\alpha,\beta)\text{and}(\alpha,\beta+1)}^u + P_{(\alpha,\beta)\text{and}(\alpha,\beta-1)}^u \right]. \quad (5.3) \end{aligned}$$

I invite the reader to derive this relation using physical considerations and compare it to the simple and powerful method of anticommuting variables.

The set of relations of eq. (5.2) along with the anti-commutator equations [eq. (2.1)] determine the model. They are

an equivalent expression of it, because A is obtainable from eqs. (5.2). The Ising-like or dimer-like problems are uniquely determined by a set of local probability relations. In field theory the equations of motion are foremost. The operator techniques used to attack such field theories may be used in statistical mechanics. I call this the operator method of local observables.

APPENDIX A

This Appendix treats the minus signs of the two dimensional close-packed dimer problem.

An isolated dimer pair between two neighboring sites r and s (see fig. 13) will be of the form $\eta_r \eta_s^\dagger + \eta_r^\dagger \eta_s$; so it has the correct sign.

To deal with a closed polygon, P , orient it counterclockwise and call the parity of P the number of minus signs which result from rule (b) of fig. 8. There are two types of polygons: type 1 (fig. 14a) and type 2 (fig. 14b). For type 1, there is a minus sign from rule (c) and no minus signs due to rule (a). Therefore, for type 1, the overall sign is opposite to the counterclockwise parity of P . For type 2, the identical conclusion is obtained using a similar approach. ¹³⁾ (which is easily verified for test examples and easily proven by induction on the area of P) says that the counterclockwise parity is $(-1)^{I+1}$ (where I is the number of interior points) if all elementary polygons (ones with no interior points) have odd parity. In fig. 14, for example, $I = 1$ and the parity is even. With the arrow assignment of fig. 6, all elementary polygons are odd parity. We conclude all polygons having an even number of interior points have the correct sign. Fortunately only these kinds of polygons occur in a covering since dimers, covering two sites at a time, cannot cover regions of an odd number of sites. Therefore, all polygons have the correct sign.

APPENDIX B

In this Appendix, I will analyze the sign problem associated with eqs. (4.3) and (4.4). The conclusion will be that the sign of a configuration of polygons is equal to the number of intersections which occur. This explains the extra minus factor in the weight of fig. (11h). I will proceed in steps: first dealing with an isolated non-self-intersecting polygon, then with one that self-intersects, and finally dealing with a multipolygonal configuration.

Consider a closed polygon, P , which does not intersect itself. I will show that its sign is positive. Choose a horizontal bond of P and proceed to the right (and eventually around the polygon). Start at the "x" and use the rules of fig. 8. When moving upward or to the right no minus signs result from rules (a) or (b) because arrows are in the correct direction and "o"'s occur before "x"'s. When moving downward or to the left, each site has a minus sign from rule (a) and a minus sign from rule (b). They cancel in pairs. Next consider what happens, when one goes around a corner. There are eight different types (see fig. 15) [two orientations times the four basic corners of fig. (10)]. They are oriented because we are moving around the polygon in a particular direction. Figure 15 summarizes the results: only corners of types d and \bar{d} lead to a minus sign. Now use the following theorem (which is easily proved by induction on the area of P): Let m_a, m_b, \dots be the number of type a, type b, etc. corners occurring in an oriented non-self-intersecting polygon, P . If P is counterclockwise oriented then

$$m_a - m_{\bar{a}} = 1 ,$$

$$m_b - m_{\bar{b}} = 1 ,$$

(B.1)

$$m_c - m_{\bar{c}} = 1 ,$$

$$m_d - m_{\bar{d}} = 1 .$$

This implies that the sign due to corners is $(-1)^{m_d}(-1)^{m_{\bar{d}}} = -1$. For clockwise oriented, P , the theorem holds with $a \leftrightarrow \bar{a}$, $b \leftrightarrow \bar{b}$, etc. Rules (a) and (b) therefore result in one minus sign which when combined with the minus sign of rule (c) gives an overall plus sign.

Now consider an oriented self-intersecting polygon, P . It may be constructed from non-intersecting ones by the pasting construction of fig. 16. The order of the operators in P is indicated in Figure 17a. When they are regrouped into the forms occurring in the non-self-intersecting polygons (Figures 17b and 17c) which "compose" P , a minus sign results for each intersection as Figure 17 illustrates.

Finally, the result holds for multipolygonal configurations because pairs of polygons can only intersect an even number of times. Summarizing, an extra minus occurs for each intersection (fig. 11h).

FOOTNOTES AND REFERENCES

- 1) S. Samuel, The Grand Partition Function in Field Theory with Applications to The Sine-Gordon Theory, to be published in Phys. Rev. D.
- 2) D. J. Chandlin, Nuovo Cimento 4, (1956) 231.
- 3) A practical application is S. Samuel, J. Math. Phys. 19, (1978) 1438.
- 4) F. A. Berezin, The Method of Second Quantization (Academic, New York, 1966). See also the appendices of reference 3.
- 5) The definition of a Pfaffian is

$$\text{Pf } A = \frac{1}{2^{N/2}} \frac{1}{(\frac{N}{2})!} \sum_{\sigma} (\text{sign } \sigma) A_{\sigma(1)\sigma(2)} A_{\sigma(3)\sigma(4)} \cdots A_{\sigma(N-1)\sigma(N)}.$$

The sum is over all permutations, σ .

- 6) For example, B. M. McCoy and T. T. Wu, The Two-Dimensional Ising Model (Harvard University Press, Cambridge, 1973). The proof that $(\text{Pf } A)^2 = \det A$ can be found on pages 47-51.
- 7) E. W. Montroll, in Brandeis University Summer Institute in Theoretical Physics, 1966, edited by M. Chrétien, E. P. Gross, and S. Deser (Gordon and Breach, New York, 1968).
- 8) H. S. Green and C. A. Hurst, Order-Disorder Phenomena (Interscience, New York, 1964).
- 9) There is an enormous amount of literature on the Ising model. See the references of reference 6 for a partial list.
- 10) See for example, pages 218-221 of reference 7.
- 11) R. J. Baxter, Annals of Physics, 70, (1972) 193.
- 12) See pages 230-248 of reference 7.
- 13) See pages 236 and 237 of reference 7.

Figure 1. The dimer operators: (a) The horizontal dimer operator,

$$\eta_{\alpha\beta}^\dagger \eta_{\alpha+1\beta}^\dagger \quad \text{and (b) The vertical dimer operator}$$

$$\eta_{\alpha\beta}^\dagger \eta_{\alpha\beta+1}^\dagger \eta_{\alpha\beta+1}^\dagger \eta_{\alpha\beta+1}^\dagger .$$

Figure 2. The two-dimensional lattice used for the Ising model:

- (a) Each site has been replaced by four sites, and
 (b) The notation used to label sites. The pair, (α, β) , labels the group and "Right", "Up", "Left", and "Down" are used to label types.

Figure 3. Bloch wall operators: (a) The horizontal dimer operator,

$$\overset{R}{\eta_{\alpha\beta}} \overset{R^\dagger}{\eta_{\alpha+1\beta}} \overset{L}{\eta_{\alpha\beta}} \overset{L^\dagger}{\eta_{\alpha+1\beta}} , \quad \text{and (b) The vertical dimer operator,}$$

$$\overset{U}{\eta_{\alpha\beta}} \overset{U^\dagger}{\eta_{\alpha\beta+1}} \overset{D}{\eta_{\alpha\beta+1}} \overset{D^\dagger}{\eta_{\alpha\beta+1}} .$$

Figure 4. The eight possibilities which can happen at a vertex.

In each case, the operator on the right will produce the dimer configuration on the left. Figures (a) - (g) represent the seven terms in eq. (3.11). To these one must add the last term which is the unity operator.

Figure 5. The lattice plane reorganized into groups of four sites each. Each unit is labelled by a pair of integers (α, β) and each of the four sites in a unit are labelled by 1, 2, 3, or 4. Sites of type 1 and 3 are called odd sites, whereas sites of type 2 and 4 are even sites.

Figure 6. (a) The eight bonds corresponding to the eight terms in eq. (4.2). Each of these operators create A-dimers and is weighted by a z^A factor. The arrows indicate the order of the η 's (see fig. 7.). (b) The B-dimer operators which make up the B-dimer action, A^B .

Figure 7. The use of arrows to indicate operator ordering. On the left-hand side is the operator, $\eta^\dagger \eta$, which equals minus the right-hand side operator, $\eta \eta^\dagger$. The arrow originates from the first anticommuting variable and terminates on the second one.

Figure 8. Sign rules. The rules for evaluating the sign of a "dimer loop" are as follows: Pick an initial "o" or "x" (here, "o" is chosen at A) and proceed around the loop (here, counterclockwise). There is a) a minus sign for each "x" occurring before an "o" (the point, B), b) a minus sign for each arrow in the opposite direction (the bond, C), and finally c) a minus sign if one begins with an "x". In this figure the sign is positive.

Figure 9. Bloch wall operators: (a) is the graphical representation of $\eta_{\alpha\beta}^\dagger \eta_{\alpha+1\beta}$ which occurs in eq. (4.4) and produces a horizontal Bloch wall; (b) is the vertical Bloch wall operator, $\eta_{\alpha\beta}^\dagger \eta_{\alpha\beta+1}$.

Figure 10. The corner operators in eq. (4.4). In all cases they occur at the (α, β) site, that is corner operators only change the direction of a curve; they do not connect neighboring sites. Although one could use labels to distinguish horizontal and vertical variables, it's easier to use the following convention: if an "o" or an "x" has a horizontal line coming into or out of it, it is a horizontal variable; on the other hand vertical variables have vertical lines flowing into or

out of them. For example, (a) involves a horizontal "x" or $\eta_{\alpha\beta}^{h\dagger}$ and a vertical "o" or $\eta_{\alpha\beta}^v$. The arrow indicates the order, so that this term is $\eta_{\alpha\beta}^{h\dagger} \eta_{\alpha\beta}^v$, the first term in A_{corner} of eq. (4.4). (b), (c), and (d) are the other three terms.

Figure 11. The eight possible configurations that can occur at a site. When disorder variables are used [eq. (3.8)], the first two columns represent corresponding spin configurations. In obtaining the weights of column 4 a (-1) factor has been included from the $(-1)^N$ of eq. (4.3). The minus signs in (b) through (g) may be eliminated because i) there are always an even number of (b) and (c) configurations and ii) corners (d) and (f) as well as (e) and (g) occur in pairs. Alternatively, one could redefine the b's and a's in eq. (4.4) to have minus signs. Configuration (h) has an extra minus sign due to reordering of anticommuting variables as described in Appendix B. The numbers in column 4 are easily obtained: For example, the b_h of (b) is obtained because a vertical bond enters and exits the vertical site and a horizontal monomer with b_h must fill the empty horizontal site.

Figure 12. Intersections. The double corners of figs. (a) and (b) are not allowed by eq. (4.3). When four lines meet at a site they must pass directly through as in fig. (c).

Figure 13. A typical dimer pair. The sign of this pair is plus

$$\text{since } \eta_R \eta_S^\dagger \eta_R^\dagger \eta_S = + \eta_R^\dagger \eta_S \eta_R \eta_S^\dagger.$$

Figure 14. The two types of polygons. Type 1 [fig. (a)] is characterized by the fact that $\eta\eta^\dagger$ occurs at each site when going counterclockwise around the polygon. Alternatively, as one goes around the curve the "x" occurs before the "o" in a given dimer. For type 2 [fig. (b)], $\eta^\dagger\eta$ products occur at each site. The parity of this polygon will be even if all elementary polygons are counterclockwise odd. The assignment of arrows in fig. (6) does make all elementary squares of odd parity.

Figure 15. The eight oriented corners and the minus sign factors associated with them.

Figure 16. The pasting construction. Polygon, P, may be obtained from two (possibly self-intersecting) polygons, P_1 and P_2 , by cutting open the corners and rejoining. There are four (two different types of pairs of corners times two orientations) possible pasting constructions.

Figure 17. How the minus sign arises. This is just a "fermion" statistics effect. The order of operators in an

intersection of P is indicated in Figure (a) and is

$$(\eta_1^\dagger \eta_3)(\eta_2^\dagger \eta_4).$$

When P is decomposed into non-intersecting polygons as in Figure 16, the order of the operators is that of (b) or (c). For case (b),

$$(\eta_1^\dagger \eta_4)(\eta_2^\dagger \eta_3) = - (\eta_1^\dagger \eta_3)(\eta_2^\dagger \eta_4),$$

that is, there is a minus sign relative to (a). For case (c), $(\eta_2^\dagger \eta_1^\dagger)(\eta_4 \eta_3)$

$$\text{is also } - (\eta_1^\dagger \eta_3)(\eta_2^\dagger \eta_4).$$

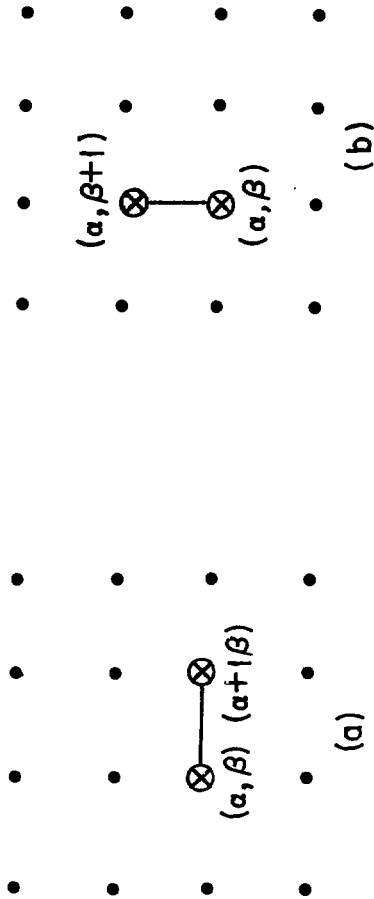


Figure 1

XBL7810-11566

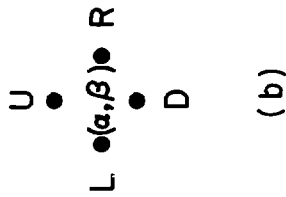
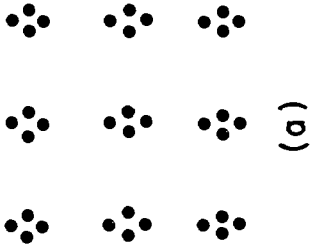


Figure 2

XBL7810-11567

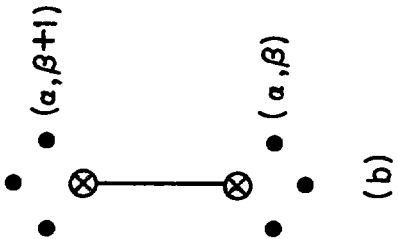
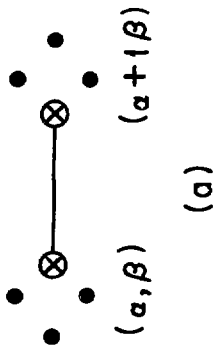


Figure 3

XBL7810-11568

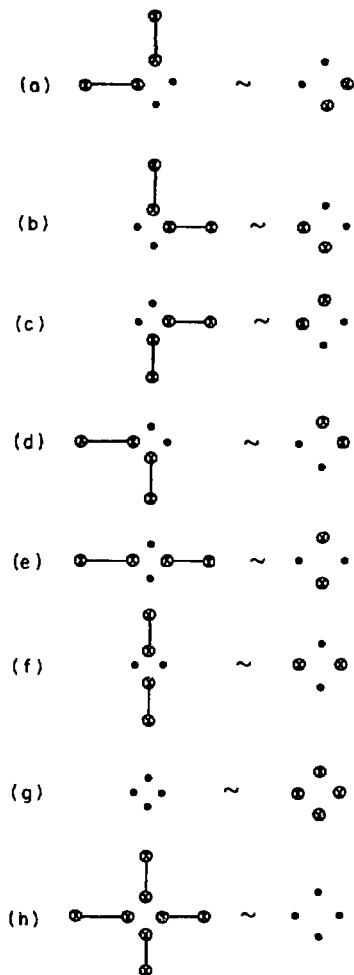


Figure 4

XBL7810-11560

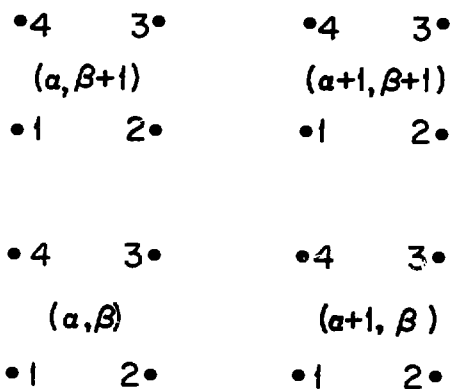


Figure 5

XBL7810 - 11564

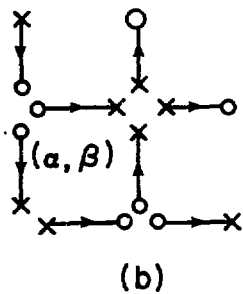
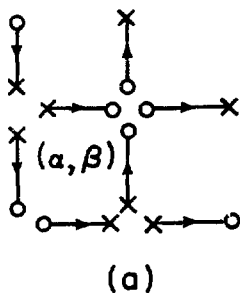


Figure 6

XBL7810 - 11565

$$x \rightarrow \bigcirc = (-) x \leftarrow \bigcirc$$

Figure 7 XBL7810-11569

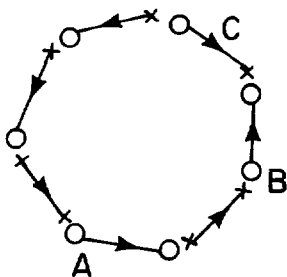


Figure 8 XBL7810-11570

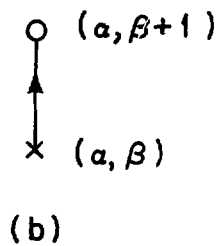
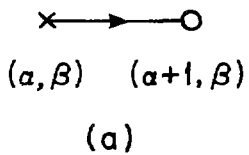


Figure 9

XBL 7810 - 11562

 (α, β)

(a)

 (α, β)

(b)

 (α, β)

(c)

 (α, β)

(d)

Figure 10

XBL7810-11563

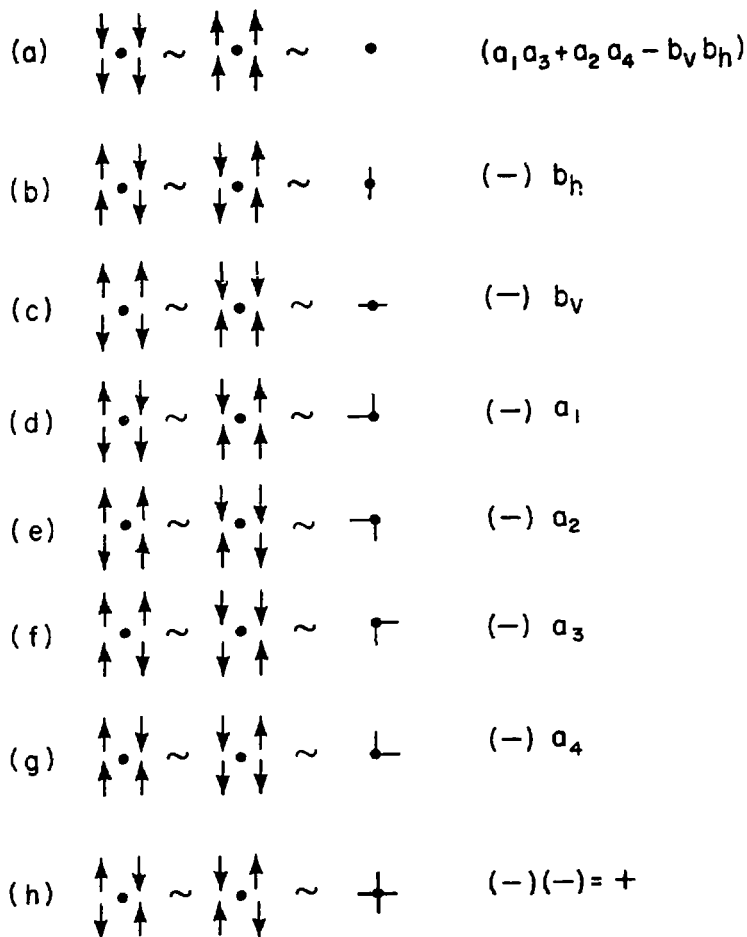


Figure 11

XBL 7810-11561

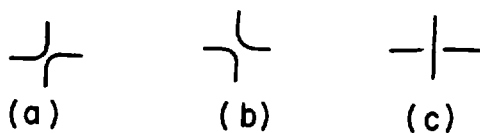


Figure 12

XBL7810-11574

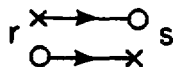
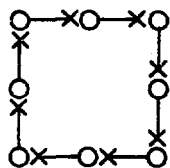
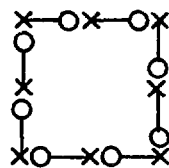


Figure 13 XBL7810-11575



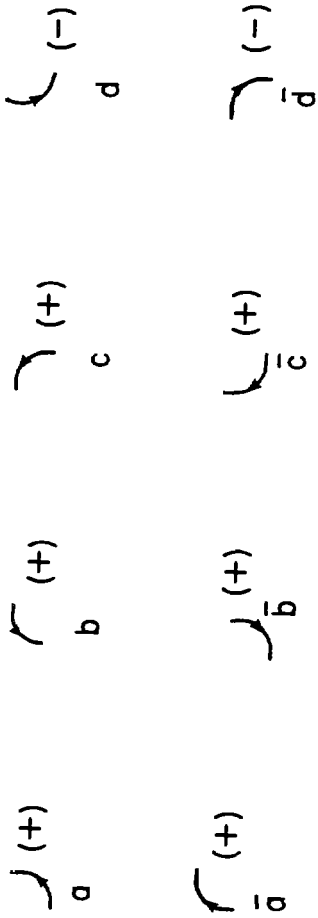
(a)



(b)

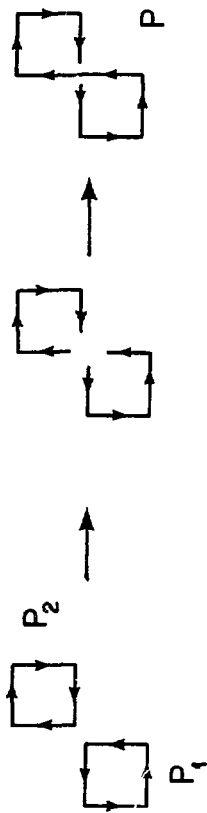
Figure 14

XBL 7810-11559



XBL 7810 - 11571

Figure 15



XBL7810-11572

Figure 16

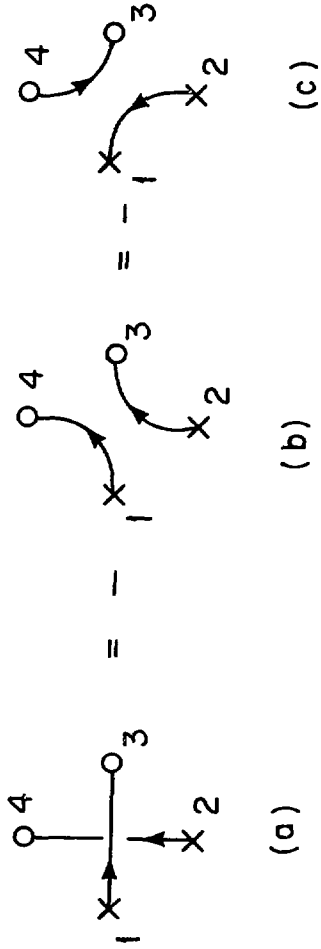


Figure 17

X BL 7810-11573

CHAPTER II

THE USE OF ANTICOMMUTING VARIABLE
INTEGRALS IN STATISTICAL MECHANICS II

I. INTRODUCTION

In the first paper¹⁾ (to be referred to as I) certain partition functions are represented as fermionic-like lattice field theories using Grassmann integrals. This allows one to use powerful field theory methods to attack statistical mechanics problems. Several models had quadratic action representations. Among these were the two-dimensional Ising model and the two-dimensional square lattice dimer problem. They are pseudo-free theories and are exactly solvable. In this paper, these two partition functions are explicitly computed (Sec. II). This is a straightforward calculation: one transforms to momentum space just as one would do with a free field theory. This partially diagonalizes the problem; it breaks up into a product of 4×4 determinants. Next, graphical methods are introduced to organize the algebra (Sec. III). They are useful because they are systematic and pictorial. Section IV considers the general class of solvable 2-dimensional close-packed dimer problems on various lattices. A set of rules are derived which quickly compute partition functions. These rules are illustrated using the square lattice and applied to the hexagonal lattice. Next, the rules are extended to general pseudo-free theories. This means that, given any quadratic action, there is a simple systematic calculational procedure. For the free fermion model anticommuting variable correlations are calculated (Sec. V). They are first considered in momentum space where the computations reduce to solving modified miniature dimer problems. The Ising model is included in the free fermion model, so that the results of Sec. V can be used to calculate spin correlations. Section VI exemplifies

th's by considering two horizontal spins. The approach generalizes so that one may, in principle, compute the vacuum expectation value of an arbitrary number of spins, although the form of the answer is cumbersome. This is because spin variables, which are the physical variables, are complicated functions of the anticommuting variables, which are the mathematical variables in terms of which computations are simple.

Paper I was a pedagogical introduction to Grassmann integral techniques. It emphasized how to use anticommuting variables and how to express partition functions as fermionic-like field theories. This paper emphasizes computational methods. It illustrates how to calculate partition functions and correlation functions. It provides graphical rules which simplify complicated calculations.

This paper considers only solvable models. They form the testing ground to see how and if the techniques work. They also form a solid foundation upon which unsolved problems may be attacked by approximation methods. The real power of anticommuting variables will come when they are applied to these unsolved models.

II. THE PARTITION FUNCTIONS FOR THE DIMER AND ISING MODELS.

In paper I, the two-dimensional Ising model was represented as a Grassmann integral over a pseudo-free fermionic-like action. A similar representation was obtained for the close-packed dimer problem. By pseudo-free action, I mean a quadratic action. Such theories are solvable by the same methods that solve free theories: transform to momentum space. This partially diagonalizes the problem because of translational invariance. What results is a product of Pfaffians of $4T \times 4T$ dimensional matrices. The variable, T , is the number of

η 's and η^\dagger 's per site. If bilinears occur as $\eta\eta^\dagger$'s the problem simplifies to a product of determinants of $T \times T$ dimensional matrices. This is why it is important that the number of variables per site not be too large.

I will always choose α to range from $-M$ to M and β to range from $-N$ to N , so that there are $(2N+1)$ rows and $(2M+1)$ columns. In the Ising model there are $(2N+1)(2M+1)$ sites, whereas in the dimer problem, there were 4 sites per (α, β) unit so that there are $4(2N+1)(2M+1)$ sites in all.

Going to momentum space means writing

$$\eta_{\alpha\beta}^r = \sum_{s,t} \frac{1}{\sqrt{2M+1}} \frac{1}{\sqrt{2N+1}} \exp\left(\frac{2\pi i\alpha s}{2M+1} + \frac{2\pi i\beta t}{2N+1}\right) a_{st}^r \quad (2.1)$$

$$\eta_{\alpha\beta}^{r\dagger} = \sum_{s,t} \frac{1}{\sqrt{2M+1}} \frac{1}{\sqrt{2N+1}} \exp\left(\frac{-2\pi i\alpha s}{2M+1} - \frac{2\pi i\beta t}{2N+1}\right) a_{st}^{r\dagger}.$$

In eq. (2.1) a_{st}^r and $a_{st}^{r\dagger}$ are an equivalent set of anticommuting variables; s ranges from $-M$ to M and β ranges from $-N$ to N . The determinant of this transformation is one. One should think in terms of the correspondence:

$$(\alpha, \beta) \leftrightarrow (x, y), \quad (2.2)$$

$$\left(\frac{2\pi s}{2M+1}, \frac{2\pi t}{2N+1}\right) \leftrightarrow (p_x, p_y).$$

The variables s and t are simply momentum variables. Equation (2.1) implies periodic boundary conditions. These conditions will always be chosen, so that one is working on a torus²).

Equation (2.1) implies the following useful formulas:

$$\sum_{\alpha\beta} \eta_{\alpha\beta}^{q\dagger} \eta_{\alpha\beta}^r = \sum_{st} a_{st}^{q\dagger} a_{st}^r ,$$

$$\sum_{\alpha\beta} \eta_{\alpha\beta}^{q\dagger} \eta_{\alpha+1\beta}^r = \sum_{st} a_{st}^{q\dagger} a_{st}^r \exp\left(\frac{2\pi i s}{2M+1}\right) ,$$

$$\sum_{\alpha\beta} \eta_{\alpha\beta}^q \eta_{\alpha+1\beta}^{r\dagger} = \sum_{st} a_{st}^q a_{st}^{r\dagger} \exp\left(\frac{-2\pi i s}{2M+1}\right) ,$$

$$\sum_{\alpha\beta} \eta_{\alpha\beta}^{q\dagger} \eta_{\alpha\beta+1}^r = \sum_{st} a_{st}^{q\dagger} a_{st}^r \exp\left(\frac{2\pi i t}{2N+1}\right) , \tag{2.3}$$

$$\sum_{\alpha\beta} \eta_{\alpha\beta}^q \eta_{\alpha\beta+1}^{r\dagger} = \sum_{st} a_{st}^q a_{st}^{r\dagger} \exp\left(\frac{-2\pi i t}{2N+1}\right) ,$$

$$\sum_{\alpha\beta} \eta_{\alpha\beta}^q \eta_{\alpha\beta}^r = \sum_{st} a_{st}^q a_{-s-t}^r ,$$

$$\sum_{\alpha\beta} \eta_{\alpha\beta}^{q\dagger} \eta_{\alpha\beta}^{r\dagger} = \sum_{st} a_{st}^{q\dagger} a_{-s-t}^{r\dagger} .$$

The variables, q and r , refer to the types of anticommuting variables. The "operator" $\exp\left(\frac{2\pi i s}{2M+1}\right)$ is like the quantum mechanical operator $\exp(i\Delta \times p_x)$ which shifts one unit in the x -direction.

Let us first solve the close-packed dimer problem. The square of the partition function has the representation given by eqs. (I.4.1) and (I.4.2). Using eqs. (2.1) and (2.3)

$$\begin{aligned}
 Z_{\text{dimer}}^2(z_h, z_v) &= \int da da^\dagger \exp(A), \\
 A &= \sum_{st} A_{st}, \quad (2.4) \\
 A_{st} &= \left[z_h \left(a_{st}^1 a_{st}^{2\dagger} + a_{st}^{4\dagger} a_{st}^3 \right) + z_v \left(a_{st}^{2\dagger} a_{st}^3 + a_{st}^{4\dagger} a_{st}^1 \right) \right. \\
 &\quad + z_h \left(a_{st}^{2\dagger} a_{st}^1 \exp\left(\frac{2\pi i s}{2M+1}\right) + a_{st}^3 a_{st}^{4\dagger} \exp\left(-\frac{2\pi i s}{2M+1}\right) \right) \\
 &\quad \left. + z_v \left(a_{st}^1 a_{st}^{4\dagger} \exp\left(\frac{2\pi i t}{2N+1}\right) + a_{st}^3 a_{st}^{2\dagger} \exp\left(-\frac{2\pi i t}{2N+1}\right) \right) \right] \\
 &\quad + \left[\text{term with } a_{st}^r \text{ and } a_{st}^{r\dagger} \text{ interchanged and} \right. \\
 &\quad \left. \text{exponents conjugated} \right].
 \end{aligned}$$

In going from paper I to eq. (2.4) I have set $z_h^A = z_h^B = z_h$ and $z_v^A = z_v^B = z_v$.

The integrals over each (s, t) can be done individually using eq. (I.2.6) yielding the determinant of the following matrix:

$$M_{st} = \begin{pmatrix} 0 & h_s & 0 & -v_t \\ -h_s & 0 & v_t & 0 \\ 0 & -v_t & 0 & -h_s \\ v_t & 0 & h_s & 0 \end{pmatrix}, \quad (2.5)$$

with

$$h_s = z_h \left(1 - \exp \frac{2\pi i s}{2M+1} \right), \quad (2.6)$$

$$v_t = z_h \left(1 - \exp \frac{2\pi i t}{2N+1} \right).$$

$$\det M_{st} = (h_s h_{-s} + v_t v_{-t})^2 = \left[(2 - 2 \cos \frac{2\pi s}{2M+1}) z_h^2 + (2 - 2 \frac{2\pi t}{2N+1}) z_v^2 \right]^2. \quad (2.7)$$

The total answer is the product of these determinants:

$$Z_{\text{dimer}}^2(z_v, z_h) = \prod_{s=-M}^M \prod_{t=-N}^N \det M_{st} = \exp \left\{ \sum_{s=-M}^M \sum_{t=-N}^N \ln(\det M_{st}) \right\}. \quad (2.8)$$

The free energy per unit site in the thermodynamic limit,

$f = -kT \ln Z$, becomes

$$-\beta f = \frac{1}{4} \int_{-\pi}^{\pi} \frac{dp_x}{(2\pi)} \int_{-\pi}^{\pi} \frac{dp_y}{(2\pi)} \ln \left[(2 - 2 \cos p_x) z_h^2 + (2 - 2 \cos p_y) z_v^2 \right], \quad (2.9)$$

which agrees with the well-known answer^{3,4,5,6}). In obtaining

eq. (2.9) sums have been replaced by integrals in the standard way

and $p_x \equiv \frac{2\pi s}{2M+1}$ and $p_y \equiv \frac{2\pi t}{2N+1}$. Finally the factor of $\frac{1}{4}$ is

due to fact that there are $4(2N+1)(2M+1)$ sites.

Now consider the two-dimensional Ising model, which can be related to the partition function for closed polygons (reviewed in Section III of paper I). The corresponding action is given by eq. (I. 4.4). In calculating the partition function, any values of b's and a's satisfying $b_h = \pm 1$, $b_v = \pm 1$, $a_1 a_3 = 1$, $a_2 a_4 = 1$, may be used. For convenience choose $b_h = b_v = 1$, $a_1 = a_3 = 1$, and $a_4 = -a_2 = i$. Equations (I. 4.4), (2.1), and (2.3) result in the action

$$A_{\text{closed polygons}} = \sum_{st} A_{st},$$

$$A_{st} = \left[z_h a_{st}^h a_{st}^h \exp\left(\frac{2\pi i s}{2M+1}\right) + z_v a_{st}^v a_{st}^v \exp\left(\frac{2\pi i t}{2N+1}\right) \right. \\ \left. + a_{st}^h a_{st}^v + a_{st}^v a_{st}^h + i a_{st}^h a_{-s-t}^v + i a_{st}^v a_{-s-t}^h + a_{st}^v a_{st}^v + a_{st}^h a_{st}^h \right]. \quad (2.10)$$

The (s,t) variables mix with (-s, -t) variables. Therefore, after doing the integrations, eq. (2.10) will result in a product of Pfaffians of 8×8 dimensional matrices. However, transforming

$$a_{st} \rightarrow a_{st}^\dagger, \quad \left(\begin{array}{c} s < 0 \\ \text{or} \\ s = 0 \ t < 0 \end{array} \right) \quad (2.11)$$

$$a_{st}^\dagger \rightarrow -a_{st},$$

for both horizontal and vertical variables, the action, (except for A_{00}) becomes of sa^\dagger form (this would not have worked for the choice $a_4 = a_2 = 1$):

$$A_{st} = \left[\begin{array}{c} h_s a_{st} h_s^\dagger + v_t a_{st} v_t^\dagger \\ - a_{st} v_{st}^\dagger - a_{st} h_{st} v_{st}^\dagger - i a_{st} h_{-s-t} v_{-s-t}^\dagger + i a_{st} v_{-s-t} h_{-s-t}^\dagger \end{array} \right], \quad (2.12)$$

where

$$h_s = 1 - z_h \exp\left(\frac{2\pi i s}{2M+1}\right), \quad (2.13)$$

$$v_t = 1 - z_v \exp\left(\frac{2\pi i t}{2N+1}\right).$$

By eq. (I. 2.6) the integration over (s,t) and $(-s, -t)$ variables is the determinant of the following matrix:

$$M_{st} = \begin{pmatrix} h_s & -1 & 0 & -i \\ -1 & v_t & i & 0 \\ 0 & -i & h_{-s} & -1 \\ i & 0 & -1 & v_{-t} \end{pmatrix} \quad (s,t) \neq (0,0). \quad (2.14)$$

$$\det M_{st} = h_s h_{-s} v_t v_{-t} - (h_s + h_{-s})(v_t + v_{-t}) + 4$$

$$= (1 + z_h^2)(1 + z_v^2) + 2(1 - z_v^2) z_h \cos \frac{2\pi s}{2M+1} + 2(1 - z_h^2) z_v \cos \frac{2\pi t}{2N+1}. \quad (2.15)$$

The $(0,0)$ integral must be done separately and gives $-(\det M_{00})^{1/2}$.

I conclude

$$\begin{aligned}
 Z_{\text{closed polygon}} &= \prod_{st} (\det M_{st})^{1/2}, \\
 &= \exp \left[\frac{1}{2} \sum_{st} \ln (\det M_{st}) \right], \\
 &\rightarrow \exp \left\{ (2N+1)(2M+1) \left(\frac{1}{2} \right) \int_{-\pi}^{\pi} \frac{dp_x}{2\pi} \int_{-\pi}^{\pi} \frac{dp_y}{2\pi} \right. \\
 &\quad \left. \times \ln \left[(1+z_h^2)(1+z_v^2) + 2(1-z_v^2)z_h \cos p_x + 2(1-z_h^2)z_v \cos p_y \right] \right\}
 \end{aligned} \tag{2.16}$$

The exponent, $1/2$, compensates for double counting (s,t) and $(-s, -t)$. In the last step of eq. (2.16) the thermodynamic limit has been taken. The angular integration variables, p_x and p_y , are simply momentum variables. Equations (2.16), (I. 3.7), and (I.3.8) [or (I. 3.9)] yield the famous Onsager result⁷⁾ for the free energy per unit site

$$\begin{aligned}
 -\beta f_{\text{Ising}} &= \frac{1}{2} \int_{-\pi}^{\pi} \frac{dp_x}{2\pi} \int_{-\pi}^{\pi} \frac{dp_y}{2\pi} \ln 4 \left[\cosh 2\beta J_v \cosh 2\beta J_h \right. \\
 &\quad \left. + \sinh 2\beta J_h \cos p_x + \sinh 2\beta J_v \cos p_y \right].
 \end{aligned} \tag{2.17}$$

III. GRAPHICAL EVALUATION OF PARTITION FUNCTIONS

In this section I will introduce a graphical method to calculate partition functions. Later, it will be extended to correlation

functions. I do this because when the number, T , of variables at a site becomes large, the evaluation of Pfaffians and determinants becomes cumbersome. It is important to have a systematic approach. I will introduce a diagrammatic method which organizes the algebraic computations. For the models dealt with so far, it will seem superfluous; however, when more complicated models are encountered, it will be quite useful. The one danger is the possibility of overlooking a graph.

Consider A_{st} in eq. (2.4). It is like a miniature dimer problem on four sites. The first and second brackets [on the left hand side of eq. (2.4)] correspond respectively to the dimers of figs. 1a and 1b. Together they form the miniature dimer problem of fig. 1c. Figure 2 gives the four possible coverings of fig. 1c and their weights. Overall signs are determined by the rules of fig. (I.8). The sum of these diagrams yields eq. (2.7) as it should.

Let us now solve the generalized closed polygon problem given by (I. 4.4), using the diagrammatic approach. This model is called the free-fermion model⁸). This problem has been solved by expressing the partition function as a product of fermion creation and annihilation operators acting on a vacuum⁹). This is the reason for the name free-fermion. The method of reference 9 is, however, different from the one used here. In particular, anticommuting variables satisfy $\{\eta^r, \eta^{s\dagger}\} = 0$ and cannot be thought of creation and annihilation operators which satisfy $\{\psi^{r\dagger}, \psi^{s\dagger}\} = \delta_{rs}$.

In momentum space the action of eq. (I. 4.4) becomes

$$\begin{aligned}
 A_{\text{free fermion}} = \sum_{s,t} & \left[z_h a_{st}^h a_{st}^{h\dagger} \exp\left(\frac{2\pi i s}{2M+1}\right) + z_v a_{st}^v a_{st}^{v\dagger} \exp\left(\frac{2\pi i t}{2N+1}\right) \right. \\
 & + a_{1st} a_{st}^{h\dagger} a_{st}^v + a_{2st} a_{st}^v a_{st}^{h\dagger} + a_{3st} a_{st}^v a_{-s-t}^h + a_{4st} a_{st}^v a_{-s-t}^h \\
 & \left. + b_h a_{st}^h a_{st}^{h\dagger} + b_v a_{st}^v a_{st}^{v\dagger} \right] .
 \end{aligned} \quad (3.1)$$

If (s,t) and $(-s, -t)$ variables are grouped together, the mixture dimer problem associated with eq. (3.1) is illustrated in fig. 3. There are nine possible coverings as fig. 4 shows. The sum of these is

$$\begin{aligned}
 L\left(\frac{2\pi s}{2M+1}, \frac{2\pi t}{2N+1}\right) = & h_s h_{-s} v_t v_{-t} - a_1 a_3 (h_s v_t + h_{-s} v_{-t}) \\
 & - a_2 a_4 (h_s v_{-t} + h_{-s} v_t) + (a_1 a_3 + a_2 a_4)^2 ,
 \end{aligned} \quad (3.2)$$

where

$$\begin{aligned}
 h_s &= b_h - z_h \exp\left(\frac{2\pi i s}{2M+1}\right) , \\
 v_t &= b_v - z_v \exp\left(\frac{2\pi i t}{2N+1}\right) ,
 \end{aligned} \quad (3.3)$$

or

$$\begin{aligned}
L\left(\frac{2\pi s}{2M+1}, \frac{2\pi t}{2N+1}\right) = & \left[\left(b_h^2 + z_h^2 - 2z_h b_h \cos \frac{2\pi s}{2M+1} \right) \left(b_v^2 + z_v^2 - 2z_v b_v \cos \frac{2\pi t}{2N+1} \right) \right. \\
& - 2(a_1 a_3 + a_2 a_4) \left(b_h - z_h \cos \frac{2\pi s}{2M+1} \right) \left(b_v - z_v \cos \frac{2\pi t}{2N+1} \right) \\
& + 2(a_1 a_3 - a_2 a_4) z_h z_v \sin \frac{2\pi s}{2M+1} \sin \frac{2\pi t}{2N+1} \\
& \left. + (a_1 a_3 + a_2 a_4)^2 \right] . \tag{3.4}
\end{aligned}$$

The partition function is

$$Z_{\text{free fermion}} = \left(\prod_{s,t} L(s,t) \right)^{1/2} , \tag{3.5}$$

which becomes in the thermodynamic limit

$$-\beta f_{\text{free fermion}} = \frac{1}{2} \int_{-\pi}^{\pi} \frac{dp_x}{2\pi} \int_{-\pi}^{\pi} \frac{dp_y}{2\pi} \ln L(p_x, p_y) , \tag{3.6}$$

where L is given by eq. (3.4). The factor of $1/2$ is due to double counting of (s,t) and $(-s, -t)$. Equations (3.4) and (3.6) agree with the known result^{8,10}).

IV. SOLVABLE TWO-DIMENSIONAL DIMER PROBLEMS

This section considers solvable two-dimensional dimer problems. By solvable, I mean solvable by the usual Pfaffian methods⁵). The models will be translated into Grassmann integral form, from which a series of graphical rules will be derived. The treatment used here does not differ from the usual Pfaffian treatment. What is gained is a simple graphical approach which allows one to rapidly solve

a dimer problem. Furthermore, the diagrammatic methods extend to any pseudo-free field theory. This section serves as a pedagogical introduction to graphical methods.

I refer the reader to the standard method of solution⁵).

There are two key points:

I. Solvability Condition. A planar dimer problem is solvable if its bonds may be oriented so that every elementary polygon is clockwise odd. Planar means it may be drawn on a piece of paper so that bonds do not cross. The bonds are then given an orientation. The direction is usually denoted by an arrow. A polygon is clockwise odd, if when traversing clockwise, one encounters an odd number of bonds oriented in the opposite direction. An elementary polygon is a non self-intersecting polygon made up of bonds which has no bonds in its interior.

II. The Method of Solution. Fix a standard B configuration which covers the lattice. Each covering (these new ones will be called A coverings) when combined with the B configuration results in a set of closed polygons and isolated dimer pairs, the partition function of which has a Pfaffian representation.

Condition I and Observation II make the problem solvable by Pfaffian methods⁵).

For every model satisfying I, the Method of Solution II can be translated into Grassmann integral form: A bond oriented from point, P, to point, Q, upon which an A-dimer may be placed corresponds to a term $\eta_P \eta_Q$ in the action (see fig. 5a). A standard B-bond between P and Q corresponds to a term $\eta_Q^\dagger \eta_P^\dagger$ (see fig. 5b). A-dimer operators are ordered with the graph orientations, whereas

B-dimer operators are ordered oppositely to the graph orientations.

The action is schematically of the form

$$A_{\text{dimer}} = \sum_{\text{A-dimers}} z_A \eta \eta + \sum_{\text{B-dimers}} \eta^\dagger \eta^\dagger . \quad (4.1)$$

The Boltzmann factors of A-dimers are z_A , whereas B-dimers have unit Boltzmann factors. It is not hard to see that this action produces the closed polygons and isolated dimer pairs used in the Method of Solution II. The signs are all positive because of Condition I. This may be proved by induction on the length of a polygon and employing Kasteleyn's theorem¹¹). Figure 6a illustrates one set of orientations on a square lattice which makes every elementary polygon clockwise odd. Figures 6b and 6c show the A-dimers and a standard B-dimer configuration consisting of horizontal dimers. It is convenient to group the sites in units of four as in fig. (I.5). The corresponding action is

$$A_{\text{dimer}} = \sum_{\alpha\beta} \left[z_h \eta_{\alpha\beta}^1 \eta_{\alpha\beta}^2 + z_v \eta_{\alpha\beta}^2 \eta_{\alpha\beta}^3 + z_h \eta_{\alpha\beta}^4 \eta_{\alpha\beta}^3 + z_v \eta_{\alpha\beta}^4 \eta_{\alpha\beta}^1 \right. \\ \left. + z_h \eta_{\alpha\beta}^2 \eta_{\alpha+1\beta}^1 + z_h \eta_{\alpha\beta}^3 \eta_{\alpha+1\beta}^4 + z_v \eta_{\alpha\beta+1}^1 \eta_{\alpha\beta}^4 + z_v \eta_{\alpha\beta}^3 \eta_{\alpha\beta+1}^2 \right] \\ + \sum_{\alpha\beta} \left[\eta_{\alpha\beta}^{2\dagger} \eta_{\alpha\beta}^{1\dagger} + \eta_{\alpha\beta}^{3\dagger} \eta_{\alpha\beta}^{4\dagger} \right] , \quad (4.2)$$

where the notation is that of Paper I (secs. III and IV). Some dimer problems satisfy

Simplifying Condition C. A graph satisfies Simplifying Condition C if vertices can be grouped into two sets (which I call odd and even) such that no two odd (or even) vertices have a bond in common.

When this condition is satisfied, transform $\eta \rightarrow \eta^\dagger$ and $\eta^\dagger \rightarrow \eta$ at all even sites. This makes the bilinears in the action of the form $\eta\eta^\dagger$, the partition function becomes a product of determinants rather than Pfaffians, the graphical rules simplify, and calculations are easier to do. Figure 7 shows the rectangular lattice after this transformation.

Graphical Rules When Condition C Holds

or

Rules When Bilinears Are of $\eta\eta^\dagger$ Form

1. Group vertices into repeating units that fill a square array. Use (α, β) to label the units and use $r = 1, 2, 3, \dots, T$ to label the different vertices within a unit. Figure (I.5) is an example for the square lattice.

2. Consider one unit, U. There are two kinds of bonds:

(a) those which are contained within U and (b) those which go from U to some other unit. Of the latter, [(b)], for every bond which goes from a type r vertex in U to a type q vertex in another unit, there is one bond which goes from a type r vertex in another unit to a type q vertex in U. Thus, they occur in pairs. Half are to be included in U and the others ignored and erased. Figure 8 illustrates this for the square lattice.

3. Keep (a) type bonds as they are. For a (b) type bond which goes from an r in U to a q in another unit, "fold" it back into U , so that it goes from r to q within U (see fig. 9). If q is on "o" located in a unit m horizontal spaces to the right and n spaces upward (m and n may be negative) multiply the bond weight by

$$\exp (i m p_x + i n p_y) . \quad (4.3)$$

If q is an "x" multiply the bond weight by the complex conjugate of eq. (4.3), that is

$$\exp (- i m p_x - i n p_y) . \quad (4.4)$$

Figure 9 illustrates this. Figure 10 shows all the weights in the square lattice example after Rule 3 has been carried out.

4. Rules 1 through 3 result in a mixture dimer problem. Solve it by finding all coverings and their weights (see fig. 11 for the square lattice). Call the sum of the diagrams $L(p_x, p_y)$. The free energy per site, f , is

$$-\beta f = \frac{1}{T} \int_{-\pi}^{\pi} \frac{dp_x}{2\pi} \int_{-\pi}^{\pi} \frac{dp_y}{2\pi} \ln L(p_x, p_y) . \quad (4.5)$$

The factor of $\frac{1}{T}$ occurs because there are T sites per unit.

Figures 12-15 illustrate the solution for the hexagonal lattice dimer problem. Figure 12 shows the lattice, the bond orientations, the units of eight vertices, and the even and odd sites. The "horizontal" direction is in the x -direction; the "vertical"

direction is in the y -direction. Both these directions are also shown in Figure 12 (one must tilt the figure a bit). There are three Boltzmann factors, z_h , z_x , and z_y , corresponding to the three directions in which bonds may point. The Boltzmann factors, the A-dimers, and the standard B-dimer configuration are shown in Figure 13. The folded-over miniature dimer problem is shown in Figure 14. The possible coverings and their values are given in Figure 15. The result is eq. (4.6) with $T = 8$ and

$$L(p_x, p_y) = \left[z_h^4 + z_x^4 \exp(-2ip_x) + z_y^4 \exp(2ip_y) + 2z_h^2 z_x^2 \exp(-ip_x) + 2z_h^2 z_y^2 \exp(ip_y) - 2z_x^2 z_y^2 \exp(ip_y - ip_x) \right]. \quad (4.6)$$

Graphical Rules When Condition C Fails

or

Rules When Bilinears Are of $\eta\eta$ and $\eta^\dagger\eta^\dagger$ Form

These rules will be exemplified by treating the square lattice dimer problem of eq. (4.2). Although Condition C is satisfied, the simplifying transformation will not be performed. Thus the action will be eq. (4.2) as it stands. Figures 6b and 6c show the A and B dimers.

1. Same as above.

2. Same as above.

3. Draw two copies of U (see fig. 16). Call them U_1 and U_2

For (a) type bonds going from r to q draw two lines: one from r in U_1 to q in U_2 and one from r in U_2 to q in U_1

(see fig. 17). For $\eta\eta$ dimers (i.e. A-dimers) of (b) type originating at an r in U and terminating at a q in another unit, again draw two lines. First draw one from r in U_1 to q in U_2 and multiply its weight by $\exp(-\text{imp}_x - \text{inp}_y)$, then draw one from r in U_2 to q in U_1 and multiply its weight by $\exp(\text{imp}_x + \text{inp}_y)$ (see fig. 18). For $\eta^\dagger\eta^\dagger$ dimers (i.e. B-dimers) do the same as for $\eta\eta$ dimers but multiply weights by the complex conjugated phase factors of the $\eta\eta$ case (see fig. 18). In all cases, if bonds are oriented from r to q they remain so, regardless of whether they go from U_1 to U_2 or U_2 to U_1 . Figure 19 shows the resulting weights for the square lattice.

4. Solve the miniature dimer problem (see fig. 20) and call the result $L(p_x, p_y)$. The free energy per unit site is

$$-\beta f = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{dp_x}{2\pi} \int_{-\pi}^{\pi} \frac{dp_y}{2\pi} \ln L(p_x, p_y). \quad (4.7)$$

Graphical Rules For A General Pseudo-Free Theory

In general, there will be $\eta\eta^\dagger$, $\eta\eta$, and $\eta^\dagger\eta^\dagger$ products. Two copies, U_1 and U_2 , of U are to be drawn. Follow the second set of rules, 1,2,3, for $\eta\eta$ and $\eta^\dagger\eta^\dagger$ products. For $\eta\eta^\dagger$ terms use rule 3 of the first set for the U_1 copy of U but for U_2 use complex conjugated phase factors. Finally, use eq. (4.7) and rule 4. Figures 3 and 4 illustrate this for the action given in eq. (I. 4.4).

V. ANTICOMMUTING VARIABLE CORRELATIONS

This section will compute the anticommuting variable correlations (or "propagators") for the free fermion model [eq. (I. 4)]. The configurations and their weights were given in fig. I. 11. In addition, there are z_h and z_v Boltzmann factors for each unit of horizontal and vertical Bloch wall.

The correlation functions will first be calculated in momentum space and then in coordinate space. It will be done graphically. The variables s and t will be used instead of p_x and p_y . The two are related by eq. (2.2).

Consider $\langle a_{st}^h a_{st}^{h\dagger} \rangle$. The operator $a_{st}^h a_{st}^{h\dagger}$ places a dimer between the "o" and "x" at the horizontal (s, t) site. Unlike $\exp(a_{st}^h a_{st}^{h\dagger})$, however, one must use it. Therefore, $Z\langle a_{st}^h a_{st}^{h\dagger} \rangle$ is related to the miniature dimer problem (MDP) of fig. 3, where one inserts a "superbond" and erases all other bonds which connect to the (s, t) horizontal sites. The result is the modified miniature dimer problem (MMDP) of fig. 21(a). To calculate $\langle a_{st}^h a_{st}^{h\dagger} \rangle$ take the value of the MMDP and divide it by the value of the MDP of fig. 3.

General Rules For Calculating Momentum Space

Correlation Functions

1. Obtain the MDP using the rules of the last section.

Since s and t variables are used rewrite, p_x and p_y in terms of s and t using the correspondence of eq. (2.2). Calculate the value of the MDP and call it $D(s, t) \equiv L(p_x, p_y) = L\left(\frac{2\pi s}{2M+1}, \frac{2\pi t}{2N+1}\right)$.

2. Let a^1 and a^2 denote two generic anticommuting variables in the MDP of rule 1. To calculate $\langle a^1 a^2 \rangle$ draw a superbond from 1 to 2 and assign it unit weight. Erase all bonds involving the 1 and 2 variables. This is the MMDP. Call its value $N(s, t)$. Then

$$\langle a^1 a^2 \rangle = N(s, t) / D(s, t) . \quad (5.1)$$

Figures 21-24 calculate the non-zero $\langle aa^\dagger \rangle$ free fermion correlations, by showing first the MMDP and then its coverings. In these figures, the upper left and upper right variables are respectively $a_{st}^v, a_{st}^{v\dagger}$ and $a_{-s-t}^v, a_{-s-t}^{v\dagger}$. The lower left and lower right pairs are $a_{st}^h, a_{st}^{h\dagger}$ and $a_{-s-t}^h, a_{-s-t}^{h\dagger}$. The bond weights are those of Figure 3. The superbonds, denoted by darker lines, have unit weight. Figure 25 shows the MMDP's for the $\langle aa^\dagger \rangle$ correlations which have no coverings. They have zero value. Figures 26-28 and Figures 29-31 calculate the non-zero $\langle aa \rangle$ and $\langle a^\dagger a^\dagger \rangle$ correlations. Finally fig. 32 shows the MMDP's for the two remaining correlations which have no coverings. The tabulated results are

$$\langle a_{st}^h a_{st}^{h\dagger} \rangle = (h_{-s}^v v_{-t} - a_1 a_3 v_t - a_2 a_4 v_{-t}) / D(s, t) \quad (\text{Fig. 21}), \quad (5.2)$$

$$\langle a_{st}^v a_{st}^{v\dagger} \rangle = (h_s^h h_{-s}^v v_{-t} - a_1 a_3 h_s - a_2 a_4 h_{-s}) / D(s, t) \quad (\text{Fig. 22}), \quad (5.3)$$

$$\langle a_{st}^h a_{st}^{v\dagger} \rangle = a_1 |h_{-s}^v v_{-t} - (a_1 a_3 + a_2 a_4)| / D(s, t) \quad (\text{Fig. 23}), \quad (5.4)$$

$$\langle a_{st}^v a_{st}^{h\dagger} \rangle = a_3 |h_{-s}^v v_{-t} - (a_1 a_3 + a_2 a_4)| / D(s, t) \quad (\text{Fig. 24}), \quad (5.5)$$

$$\langle a_{st}^v a_{-s-t}^{h\dagger} \rangle = 0 \quad (\text{Fig. 25(a)}), \quad (5.6)$$

$$\langle a_{st}^h a_{-s-t}^{v\dagger} \rangle = 0 \quad (\text{Fig. 25(b)}), \quad (5.7)$$

$$\langle a_{st}^h a_{-s-t}^{h\dagger} \rangle = 0 \quad (\text{Fig. 25(c)}), \quad (5.8)$$

$$\langle a_{st}^v a_{-s-t}^{v\dagger} \rangle = 0 \quad (\text{Fig. 25(d)}), \quad (5.9)$$

$$\langle a_{st}^h a_{-s-t}^h \rangle = a_1 a_2 (v_t - v_{-t}) / D(s, t) \quad (\text{Fig. 26}), \quad (5.10)$$

$$\langle a_{st}^v a_{-s-t}^v \rangle = a_2 a_3 (h_{-s} - h_s) / D(s, t) \quad (\text{Fig. 27}), \quad (5.11)$$

$$\langle a_{st}^v a_{-s-t}^h \rangle = a_2 [(a_1 a_3 + a_2 a_4) - h_s v_{-t}] / D(s, t) \quad (\text{Fig. 28}), \quad (5.12)$$

$$\langle a_{st}^v a_{st'}^h \rangle = 0 \quad (\text{Fig. 32(a)}), \quad (5.13)$$

$$\langle a_{st}^{h\dagger} a_{-s-t}^{h\dagger} \rangle = a_3 a_4 (v_t - v_{-t}) / D(s, t) \quad (\text{Fig. 29}), \quad (5.14)$$

$$\langle a_{st}^{v\dagger} a_{-s-t}^{v\dagger} \rangle = a_1 a_4 (h_{-s} - h_s) / D(s, t) \quad (\text{Fig. 30}), \quad (5.15)$$

$$\langle a_{st}^{v\dagger} a_{-s-t}^{h\dagger} \rangle = a_4 [(a_1 a_3 + a_2 a_4) - h_s v_{-t}] / D(s, t) \quad (\text{Fig. 31}), \quad (5.16)$$

$$\langle a_{st}^{v\dagger} a_{st'}^{h\dagger} \rangle = 0 \quad (\text{Fig. 32(b)}), \quad (5.17)$$

where h_s , v_t , and $D(s, t)$ are given by eqs. (3.3) and (3.4).

Of course, correlations involving (s, t) and (s', t') variables vanish if neither $(s, t) \neq (s', t')$ nor $(s, t) = (-s', -t')$.

To obtain coordinate space correlations, use eq. (2.1) to express η 's in terms of a 's, and then use the results of eqs. (5.3)-(5.18). The thermodynamic limit can be taken and the correlations are

$$\langle \eta_{\alpha\beta}^h \eta_{\alpha'\beta'}^{\dagger} \rangle = \int_{-\pi}^{\pi} \frac{dp_x}{2\pi} \int_{-\pi}^{\pi} \frac{dp_y}{2\pi} \exp \left[i(\alpha - \alpha')p_x + i(\beta - \beta')p_y \right] \times \quad (5.18)$$

$$\left[h(-p_x)v(p_y)v(-p_y) - a_1 a_3 v(p_y) - a_2 a_4 v(-p_y) \right] / L(p_x, p_y),$$

$$\langle \eta_{\alpha\beta}^v \eta_{\alpha'\beta'}^{\dagger} \rangle = \int_{-\pi}^{\pi} \frac{dp_x}{2\pi} \int_{-\pi}^{\pi} \frac{dp_y}{2\pi} \exp \left[i(\alpha - \alpha')p_x + i(\beta - \beta')p_y \right] \times \quad (5.19)$$

$$\left[h(p_x)h(-p_x)v(-p_y) - a_1 a_3 h(p_x) - a_2 a_4 h(-p_x) \right] / L(p_x, p_y),$$

$$\langle \eta_{\alpha\beta}^h \eta_{\alpha'\beta'}^v \rangle = \int_{-\pi}^{\pi} \frac{dp_x}{2\pi} \int_{-\pi}^{\pi} \frac{dp_y}{2\pi} \exp \left[i(\alpha - \alpha')p_x + i(\beta - \beta')p_y \right] \times$$

$$a_1 \left[h(-p_x)v(-p_y) - (a_1 a_3 + a_2 a_4) \right] / L(p_x, p_y), \quad (5.20)$$

$$\langle \eta_{\alpha\beta}^v \eta_{\alpha'\beta'}^{\dagger} \rangle = \int_{-\pi}^{\pi} \frac{dp_x}{2\pi} \int_{-\pi}^{\pi} \frac{dp_y}{2\pi} \exp \left[i(\alpha - \alpha')p_x + i(\beta - \beta')p_y \right] \times$$

$$a_3 \left[h(-p_x)v(-p_y) - (a_1 a_3 + a_2 a_4) \right] / L(p_x, p_y), \quad (5.21)$$

$$\langle \eta_{\alpha\beta}^h \eta_{\alpha'\beta'}^h \rangle = \int_{-\pi}^{\pi} \frac{dp_x}{2\pi} \int_{-\pi}^{\pi} \frac{dp_y}{2\pi} \exp \left[i(\alpha - \alpha')p_x + i(\beta - \beta')p_y \right] \times$$

$$a_1 a_2 \left[v(p_y) - v(-p_y) \right] / L(p_x, p_y), \quad (5.22)$$

$$\langle \eta_{\alpha\beta}^v \eta_{\alpha'\beta'}^v \rangle = \int_{-\pi}^{\pi} \frac{dp_x}{2\pi} \int_{-\pi}^{\pi} \frac{dp_y}{2\pi} \exp \left[i(\alpha - \alpha')p_x + i(\beta - \beta')p_y \right] \times \\ a_2 a_3 \left[h(-p_x) - h(p_x) \right] / L(p_x, p_y), \quad (5.23)$$

$$\langle \eta_{\alpha\beta}^v \eta_{\alpha'\beta'}^h \rangle = \int_{-\pi}^{\pi} \frac{dp_x}{2\pi} \int_{-\pi}^{\pi} \frac{dp_y}{2\pi} \exp \left[i(\alpha - \alpha')p_x + i(\beta - \beta')p_y \right] \times \\ a_2 \left[(a_1 a_3 + a_2 a_4) - h(p_x)v(-p_y) \right] / L(p_x, p_y), \quad (5.24)$$

$$\langle \eta_{\alpha\beta}^{h\dagger} \eta_{\alpha'\beta'}^{h\dagger} \rangle = \int_{-\pi}^{\pi} \frac{dp_x}{2\pi} \int_{-\pi}^{\pi} \frac{dp_y}{2\pi} \exp \left[i(\alpha' - \alpha)p_x + i(\beta' - \beta)p_y \right] \times \\ a_3 a_4 \left[v(p_y) - v(-p_y) \right] / L(p_x, p_y), \quad (5.25)$$

$$\langle \eta_{\alpha\beta}^{v\dagger} \eta_{\alpha'\beta'}^{v\dagger} \rangle = \int_{-\pi}^{\pi} \frac{dp_x}{2\pi} \int_{-\pi}^{\pi} \frac{dp_y}{2\pi} \exp \left[i(\alpha' - \alpha)p_x + i(\beta' - \beta)p_y \right] \times \\ a_1 a_4 \left[h(-p_x) - h(p_x) \right] / L(p_x, p_y), \quad (5.26)$$

$$\langle \eta_{\alpha\beta}^{v\dagger} \eta_{\alpha'\beta'}^{h\dagger} \rangle = \int_{-\pi}^{\pi} \frac{dp_x}{2\pi} \int_{-\pi}^{\pi} \frac{dp_y}{2\pi} \exp \left[i(\alpha' - \alpha)p_x + i(\beta' - \beta)p_y \right] \times \\ a_4 \left[(a_1 a_3 + a_2 a_4) - h(p_x)v(-p_y) \right] / L(p_x, p_y), \quad (5.27)$$

where

$$h(p_x) = b_h - z_h \exp(ip_x), \quad (5.28)$$

$$v(p_y) = b_v - z_v \exp(ip_y),$$

and L is given by eq. (3.4). Equations (5.18), (5.19), (5.20), (5.21), (5.22), (5.23), (5.24), (5.25), (5.26), and (5.27) are respectively obtained from eqs. (5.2), (5.3), (5.4), (5.5), (5.10), (5.11), (5.12), (5.14), (5.15), and (5.16) by replacing h_s and v_t

by the corresponding momentum valued functions of eq. (5.28). The factors $\exp [i(\alpha - \alpha')p_x]$ and $\exp [i(\beta - \beta')p_y]$ in eqs. (5.18) - (5.24) are translation operators. Equations (5.25) - (5.27) have conjugated translation factors because daggered variables are involved.

Equations (5.18) - (5.27) are the coordinate-space anticommuting variable correlation functions for the free fermion model.

VI. THE ISING MODEL CORRELATION FUNCTIONS

This section will calculate the correlation function of two spin variables in the same row. It will be compared to the known result as a check on anticommuting variable techniques. Two horizontal spins are chosen for illustrative purposes only. The approach extends to an arbitrary pair; in fact, the vacuum expectation value of several σ 's can be computed. The only drawback is the cumbersome form of the answer: a Pfaffian of (in general) large size. In short, everything you ever wanted to know about the Ising model is expressible as a Pfaffian.

We will need the free fermion anticommuting variable correlations [eqs. (5.18) - (5.27)]. Bond variables will be used, in which case the Ising model is related to the free fermion (or closed - polygon) partition function by eqs. (I. 3.7) and (I. 3.9), when

$$a_1 = a_2 = a_3 = a_4 = b_v = b_h = -1. \quad (6.1)$$

The weights of configurations are given in fig. I.11. These values must be used (as opposed to the less restrictive conditions $a_1 a_3 = a_2 a_4 = b_v^2 = b_h^2 = 1$) because correlation functions, unlike the partition function, need not have the same number of a_1 and a_3

type corners, a_2 and a_4 type corners, etc. This is obvious from eqs. (5.18) - (5.27) where correlations are not simply functions of $a_1 a_3, a_2 a_4$, etc.

Spin variable correlation functions can be considered as partition functions on a defective lattice⁵). I refer the reader to reference 5, p. 248 - 257. This means that spin correlations are (up to multiplicative constants) the partition functions of Ising models with modified Bloch wall Boltzmann factors along selected paths. For example, $Z_{\text{Ising}}(\sigma_{1,0} \sigma_{m+1,0})$ is z_h^m times the Ising model with the usual z_h and z_v Boltzmann factors for all Bloch walls except for the horizontal ones between $(1,0)$ and $(m+1,0)$ where z_h^{-1} is the Boltzmann factor. This defective lattice partition function is obtained by replacing

$$\begin{aligned} & \exp\left(\sum_{\alpha=1}^m z_h \eta_{\alpha 0}^{h\dagger} \eta_{\alpha+1 0}^h\right) \quad \text{by} \quad \exp\left[\sum_{\alpha=1}^m z_h \eta_{\alpha 0}^{h\dagger} \eta_{\alpha+1 0}^h + \sum_{\alpha=1}^m (z_h^{-1} - z_h) \eta_{\alpha 0}^{h\dagger} \eta_{\alpha+1 0}^h\right] \\ & = \exp\left(\sum_{\alpha=1}^m z_h \eta_{\alpha 0}^{h\dagger} \eta_{\alpha+1 0}^h\right) \prod_{\alpha=1}^m \left[1 + (z_h^{-1} - z_h) \eta_{\alpha 0}^{h\dagger} \eta_{\alpha+1 0}^h\right], \quad \text{so that} \end{aligned}$$

$$\langle \sigma_{1,0} \sigma_{m+1,0} \rangle = \left\langle \prod_{\alpha=1}^m \left[s_h + (1 - z_h^2) \eta_{\alpha 0}^{h\dagger} \eta_{\alpha+1 0}^h \right] \right\rangle \quad (6.2)$$

Equation (6.2) typifies how spin variable correlations are related to anticommuting variable correlations. Equation (6.2) can be generalized to the case when the left hand side is the vacuum expectation value of several σ 's .

For pseudo-free theories, the following formulas are useful:

$$\langle \eta_1 \eta_2 \cdots \eta_m \rangle = \text{Pf } M_{ij} \quad (\text{for } m \text{ even}), \quad (6.3)$$

where

$$M_{ij} = \langle \eta_i \eta_j \rangle. \quad (6.4)$$

If $\langle \eta_i \eta_j \rangle = \langle \eta_i^\dagger \eta_j^\dagger \rangle = 0$, then

$$\langle \eta_1^\dagger \eta_1 \eta_2^\dagger \eta_2 \cdots \eta_m^\dagger \eta_m \rangle = \det M_{ij}, \quad (6.5)$$

where

$$M_{ij} = \langle \eta_i^\dagger \eta_j \rangle. \quad (6.6)$$

These formulas are the analogues of Wick's expansion. In eq. (6.3) one sums over all pairings of η 's, the sign of which is determined by how many permutations are required to get the η 's in paired form.

The vacuum expectation value of an arbitrary product of spins is expressible as a linear combination of anticommuting variable correlations. These vacuum expectation values can be computed using eqs. (5.18) - (5.27) and eq. (6.3). I will demonstrate this for two horizontal spins.

Equations (5.22) and (5.25) imply $\langle \eta_{\alpha 0}^h \eta_{\beta 0}^h \rangle = \langle \eta_{\alpha 0}^{h\dagger} \eta_{\beta 0}^{h\dagger} \rangle = 0$ for all α and β . Apply eq. (6.5) to (6.2). The z_h term of

$|z_h + (1 - z_h^2) \eta_{\alpha 0}^{h\dagger} \eta_{\alpha+1}^h |$ in eq. (6.2) multiplies the same factor as the term in the Wick expansion obtained by contracting

$\eta_{\alpha 0}^{h\dagger}$ with $\eta_{\alpha+1 0}^h$. Therefore

$$\langle \sigma_{1,0}^{\alpha} \sigma_{m+1,0} \rangle = \det M_{ij}, \quad (6.7)$$

where

$$M_{ij} = z_h \delta_{ij} + (1 - z_h^2) \langle \eta_{i0}^{h\dagger} \eta_{j+1 0}^h \rangle \quad (6.8)$$

$$= \int_{-\pi}^{\pi} \frac{dp_x}{2\pi} \int_{-\pi}^{\pi} \frac{dp_y}{2\pi} \exp \left[i p_x (j - i) \right] \quad (6.8)$$

$$\left\{ z_h - (1 - z_h^2) \exp(ip_x) \left| h(p_x) v(p_y) v(-p_y) - v(p_y) v(-p_y) \right| \right\} / L(p_x, p_y).$$

In obtaining eq. (6.8), eq. (5.18) has been used. Equations (6.7) and (6.8) express the correlation function of two horizontal spins as a Toeplitz determinant, as is usually done and yields the correct result^{4,5}).

To calculate the vacuum expectation value of a product of spin variables, proceed analogously. It will be equivalent to an Ising model on a defective lattice. When expressed in terms of anticommuting variables, it will result in an expression of the form

$$(6.9)$$

$$\langle \prod \sigma \text{'s} \rangle = \langle (c_{12} + d_{12} \eta_1 \eta_2) (c_{34} + d_{34} \eta_3 \eta_4) \cdots (c_{2m-1 2m} + d_{2m-1 2m} \eta_{2m-1} \eta_{2m}) \rangle$$

In eq. (6.9) η_i denotes an anticommuting variable such as

$\eta_{\alpha\beta}^h$, $\eta_{\alpha\beta}^{h\dagger}$, $\eta_{\alpha\beta}^v$, or $\eta_{\alpha\beta}^{v\dagger}$. The variables $c_{i i+1}$ and $d_{i i+1}$ are constants determined by the defective lattice. For convenience

write $d_{i+1} = d_i d_{i+1}$; any values of d_i satisfying this will do. Wick's expansion along with eq. (6.3) tells us that eq. (6.9) is

$$\langle \prod \sigma' \rangle = \text{Pf } M_{ij}, \quad (6.10)$$

where

$$M_{ij} = \begin{cases} d_i d_j \langle \eta_i \eta_j \rangle + \delta_{i+1, j} c_{i+1}, & i \text{ odd} \\ d_i d_j \langle \eta_i \eta_j \rangle - \delta_{i-1, j} c_{i-1}, & i \text{ even} . \end{cases} \quad (6.11)$$

The $\langle \eta \eta \rangle$ correlations are given in eqs. (5.18) - (5.28).

In principle, all Ising model spin correlations may be calculated using the above method. The reason they result in such cumbersome expressions is the following: The variables which solve the Ising model are the η 's. They might be called the mathematical variables because they represent it as a pseudo-free field theory. Correlation functions of anticommuting variables are simple to compute. Contrast this with the spin variables. They are the physical variables. They are, however, complicated functions of the mathematical variables, the η 's, which means that spin variable computations result in cumbersome expressions. In conclusion, there are two types of variables, spin variables which have a simple physical interpretation but are mathematically awkward to work with and η variables which do not have as simple a physical interpretation but are easy to work with mathematically.

VII. SUMMARY

Here is a summary of these first two papers. The focus of attention was solvable two-dimensional statistical mechanics models, in particular, the Ising model, the free-fermion model, and the close-packed dimer problems. The partition functions were expressed as integrals over anticommuting variables. In this form they resembled fermionic field theories. The solvable models had quadratic actions, which were computed by using free field theory techniques. More importantly, a series of graphical rules were derived which allowed one to compute partition functions and anticommuting variable correlation functions by solving miniature dimer problems. This provided a quick and simple graphical calculational approach. Many models can be solved by drawing a few diagrams. Finally, I showed how to calculate the vacuum expectation value of an arbitrary number of Ising spin variables.

For the most part, there are no new results. What has been gained is a powerful reorganization of old methods. Abstruse Pfaffian techniques have been rewritten as a set of simple graphical rules so that calculations are straightforward and systematic. The Grassmann integral has formulated the problem in terms of a field theory where powerful field theory methods have been applied.

These first two papers have dealt with solvable models. One need only add a term,

$$\sum_{\alpha\beta} \Delta_{\alpha\beta}^h h^\dagger v v^\dagger, \quad (7.1)$$

to the free-fermion action of eq. (I. 4.4) to break the free-fermion constraint and obtain the general eight vertex model⁸). This model is unsolved. It is an interacting field theory. The approximation methods used for interacting field theories can be applied to it. Here is where the real power of anticommuting variables is. Most interesting statistical mechanics problems are not solvable; an example is the 3-d Ising model¹²). It is important to have viable approximation schemes. Such schemes will be obtained via Grassmann integrals. Furthermore, they will be, in general, systematic and simple.

In short, these first two papers have formed a testing ground for anticommuting variable techniques. They formed a solid foundation of solvable models upon which unsolvable models can be approached.

REFERENCES

1. S. Samuel, The Use of Anticommuting Integrals in Statistical Mechanics I. References to equations and figures in this paper will be prefixed by a I, e.g. eq. (I. 1.1) and fig. I.1 refer to equation (1.1) and fig. 1 of reference 1.
2. When this is done, some terms have the incorrect sign (those involving loops around the torus). This difficulty can be overcome by standard methods (see pages 61 to 67 of reference 4). However, in the thermodynamic limit, such configurations will be a zero measure effect and can be ignored.
3. H. N. V. Temperley and M. E. Fisher, *Phil. Mag.* 6, (1961) 1061.
M. E. Fisher, *Phys. Rev.* 124, (1961) 1664.
4. See, for example, B. M. McCoy and T. T. Wu, The Two-Dimensional Ising Model (Harvard University Press, Cambridge, 1973).
5. See, for example, E. W. Montroll, Brandeis University Summer Institute in Theoretical Physics, 1966, edited by M. Chrétien, E. P. Gross, and S. Deser (Gordon and Breach, New York, 1968).
6. See, for example, H. S. Green and C. A. Hurst, Order-Disorder Phenomena (Interscience, New York, 1964).
7. L. Onsager, *Phys. Rev.* 65, (1964) 117. Textbook derivations are given in references 4,5, and 6.
8. C. Fan and F. Y. Wu, *Phys. Rev.* B2, (1970) 725.
9. See reference 6, Chapter 4.
10. Reference 6, Sec. 5.3.
11. P. M. Stealey, *J. Math. Phys.* 4, (1963) 287. A pedagogical version is given in reference 5.
12. The techniques used here extend to drawing closed surfaces in three

dimensions. Thus, the 3-d Ising model has an interesting interacting fermionic-like field theory representation.

S. Samuel, to be published.

Figure 1. The Miniature Dimer Problem. Figures (a) and (b) are the graphical representation of the first and second terms in A_{st} of eq. (2.4). The h_s, v_t , etc. (eq. 2.6) factors are the weights of the dimers. The sum of (a) and (b) form the miniature dimer problem of (c).

Figure 2. The Four Possible Coverings. The weights of these diagrams are: (a) $(v_t v_{-t})^2$, (b) $(h_s h_{-s})^2$, (c) $(h_s h_{-s} v_t v_{-t})$, (d) $(h_s h_{-s} v_t v_{-t})$. The sum yields eq. (2.7).

Figure 3. The Miniature Dimer Problem for the Free-Fermion model. The upper left "o" and "x" are $a_{st}^v, a_{st}^{v\dagger}$; the lower left are $a_{st}^h, a_{st}^{h\dagger}$; the upper right are $a_{-s-t}^v, a_{-s-t}^{v\dagger}$; the lower right are $a_{-s-t}^h, a_{-s-t}^{h\dagger}$. The weights of bonds are as indicated with h_s and v_t given by eq. (3.3).

Figure 4. The Possible Coverings of fig.3. The arrows are shown to aid in determining the sign [use rules of fig. (I.8)]. The values of these diagrams are (a) $(h_s h_{-s} v_t v_{-t})$, (b) $(-a_1 a_3 h_s v_t)$, (c) $(-a_1 a_3 h_{-s} v_{-t})$, (d) $(a_1 a_3 a_1 a_3)$, (e) $(-a_2 a_4 h_s v_{-t})$, (f) $(-a_2 a_4 h_{-s} v_t)$, (g) $(a_1 a_2 a_3 a_4)$, (h) $(a_1 a_2 a_3 a_4)$, and (i) $(a_2 a_4 a_2 a_4)$.

Figure 5. The A and B Dimer Operators. In (a) is a typical bond oriented from P to Q. In the action will correspond the term $\eta_P \eta_Q$ as in (b). If a standard B-dimer lies on this bond then there is a term $\eta_Q^\dagger \eta_P^\dagger$ as in (c). The A-dimers are associated with $\eta \eta$ products, whereas

B-dimers are associated with $\eta^\dagger \eta^\dagger$ products.

Figure 6. Square Lattice Dimer Problem. Figure (a) shows the orientations which make every elementary square clockwise odd. Figure (b) represents the A-dimer operators and fig. (c) is the standard B-dimer configuration consisting of horizontal dimers.

Figure 7. The Simplifying Transformation. Condition \cup holds for the square lattice of fig. 6. After the transformation $\eta^\dagger \rightarrow \eta$ at even sites, the dimer operators of figs. 6a and 6b become those shown here. The B-dimers are drawn above the A-dimers.

Figure 8. Illustration of Rule 2. Figure (a) shows the (α, β) unit. There are two B-dimers and four A-dimers entirely contained in (α, β) . There are eight A-dimers which connect sites in (α, β) to sites in nearby units. They occur in pairs. For example, the upper right A-dimer, $\eta_{\alpha\beta}^3 \eta_{\alpha\beta+1}^{2\dagger}$, has a partner, the lower right A-dimer, $\eta_{\alpha\beta-1}^3 \eta_{\alpha\beta}^{2\dagger}$. Rule 2 erases one bond from each pair. Figure (b) is an example of what results.

Figure 9. Rule 3 for $\eta \eta^\dagger$ Products. Figure (a) shows the two dimers of fig. 8b which start in the (α, β) unit at sites 2 and 3 and go to the sites 1 and 4 of the $(\alpha + 1, \beta)$ unit. Rule 3 says to "fold" these back into the (α, β) unit as shown in (b). Let "o" and "x" correspond to the anticommuting variables a and a^\dagger . Then the $a_2^\dagger a_1$ bond weight gets multiplied by $\exp(ip_x)$ whereas the $a_3 a_4^\dagger$ weight gets multiplied by $\exp(-ip_x)$.

Figure 10. The Weights for the Square Lattice. Rule 3 applied to fig. 8b results in this figure. The weights of the B-dimers remains 1 as indicated. The A-dimer weights have contributions from (a) type bonds as well as (b) types. When added they result in the factors

$$h(p_x) = z_h [1 - \exp(ip_x)], \quad v(p_y) = z_v [1 - \exp(ip_y)], \text{ etc.}$$

Figure 11. The Two Coverings of Figure 10. The value of (a) is $h(p_x)h(-p_x) = z_h^2(2 - 2 \cos p_x)$. The value of (b) is $v(p_y)v(-p_y) = z_v^2(2 - 2 \cos p_y)$. The sum of these is $L(p_x, p_y)$. When put into eq. (4.5), the free energy per site is obtained.

Figure 12. The Hexagonal Dimer Problem. This is the hexagonal lattice. The above bond orientation make every elementary hexagon clockwise odd. The units are outlined by dotted lines. There are eight sites in each, and (α, β) label them. This lattice satisfies Simplifying Condition C; the odd sites are denoted by larger dots. The x-direction is northeast and the y-direction is northwest as indicated.

Figure 13. The Dimer Operators. Figure (a) shows the A-dimers and their weights. Only half of the "external" dimers have been kept in accord with rule 3. Figure (b) shows the B-dimers. Their weights are unity. If this B configuration is chosen in every unit, then every site is covered by a B-dimer.

Figure 14. The Miniature Dimer Problem. Rule 3 applied to Figure 13 results in this miniature dimer problem with the indicated bond weights.

Figure 15. The Coverings of Figure 14. There are nine coverings. Their values are (a) (z_h^4) , (b) $(z_x^4 \exp(-2ip_x))$, (c) $(z_y^4 \exp(2ip_y))$, (d) $(z_h^2 z_x^2 \exp(-ip_x))$, (e) $(z_h^2 z_x^2 \exp(-ip_x))$, (f) $(z_h^2 z_y^2 \exp(ip_y))$, (g) $(z_h^2 z_y^2 \exp(ip_y))$, (h) $(-z_x^2 z_y^2 \exp(ip_y - ip_x))$ and (i) $(-z_x^2 z_y^2 \exp(ip_y - ip_x))$. The sum of these values gives the $L(p_x, p_y)$ of eq. (4.6).

Figure 16. The Two Copies. For the square lattice, U consists of four sites. Rule 3 says to draw two copies of U . These are labelled U_1 and U_2 . The different sites within each U_1 have been numbered 1, 2, 3, and 4. One should think of U_1 as representing (s, t) variables and U_2 as representing $(-s, -t)$ variables.

Figure 17. The (a)-Type Bonds. In Figure (a), there is an A-dimer from fig. 6b and a B-dimer from fig. 6c. Each of these results in two dimers, one from U_1 to U_2 and one from U_2 to U_1 as (b) indicates. The orientation remains the same, so that the A-dimer in U which goes from 4 to 3, still goes from 4 to 3 in both cases in Figure (b).

Figure 18. The (b)-type Bonds. Figure (a) shows one $\eta\eta$ (b)-type bond and one $\eta^\dagger\eta^\dagger$ (b)-type bond. Although the latter does not occur in the standard B configuration of

fig. 6c, it has been put in here for illustrative purposes. If U is the (α, β) unit then the two bonds go from the (α, β) unit to the $(\alpha + 1, \beta)$ unit. Both give rise to two dimers in (b) the weights of which get multiplied by the indicated phase factors.

Figure 19. The Resulting Bond Weights. Figure (a) shows the resulting A-dimers and their bond weights. Figure (b) shows the B-dimers. Their weights are all unity. Here, $h(p_x) = z_h[1 - \exp(ip_x)]$ and $v(p_y) = z_v[1 - \exp(ip_y)]$. When superimposed (a) and (b) give rise to a miniature dimer problem.

Figure 20. The Coverings of Figure 19. There are four coverings of Figure 19. Their values are (a) $[h(p_x)h(-p_x)]^2$, (b) $[v(p_y)v(-p_y)]^2$, (c) $[h(p_x)h(-p_x)v(p_y)v(-p_y)]$, and (d) $[h(p_x)h(-p_x)v(p_y)v(-p_y)]$.

Figure 21. The $\langle a_{st}^h a_{st}^{h\dagger} \rangle$ correlation. Figure (a) is the MMDP. Figures (b), (c), and (d) are the coverings. Their values are (b) $(h_{-s} v_{-t} v_{-t})$, (c) $(-a_1 a_3 v_t)$, and (d) $(-a_2 a_4 v_{-t})$.

Figure 22. The $\langle a_{st}^v a_{st}^{v\dagger} \rangle$ correlation. Figure (a) is the MMDP. Figures (b), (c), and (d) are the coverings. Their values are (b) $(h_s h_{-s} v_{-t})$, (c) $(-a_1 a_3 h_s)$, and (d) $(-a_2 a_4 h_{-s})$.

Figure 23. The $\langle a_{st}^h a_{st}^{v\dagger} \rangle$ correlation. Figure (a) is the MMDP. Figures (b), (c) and (d) are the coverings. Their values are (b) $(-a_1 a_2 a_4)$, (c) $(-a_1 a_1 a_3)$, and (d) $(a_1 h_{-s} v_{-t})$.

Figure 24. The $\langle a_{st}^v a_{st}^{h\dagger} \rangle$ correlation. Figure (a) is the MMDP. Figures (b), (c), and (d) are the coverings. Their values are (b) $(-a_2 a_3 a_4)$, (c) $(-a_1 a_3 a_3)$, and (d) $(a_3 h_{-s} v_{-t})$.

- Figure 28. The zero $\langle aa^\dagger \rangle$ correlations. Figures (a), (b), (c), and (d) are the MMDP's for the $\langle a_{st}^v a_{-s-t}^h \rangle$, $\langle a_{st}^h a_{-s-t}^v \rangle$, $\langle a_{st}^h a_{-s-t}^h \rangle$, and $\langle a_{st}^v a_{-s-t}^v \rangle$ correlations. None of these MMDP's have any coverings.
- Figure 29. The $\langle a_{st}^h a_{-s-t}^h \rangle$ correlations. Figure (a) is the MMDP. Figures (b) and (c) are the two coverings. Their values are (b) $(a_1 a_2 v_t)$ and (c) $(-a_1 a_2 v_{-t})$.
- Figure 30. The $\langle a_{st}^v a_{-s-t}^v \rangle$ correlation. Figure (a) is the MMDP. Figures (b) and (c) are the two coverings. Their values are (b) $(-a_2 a_3 h_s)$ and (c) $(a_2 a_3 h_{-s})$.
- Figure 31. The $\langle a_{st}^v a_{-s-t}^h \rangle$ correlation. Figure (a) is the MMDP. Figures (b), (c), and (d) are the coverings. Their values are (b) $(a_2 a_2 a_4)$, (c) $(a_1 a_2 a_3)$, and (d) $(-a_2 h_s v_{-t})$.
- Figure 29. The $\langle a_{st}^h a_{-s-t}^h \rangle$ correlations. Figure (a) is the MMDP. Figures (b) and (c) are the coverings. Their values are (b) $(a_3 a_4 v_t)$ and (c) $(-a_3 a_4 v_{-t})$.
- Figure 30. The $\langle a_{st}^v a_{-s-t}^v \rangle$ correlation. Figure (a) is the MMDP. Figures (b) and (c) are the coverings. Their values are (b) $(a_1 a_4 h_{-s})$ and (c) $(-a_1 a_4 h_s)$.
- Figure 31. The $\langle a_{st}^v a_{-s-t}^h \rangle$ correlation. Figure (a) is the MMDP. Figures (b), (c), and (d) are the coverings. Their values are (b) $(a_2 a_4 a_4)$, (c) $(a_1 a_3 a_4)$, and (d) $(-a_4 h_s v_{-t})$.
- Figure 32. The $\langle a_{st}^v a_{st}^h \rangle$ and $\langle a_{st}^v a_{st}^v \rangle$ correlations. Figures (a) and (b) are the MMDP's. Neither has a covering

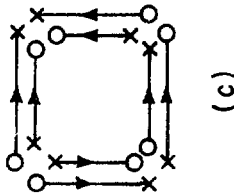
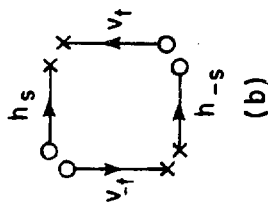
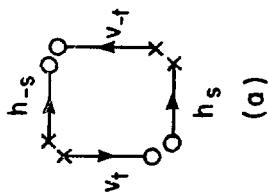


Figure 1

XBL 7810 - 11589

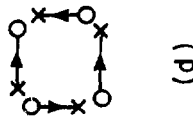
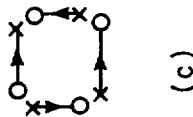
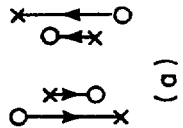
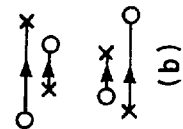


Figure 2

XBL 7810 - 11590

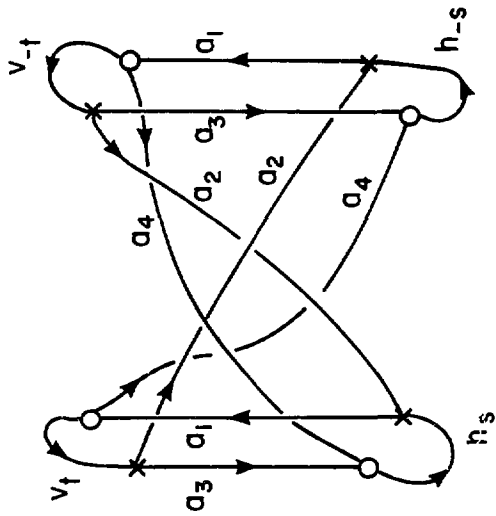


Figure 3 XBL 7810-11598

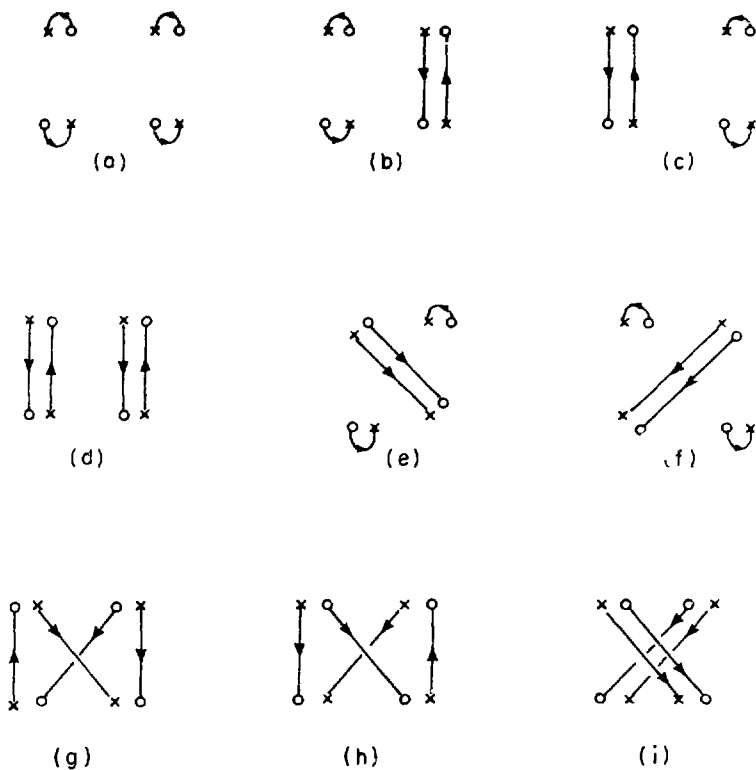
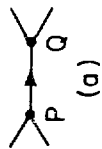
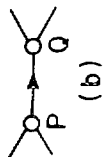


Figure 4

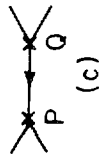
XBL7810 - 6608



(a)



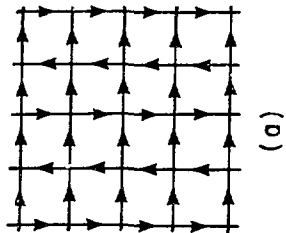
(b)



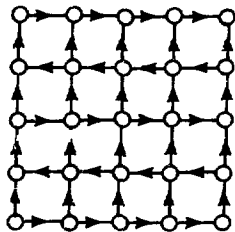
(c)

Figure 5

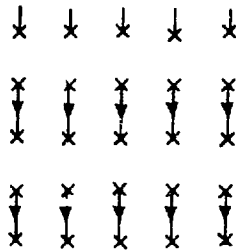
XB 7810-11593



(a)



(b)



(c)

Figure 6

XL 7810-11594

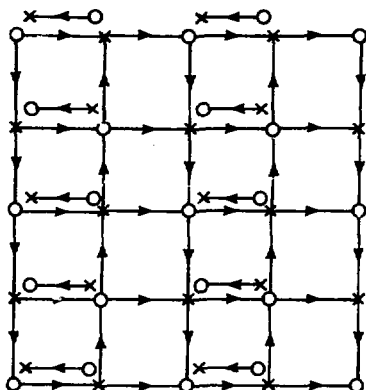


Figure 7 xBL7810-11591

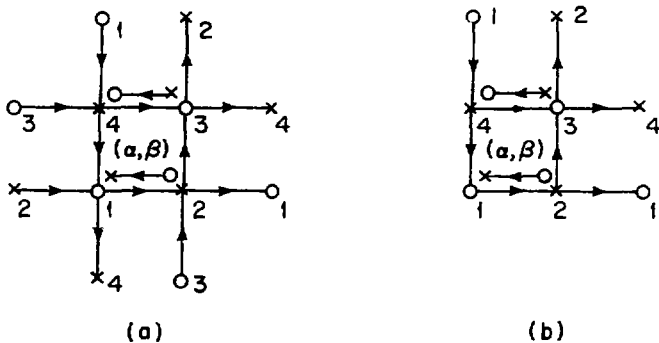


Figure 8

xBL7810-11592

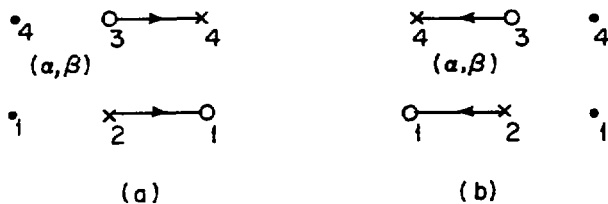


Figure 9

XBL7810-11581

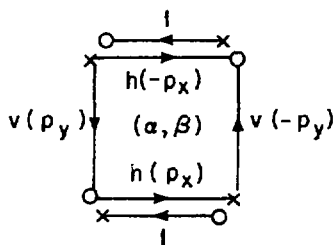


Figure 10

XBL7810-11582

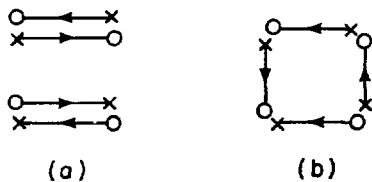


Figure 11

XBL7810-11583

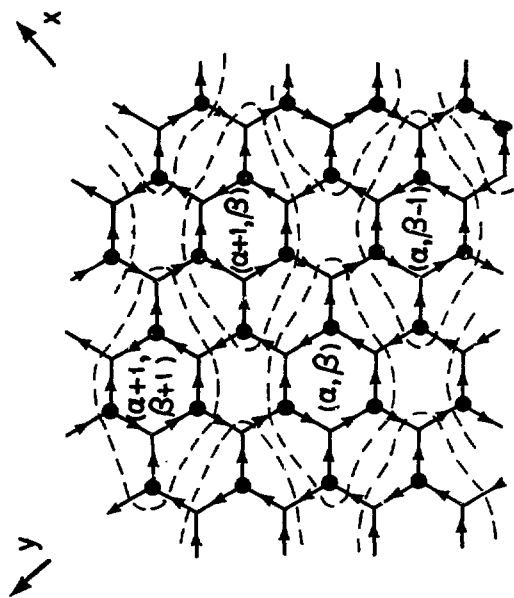
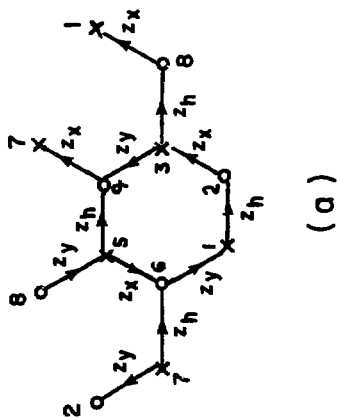
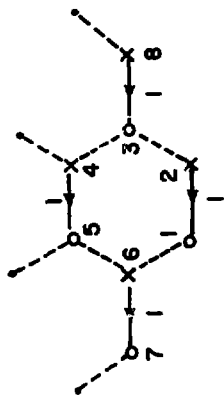


Figure 12 XBL 7810 - 6620



(a)



(b)

Figure 13

XBL7810-6619

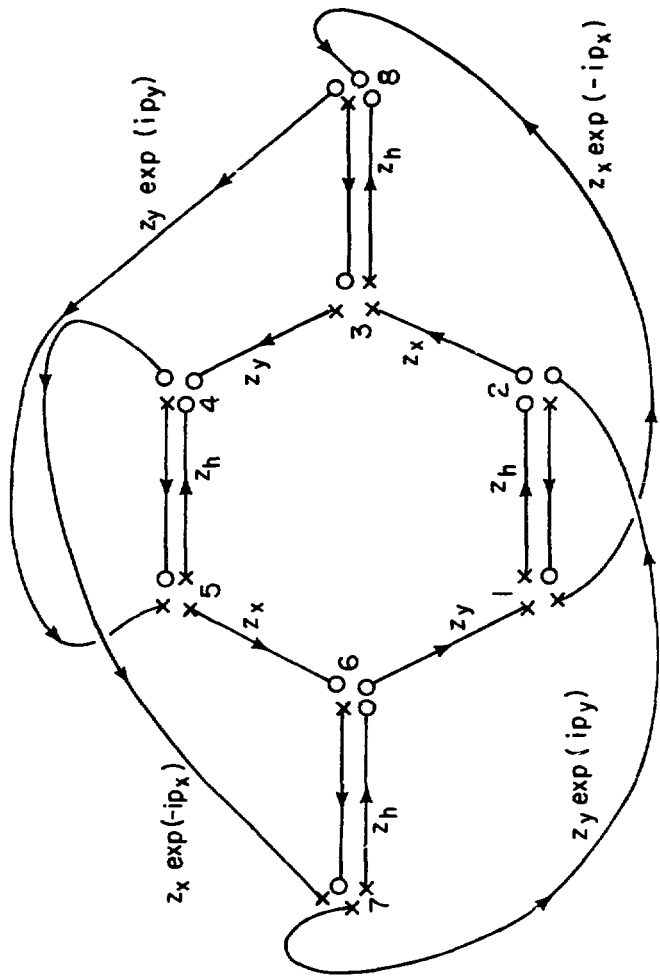


Figure 14

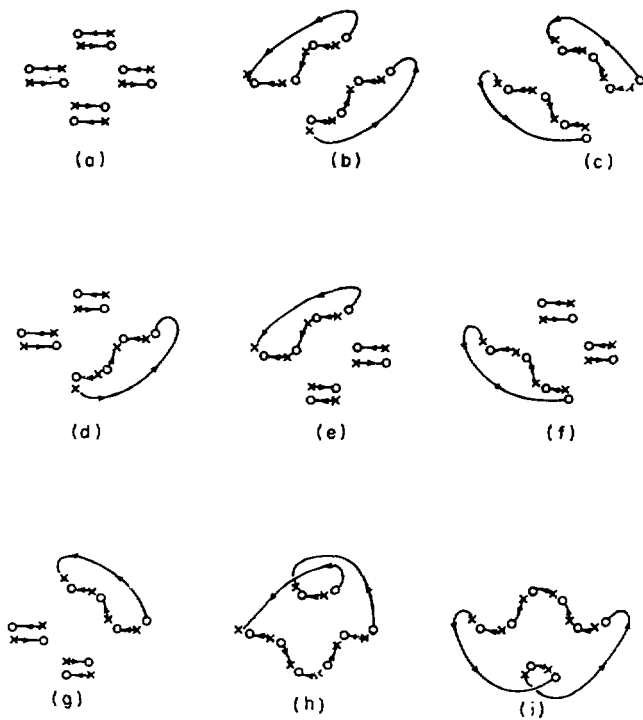


Figure 15

XBL7810-11584

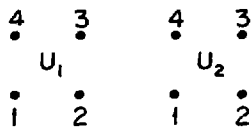


Figure 16

XBL7810-11595

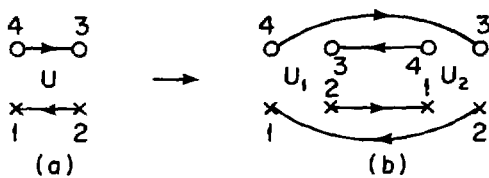


Figure 17

XBL7810-11596

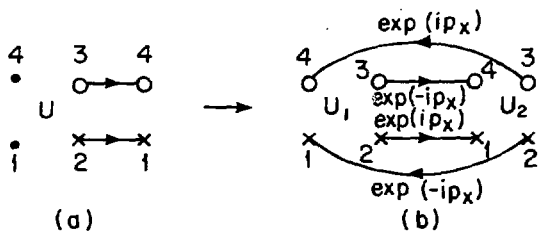


Figure 18

XBL7910-11597

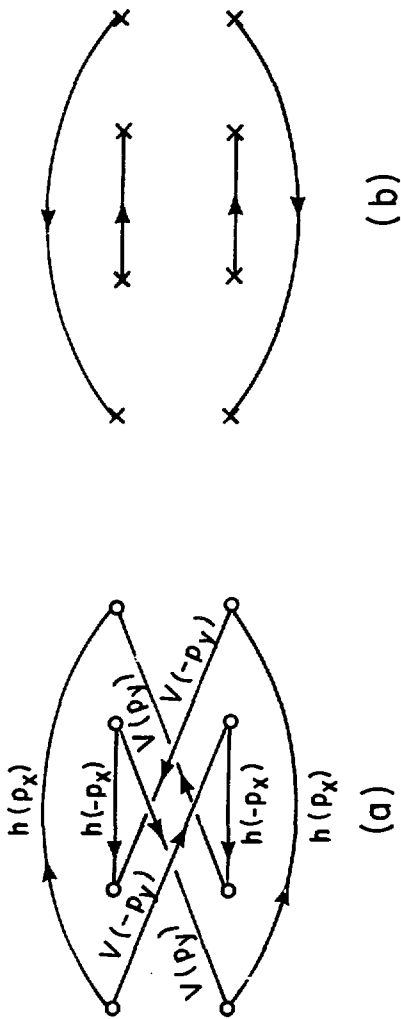


Figure 19

XBL7810-6618

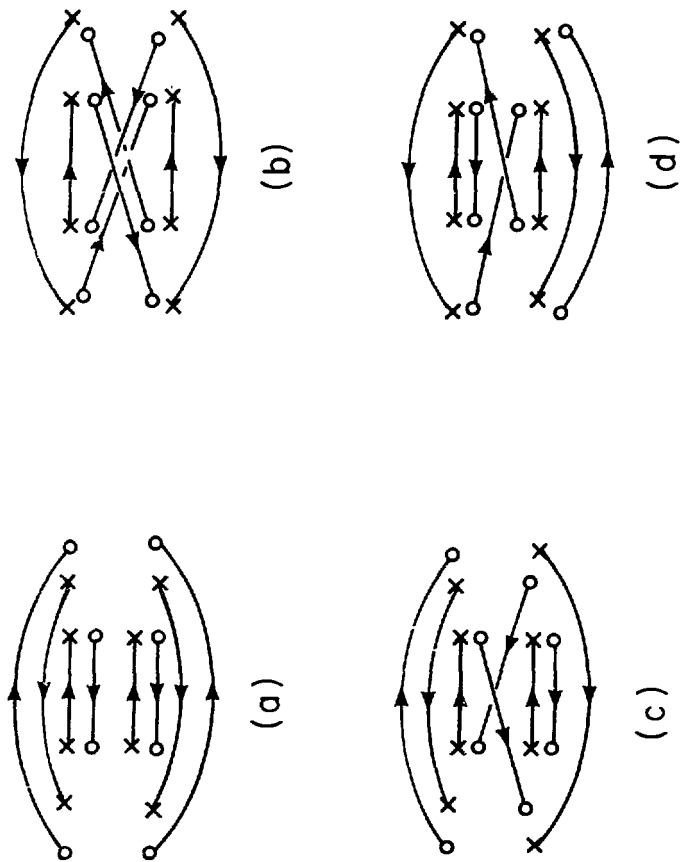


Figure 20

XBL7810-6517



Figure 21 XBL7810-6611

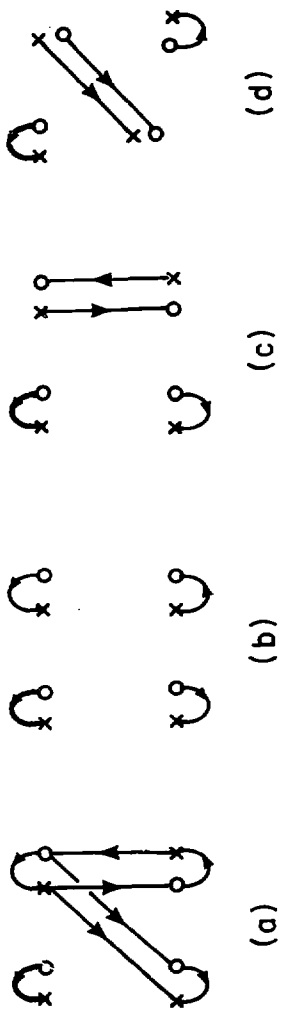


Figure 22

XBL7810 - 6610

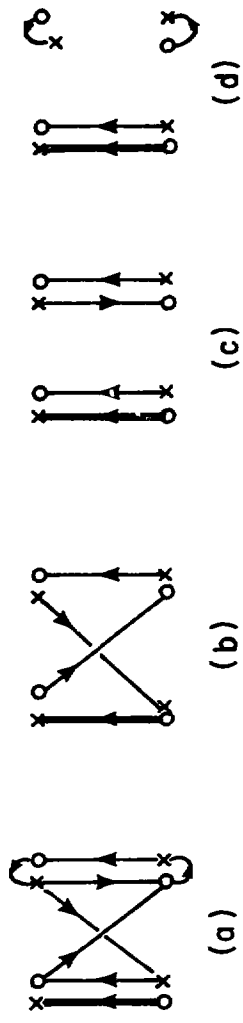
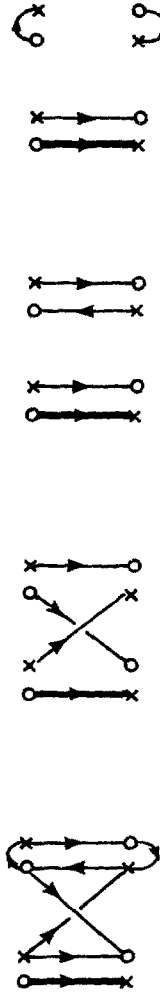
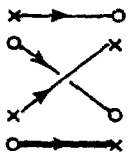


Figure 23

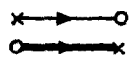
XBL 7810 - 6609



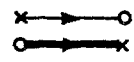
(a)



(b)



(c)



(d)

Figure 24

XBL7810-6613

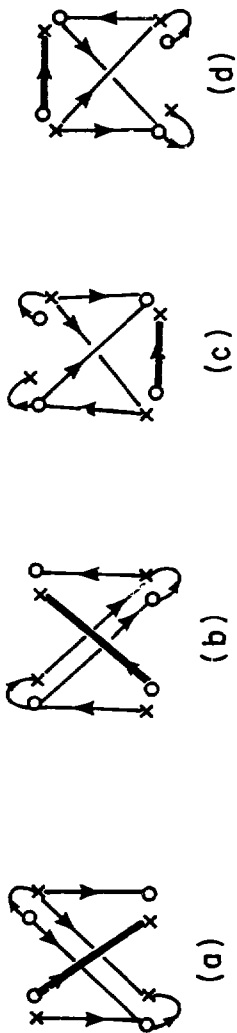


Figure 25

XBL7810-6616

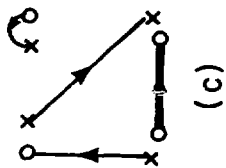
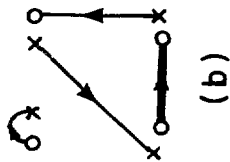
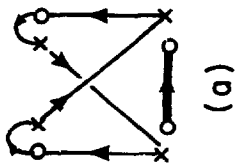
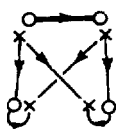
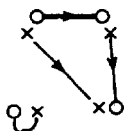


Figure 26

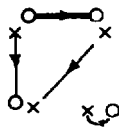
XBL7810-6621



(a)



(b)



(c)

Figure 27

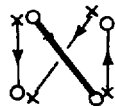
XBL7810-11585



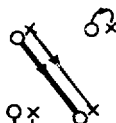
(a)



(b)



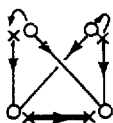
(c)



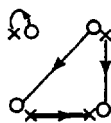
(d)

Figure 28

XBL7810-11586



(a)



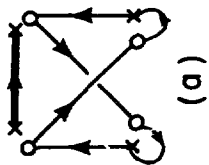
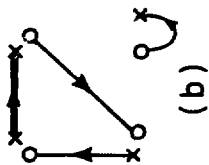
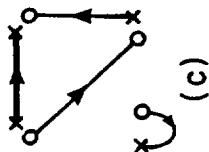
(b)



(c)

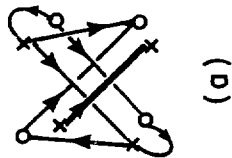
Figure 29

XBL7810-11587

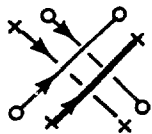


XBL7810-6614

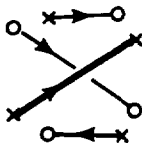
Figure 30



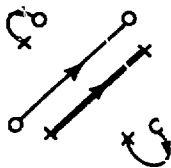
(a)



(b)



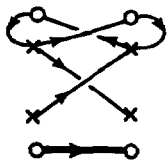
(c)



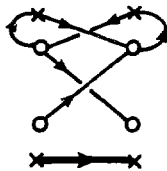
(d)

Figure 31

XBL7810 - 6615



(a)



(b)

Figure 32 XBL7810 - 6612

CHAPTER III

THE PSEUDO-FREE 128 VERTEX MODEL

INTRODUCTION

Two fundamental papers (Samuel 1978 a, b) (to be referred to as I and II) have recently developed a new approach to attacking Ising-like spin models and ferroelectric systems. This paper will use the new methods to solve a new model called the pseudo-free 128 vertex model.

An enormous number of statistical mechanics problems have graphical representations. This means that the partition function is a sum over graphical configurations appropriately weighted by Boltzmann factors. Papers I and II show that it is sometimes possible to find a lattice fermionic-like field theory which reproduces the graphical configurations with the correct weights. The field theory is written in path integral form. The path integral for fermionic systems is an anticommuting variable one. Anticommuting variables provide a powerful new approach to statistical mechanics problems. References I and II were devoted to developing their application to interesting systems. These two papers were pedagogical. They reviewed the theory of anticommuting variables and developed ways of expressing partition functions in terms of them. Graphical methods were introduced in II that quickly calculate partition functions and anticommuting variable correlation functions. A whole class of solvable models were resolved using the new methods as a check that they did indeed work.

This paper is concerned with the pseudo-free 128 vertex model. It has 32 free parameters and encompasses a wide range of systems. A close relative is the 128 + 8 pseudo-free model. It is

even more general with 40 parameters. It is also solved in this paper.

Papers I and II systematically discussed the anticommuting variable techniques. For this reason few details of the 128 vertex model calculations are given. The model and the results are simply presented. How to overcome various difficulties such as the sign problems, how to get vertex weight factors, etc. are straight forward. It is suggested that the reader consult references I and II.

Section II gives a brief description of the pseudo-free 128 vertex model, Sec. III calculates its partition function, and Sec IV treats the 128 + 8 pseudo-free vertex model. Finally Appendix A discusses the minus sign problem due to anticommuting variable reorderings.

It should be mentioned that, in principle, these models can be solved using the Pfaffian methods. As noted in reference I, the integral over a quadratic action is always a Pfaffian. The anticommuting variables have the advantage of easily determining minus sign factors, of systematically organizing algebra, and of establishing directly a connection with field theory. References to Pfaffian methods can be found in I & II.

II THE MODEL

Ising models are, in general, related to closed polygon partition functions (CPPF's) where sides may overlap but cannot intersect. In such a CPPF, one sums over closed polygons weighting the sides by "Bloch wall" Boltzmann factors. The two-dimensional Ising model thus has such a representation[†]. The Ising model is not

† See the references in I and II.

the most general model which is easily solvable. The corners of polygons may also be weighted, resulting in the so-called free-fermion model described by the action of equation (I. 4.4) whose weights are given in figure I. 11. Let $W_{(p)}$ be the weight of figure I. 11p. Then, the following constraint, known as the free-fermion constraint is satisfied: $W_{(a)}W_{(h)} + W_{(b)}W_{(c)} = W_{(d)}W_{(f)} + W_{(e)}W_{(g)}$. Thus, although the free-fermion model is not the most general eight-vertex model, it is the most general easily solvable model.

Slightly more complicated than the basic Ising model would be to include one set of diagonal next nearest neighbor interactions. Such a system is equivalent to the Ising model on a triangular lattice. It is again related to a CPPF. By weighting corners as well as sides, a free-fermion generalization, known as the pseudo-free 32 vertex model (Satto and Wu 1975) is obtained. They have solved this model and discussed some of its interesting submodels and critical phenomenon.

When both next nearest neighbor interactions are included, the Ising model cannot be solved. Spins sit on the sites of a square lattice (figure 1a). Bonds are drawn between sites which interact (figure 1b). The four directions inclined, horizontal, diagonal, and vertical, are respectively denoted by "i", "h", "d", and "v" as shown in figure 2. The polygons of the corresponding CPPF are drawn on the lattice of figure 1a using the bonds of figure 1b. The number of polygons is arbitrary. Although edges may intersect (figure 3a), they are not allowed to overlap (figure 3b). Weighting the corners of polygons results in a more general CPPF. The most general, easily solvable CPPF of this form is

the pseudo-free 128 vertex model. It is the free-fermion generalization of the next nearest neighbor Ising model.

It has 32 parameters which may be varied independently. It is thus a very general model. For example, it includes the pseudo-free 32 vertex model, which as Sacco and Wu (1975) noted, contains interesting models as subcases. Many new models are contained in the pseudo-free 128 vertex model.

As its name implies 128 configurations can happen at a site. This is to be compared to the eight vertex model where there are only eight. Of course, the solvable pseudo-free 128 vertex model does not assign arbitrary weights to all 128 configurations, only about one fourth of these are independent. The rest are determined by "free-fermion constraints". Vertex models are related to ferroelectric systems. From this point of view the pseudo-free 128 vertex model can be considered as a very general ferroelectric model.

As discussed in references I & II, the partition function can be written as an anticommuting variable integral over an action, A_{wall}^{128} . The action consists of three pieces, A_{wall}^{128} , A_{corner}^{128} , and A_{monomer}^{128} . They are given by

$$A_{\text{wall}}^{128} = \sum_{\alpha\beta} \left[z_i n_{\alpha\beta}^{\dagger i} n_{\alpha+1\beta-1}^i + z_h n_{\alpha\beta}^{\dagger h} n_{\alpha+1\beta}^h + z_d n_{\alpha\beta}^{\dagger d} n_{\alpha+1\beta+1}^d + z_v n_{\alpha\beta}^{\dagger v} n_{\alpha\beta+1}^v \right], \quad (2.1)$$

$$A_{\text{corner}}^{128} = \sum_{\alpha\beta} \sum_{(f,g) \in S} \left[c_{fg}^1 n_{\alpha f}^{\dagger} n_{\alpha\beta}^g + c_{gf}^2 n_{\alpha\beta}^{\dagger} n_{\alpha f}^g + c_{gf}^3 n_{\alpha\beta}^{\dagger} n_{\alpha f}^g + c_{gf}^4 n_{\alpha\beta}^{\dagger} n_{\alpha f}^g \right], \quad (2.2)$$

$$A_{\text{monomer}}^{128} = \sum_{\alpha\beta} \left[b_i \eta_{\alpha\beta}^i \eta_{\alpha\beta}^{i\dagger} + b_h \eta_{\alpha\beta}^h \eta_{\alpha\beta}^{h\dagger} + b_d \eta_{\alpha\beta}^d \eta_{\alpha\beta}^{d\dagger} + b_v \eta_{\alpha\beta}^v \eta_{\alpha\beta}^{v\dagger} \right]. \quad (2.3)$$

The η 's are anticommuting variables. There are four types at each site: inclined, horizontal, diagonal, and vertical ones. In addition, there is a daggered and undaggered version of each. The α and β label sites, that is, (α, β) are a site's cartesian coordinates.

The terms in (2.1) have the graphical representation of figure 4. The conventions established in references I & II are used: daggered variables and undaggered variables correspond to "x"'s and "o"'s, the direction of a line entering a variable determines whether it is an inclined, horizontal, diagonal, or vertical type, and arrows denote the order of bilinears. The constants, z_i , z_h , z_d , and z_v , are the Bloch wall Boltzmann factors. Each inclined, horizontal, diagonal, or vertical unit of wall is weighted by z_i , z_h , z_d , or z_v .

In equation (2.2), S is the following set of ordered pairs:

$$S = \{(i,h), (i,d), (i,v), (h,d), (h,v), (d,v)\}. \quad (2.4)$$

The set, S , is used so that equation (2.2) can be written concisely. The constants, c_{fg}^{ℓ} ($\ell = 1, 2, 3, 4$ and $(f, g) \in S$), allow corners to be weighted. Like the z 's, their values are at one's disposal. There are 24 of them. The terms in (2.2) correspond to those of figure 5. It is useful to define

$$\begin{aligned}
 c_{gf}^1 &\equiv c_{gf}^3, \\
 c_{fg}^2 &\equiv -c_{gf}^2, \\
 c_{fg}^3 &\equiv c_{gf}^1, \\
 c_{fg}^4 &\equiv -c_{gf}^4,
 \end{aligned}
 \tag{2.5}$$

for $(f,g) \in S$. Then,

$$A_{\text{corner}}^{128} = \sum_{\alpha\beta} \sum_{fg} \left[c_{fg}^1 \eta_{\alpha\beta}^{f\dagger} \eta_{\alpha\beta}^g + \frac{1}{2} c_{gf}^2 \eta_{\alpha\beta}^{g\dagger} \eta_{\alpha\beta}^{f\dagger} + \frac{1}{2} c_{gf}^4 \eta_{\alpha\beta}^g \eta_{\alpha\beta}^f \right], \tag{2.6}$$

where the sum is over distinct f and g among the set $\{i, h, d, v\}$.

Equation (2.3) contains the monomer terms and the remaining four free parameters, b_i , b_h , b_d , and b_v .

In a functional integral these three actions draw polygons, A_{wall}^{128} draws the walls, A_{corner}^{128} forms corners, and A_{monomer}^{128} fills unfilled sites. The integral is an anticommuting variable one over the action, A_{128} :

$$A_{128} = A_{\text{wall}}^{128} + A_{\text{corner}}^{128} + A_{\text{monomer}}^{128}. \tag{2.7}$$

The pseudo-free 128 vertex model is a fermionic-like pseudo-free field theory.

By expanding the action, the CPPF configurations are obtained. Table 1 shows the weights of each vertex configuration after Bloch wall Boltzmann factors have been extracted. It turns out that the overall sign of a vertex weight is determined by the number of line intersections as figure 6 illustrates. The total weight of any

polygonal configuration is the product of table 1 vertex weights at each site times the Bloch wall Boltzmann factors, z_f ($f = i, h, d, v$), for each unit of wall. The first page of table 1 has the configurations where six edges enter a site; pages two through four contain configurations with four lines entering; and page five has those where two edges enter. The two remaining configurations, those with zero or eight lines entering (boxes 127 and 128), are placed at the top of page two.

One must be careful of minus signs which result from reordering the anticommuting variables. Appendix A proves that the overall sign of a closed non self-intersecting polygon is plus. The overall sign for intersecting polygons is $(-1)^I$, where I is the number of intersections. For intersections which occur at a vertex the minus sign factors have been included in the weights of table 1. There are, however, intersections which do not occur at a vertex (see figure 7). An additional minus sign factor must be included for each of these types of intersections.

The vertex weights are expressed in terms of the following coefficients:

$$c_{ef;g}^1 \equiv c_{eg}^1 c_{gf}^1 - c_{ge}^2 c_{fg}^4, \quad (2.8)$$

$$c_{ef;g}^2 \equiv c_{eg}^1 c_{gf}^2 - c_{ge}^2 c_{fg}^1, \quad (2.9)$$

$$c_{ef;g}^4 \equiv c_{eg}^4 c_{gf}^1 - c_{ge}^1 c_{fg}^4, \quad (2.10)$$

$$c_{ef;g}^{-k} \equiv b_g c_{ef}^k + c_{ef;g}^k, \quad (2.11)$$

$$c_{ef;gj}^1 \equiv c_{eg}^1 c_{gf;j}^1 - c_{ge}^2 c_{fg;j}^4 + c_{ej}^1 c_{jf;g}^1 - c_{je}^2 c_{fj;g}^4, \quad (2.12)$$

$$c_{ef;gj}^2 \equiv c_{eg}^1 c_{gf;j}^2 - c_{ge}^2 c_{fg;j}^1 + c_{ej}^1 c_{jf;g}^2 - c_{je}^2 c_{fj;g}^1, \quad (2.13)$$

$$c_{ef;gj}^4 \equiv c_{eg}^4 c_{gf;j}^1 - c_{ge}^1 c_{fg;j}^4 + c_{ej}^4 c_{jf;g}^1 - c_{je}^1 c_{fj;g}^4, \quad (2.14)$$

$$\bar{c}_{ef;gj}^{-\ell} \equiv c_{ef}^{\ell} \bar{F}_{gj} + b_j c_{ef;g}^{\ell} + b_g c_{cf;j}^{\ell} + c_{ef;gj}^{\ell}, \quad (2.15)$$

$$F_{fg} \equiv -c_{fg}^1 c_{gf}^1 + c_{fg}^2 c_{gf}^4, \quad (2.16)$$

$$\bar{F}_{fg} \equiv b_f b_g + F_{fg}, \quad (2.17)$$

$$F_{efg} \equiv -c_{ef}^1 c_{fe;g}^1 + c_{ef}^2 c_{fe;g}^4 - c_{eg}^1 c_{ge;f}^1 + c_{eg}^2 c_{ge;f}^4, \quad (2.18)$$

$$\bar{F}_{efg} \equiv b_e b_f b_g + b_e F_{fg} + b_f F_{eg} + b_g F_{ef} + F_{efg}, \quad (2.19)$$

$$F_{ihdv} = \left[\begin{aligned} & -c_{id;h}^1 c_{di;v}^1 + c_{id;h}^2 c_{di;v}^4 \\ & -c_{iv;d}^1 c_{vi;h}^1 + c_{iv;d}^2 c_{vi;h}^4 \\ & -c_{ih;v}^1 c_{hi;d}^1 + c_{ih;v}^2 c_{hi;d}^4 \\ & -c_{iv;h}^1 c_{vi;d}^1 + c_{iv;h}^2 c_{vi;d}^4 \\ & -c_{ih;d}^1 c_{hi;v}^1 + c_{ih;d}^2 c_{hi;v}^4 \\ & -c_{id;v}^1 c_{di;h}^1 - c_{id;v}^2 c_{di;h}^4 \end{aligned} \right]. \quad (2.20)$$

In equations (2.8) - (2.19), each $e, f, g,$ and j stands for any of the $i, h, d,$ and v . All subscripts must be distinct. In (2.11) and (2.15) $\ell = 1, 2,$ or 4 .

The coefficients satisfy the following symmetry properties: the c^2 's, c^4 's, \bar{c}^2 's, and \bar{c}^4 's are antisymmetric in the two indices before the semicolon and symmetric in the indices after the semicolon. For example, $c_{ef;g}^2 = -c_{fe;q}^2$, $c_{ef;gj}^2 = -c_{fe:gj}^2 = c_{ef;jg}^2 = -c_{fe;gj}^2$. The F 's and \bar{F} 's are completely symmetric in their indices.

They have the following interpretation. Corners can combine to fill the anticommuting variable sites. F_{fg} (respectively, F_{efg} and F_{ihdv}) is the weight which results in filling the f and g (e, f, g and all) sites by using two (three and four) corners. F_{ihdv} excludes terms in which two pairs are filled separately, i.e. there is no term proportional to $F_{ih}F_{dv}$. \bar{F}_{fg} (respectively, \bar{F}_{efg}) is the way f, g (e, f, g) sites can be filled by using monomers and corners.

Likewise, two corners can combine to form a third. $c_{ef;g}^l$ (respectively, $c_{ef;gj}^l$) is the way two (three) corners combine to form a c_{ef}^l corner and in the process use up the g (g and j) variables. $\bar{c}_{ef;g}^l$ (respectively, $\bar{c}_{ef;gj}^l$) is the way a c_{ef}^l corner can be formed, in which g (g and j) sites get filled, by using both monomers and corners.

All the definitions of functions in table 1 have been supplied except for the weight, w_{127} , of box 127. It is

$$\begin{aligned}
 w_{127} \equiv \bar{F}_{ihdv} \equiv & \left[(b_i b_h b_v b_d) + (b_i b_h F_{dv} + b_d b_i F_{hv} \right. \\
 & + b_v b_i F_{dh} + b_d b_h F_{iv} + b_v b_h F_{id} + b_v b_d F_{ih}) \\
 & + (b_i F_{hdv} + b_h F_{idv} + b_d F_{ivh} + b_v F_{idh}) \\
 & \left. + (F_{ih} F_{dv} + F_{id} F_{hv} + F_{iv} F_{dh}) + (F_{ihdv}) \right] .
 \end{aligned} \tag{2.21}$$

Table 1 along with figure 7, essentially defines the model.

II THE SOLUTION

The partition function can be related to a miniature dimer problem using the methods developed in II. If one, then, interchanges dagger and undaggered variables for $(-s, -t)$ variables, a determinant is obtained.

Define

$$\begin{aligned}
 i(p_x, p_y) &= b_i - z_i \exp(ip_x - ip_y) , \\
 h(p_x) &= b_h - z_h \exp(ip_x) , \\
 d(p_x, p_y) &= b_d - z_d \exp(ip_x + ip_y) , \\
 v(p_y) &= b_v - z_v \exp(ip_y) .
 \end{aligned} \tag{3.1}$$

Let D be the following 8×8 diagonal matrix:

$$D(p_x, p_y) = \begin{pmatrix} i(p_x, p_y) & & & & & \\ & h(p_x) & & & & \\ & & d(p_x, p_y) & & & \\ & & & v(p_y) & & \\ & & & & i(-p_x, -p_y) & \\ & & & & & h(-p_x) \\ & & & & & & d(-p_x, -p_y) \\ & & & & & & & v(-p_y) \end{pmatrix} \quad (3.2)$$

Let C^1 , C^2 , and C^4 be the following 4×4 arrays of numbers:

$$C^1 = \begin{pmatrix} 0 & c_{ih}^1 & c_{id}^1 & c_{iv}^1 \\ c_{hi}^1 & 0 & c_{hd}^1 & c_{hv}^1 \\ c_{di}^1 & c_{dh}^1 & 0 & c_{dv}^1 \\ c_{vi}^1 & c_{vh}^1 & c_{vh}^1 & 0 \end{pmatrix} \quad (3.3)$$

$$C^2 = \begin{pmatrix} 0 & c_{ih}^2 & c_{id}^2 & c_{iv}^2 \\ c_{hi}^2 & 0 & c_{hd}^2 & c_{hv}^2 \\ c_{di}^2 & c_{dh}^2 & 0 & c_{dv}^2 \\ c_{vi}^2 & c_{vh}^2 & c_{vd}^2 & 0 \end{pmatrix} \quad (3.4)$$

$$C^4 = \begin{pmatrix} 0 & c_{ih}^4 & c_{id}^4 & c_{iv}^4 \\ c_{hi}^4 & 0 & c_{hd}^4 & c_{hv}^4 \\ c_{di}^4 & c_{dh}^4 & 0 & c_{dv}^4 \\ c_{vi}^4 & c_{vh}^4 & c_{vd}^4 & 0 \end{pmatrix} \quad (3.5)$$

Let $[C^1]^t$ denote the 4×4 matrix which is the transpose of C^1 . Define the 8×8 matrix, $M(p_x, p_y)$, by

$$M(p_x, p_y) = D(p_x, p_y) + \begin{pmatrix} -[C^1]^t & [C^4] \\ [C^2] & [C^1] \end{pmatrix}, \quad (3.6)$$

and set

$$L(p_x, p_y) = \text{Det } M(p_x, p_y), \quad (3.7)$$

where Det stands for the determinant. The partition function for the pseudo-free 128 vertex model, Z_{128} , in the thermodynamic limit,

is

$$Z_{128} = \exp \left[T \frac{1}{2} \int_{-\pi}^{\pi} \frac{dp_x}{2\pi} \int_{-\pi}^{\pi} \frac{dp_y}{2\pi} \ln L(p_x, p_y) \right], \quad (3.8)$$

where T is the total number of sites. The free energy per site, f_{128} , is

$$-\beta f_{128} = \frac{1}{2} \int_{-\pi}^{\pi} \frac{dp_x}{2\pi} \int_{-\pi}^{\pi} \frac{dp_y}{2\pi} \ln L(p_x, p_y), \quad (3.9)$$

where β is the inverse temperature.

For particular models where the z 's, c 's, and b 's take on certain values, the determinant in (3.7) can be evaluated by using computers. One can then obtain the free energy by using (3.9). Other physically interesting quantities such as the energy per site and the specific heat can be obtained by taking derivatives with respect to β .

IV THE 128 + 8 PSEUDO-FREE VERTEX MODEL

Closely related to the pseudo-free 128 vertex model is the 128 + 8 pseudo-free vertex model. Append to the lattice of figure 1 the points where inclined and diagonal bonds cross, that is, sites with half-integer cartesian coordinates. Figure 8a shows the original sites (the round ones) and the new half-integer sites (the square ones). The terms, round and square, or, integer and half-integer, will be used to distinguish the two types of sites. For round sites, bonds are drawn to the four nearest neighbor round sites and the nearest neighbor square sites, but, for square sites, bonds are drawn only to the four nearest neighbor round sites (figure 8b). What is the most general easily solvable closed polygon partition function which can be drawn on the lattice of figure 8b? The answer is the 128 + 8 pseudo-free vertex model. This CPPF is

required to have properties similar to the 128 vertex model: any number of polygons are allowed; they must be drawn on the lattice of figure 8b, sides can intersect but cannot overlap; and the corners and sides are weighted by various factors. This CPPF is generated by using an anticommuting variable integral over an action, A_{128+8} . The action again consists of three pieces: one that draws the walls, A_{wall}^{128+8} ; one that forms corner, $A_{\text{corner}}^{128+8}$; and one that fills unfilled anticommuting variable sites, $A_{\text{monomer}}^{128+8}$.

$$\begin{aligned}
 A_{\text{wall}}^{128+8} = \sum_{\alpha\beta} & \left[z_d' n_{\alpha\beta}^d n_{\alpha+1\beta+1}^d + z_d'' n_{\alpha+1\beta+1}^d n_{\alpha+1\beta+1}^d \right. \\
 & + z_i' n_{\alpha\beta+1}^i n_{\alpha+1\beta+1}^i + z_i'' n_{\alpha+1\beta+1}^i n_{\alpha+1\beta}^i + z_h n_{\alpha\beta}^h n_{\alpha+1\beta}^h + (4.1) \\
 & \left. z_v n_{\alpha\beta}^v n_{\alpha\beta+1}^v \right].
 \end{aligned}$$

The z_h and z_v wall operators are shown in figure 4, while the z_i' , z_i'' , z_d' , z_d'' wall operators are shown in figure 9. The weights of the two different kinds of diagonal bonds have been chosen independently; hence the two parameters z_d' and z_d'' . The same goes for inclined bonds.

The corner action consists of a piece, A_{corner}^{128} , identical to (2.2), and a piece that forms corners at square sites:

$$\begin{aligned}
 A_{\text{corner}}^{128+8} &= A_{\text{corner}}^{128} + A_{\text{corner}}^8, \\
 A_{\text{corner}}^8 &= \sum_{\alpha\beta} \left[c^1 n_{\alpha+\frac{1}{2}\beta+\frac{1}{2}}^i n_{\alpha+\frac{1}{2}\beta+\frac{1}{2}}^d + c^2 n_{\alpha+\frac{1}{2}\beta+\frac{1}{2}}^d n_{\alpha+\frac{1}{2}\beta+\frac{1}{2}}^i \right. \\
 &\quad \left. + c^3 n_{\alpha+\frac{1}{2}\beta+\frac{1}{2}}^d n_{\alpha+\frac{1}{2}\beta+\frac{1}{2}}^i + c^4 n_{\alpha+\frac{1}{2}\beta+\frac{1}{2}}^d n_{\alpha+\frac{1}{2}\beta+\frac{1}{2}}^i \right]. \quad (4.2)
 \end{aligned}$$

The round corner operators are shown in figure 5, while the square corner ones are shown in figure 10.

Finally, the monomer action consists of a piece, A_{monomer}^{128} , which fills round anticommuting variable sites, and a piece which fills square sites:

$$\begin{aligned}
 A_{\text{monomer}}^{128+8} &= A_{\text{monomer}}^{128} + A_{\text{monomer}}^8, \\
 A_{\text{monomer}}^8 &= \sum_{\alpha\beta} \left[m_i n_{\alpha+\frac{1}{2}\beta+\frac{1}{2}}^i n_{\alpha+\frac{1}{2}\beta+\frac{1}{2}}^i + m_d n_{\alpha+\frac{1}{2}\beta+\frac{1}{2}}^d n_{\alpha+\frac{1}{2}\beta+\frac{1}{2}}^d \right], \quad (4.3)
 \end{aligned}$$

where A_{monomer}^{128} is given in (2.3).

At round sites there are four kinds of anticommuting variables: inclined, horizontal, diagonal, and vertical, whereas at square sites there are only two kinds: inclined and diagonal.

The result is a vertex model with two kinds of vertices: square and round. The weights of the round vertices are the same as for the pseudo-free 128 vertex model and are given in table 1. The weights of the square vertices are the same as the pseudo-free eight vertex model (i.e. free-fermion model) and are given in table 2. All wall

weights have been extracted, so that the total weight is the vertex weights times the wall weights. If $m_i = m_d = 1$ and $c^1 = c^2 = c^3 = c^4 = 0$, the pseudo-free 128 vertex model is obtained along with the minus sign factor of figure 7.

In Appendix A, it is proven that non self-intersecting polygons have no overall minus signs due to reorderings of anti-commuting variables. For intersecting polygons, a (-1) results for each intersection. These minus sign factors have been absorbed into the weights of tables 1 and 2.

The 128+8 pseudo-free model has 40 parameters. The anti-commuting variable integrals over square sites can be performed since they do not couple to each other. The result is

$$\prod_{\alpha\beta} \left[f + m_i z_d^1 z_d^2 \tilde{n}_{\alpha\beta}^d \tilde{n}_{\alpha+1\beta+1}^d + m_d z_i^1 z_i^2 \tilde{n}_{\alpha\beta+1}^i \tilde{n}_{\alpha+1\beta}^i \right. \\ \left. + z_i^1 c^1 z_d^1 \tilde{n}_{\alpha\beta+1}^i \tilde{n}_{\alpha+1\beta+1}^d + z_d^1 c^2 z_i^1 \tilde{n}_{\alpha\beta}^d \tilde{n}_{\alpha\beta+1}^i + z_d^1 c^3 z_i^1 \tilde{n}_{\alpha\beta}^d \tilde{n}_{\alpha+1\beta}^i \right. \\ \left. + z_d^1 c^4 z_i^1 \tilde{n}_{\alpha+1\beta+1}^d \tilde{n}_{\alpha+1\beta}^i + z_i^1 z_i^2 z_d^1 z_d^2 \tilde{n}_{\alpha\beta}^d \tilde{n}_{\alpha+1\beta+1}^d \tilde{n}_{\alpha\beta+1}^i \tilde{n}_{\alpha+1\beta}^i \right], \quad (4.4)$$

which can be written as

$$f^T \exp \left[\sum_{\alpha\beta} \left(z_d^1 \tilde{n}_{\alpha\beta}^d \tilde{n}_{\alpha+1\beta+1}^d + z_i^1 \tilde{n}_{\alpha\beta+1}^i \tilde{n}_{\alpha+1\beta}^i \right. \right. \\ \left. \left. + k_{id}^1 \tilde{n}_{\alpha\beta}^i \tilde{n}_{\alpha+1\beta}^d + k_{di}^2 \tilde{n}_{\alpha\beta}^d \tilde{n}_{\alpha\beta+1}^i + k_{di}^3 \tilde{n}_{\alpha\beta}^d \tilde{n}_{\alpha+1\beta}^i + k_{di}^4 \tilde{n}_{\alpha\beta}^d \tilde{n}_{\alpha\beta-1}^i \right) \right], \quad (4.5)$$

where T is the total number of (square) sites and

$$\begin{aligned}
 f &\equiv m_i m_d - c^1 c^3 - c^2 c^4, \\
 z_d &\equiv z_d^1 z_d^2 m_i / f, \\
 z_i &\equiv z_i^1 z_i^2 m_d / f, \\
 k_{id}^1 &\equiv z_i^1 z_d^1 c^1 / f, \\
 k_{di}^2 &\equiv z_d^1 z_i^2 c^2 / f, \\
 k_{di}^3 &\equiv z_d^2 z_i^1 c^3 / f, \\
 k_{di}^4 &\equiv z_d^2 z_i^2 c^4 / f.
 \end{aligned}
 \tag{4.6}$$

It is useful to define

$$\begin{aligned}
 k_{di}^1 &\equiv k_{di}^3, \\
 k_{id}^2 &\equiv -k_{di}^2, \\
 k_{id}^4 &\equiv -k_{di}^4, \\
 \ln f &= -\beta f_{128+8}^0.
 \end{aligned}
 \tag{4.7}$$

The k_{fg}^{ℓ} ($\ell = 1, 2, 3$, or 4 ; $f, g = i$ or d) terms in (4.5) have the pictorial representation given in figure 11. The resulting anticommuting variable action is the same as for the pseudo-free 128 vertex model except for the four k_{fg}^{ℓ} terms, and the fact that z_d and z_i are related to square site parameters via equation (4.6).

Let $D(p_x, p_y)$, C^1 , C^2 , and C^4 be the same matrices as in equations (3.2), (3.3), (3.4), and (3.5). Define

$$K^1(p_x) = \begin{pmatrix} 0 & 0 & k_{id}^1 \exp(-ip_x) & 0 \\ 0 & 0 & 0 & 0 \\ k_{di}^1 \exp(-ip_x) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad (4.8)$$

$$K^2(p_y) = \begin{pmatrix} 0 & 0 & k_{id}^2 \exp(ip_y) & 0 \\ 0 & 0 & 0 & 0 \\ k_{di}^2 \exp(-ip_y) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad (4.9)$$

$$K^4(p_y) = \begin{pmatrix} 0 & 0 & k_{id}^4 \exp(-ip_y) & 0 \\ 0 & 0 & 0 & 0 \\ k_{di}^4 \exp(ip_y) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad (4.10)$$

Let $[K^1(p_x)]^\dagger$ denote the hermitian conjugate of $K^1(p_x)$, i.e., $[K^1(p_x)]^\dagger = [K^1(-p_x)]^t$. Let

$$M(p_x, p_y) = D(p_x, p_y) + \begin{pmatrix} - [C^{1t} + K^1(p_x)]^\dagger & [C^4 + K^4(p_y)] \\ [C^2 + K^2(p_y)] & [C^1 + K^1(p_x)] \end{pmatrix} \quad (4.11)$$

$$L(p_x, p_y) = \text{Det } M(p_x, p_y) \quad (4.12)$$

Then, the free energy per unit site, f_{128+8} (that is, per round and square site pair) is

$$-\beta f_{128+8} = -\beta f_{128+8}^0 + \frac{1}{2} \int_{-\pi}^{\pi} \frac{dp_x}{2\pi} \int_{-\pi}^{\pi} \frac{dp_y}{2\pi} \ln L_{128+8}(p_x, p_y) \quad (4.13)$$

where f_{128+8}^0 is given in (4.7)

V CONCLUSION

Two new statistical mechanics models have been solved. They are solvable via the Pfaffian method although this paper solves them using the anticommuting variables.

The next step is to determine the physics of these models, in particular, the critical phenomenon. Because of the 8×8 determinants in equations (3.7) and (4.12), this will be quite tedious. The use of computers to evaluate these determinants will probably be necessary. One can say, however, that there will be multiple phase transitions with Ising-like logarithmically divergent specific heat. This is because one submodel, the pseudo-free 32 vertex model, is known to have such multiple phase transitions (Sacco and Wu 1975)

APPENDIX A. Overall Minus Signs: The Non-Self Intersecting Polygons

This Appendix will prove that there are no overall minuses created by reorderings of anticommuting variables for a non-self intersecting polygon drawn on the $128+8$ lattice of figure 8b. This also proves the result for the pseudo-free 32 vertex and pseudo-free 128 vertex models since any polygon drawn on their lattice can be drawn on the $128+8$ lattice and the same kinds of bilinear operators are used.

The proof is similar to that for the free-fermion model, which was given in Appendix B of I and to which the reader is referred. Extensive use will be made of the sign rules (a), (b), and (c) of

figure 8 of reference 1. The proof proceeds via induction on the area of a polygon. Any polygon can be built from the four elementary triangles of figure 12 (see figure 13). These are the polygons of minimum area. Figure 14 starts the induction process by proving that these have an overall plus sign.

As required by the sign rules, the polygon is given an orientation. Choose the starting point to be on "x". Move around the polygon and count the number of minus signs due to rules (a) and (b). When moving in the positive directions of figure 2, no minus signs occur because "x"'s are after "o"'s and arrows point in the correct directions. When moving in the negative directions, there is a minus sign factor because "x"'s occur before "o"'s, but, in addition, there is a minus sign factor because arrows point in the wrong direction. Moving in straight lines causes no minuses. Next consider corners. There are 56 different corners; the 28 types of figures 5 and 10 are multiplied by two orientations. Figure 15 summarizes the results. The corners of figure 15 create a minus sign and all others do not. The easy way to find the overall minus sign is to count the number of figure 15 corners in an oriented polygon. If the number is odd, then the extra minus due to rule (c) makes the overall sign positive.

The elementary triangles can be attached to polygons in 24 different ways: each of the four elementary triangles can attach one side or two sides in three ways. All twenty-four are illustrated in figure 16. Each of these results in several cases depending on the neighboring structure where the triangle is joined. In total, there are 480 different cases to consider. These are all shown in figure 16. It is found that the addition of an elementary

polygon creates zero or two minus factors or removes two minus factors. This implies that the overall minus sign factor due to corners is the same as for the elementary triangles, namely minus. The number of corner minuses is odd. When combined with the rule (c) minus, the claim is proved: a non self intersecting polygon has no minus signs due to reorderings of anticommuting variables.

REFERENCES

Samuel S 1978a The Use of Anticommuting Integrals in Statistical Mechanics I LBL preprint 8217.

Samuel S 1978b The Use of Anticommuting Integrals in Statistical Mechanics II LBL preprint 8300.

Sacco J E and Wu F Y 1975 J. Phys. A8 1780.

Figure 1. The Square Lattice.

Figure 2. The Four Directions.

Figure 3. Allowed and Forbidden Configurations. The sides of polygons may intersect as in figure (a) but cannot overlap as in figure (b).

Figure 4. The Wall Operators.

Figure 5. The Twenty-Four Corner Operators.

Figure 6. Overall Minus Signs. The configurations in boxes 1,4, 30, and 128 of table 1 are reproduced here. They have been redrawn so that the intersections can be seen. If the number of intersections is even the overall sign is positive, while an odd number of intersections yields a negative sign. Boxes 1, 4, 30, and 128 have respectively one, three, zero, and six intersections; hence boxes 1 and 4 have an overall

minus sign, while boxes 30 and 128 do not.

Figure 7. Extra Minus Sign. An extra minus sign factor results when any two sides intersect between lattice sites. This figure is an example in which this happens. The weight of this polygon is the product of Bloch wall Boltzmann factors, the product of table 1 vertex factors, times an extra minus one: $\left[z_i z_d z_h z_h \right] \times \left[(\text{box } 105) (\text{box } 112) (\text{box } 117) (\text{box } 126) \right] \times [-1]$.

Figure 8. (a) The 128 + 8 Vertex Model Lattice. (b) The Bonds in the 128 + 8 Vertex Model. The sites in figure 1a are the round ones here. In addition, sites have been added at the points with half integer cartesian coordinates (the square sites).

Figure 9. The Diagonal and Inclined Wall Operators. A square site has bonds connecting to the four nearest neighbor round sites. This figure shows the four wall operators which produce these bonds. Each of the four have been assigned a separate weight.

Figure 10. The Four Corner Operators at a Square Site.

Figure 11. The k_{fg}^{ℓ} Operators. After square site integrals have been performed, the 128 + 8 vertex model becomes the 128 vertex model with the addition of these four terms.

Figure 12. The Four Elementary Triangles of the Lattice of figure 8b.

Figure 13. Building Up a Polygon From Elementary Triangles. The polygon of figure (b) is obtained from the polygon of figure (a) by attaching the elementary triangle of figure 12b.

Figure 14. The Overall Sign of the Elementary Triangles. The sign is determined by the sign rules of figure 8 of reference I. Begin at the x near the point, Λ , and proceed counterclockwise around the triangle. The minuses due to rules (a) and (b) are shown here. In each figure there are an odd number of them. In addition there is a minus due to rule (c). Thus the overall sign of each of the four elementary triangles is plus.

Figure 15. The Oriented Corners Which Create a Minus Sign. Figures (a) through (l) (respectively, figures (m) and (n)) show the round (square) vertex corners which create a minus sign because of anticommuting variable reordering.

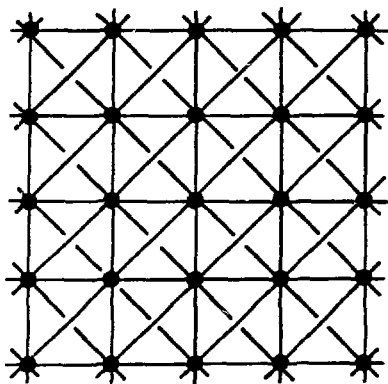
Figure 16. The 480 Cases. Here are the 480 cases which must be considered in the induction step. Each of the 12 boxes shows two of the 24 ways of appending an elementary triangle. In the left half of a box one side is joined, while in the right half two sides are joined. The joining triangle is the one formed by the solid and dotted edges. Only the neighboring structure of the polygon, to which the elementary triangle is being attached, is shown. When this triangle is attached to a configuration on the left, a configuration on the right results (see figure 17a, which is an example for box 1), and when this triangle is attached to a configuration on the right a configuration on the left results (see figure 17b, which is an example for box 1 and figure 13 which is an example for box 7).

An arrow on a line indicates that when the orientation is in that direction then one of the figure 15 corners is involved and a minus factor is present. Box 1 shows that the corner minus sign structure is unchanged in the joining process. Sometimes the process creates (or removes) a figure 15 corner, however another one is always created or removed at one of the two other vertices (see figure 17c, which is a subcase of box 6). By inspecting these boxes, corner minus sign factors are seen to be created or removed in pairs so that the overall minus sign factor is unchanged.

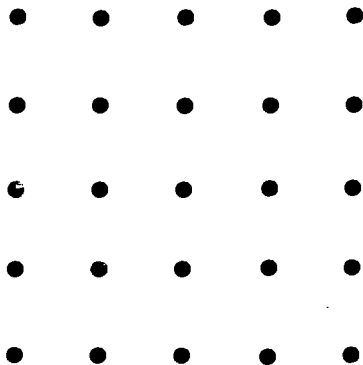
Figure 17. Examples of the Figure 16 Induction Step. Figure (a) is an example of going from a box 1 left configuration to a box 1 right configuration. Figure (b) shows a box 1 right configuration going to a box 1 left configuration. The arrows denote the location of a figure 15 corner when traversing the polygons in a counter-clockwise direction. In figures (a) and (b) no new figure 15 corners are created. Figure (c) is an example of a box 6 transformation where two extra figure 15 corners are created, when the polygon is oriented in the clockwise direction.

Table 1. The Weights of the Vertex Configurations of The Pseudo-Free 128 Vertex Model. The Bloch wall Boltzmann factors have been extracted. The weights are expressed directly in terms of the parameters of the action [equations (2.1)-(2.3)] or via the functions in equations (2.8)-(2.21).

**Table 2. The Weights of the Square Vertices in the $128 + 8$
Pseudo-Free Vertex Model.**



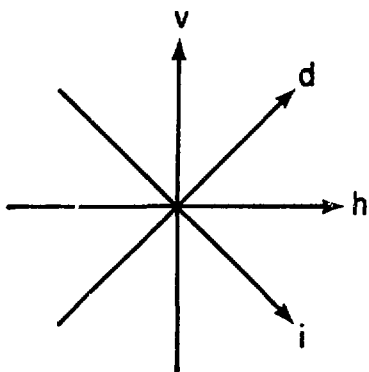
(b)



(a)

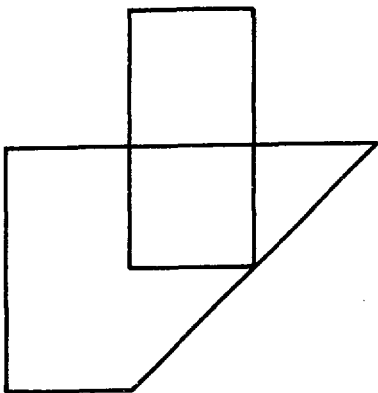
XBL 792-465

Figure 1

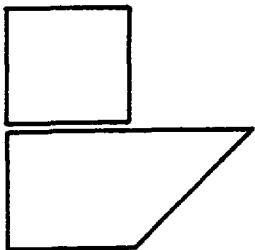


XBL 792-466

Figure 2



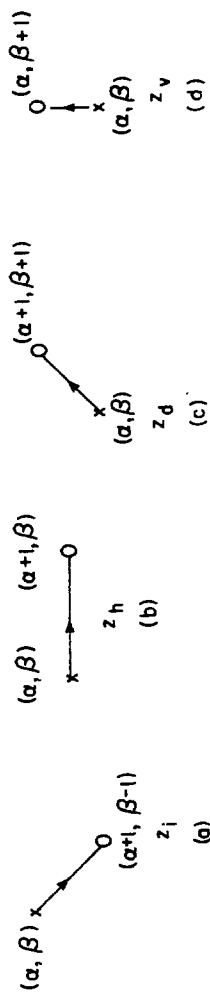
(a)



(b)

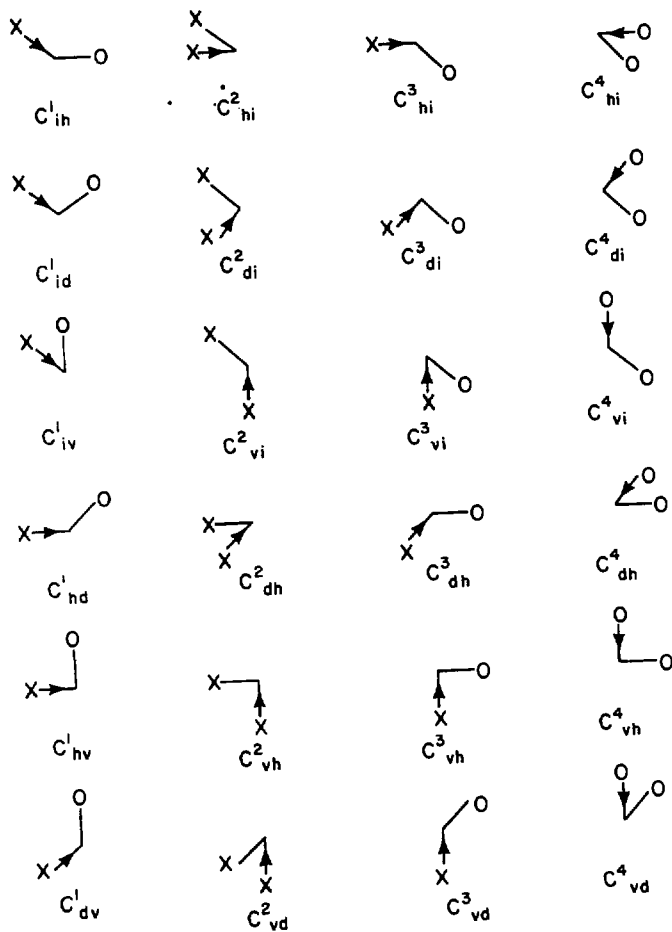
XBL 792 - 467

Figure 3



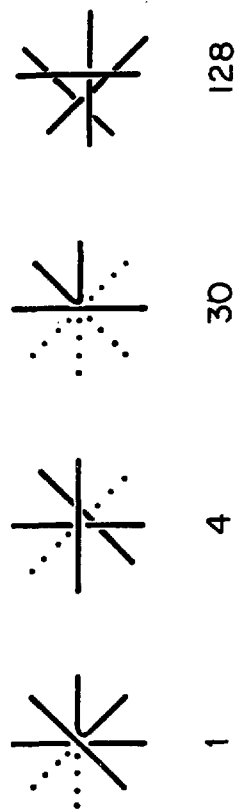
XBL 792-468

Figure 4



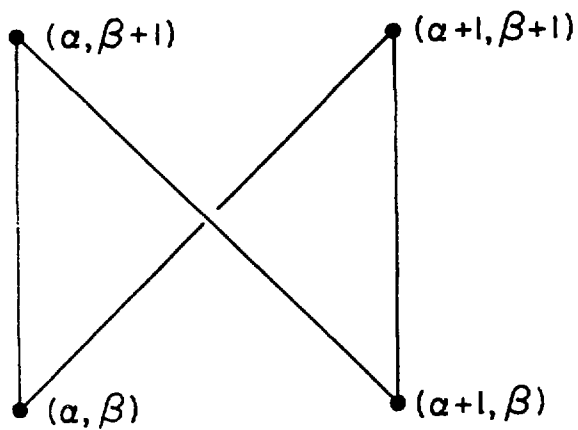
XBL 792-457

Figure 5



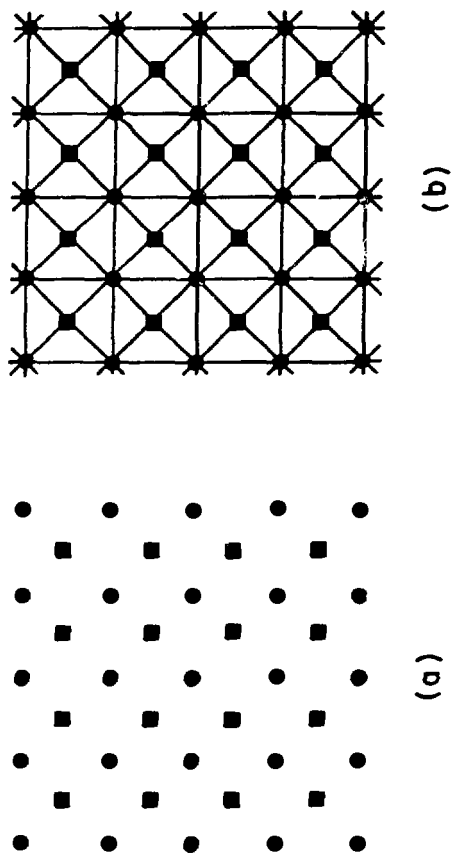
XBL 792 - 469

Figure 6



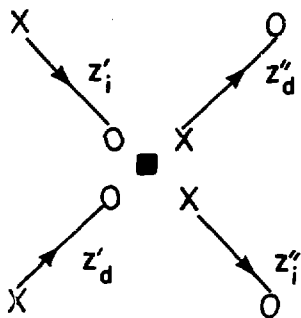
XBL 792-470

Figure 7



XBL 792-471

Figure 8



XBL 792-472

Figure 9



c^1

(a)



c^2

(b)



c^3

(c)



c^4

(d)

Figure 10

XBL 792 -473

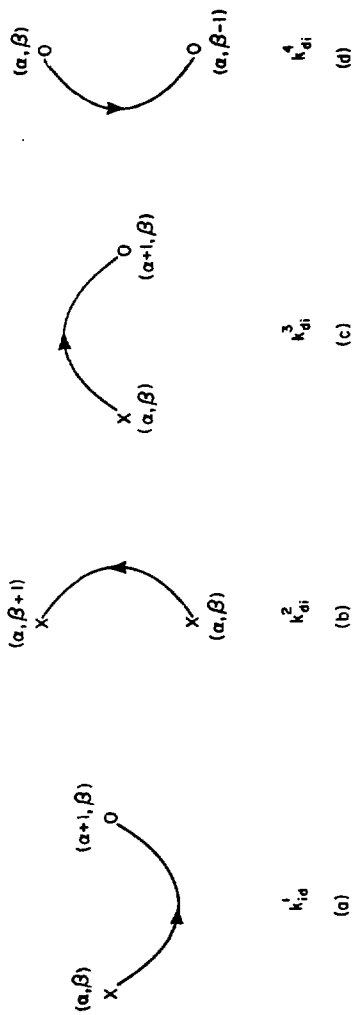
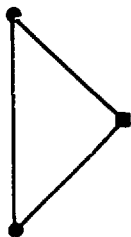


Figure 11

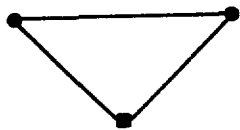
XBL 792-474



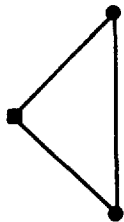
(d)



(c)



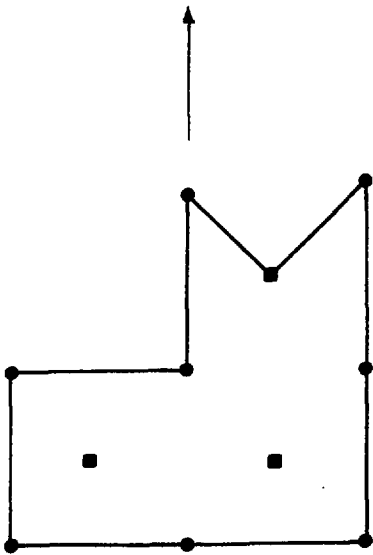
(b)



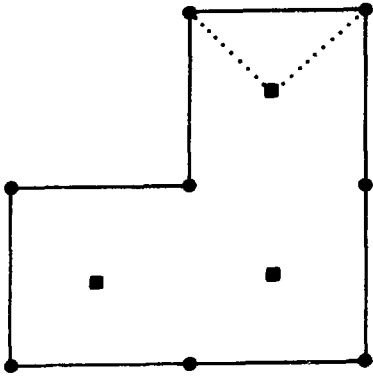
(a)

Figure 12

XBL 792-475



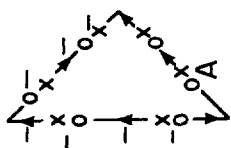
(a)



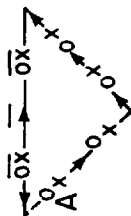
(b)

XBL 792-476

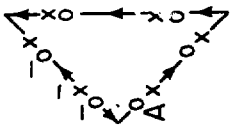
Figure 13



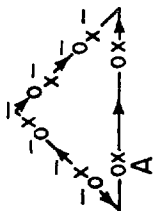
(d)



(c)



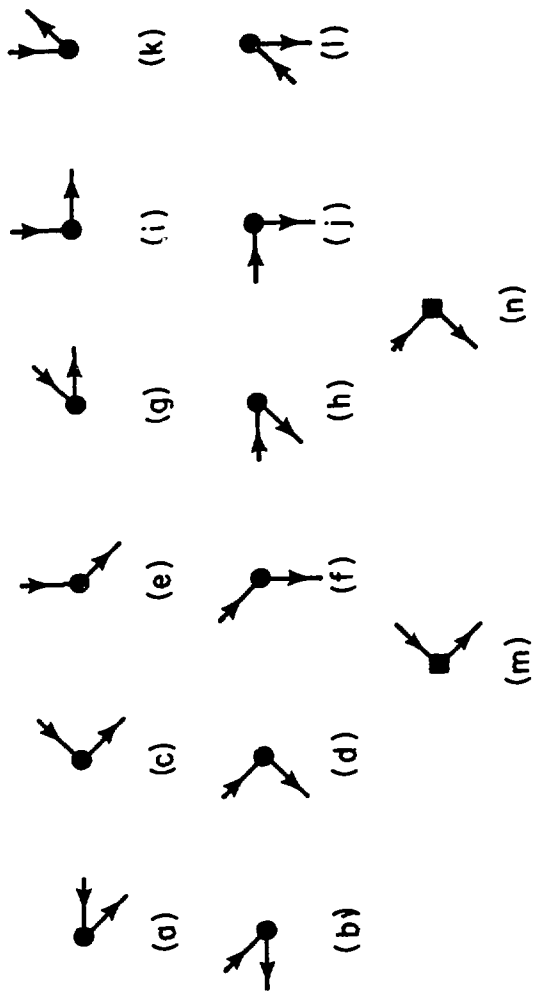
(b)



(a)

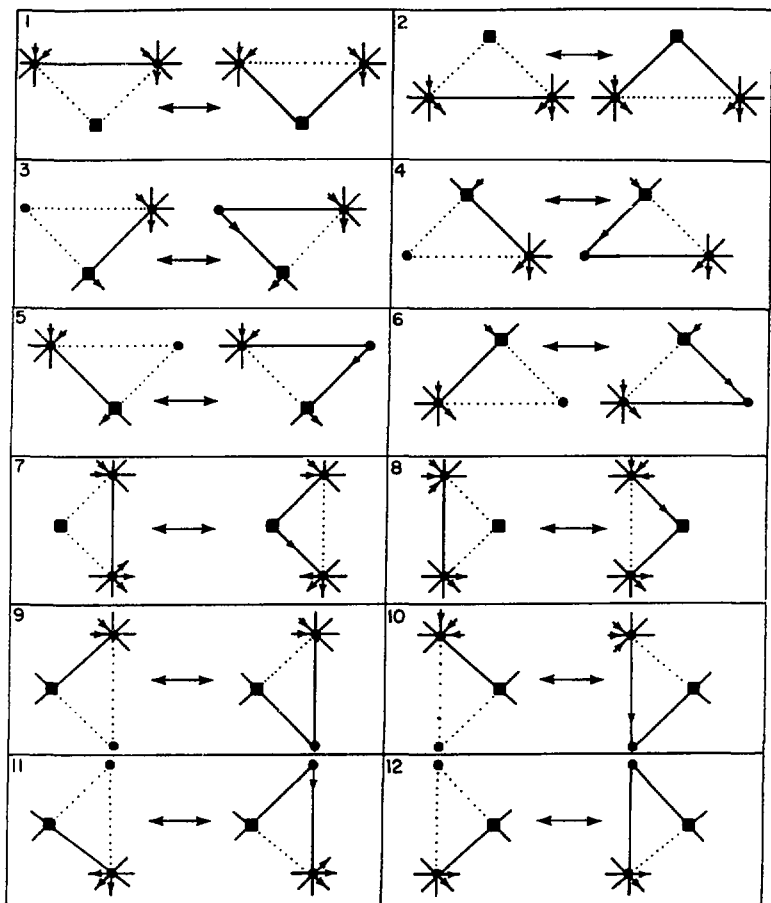
XBL 792 - 478

Figure 14



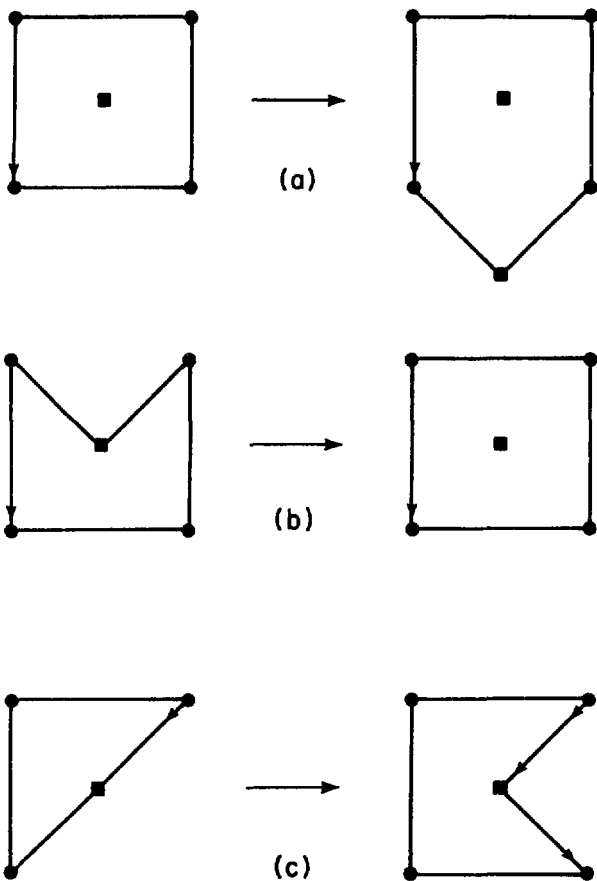
XBL 792-464

Figure 15



XBL 792-458

Figure 16



XBL 792-477

Figure 17

| | | | |
|----------------------|-----------------------|-----------------------|-----------------------|
| 1 $-C_{hi}^4$ | 8 $-C_{dh}^4$ | 15 C_{id}^1 | 22 $-b_v$ |
| 2 C_{di}^4 | 9 C_{vh}^4 | 16 $-C_{hd}^1$ | 23 $-C_{hi}^2$ |
| 3 $-C_{vi}^4$ | 10 $-C_{ih}^1$ | 17 $-b_d$ | 24 C_{di}^2 |
| 4 $-b_i$ | 11 $-b_h$ | 18 $-C_{vd}^3$ | 25 $-C_{vi}^2$ |
| 5 $-C_{hi}^3$ | 12 $-C_{dh}^3$ | 19 $-C_{iv}^1$ | 26 $-C_{dh}^2$ |
| 6 C_{di}^3 | 13 C_{vh}^3 | 20 C_{hv}^1 | 27 C_{vh}^2 |
| 7 $-C_{vi}^3$ | 14 C_{vd}^4 | 21 $-C_{dv}^1$ | 28 $-C_{vd}^2$ |

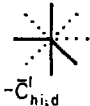
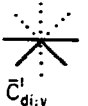
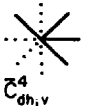


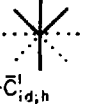



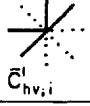
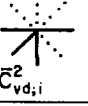
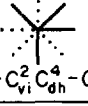
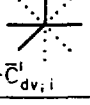
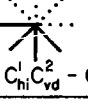
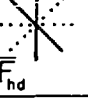
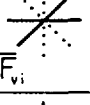
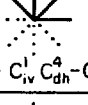
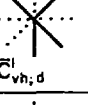
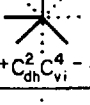
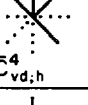
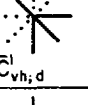
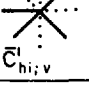
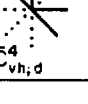
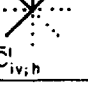
XBL 792-460

Table 1

| | | |
|---|---|---|
| 127 W_{127} | 35 $-\bar{C}_{vi;h}^4$ | 43 $C_{dh}^1 C_{vi}^1 + C_{vd}^2 C_{hi}^4 - C_{di}^1 C_{vh}^1$ |
| 128 i | 36 $C_{dv}^1 C_{hi}^4 + C_{di}^1 C_{vh}^4 - C_{dh}^1 C_{vi}^4$ | 44 $\bar{C}_{vd;i}^4$ |
| 29 $C_{vd}^4 C_{hi}^4 + C_{vi}^4 C_{dh}^4 - C_{vh}^4 C_{di}^4$ | 37 $\bar{C}_{hi;v}^4$ | 45 $C_{hv}^1 C_{di}^4 + C_{hi}^1 C_{vd}^4 - C_{hd}^1 C_{vi}^4$ |
| 30 $\bar{C}_{dh;i}^4$ | 38 $-\bar{F}_{ih}$ | 46 $-\bar{C}_{vi;d}^4$ |
| 31 $\bar{C}_{di;h}^4$ | 39 $-\bar{C}_{dh;i}^1$ | 47 $-\bar{C}_{di;v}^4$ |
| 32 $\bar{C}_{hi;d}^4$ | 40 $\bar{C}_{vh;i}^1$ | 48 $-\bar{C}_{hd;i}^1$ |
| 33 $C_{dh}^4 C_{vi}^1 + C_{vd}^1 C_{hi}^4 - C_{vh}^1 C_{di}^4$ | 41 $-\bar{C}_{di;h}^1$ | 49 $-\bar{F}_{di}$ |
| 34 $-\bar{C}_{vh;i}^4$ | 42 $\bar{C}_{vi;h}^1$ | 50 $-\bar{C}_{vd;i}^1$ |

























XBL 792-461

Table 1 (Continued)

| | | | | | |
|----|---|----|--|----|--|
| 51 |  $-\bar{C}_{hi;d}^1$ | 59 |  $\bar{C}_{di;v}^1$ | 67 |  $\bar{C}_{dh;v}^4$ |
| 52 |  $C_{vh}^2 C_{di}^4 + C_{hd}^1 C_{vi}^1 - C_{hi}^1 C_{vd}^1$ | 60 |  $\bar{C}_{dh;i}^2$ | 68 |  $-\bar{C}_{id;h}^1$ |
| 53 |  $\bar{C}_{vi;d}^1$ | 61 |  $-\bar{C}_{vh;i}^2$ | 69 |  $-\bar{C}_{ih;d}^1$ |
| 54 |  $\bar{C}_{hv;i}^1$ | 62 |  $\bar{C}_{vd;i}^2$ | 70 |  $C_{id}^1 C_{vh}^1 + C_{vi}^2 C_{dh}^4 - C_{ih}^1 C_{vd}^1$ |
| 55 |  $-\bar{C}_{dv;i}^1$ | 63 |  $C_{dh}^2 C_{vi}^1 + C_{hi}^1 C_{vd}^2 - C_{di}^1 C_{vh}^2$ | 71 |  $-\bar{F}_{hd}$ |
| 56 |  $-\bar{F}_{vi}$ | 64 |  $C_{ih}^1 C_{vd}^4 + C_{iv}^1 C_{dh}^4 - C_{id}^1 C_{vh}^4$ | 72 |  $-\bar{C}_{vh;d}^1$ |
| 57 |  $C_{hv}^1 C_{di}^1 + C_{dh}^2 C_{vi}^4 - C_{hi}^1 C_{dv}^1$ | 65 |  $\bar{C}_{vd;h}^4$ | 73 |  $-\bar{C}_{vh;d}^1$ |
| 58 |  $-\bar{C}_{hi;v}^1$ | 66 |  $\bar{C}_{vh;d}^4$ | 74 |  $\bar{C}_{iv;h}^1$ |

XBL 792-463

Table 1 (Continued)

| | | | | | |
|---|--|---|--|---|--|
| 75 |  | 83 |  | 91 |  |
| $C_{dh}^1 C_{iv}^1 + C_{di}^2 C_{vh}^4 - C_{dv}^1 C_{ih}^1$ | | $\bar{C}_{vd;h}^2$ | | $C_{hi}^2 C_{vd}^1 + C_{vh}^2 C_{id}^1 - C_{hd}^1 C_{vi}^2$ | |
| 76 |  | 84 |  | 92 |  |
| $-\bar{C}_{ih;v}^1$ | | $C_{hd}^1 C_{iv}^1 + C_{hi}^2 C_{vd}^4 - C_{hv}^1 C_{id}^1$ | | $-\bar{C}_{vi;d}^2$ | |
| 77 |  | 85 |  | 93 |  |
| $-\bar{C}_{dv;h}^1$ | | $\bar{C}_{iv;d}^1$ | | $\bar{C}_{vh;d}^2$ | |
| 78 |  | 86 |  | 94 |  |
| $-\bar{F}_{hv}$ | | $\bar{C}_{id;v}^1$ | | $C_{dh}^2 C_{iv}^1 + C_{hi}^2 C_{dv}^1 - C_{hv}^1 C_{di}^2$ | |
| 79 |  | 87 |  | 95 |  |
| $-\bar{C}_{dh;v}^1$ | | $-\bar{C}_{hv;d}^1$ | | $\bar{C}_{hi;v}^2$ | |
| 80 |  | 88 |  | 96 |  |
| $\bar{C}_{di;h}^2$ | | $-\bar{C}_{nd;v}^1$ | | $-\bar{C}_{di;v}^2$ | |
| 81 |  | 89 |  | 97 |  |
| $-\bar{C}_{vi;h}^2$ | | $-\bar{F}_{vd}$ | | $\bar{C}_{dh;v}^2$ | |
| 82 |  | 90 |  | 98 |  |
| $C_{di}^2 C_{vh}^1 + C_{ih}^1 C_{vd}^2 - C_{dh}^1 C_{vi}^2$ | | $\bar{C}_{hi;d}^2$ | | $C_{hi}^2 C_{vd}^2 + C_{vi}^2 C_{dh}^2 - C_{di}^2 C_{vh}^2$ | |


XBL 792-462

Table 1 (Continued)

| | | | |
|--------------------------------|--------------------------------|--------------------------------|--------------------------------|
| 99 $\bar{C}_{hi;vd}^2$ | 106 $\bar{C}_{dh;vi}^2$ | 113 $\bar{C}_{di;hv}^1$ | 120 \bar{F}_{ihd} |
| 100 $\bar{C}_{di;hv}^2$ | 107 $\bar{C}_{vh;di}^2$ | 114 $\bar{C}_{dh;iv}^1$ | 121 $\bar{C}_{hi;dv}^4$ |
| 101 $\bar{C}_{vi;hd}^2$ | 108 $\bar{C}_{hi;vd}^1$ | 115 \bar{F}_{vhi} | 122 $\bar{C}_{di;vh}^4$ |
| 102 \bar{F}_{vdh} | 109 \bar{F}_{ivd} | 116 $\bar{C}_{dv;ih}^1$ | 123 $\bar{C}_{vi;dh}^4$ |
| 103 $\bar{C}_{ih;dv}^1$ | 110 $\bar{C}_{hc;iv}^1$ | 117 $\bar{C}_{vi;hd}^1$ | 124 $\bar{C}_{dh;vi}^4$ |
| 104 $\bar{C}_{id;hv}^1$ | 111 $\bar{C}_{hv;id}^1$ | 118 $\bar{C}_{vh;id}^1$ | 125 $\bar{C}_{vh;di}^4$ |
| 105 $\bar{C}_{iv;dh}^1$ | 112 $\bar{C}_{vd;hi}^2$ | 119 $\bar{C}_{vd;ih}^1$ | 126 $\bar{C}_{vd;hi}^4$ |

XBL 792-459

Table 1 (Continued)

(a)  $(m_i m_d - c^1 c^3 - c^2 c^4)$

(b)  m_i

(c)  m_d

(d)  c^1

(e)  c^2

(f)  c^3

(g)  c^4

(h)  -1

XBL 792-456

Table 2

PART II

APPLICATIONS OF STATISTICAL MECHANICS
TO FIELD THEORY

CHAPTER IV

THE GRAND PARTITION FUNCTION IN FIELD THEORY
WITH APPLICATIONS TO THE SINE-GORDON THEORY

I. INTRODUCTION

This paper will employ a technique, known as gaussian integration,¹ by which certain field theories are identified with a gas of interacting particles. Originally the method was used to re-write a partition function in a field-theoretic way. Field theory techniques were then used to obtain results for the statistical mechanical system. The method can also be applied to a grand partition function.² This makes it possible to go from a thermodynamic system where particle number is not fixed to a field theory representation, which simplifies the analysis of the thermodynamic system. The idea of this paper is to reverse the process: In some cases the analogue field theory is a relativistic one. An example is the sine-Gordon which is equivalent to a neutral Coulomb gas.³ The vacuum expectation value of $e^{i\phi_1} = e^{i\int_0^{2\pi} f \cos \beta \phi dx}$, which is a sum of vacuum bubble diagrams, is equal to the grand partition function for such a Coulomb system, λ_0 playing the role of the absolute activity and β playing the role of the inverse temperature. In these cases one can analyze the field theory by using the underlying statistical mechanical analogue. Knowledge of the sine-Gordon will yield information about the Coulomb plasma. Likewise, one may use the Coulomb plasma to gain information about the sine-Gordon. This is the plan of this paper. It enables one to use one's intuition of the Coulomb plasma to obtain field theoretic results.

Some of the results of this paper have appeared in the mathematics literature.^{3,4,5} The author feels these are worth repeating since such

mathematical presentations are not accessible to most physicists. This paper stresses simple, physical, and intuitive methods of derivation.

The paper is organized as follows: Section II reviews the gaussian representation method. The system can have several different types of particles with different activities. Various interparticle potentials may also be used. The physical significance of the smearing of fields is discussed. Section III checks perturbatively the results of Section II. This check gives one insight into the statistical mechanical—field theoretic analogy. In particular, the Feynman diagrams have a simple physical description in terms of the underlying thermodynamic system. This correspondence is outlined in Table 1. From Section IV onward the main concern of the paper is the two dimensional sine-Gordon. Section IV introduces the sine-Gordon field theory and discusses its infrared singular nature which in the Coulomb analogue model forces strict neutrality. Section V determines the phases of the sine-Gordon. At low temperatures there is a dipole gas, whereas at high temperatures there is a plasma phase. The impact on the existence of solitons is discussed. Section VI shows how non-linear σ -model is equivalent to the sine-Gordon. In Section VII the renormalization is performed to all orders in λ_R and βq^2 , when βq^2 is small. This shows that the theory is well defined. Other aspects of renormalization are also dealt with. In Section VIII remaining ideas are discussed, most of which depend heavily on the Coulomb gas analogy. Most important is the vacuum structure and its effect on the theory. Charge screening and shielding, fractional

charges, the effects of infrared divergences on Feynman rules, and the sine-Gordon solitons are discussed. Section IX is the summary. There, the main results are simply enumerated. The paper concludes with a comment on vacuum gases as models for hadrons.

II. GAUSSIAN REPRESENTATION

The partition function for a system of interacting particles may be represented as a field theory. The technique, known as gaussian representation, is well-known in statistical mechanics.^{1,2} In certain cases, the resulting field theory is a relativistic field theory. These are of particular interest since it allows one to think in terms of the underlying statistical mechanical system. This association provides physical insight into the field theory which one usually does not have, thus allowing for the extraction of the interesting physical effects.

This section will review the gaussian representation method along with comments concerning the use of different potentials, the smearing of fields, and various other technicalities which occur in passing from the grand partition function to the field theory. For simplicity the gaussian representation method will be first applied to a specific example: the Coulomb plasma in three dimensions. This was actually done by Polyakov⁶ in analyzing a three dimensional instanton confinement mechanism. His instantons were monopoles interacting via a potential which was Coulomb-like for large distances and mitigated for short distances. Because of the softened short distance behavior the Polyakov model has a natural renormalization prescription. This will be obvious later on. For the true Coulomb gas there is no natural renormalization and the grand partition function will be ultraviolet singular. For the present, ignore the bad short distant behavior and any infinities which result from the use of the bare $\frac{1}{r}$ potential.

Consider the grand partition function, \mathcal{Z} , for a plasma containing an arbitrary number of positive charges (ions) and negative charges (electrons) interacting via a Coulomb potential at a temperature, $\frac{1}{\beta}$, and having an absolute activity, λ_0 :

$$\mathcal{Z} = \sum_{p=0}^{\infty} \frac{\lambda_0^p}{p!} \sum_{i=0}^{\infty} \frac{\lambda_0^i}{i!} \int_V d^3R_1 \cdots d^3R_p \int_V d^3X_1 \cdots d^3X_i$$

$$\times \exp \left[-\frac{\beta q^2}{2} \left(\sum_{\ell \neq m}^p \frac{1}{|\vec{R}_\ell - \vec{R}_m|} + \sum_{\ell \neq m}^i \frac{1}{|\vec{X}_\ell - \vec{X}_m|} - \sum_{\ell=1}^p \sum_{m=1}^i \frac{1}{|\vec{R}_\ell - \vec{X}_m|} \right) \right] \quad (2.1)$$

In equation (2.1) the charge is q on both the ion and the electron. Both species have the same activity so that although the system need not be neutral, only configurations which are nearly neutral should contribute to \mathcal{Z} as we know from physical considerations. λ_0 being the activity is related to the chemical potential, μ_0 , by $\lambda_0 = e^{\beta \mu_0}$. In Eq. (2.1) V is the volume of interest, i.e. the charges are confined to the region, V . Of course, \mathcal{Z} does not exist because of the infinity resulting when a $+$ approaches a $-$. One can introduce repulsive cores; alternatively one can smear the charges a bit. The latter procedure is more natural since, as will be shown later, the smearing of charges corresponds to the smearing of fields, a practice which naturally occurs in the rigorous mathematical treatment of field theories.

Keeping track of combinatorics, (2.1) may be rewritten as

$$\begin{aligned}
 \mathcal{J} &= \sum_{n=0}^{\infty} \frac{\lambda_0^n}{n!} \sum_{q_1=\pm q} \cdots \sum_{q_n=\pm q} \int_V d^3R_1 \cdots d^3R_n \\
 &\times \exp \left\{ -\frac{\beta}{2} \sum_{\ell \neq m} \frac{q_\ell q_m}{|\vec{R}_\ell - \vec{R}_m|} \right\}. \quad (2.2)
 \end{aligned}$$

Consider

$$\begin{aligned}
 &W(R_1, \dots, R_n, q_1, \dots, q_n) \\
 &= \frac{\int \mathcal{D}_X e^{-\beta \int_{\mathbb{R}^3} \frac{\nabla_X \cdot \nabla_X}{8\pi} d^3R + i\beta \int_V \rho(R) \chi(R) d^3R}}{\int \mathcal{D}_X e^{-\beta \int_{\mathbb{R}^3} \frac{\nabla_X \cdot \nabla_X}{8\pi} d^3R}}, \quad (2.3)
 \end{aligned}$$

with

$$\rho(\vec{R}; \vec{R}_1, \dots, \vec{R}_n; q_1, \dots, q_n) = \sum_{\ell=1}^n q_\ell \delta^3(\vec{R} - \vec{R}_\ell). \quad (2.4)$$

Equation (2.3) may be evaluated using the usual rules for gaussian functional integrals:

$$\begin{aligned}
W &= e \int \rho(\vec{R}) \frac{\beta\pi}{4} \frac{1}{v^2} (\vec{R}-\vec{R}') \rho(\vec{R}') d^3R d^3R' \\
&= e \left[-\frac{\beta}{2} \int \rho(\vec{R}) \frac{1}{|\vec{R}-\vec{R}'|} \rho(\vec{R}') \right. \\
&\quad \left. - \frac{\beta}{2} \sum_{\ell \neq m}^n \frac{q_\ell q_m}{|\vec{R}_\ell - \vec{R}_m|} \right] - \beta \text{ (self energy term)}. \quad (2.5)
\end{aligned}$$

The self-energy term is $-\frac{q^2 n}{2} \frac{1}{|\vec{0}|}$ and is infinite. This infinity is made finite by smearing the charges (which is equivalent to smearing fields) or is completely eliminated by normal ordering the final lagrangian of Eq. (2.7) as is revealed in perturbation theory.

Notice that Eq. (2.5) contains the factor present in the integrand of Eq. (2.3). Thus, up to self energy terms,

$$\begin{aligned}
Z &= \frac{1}{N} \iint \mathcal{D}\chi e^{-\beta \int \frac{\nabla\chi \cdot \nabla\chi}{8\pi}} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \sum_{q_1=\pm q} \dots \sum_{q_n=\pm q} \\
&\quad \times \int_V d^3R_1 \dots d^3R_n e^{i\beta \sum_{\ell=1}^n \chi(R_\ell) q_\ell} \\
&= \frac{1}{N} \iint \mathcal{D}\chi e^{-\beta \int \frac{\nabla\chi \cdot \nabla\chi}{8\pi}} \times
\end{aligned}$$

$$\times \sum_{n=0}^{\infty} \frac{\lambda_0^n}{n!} \left(\sum_{q_0=\pm q} \int d^3R e^{i\beta\chi(R)q_0} \right)^n \quad (2.6)$$

or

$$\mathcal{Z} = \frac{1}{N} \iint \mathcal{D}\chi e^{-\beta \int_{R^3} \frac{\nabla\chi \cdot \nabla\chi}{8\pi} + 2\lambda_0 \int_V \cos[\beta q\chi(R)] d^3R} \quad (2.7)$$

$$\text{where } N = \iint \mathcal{D}\chi e^{-\beta \int_{R^3} \frac{\nabla\chi \cdot \nabla\chi}{8\pi}}$$

Equation (2.7) is the fundamental gaussian representation of the grand canonical sum for a Coulomb plasma. \mathcal{Z} has been expressed as the field theoretic $\langle \exp(2\lambda_0 \int \cos\beta q\chi) \rangle$ where the bracket represents an average with respect to the free massless Euclidean functional measure in three dimensions. The corresponding lagrangian is the sine-Gordon and hence one has the result that the sine-Gordon field theory is equivalent to the Coulomb plasma.

χ , in some sense, represents a coarse-grained Coulomb potential. The equation of motion for χ is

$$\nabla^2 \chi = 4\pi(2\lambda_0 q) \sin\beta q\chi. \quad (2.8)$$

Let $\phi = ix$. ϕ then satisfies

$$\nabla^2 \phi = -4\pi(\lambda_0 q) (-2\sinh\beta q\phi), \quad (2.9)$$

which is the well-known Debye-Hückel equation.⁷ The Debye-Hückel equation is usually derived by assuming that the Coulomb potential, ϕ , satisfies $\nabla^2 \phi(\chi) = -4\pi\rho(\chi)$ where $\rho(\chi)$ is the local charge density and hence equal to a mean charge density, n_0 , times the Boltzmann factor for a plus charge to be at x ($\exp\{-\beta q\phi(\chi)\}$) minus the Boltzmann factor for a minus charge to be at x ($\exp\{\beta q\phi(\chi)\}$). For high temperatures $\lambda_0 q = n_0$ so that the Debye-Hückel derivation yields the same result as gaussian integration. The Debye-Hückel derivation is, at best, heuristic. For example, it is not clear why one should use the Boltzmann factors, $\exp\{\pm\beta q\phi(\chi)\}$, rather than the probability factors, $\exp\{\pm\beta q\phi(\chi)\}/[\exp\{\beta q\phi(\chi)\} + \exp\{-\beta q\phi(\chi)\}]$. Gaussian integration eliminates this guesswork. It tells us that the correct charge density factor is $\lambda_0 q$ when the Boltzmann factor $\exp\{-\beta q\phi(\chi)\} - \exp\{\beta q\phi(\chi)\}$ is used.

In the above example the equation of motion of the field theory corresponds to the Debye-Hückel equation of the Coulomb plasma. The gaussian representation method applies to systems interacting via arbitrary two-body potentials. Using the field theory representations of these statistical mechanical systems, one can obtain the analogue of the Debye-Hückel equation by looking at the corresponding field theory equations of motion.

The previous derivation may be generalized in several different ways. First of all, it doesn't depend on the dimension. One merely

replaces the integrals in the action by d-dimensional integrals. For example, if the integrals in equation (2.7) were replaced by d-dimensional integrals then one would obtain the Coulomb gas in d-dimensions (the interparticle Coulomb potential would be $\frac{1}{d-2} \frac{4\pi}{V(S^{d-1})} \frac{q_1 q_2}{r^{d-2}}$, and if one replaced the 8π factors by

$2V(S^{d-1})$ then the interparticle potential would be $\frac{1}{d-2} \frac{q_1 q_2}{r^{d-2}}$).

Here $V(S^{d-1})$ is the volume of the d-1 dimensional sphere.

Secondly, one can use other potentials such as the Yukawa $\exp(-mr)/r$. Consider a potential $V(r, r')$. Let H_0 be the inverse of V , so that $H_0(r, r') \equiv \langle r | H_0 | r' \rangle$ satisfies $\int f(r) H_0(r, r') V(r', r'') g(r'') d^3 r' d^3 r'' = \int f(r) g(r) d^3 r$ (for reasonable f and g). One needs to assume that $\int f(r) H_0(r, r') f(r') d^3 r d^3 r' \geq 0$ for all reasonable f . The partition function for particles interacting via V with activity and inverse temperature respectively λ_0 and β is

$$Z = \sum_n \frac{\lambda_0^n}{n!} \int d^3 R_1 \dots d^3 R_n e^{-\frac{\beta}{2} \sum_{i \neq j} V(R_i, R_j)}, \quad (2.10)$$

which may be expressed in terms of functional integrals using the gaussian integration method:

$$Z = \frac{1}{N} \int \mathcal{D}\chi e^{-\frac{\beta}{2} \int d^3r d^3r' \chi(r) H_0(r, r') \chi(r')} \times e^{\lambda_0 \int_V e^{i\chi(r)} \beta d^3r} \quad (2.11)$$

$$\text{where } N = \int \mathcal{D}\chi e^{-\frac{\beta}{2} \int d^3r d^3r' \chi(r) H_0(r, r') \chi(r')}$$

$$\text{For example, if } V = \frac{e^{-mr}}{r}, H_0(r, r') = \frac{1}{4\pi} \int e^{ik \cdot (r-r')} (k^2 + m^2) \frac{d^3k}{(2\pi)^3}$$

and

$$Z = \frac{1}{N} \int \mathcal{D}\chi e^{-\beta \int \frac{(\nabla\chi \cdot \nabla\chi + m^2\chi^2)}{8\pi} + \lambda_0 \int_V e^{i\beta\chi}} \quad (2.12)$$

Lastly, one can have a gas of several particles with different charges, $q^{(1)}, q^{(2)}, \dots, q^{(m)}$ and activities $\lambda_0^{(1)}, \dots, \lambda_0^{(m)}$.

In the Yukawa case one would call the q 's quanta rather than charges.

The two body potential would be $\frac{q_1 q_2}{r}$ in the case of the Coulomb gas, $q_1 q_2 \frac{e^{-mr}}{r}$ in the case of a Yukawa gas, and $q_1 q_2 V(r)$ in the case of a general gas where $V(r)$ represents the basic potential between two positive unit quanta. The grand canonical sum is

$$\begin{aligned}
 \mathcal{Z} = & \sum_{\ell_1=0}^{\infty} \cdots \sum_{\ell_m=0}^{\infty} \frac{[\lambda_o^{(1)}]_{\ell_1}^{\ell_1}}{\ell_1!} \cdots \frac{[\lambda_o^{(m)}]_{\ell_m}^{\ell_m}}{\ell_m!} \\
 & \times \int_V d^3x_{\ell_1}^{(1)} \cdots d^3x_{\ell_1}^{(1)} \int_V d^3x_{\ell_1}^{(2)} \cdots d^3x_{\ell_2}^{(2)} \\
 & \cdots \int_V d^3x_{\ell_1}^{(m)} \cdots d^3x_{\ell_m}^{(m)} e^{-\beta U}, \quad (2.13)
 \end{aligned}$$

where ℓ_i is the number of the i^{th} specie present, $x_j^{(i)}$ is the coordinate of the j^{th} particle of the i^{th} specie, and U is the sum of potential energy terms between all pairs of particles:

$$U = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^{\ell_i} \sum_{\ell=1}^{\ell_j} q^{(i)} q^{(j)} v(x_k^{(i)}, x_{\ell}^{(j)}) - \text{self-energy}, \quad (2.14)$$

The same derivation as before works with the ρ in Eq. (2.4) replaced by

$$\rho(\vec{r}) = \sum_{i=1}^m \sum_{\ell=1}^{\ell_i} q^{(i)} \delta^3(\vec{x}_{\ell}^{(i)} - \vec{r}), \quad (2.15)$$

and the $\frac{\nabla(\cdot) \cdot \nabla(\cdot)}{4\pi}$ replaced by H_o . \mathcal{Z} becomes

$$\mathcal{Z} = \frac{1}{N} \iint \mathcal{D}\chi e^{-\frac{\beta}{2} (\chi, H_o \chi) + \int_V \sum_{i=1}^m \lambda_o^{(i)} e^{i\beta q^{(i)} \chi(r)} d^3r}. \quad (2.16)$$

In particular, when $m=2$, $\lambda_0^{(1)} = \lambda_0^{(2)} = \lambda_0$, $q^{(1)} = -q^{(2)} = q$, and the Coulomb potential is used, Eq. (2.16) reduces to Eq. (2.7). By choosing $\lambda_0^{(1)} = \lambda_0^+$ different from $\lambda_0^{(2)} = \lambda_0^-$ one can deal with a Coulomb plasma of ions and electrons with an excess of ions (or electrons). Finally, for a neutral system of Yukawa particles with quanta $+$ and $-q$ one obtains

$$\mathcal{L} = \frac{1}{N} \iint d\chi e^{-\beta \int \frac{(\nabla\chi \cdot \nabla\chi + m^2 \chi^2)}{8\pi}} + 2\lambda_0 \int_V \cos\beta q\chi}, \quad (2.17)$$

whose underlying lagrangian is the "massive" sine-Gordon,

$\mathcal{L} = \beta \frac{\nabla\chi \cdot \nabla\chi}{8\pi} + \beta \frac{m^2 \chi^2}{8\pi} - 2\lambda_0 \cos\beta q\chi$. In two dimensions the massive Schwinger model at zero Coleman angle is equivalent to the massive sine-Gordon⁸ and hence is equivalent to a neutral Yukawa gas.⁵

For both Coulomb and Yukawa gases, singularities occur when opposite charges approach each other. In addition there are self-interaction infinities. The self-energy terms can be eliminated by normal ordering the potential which is equivalent to absorbing the infinity into λ_0 as will be shown in the next section. This is well-known to sine-Gordon theorists. The singularity resulting from plus-minus short distance interaction is not so simply eliminated. One convenient possibility is to smear the point charges. This is a reasonable procedure since point charges never exist anyway. Replace a point charge at R_i by a charge distribution $f(r-R_i)$. Hence $\int f(x) d^3x = 1$ and $f(x)$ is peaked

about $x=0$. ρ in Eq. (2.4) would be replaced by

$$\rho_{\text{ smeared}}(\vec{r}) = \sum_{t=1}^n q_t f(\vec{r}-\vec{R}_t). \quad (2.16)$$

The limit $f(x) \rightarrow \delta^3(x)$ reproduces the point charge distribution. The effect on the field theory representation is to replace $\exp(i\beta q_X(R))$ by $\exp(i\beta q_X(\vec{r})f(\vec{r}-\vec{R})) \equiv \exp(i\beta(\chi * f)(\vec{R}))$ so that equation (2.7) becomes

$$\mathcal{Z}_{\text{ smeared}} = \frac{1}{N} \int \mathcal{D}\chi e^{-\beta \int \frac{\nabla\chi \cdot \nabla\chi}{2} + 2\lambda_0 \int_V \cos\beta q(\chi * f)(\vec{R}) d^3R} \quad (2.19)$$

and equation (2.16) is replaced by

$$\mathcal{Z}_{\text{ smeared}} = \frac{1}{N} \int \mathcal{D}\chi e^{-\frac{\beta}{2} (\chi, H_0 \chi) + \int_V \sum_{i=1}^m \lambda_0^{(i)} e^{i\beta q^{(i)}(\chi * f)}} \quad (2.20)$$

This type of smearing is necessary in mathematical field theory where fields are distributions and must always be smeared with test functions. In these models the smearing of fields is natural since it corresponds to the smearing of point charges. The self-energy term also becomes finite and is equal to

$$\frac{q^2}{2} \int \rho(\vec{r}) \frac{e^{-m|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|} \rho(\vec{r}') d^3r d^3r' \quad (2.21)$$

for the Yukawa case. The self-energy in the Coulomb case is given by Eq. (2.21) with $m=0$ and in the general case is

$$\frac{q^2}{2} \int \rho(\vec{r}) V(\vec{r}, \vec{r}') \rho(\vec{r}') d^3r d^3r'. \quad (2.22)$$

III. PERTURBATIVE VERIFICATION

The purpose of this section is twofold. First, the formal gaussian representation is verified in perturbation theory. It is checked to third order in λ_0^3 for the Coulomb plasma model (sine-Gordon field theory) in three dimensions. All orders in β are resummed to give the first few terms of the grand partition function. Thus perturbation theory when rearranged does indeed give the grand canonical sum. The second purpose of this section is to set up a correspondence between perturbative Feynman diagrams and the statistical mechanical system. This is done in the latter part of this section and the results are summarized in Table 1.

$\langle \exp[2\lambda_0 \int \cos \beta q \chi] \rangle$ is the sum of vacuum bubble diagrams. To obtain the Feynman rules one could rescale χ so that $S_0 = \frac{1}{2} \int \nabla \chi \cdot \nabla \chi \, d^3R$, however when not rescaled $S_0 (= \int \frac{\nabla \chi \cdot \nabla \chi}{8\pi} \, d^3R)$ acts like the electrostatic energy of the system. To retain this physical meaning the Feynman rules will be listed without χ rescaled. For bubble diagrams they are

a) Draw all topologically distinct vacuum bubbles (connected or disconnected) with vertices of an arbitrary even order (including zero order). Order, here, refers to the number of lines attached to a vertex.

b) For each vertex associate a factor $(2\lambda_0) \int_V d^3r_i$. i refers to the i^{th} vertex.

c) For each vertex of order $2n$ associate a factor of $(-\beta^2 q^2)^n$.

- d) For each propagator associate a factor of $\frac{1}{\beta} \frac{1}{|r_i - r_j|} .^9$
- e) Put in a factor $\frac{1}{k!}$ for each pair of vertices connected by k lines. See Figure 1a.
- f) Put in a factor of $\frac{1}{(2k)!} = \frac{1}{2^k k!}$ for each k -fold self-energy tadpole. See Figure 1b.
- g) Put in a factor of $[(\text{order of symmetry group of graph})!]^{-1}$.
- h) The empty graph is to be included and contributes unity.

Equivalent rules for a) and g) are:

- a) Draw all bubble graphs (topologically distinct or not), that is label the vertices and treat them as distinguishable.
- g) Put in a factor of $[\text{number of vertices}]^{-1}$.

The effect of the self-energy tadpoles is to renormalize λ_0 . Any graph can be drawn as a graph without tadpoles plus tadpoles adjoined. Consider the effect of adding an arbitrary number of tadpoles to a "bare" vertex (see Figure 2). The following factor will multiply the "tadpoleless" vertex:

$$\sum_{n=0}^{\infty} (-\beta^2 q^2)^n \left(\frac{1}{n!} \frac{1}{2^n} \right) \left(\frac{1}{\beta |\vec{0}|} \right)^n = e^{-\frac{1}{2} \beta q^2 \frac{1}{|\vec{0}|}} . \quad (3.1)$$

If the smeared interaction $2\lambda_0 \cos(\beta q x * f)$ is used Eq. (3.1) becomes

$$e^{-\frac{1}{2} \beta q^2 \int f(r) \frac{1}{|r-r'|} f(r') d^3r d^3r'} . \quad (3.2)$$

Comparing this with equations (2.5) and (2.21 with $m=0$), one sees that the effect of tadpoles is to multiply each vertex by $e^{-\beta}$ self-energy. Hence rules b) and f) are modified to

b) For each vertex associate a factor $2\lambda_R \int_V d^3r_i$ with λ_R the renormalized activity and $\lambda_R = \lambda_0 e^{-\beta \text{self-energy}}$.

f) Do not include self-energy tadpole diagrams.

Alternatively one may use $2\lambda_R \cos \beta q x$ as the interaction density. Normal ordering corresponds to a renormalization of λ_0 . This fact, well known to sine-Gordon theorists, actually holds for any interaction which can be represented in gaussian form. One can also see this in the grand canonical sum (Eq. [2.2]) where the self-energy terms would simply factor out to multiply λ_0^n by $e^{-\beta n \text{S.E.}}$, i.e.

$$\lambda_0 \rightarrow \lambda_0 e^{-\beta \text{self-energy}} = \lambda_R.$$

It will now be checked that to order λ_R^3 what perturbation theory reproduces the \mathcal{Z} of Eq. (2.2). To zero'th order the empty diagram contributes 1 and \mathcal{Z} begins with 1. The diagrams of order λ_0 are shown in Figure 3. They contribute $2\lambda_R \int_V d^3r = 2\lambda_R V$ which equals

$\sum_{q_1=-iq} \lambda_R \int_V d^3r$. The diagrams of order λ_R^2 are shown in Figure 4. They

sum to

$$\lambda_{R^2}^2 =$$

$$\frac{1}{2} (2\lambda_R)^2 \sum_{n=0}^{\infty} \frac{1}{(2n)!} (-\beta^2 q^2)^{2n} \int_V d^3R_1 d^3R_2 \frac{1}{(\beta |R_1 - R_2|)^{2n}}$$

$$= \lambda_R^2 \int_V d^3R_1 d^3R_2 \left(e^{-\beta q^2 / |R_1 - R_2|} + e^{\beta q^2 / |R_1 - R_2|} \right). \quad (3.3)$$

The $\frac{1}{2}$ factor multiplying the expression comes from the symmetry factor of rule g. (3.3) equals

$$\frac{\lambda_R^2}{2!} \sum_{q_1 = \pm q} \sum_{q_2 = \pm q} \int d^3R_1 d^3R_2 e^{-\beta \frac{q_1 q_2}{|R_1 - R_2|}}, \quad (3.4)$$

which is the second order term of Eq. (2.2).

The diagrams of order λ_R^3 are shown in Figure 5. They separate into two classes: those with an even number of propagators between vertices (Figure 5a), and those with an odd number of propagators between vertices (Figure 5b). The even case contributes

$$\begin{aligned}
\lambda_{R_3}^{3_2}(e) &= \\
& \frac{(2\lambda_R)^3}{3!} \sum_{l=0}^{\infty} \frac{1}{(2l)!} \sum_{m=0}^{\infty} \frac{1}{(2m)!} \sum_{n=0}^{\infty} \frac{1}{(2n)!} \\
& \times \int d^3R_1 d^3R_2 d^3R_3 (-b^2 q^2)^{2l+2m+2n} \\
& \times \left(\frac{1}{b|R_1-R_2|} \right)^{2l} \left(\frac{1}{b|R_2-R_3|} \right)^{2m} \left(\frac{1}{b|R_3-R_1|} \right)^{2n}. \quad (3.5)
\end{aligned}$$

l is the number of propagators between vertex 1 and 2, m is the number between 2 and 3, and n is the number between 3 and 1.

Summation of (3.5) yields

$$\begin{aligned}
\lambda_{R_3}^{3_2}(e) &= \\
& \frac{\lambda_R^3}{3!} \int d^3R_1 d^3R_2 d^3R_3 \left(e^{\frac{\beta q^2}{|R_1-R_2|}} + e^{\frac{-\beta q^2}{|R_1-R_2|}} \right) \\
& \times \left(e^{\frac{\beta q^2}{|R_2-R_3|}} + e^{\frac{-\beta q^2}{|R_2-R_3|}} \right) \\
& \times \left(e^{\frac{\beta q^2}{|R_3-R_1|}} + e^{\frac{-\beta q^2}{|R_3-R_1|}} \right). \quad (3.6)
\end{aligned}$$

The odd case is similar and gives

$$\lambda_{R^3}^{3z^{(o)}} = \frac{-\lambda_R^3}{3!} \int d^3R_1 d^3R_2 d^3R_3 \left(e^{\frac{\beta q^2}{|R_1-R_2|}} - e^{\frac{-\beta q^2}{|R_1-R_2|}} \right) \\ \times \left(e^{\frac{\beta q^2}{|R_2-R_3|}} - e^{\frac{-\beta q^2}{|R_2-R_3|}} \right) \left(e^{\frac{\beta q^2}{|R_3-R_1|}} - e^{\frac{-\beta q^2}{|R_3-R_1|}} \right) . \quad (3.7)$$

Equations (3.6) and (3.7) sum to give

$$\lambda_{R^3}^{3z^2} = \lambda_R^3 \left(Z_3^{(v)} + Z_3^{(c)} \right) = \frac{2\lambda_R^3}{3!} \int d^3R_1 d^3R_2 d^3R_3 \\ \times \left\{ \exp-\beta q^2 \left(\frac{1}{|R_1-R_2|} + \frac{1}{|R_2-R_3|} + \frac{1}{|R_3-R_1|} \right) \right. \\ + \exp\beta q^2 \left(\frac{-1}{|R_1-R_2|} + \frac{1}{|R_2-R_3|} + \frac{1}{|R_3-R_1|} \right) \\ + \exp\beta q^2 \left(\frac{1}{|R_1-R_2|} - \frac{1}{|R_2-R_3|} + \frac{1}{|R_3-R_1|} \right) \\ \left. + \exp\beta q^2 \left(\frac{1}{|R_1-R_2|} + \frac{1}{|R_2-R_3|} - \frac{1}{|R_3-R_1|} \right) \right\} \quad (3.8)$$

and this agrees with the third order term in Eq. (2.2).

Doubtlessly, perturbation theory reproduces the grand partition function to all orders in λ_R for this particular example, the Coulomb plasma in three dimensions. If another interaction had been used, or if several charges of different activities had been used, perturbation theory would have reproduced \mathcal{Z} . Of course, the Feynman rules would have to be modified. In particular Rule d would have to be replaced by

d) For each propagator associate a factor of $\frac{1}{g} V(r_i, r_j)$.

Particles of different activities would lead to vertices which would have to be distinguished. The appropriate λ 's would have to be associated with the appropriate vertices, etc. Finally, the types of graphs and the factors associated with a particular vertex order would be different. The Feynman rules for other theories are thus easily obtained by modifying the rules presented in this paper.

A simplification can be made. There are zero order vertices (vertices to which no lines are attached) in the Feynman graphs because the interaction $2\lambda_R \cos \beta q x$: when Taylor expanded begins with $2\lambda_R$. One can rewrite $2\lambda_R \cos \beta q x$: [or $\lambda^{(i)} e^{i\beta q^{(i)} x}$] as $2\lambda_R (\cos \beta q x - 1) + 2\lambda_R$ [similarly for $\lambda^{(i)} e^{i\beta q^{(i)} x}$ type terms]. Eq. (2.7) can be rewritten as

$$Z = \frac{1}{N} e^{2\lambda_R V} \int D\phi e^{-\int \frac{V_R \cdot \nabla^2 \phi}{6\pi} + 2\lambda_R \int (\cos \phi - 1)} \quad (3.9)$$

As a result the statement, "Don't include zero order graphs but multiply all graphs by $e^{2\lambda_R V}$ " is added to Feynman rule a.

We know from perturbation theory that the contributions of graphs may be expressed in the form

$$Z = \exp \left[\sum \text{connected graphs} \right] = \exp \left[2\lambda_R V + \left(\sum_{k=2}^{\infty} \lambda_R^k b_k \right) V \right], \quad (3.10)$$

where b_k = the connected graphs of order λ_R^k . The volume factor is put into the definition of the b_k because vacuum bubble diagrams are proportional to V due to translational invariance (actually this is not quite correct because of boundary effects, but is approximately true in the large volume limit). The vacuum energy per unit area (three dimensions=two space + one time), ξ , for the field theory is

$$\mathcal{E} = \sum_{\ell=1}^{\infty} b_{\ell} \lambda_R^{\ell} \quad (b_1 \neq 2). \quad (3.11)$$

The b_{ℓ} have an important connection with statistical mechanics. They are the cluster integrals of the Mayer expansion.¹⁰ The thermodynamic properties are determined by these b_{ℓ} . In particular

$$\langle \rho \rangle = \sum_{\ell=1}^{\infty} \ell b_{\ell} \lambda_R^{\ell} \quad (3.12a)$$

$$\beta \langle p \rangle = \sum_{\ell=1}^{\infty} b_{\ell} \lambda_R^{\ell}, \quad (3.12b)$$

where ρ is the density of particles, $\langle \rho \rangle = \frac{\langle N \rangle}{V}$, and p is the pressure. When λ_R is expressed in terms of $\langle \rho \rangle$ via Eq. (3.12a) and substituted into (3.12b) the equation of state is obtained.

$\beta \langle p \rangle$ of the statistical mechanical system is equal to the vacuum energy density, \mathcal{E} , of the field theory.

The small parameter in the Mayer expansion is the function, $e^{-\beta V(r)} - 1$. Often $V(r)$ is short ranged so that $e^{-\beta V(r)} - 1$ is non-zero only in a small region compared to V . Such a case occurs in the very massive ($m \gg \frac{1}{V}$) Yukawa gas. An immediate application would be to the massive Schwinger model in two dimensions at zero Coleman angle, since, as previously noted, this is equivalent to

to the massive sine-Gordon lagrangian (see Eq. [2.17]). The charge, e , of the Schwinger model is related to the mass, m , of the massive sine-Gordon by $\frac{m^2 \hbar}{2} = e^2$. Hence one may obtain results in the strong coupling limit of the massive Schwinger model by using the Mayer cluster expansion applied to a Yukawa gas.¹¹

In performing this check to order λ_R^3 one notices a correspondence between diagrammatic perturbation theory and the Coulomb gas. The vertices of Feynman diagrams are the ions and electrons of the plasma. In the Yukawa gas case, they would be the quanta and in the potential $[V(r)]$ case, one might call them the molecules. Up to a temperature factor the propagators represent the interactions. The number of interactions a particle undergoes is the same as the order of the vertex. Pairs of particles may undergo arbitrarily many interactions and when summed these give the Boltzman factors. An external vertex at x corresponds to fixing a molecule in the gas at x . In field theoretic language it is the vacuum expectation value of the operator, $\phi_q(x) \equiv e^{iBq\chi(x)}$. This operator may be interpreted as producing a charge, q , at x . Diagrams with several fixed external vertices are related to the correlation functions used in statistical mechanics.^{7,10} In field theory they are the Green's functions of the $\phi_q(x)$ fields. Finally writing

$$g = \sum_{N=0}^{\infty} \lambda_R^N Z_N \quad (\text{with } Z_N \text{ the partition function for } N$$

interacting particles), one sees that the N^{th} order diagrams yield

the N -particle partition function. Thus there is a complete correspondence between Feynman diagrams and the Coulomb plasma. This correspondence is summarized in Table 1.

There is a sense in which the usual polynomial field theories (such as $g \frac{\chi^4}{4!}$) are Coulomb or Yukawa plasmas. In a high temperature limit consider the sine-Gordon Lagrangian whose underlying statistical mechanical system is the Coulomb plasma. The potential, $V = -2\lambda_R \cos \sqrt{4\pi\beta} \phi$: (χ has been rescaled to eliminate the β dependence in $H_0 (= \frac{1}{2} \int \nabla\chi \cdot \nabla\chi$), in this limit may be expanded in a Taylor series; $V = 2\lambda_R (-1 + 4\pi\beta q^2 : \frac{\chi^2}{2} : - (4\pi\beta q^2)^2 : \frac{\chi^4}{4!} : + \dots)$. If the system is "hot" and χ does not fluctuate violently from 0, then $V \approx -2\lambda_R + \lambda_R 8\pi\beta q^2 : \frac{\chi^2}{2} : - 2\lambda_R (4\pi\beta q^2)^2 : \frac{\chi^4}{4!} :$ and hence one has a massive χ^4 theory with mass equal to $8\pi\lambda_R \beta q^2$ and a small negative χ^4 -coupling constant, g , equal to $-32\pi^2 \beta^2 q^4 \lambda_R$. Perturbation theory would be useful in this high temperature limit since the coupling constant is small. Allowing arbitrary charges and activities the potential, V , becomes

$$V = - \sum_i \lambda^{(i)} e^{i\beta q^{(i)} \chi} \quad (3.13)$$

By adjusting the $\lambda^{(i)}$ and $q^{(i)}$ one may obtain better approximations to polynomial self-coupled field theories. In fact Eq. (3.13) is almost a Fourier transform. Unfortunately the $\lambda^{(i)}$ must be positive to retain their physical meaning. This restriction ruins the possibility of exact approximation.

IV. THE TWO DIMENSIONAL SINE-GORDON

The last two sections have presented general methods and techniques. A specific example, the two dimensional sine-Gordon theory,¹² will be, for the most part, the subject of the rest of this paper. As previously noted, this field theory is equivalent to a two dimensional Coulomb gas. The interparticle potential is a logarithmic one:

$$V(r) = -2q_1q_2 \ln \frac{|R_1 - R_2|}{a_0}, \quad (4.1)$$

with a_0 arbitrary. Eq. (4.1) is also the interaction between two parallel lines of charge, one with a charge per unit length of q_1 and one of charge per unit length of q_2 . Thus one may view the charges, q , as the charge densities in wires in the usual three dimensional world. The wires are restricted to be perpendicular to a two dimensional sheet. Another equivalent model is to replace the charged wires by currents. The magnetic interaction leads to the same logarithmic potential, $V = -2I_1I_2 \ln \frac{r}{a_0}$. The q 's would then be the currents, I .

The two dimensional Coulomb plasma differs from the three dimensional version in one important way. Due to infrared divergences a smeared charge distribution has infinite energy unless it is neutral. Consider such a charge distribution, ρ , restricted to a finite

region, The electric field goes like $Q_T^2 \frac{\vec{r}}{r^2}$ for large r . Therefore the energy density $\frac{\nabla\phi \cdot \nabla\phi}{8\pi}$ goes like $\frac{Q_T^2}{8\pi} \frac{1}{r^2}$ and hence the

total energy diverges logarithmically unless the total charge, Q_T , is zero. In dealing with this two dimensional Coulomb gas one has two choices. The first is to require total neutrality. Z would become

$$Z = \sum_{n=0}^{\infty} \frac{\lambda_0^{2n}}{n!n!} \int d^2R_1 \cdots d^2R_n \int d^2X_1 \cdots d^2X_n \exp \left\{ 8q^2 \sum_{i,j} \left(\ln \frac{|R_i - R_j|}{a_0} + \ln \frac{|X_i - X_j|}{a_0} - 2 \ln \frac{|R_i - X_j|}{a_0} \right) \right\}. \quad (4.2)$$

Z is independent of a_0 as long as self-energies are retained.

Z still corresponds to the sine-Gordon field theory because non-neutral plasmas do not contribute to the functional integral, however naive perturbation theory is incorrect. A correct way of obtaining the Feynman rules is to use the massive sine-Gordon lagrangian and take the limit $m \rightarrow 0$. As soon as $m^2 V \ll 1$ the massive propagator becomes $2 \ln |mr|$. Perturbation theory when rearranged and partially summed gives for Z

$$\begin{aligned}
 \mathcal{Z} = & \lim_{m \rightarrow 0} \sum_{n=0}^{\infty} \frac{\lambda_0^n}{n!} \sum_{q_1 = \pm q} \cdots \sum_{q_n = \pm q} \int d^2 R_1 \cdots d^2 R_n \\
 & \times \exp \left\{ \beta \sum_{i,j} q_i q_j \ln m |R_i - R_j| \right\}. \quad (4.3)
 \end{aligned}$$

The non-neutral sums in Eq. (4.3) are proportional to $m^{\beta Q_T^2}$ and hence vanish as $m \rightarrow 0$. The Feynman rules should use the propagator of Eq. (4.2) in the limit where a_0 goes to infinity. The technique of adding a mass term to the sine-Gordon lagrangian and letting the mass go to zero is not new. S. Coleman¹³ used it in a paper showing the equivalence of the sine-Gordon with the massive Thirring model. It has a physical meaning since it demands neutrality of the plasma system. This paper will use this version of handling the infrared divergence. Total neutrality will always be maintained.

An alternative approach is to enclose the system in a grounded conducting casing. If there is an excess charge within V then an equal and opposite charge will appear on the conductor. In calculating the partition function one integrates $\nabla\phi \cdot \nabla\phi$ only over the volume, V , since the conductor causes $\nabla\phi$ to be zero outside V . The gaussian representation of \mathcal{Z} would be modified to

$$\mathcal{Z} = \frac{1}{N} \iint \mathcal{D}\chi e^{-\beta \int_V \frac{\nabla\chi \cdot \nabla\chi}{8\pi} + 2\lambda_0 \int_V \cos\beta q\chi}. \quad (4.4)$$

The equation of motion, $\frac{\beta}{4\pi} \nabla^2 \chi - 2\lambda_0 \beta q \sin \beta q \chi = 0$, is not valid because of surface terms due to integration by parts. These surface terms represent dynamical degrees of freedom and must be quantized. In principle this can be done using the techniques of M. Halpern and P. Senjanovic.¹⁴

V. THE PHASES OF THE SINE-GORDON

This section will review the work done on the two dimensional Coulomb gas^{5,15} and relate it to the work done on the sine-Gordon field theory.^{3,4,5,13,16,17} In particular, the phase of the system will be determined. Coleman has shown that a vacuum instability occurs when βq^2 gets too large.¹³ This corresponds to a phase transition in the Coulomb system.⁵

Because the works of others will be referred to and because people have used different variables to denote the parameters of the sine-Gordon equation, there is some notational confusion. For example, the β that Coleman uses is not the inverse temperature. When confusion is possible I will subscript letters with the authors initials. For example, the β of Hauge and Hemmer is $\frac{1}{2}$ the β used in this paper so that $\beta_{H.H.}$ will refer to their inverse temperature. This paper, for the most part, conforms with the notation of Kosterlitz and Thouless. Table 2 provides the relations between the parameters of this paper and the parameters which others use.

The method of Kosterlitz and Thouless will be used to determine the phases of the Coulomb system when λ_R is small. For $\beta q^2 \ll 1$, λ_R corresponds to the density, so that small λ_R means a dilute system. In fact, at $\beta q^2 = 0$,

$$Z = \sum_{n=0}^{\infty} \frac{(\lambda_R v)^{2n}}{n! n!} . \text{ Consider the situation where}$$

$\lambda_R V \gg 1$ even though $\lambda_R \ll 1$. g has a maximum contribution for $n \gg 1$.
Replacing the sum by an integral and using Stirling's formula,

$$\begin{aligned}
 g &\approx \int e^{2n \log \lambda_R V + 2n - 2n \log n - \log n} \frac{dn}{2\pi} \\
 &\approx \sqrt{\frac{-1}{2\pi f''(n_{\max})}} e^{f(n_{\max})} \\
 &\approx \sqrt{\frac{\lambda_R V}{4\pi}} \exp\{2\lambda_R V - \log(\lambda_R V)\} \\
 &= \sqrt{\frac{1}{4\pi\lambda_R V}} e^{2\lambda_R V}, \tag{5.1}
 \end{aligned}$$

where $f(n) = 2n \log \lambda_R V + 2n - 2n \log n - \log n$ and $n_{\max} e^{\frac{1}{2n_{\max}}} = \lambda_R V$. $n_{\max} \approx \lambda_R V$ for $\lambda_R V$ large. The integral has been approximated by Laplace's method. $\langle N \rangle = \lambda_R \frac{2}{2\lambda_R} \log g \approx 2\lambda_R V$. $\lambda_R = \frac{\langle N \rangle}{2V}$ which shows λ_R is indeed a density. The limits $\lambda_R V \gg 1$ but $\lambda_R \ll 1$ correspond to a situation where many ions are present but the density is small which is the proper statistical limit (in the limit $\lambda_R V \ll 1$, $\langle N \rangle \approx \frac{1}{2} \lambda_R^2 V^2$ so $\langle N \rangle \ll 1$ which is undesirable).

When λ_R is small one can calculate the mean square distance between an ion and an electron by assuming that the other charges in the plasma may be neglected. In fact the exact expression for

the mean square distance is

$$\langle r^2 \rangle = \frac{\sum_{n=0}^{\infty} \frac{\lambda_0^{2n}}{n!n!} \int d^2x_1 \dots d^2x_N \int d^2R_1 \dots d^2R_N (x_1 - R_1)^2 e^{-\beta U_N}}{\sum_{N=1}^{\infty} \frac{\lambda_0^{2N}}{n!n!} \int d^2x_1 \dots d^2x_N \int d^2R_1 \dots d^2R_N e^{-\beta U_N}}, \quad (5.2)$$

where U_N is the energy of the configuration (see Eq. [4.2]). The $N=1$ term gives for the mean square distance

$$\langle r^2 \rangle \approx \frac{\int_{r_0}^R r dr e^{-2\beta q^2 \ln r^2}}{\int_{r_0}^R r dr e^{-2\beta c^2 \ln r}} = \frac{\left(R^{4-2\beta q^2} - r_0^{4-2\beta q^2} \right) \left(2-2\beta q^2 \right)}{\left(4-2\beta q^2 \right) \left(R^{2-2\beta q^2} - r_0^{2-2\beta q^2} \right)}, \quad (5.3)$$

where r_0 is an ultraviolet cutoff introduced to make $\langle r^2 \rangle$ well defined for $\beta q^2 \geq 1$. In fact if the charges were not point charges but "ringlets" of charge densities of radius, r_0 , Eq. (5.3) would be the mean square distance between ringlets. In using Eq. (5.3) I am not implying that the $N=1$ term dominates. It is obvious from the above discussion that a large value of N dominates. Using the dominate term is, of course, like using a partition function in lieu of the grand sum. Eq. (5.3) is inaccurate in the region $r > \sqrt{\frac{V}{\langle N \rangle}}$

and the integrals should probably be cut off at such a value (which is still a large number in the dilute gas approximation). In Eq. (5.3) I have neglected "edge effects" which occur if one of the charges is near the boundary of the volume. Eq. (5.3) is calculated on the basis that one of the charges is at the center of the volume.

Eq. (5.3) yields the following result for $\beta q^2 < 1$:

$$\langle r^2 \rangle \approx \frac{R^2(1-\beta q^2)}{(2-\beta q^2)} \quad (\lambda_R \text{ small}). \quad (5.4)$$

The fluctuation in the distance between charges according to Eq. (5.4) is large and hence for $\beta q^2 < 1$ the Coulomb system is in the plasma phase, that is the electrons and ions do not pair up to form dipoles. At $\beta q^2 = 1$ the same is still true since

$$\langle r^2 \rangle \approx \frac{R^2}{2 \ln \frac{R}{r_0}} \quad (\lambda_R \text{ small}). \quad (5.5)$$

In the region where $1 < \beta q^2 < 2$ $\langle r^2 \rangle$ is renormalization dependent, i.e. $\langle r^2 \rangle$ depends on the parameter r_0 :

$$\begin{aligned} \langle r^2 \rangle &\approx \frac{\beta q^2 - 1}{2 - \beta q^2} R^2 \left(\frac{r_0}{R} \right)^{2\beta q^2 - 2} \\ &= \frac{\beta q^2 - 1}{2 - \beta q^2} r_0^2 \left(\frac{R}{r_0} \right)^{4 - 2\beta q^2}. \end{aligned} \quad (5.6)$$

Eq. (5.6) shows that although $\langle r^2 \rangle$ is not of the order of the size of the system, it is still much greater than r_0^2 and hence a dipole

collapse has not yet occurred. Although some dipoles may exist, the preponderance of ions and electrons are still unbound and thus the Coulomb plasma phase still occurs. Of course as λ_R gets bigger the above conclusions may no longer be correct since Eq. (5.3) is based on the dilute gas approximation. It's conceivable that if λ_R gets large enough a phase transition might occur.

Finally for $\beta q^2=2$ and $\beta q^2>2$ the dipole phase occurs since

$$\langle r^2 \rangle \approx 2r_0^2 \ln \frac{R}{r_0} \quad (\beta q^2=2)$$

$$\langle r^2 \rangle \approx \frac{\beta q^2 - 1}{\beta q^2 - 2} r_0^2 \quad (\beta q^2 > 2). \quad (5.7)$$

A phase transition occurs around $\beta q^2=2$ for small λ_R . The nature of the phase transition is simple: as βq^2 increases the free ions and electrons of the Coulomb plasma collapse to form dipoles, and a new gas of weakly interacting dipoles is formed. This phase transition has been examined in more detail by Kosterlitz and Thouless who find a divergence in the polarizability in going from high β to low β (i.e. from dipoles to the plasma). This divergence is understandable since as the temperature increases the average separation between a plus and minus forming a dipole increases. This causes the dipole moment to increase and as a result the polarizability of the system is greatly enhanced. Using reasonable methods they obtain that $q^2 \beta_{\text{critical}} \approx 2$.

This phase transition has an important relation to the sine-

Gordon field theory. The point $\beta q^2=2$ corresponds to $\beta_C^2 = 8\pi$. It was precisely at this point that Coleman found a vacuum instability. One can now understand this instability from the Coulomb point of view: it is precisely a phase transition from an ion plasma to a dipole gas.

Solitons are known to exist as solutions to the sine-Gordon equation.¹⁸ The way the critical temperature varies as λ_R varies is important since it may affect the number and stability of soliton-antisoliton bound states. Luther¹⁷ has proven quantum mechanically stable bound states occur for $n=1,2,\dots, \frac{\beta q^2}{2-\beta q^2}$ ($0 < \beta q^2 < 2$) ($n=0$ always exists and is the usual soliton) with masses of the form

$$m_n = C(\beta) \sin \frac{n\pi}{2} \frac{\beta q^2}{2-\beta q^2}, \quad (5.3)$$

where $C(\beta)$ is a temperature-dependent renormalized constant. The renormalization of C depends on the lattice spacing and the x-y anisotropy (Luther used the spin $\frac{1}{2}$ x-y-z lattice chain to obtain the above results). This seems to indicate the number of bound states does not depend on λ_R .

There are three possible phase diagrams which might occur. These are shown in Figure 6. In these diagrams the pressure, p , or the density, ρ , may be substituted for λ_R if the equation of state is known. One strongly suspects that λ_R increases monotonically as p or ρ increases for fixed volume and temperature. If Figure (6a) is the situation, the dipole phase (λ_R very large) would probably prevent the solitons from existing even though βq^2 is less

than 2. Both (a) and (c) cases require an additional parameter enter the theory, since λ_R , being the only dimensional constant, has no dimensional quantity to set the scale.¹⁹ It's therefore meaningless to plot λ_R versus βq^2 . Indications (see section 7) are that for $\beta q^2 \gg 1$ a cutoff must be introduced due to ultraviolet singularities. If this cutoff can be removed without introducing a new dimension into the theory (i.e. there is no dimensional transmutation) then the only possibility would be (b) and solitons of arbitrary n occur for βq^2 sufficiently close to 2. Kosterlitz and Thouless obtained a curve similar to (c). They got $\beta_{\text{critical}} q^2 = 2(1 + c\lambda_R)$ with $c \approx 1.3\pi$. They used a cutoff in their potential ($U_{\text{K.T.}}(r) = -q_1 q_2 \ln \frac{r}{r_0} + 2\mu$ for $r > r_0$ and $U(r) = 0$ for $r < r_0$). The reason for their result is simple: they view the Coulomb system from the dipole side of $\beta_{\text{critical}} q^2$. Their first approximation was to neglect the effects of all other dipoles in calculating $\langle r^2 \rangle$, the mean distance squared between the plus and minus constituents of the dipole. They obtained $\beta_{\text{critical}} q^2 = 2$. What corrections result if other dipoles are taken into consideration? Basically, it will be easier to separate the plus and minus constituents because dipoles will interpolate to reduce the potential. Consequently, each charge is partially screened and it will be easier to pull them apart. The presence of dipoles lowers the temperature at which the phase occurs (in other words $\beta_{\text{critical}} q^2$ increases with λ_R). They expressed the screening in terms of an effective dielectric constant $\epsilon(r)$ which depended on the separation of the plus and minus. If possibility (c) occurs as Kosterlitz and Thouless have predicted

one would expect the number of states to be different than is presently predicted (a natural guess would be $n=1,2,\dots$,

$$\frac{\beta}{\beta_{\text{critical}} - \beta} \text{ and correspondingly } m_n = C(\beta) \sin \frac{n\pi}{2} \left[\frac{\beta}{\beta_{\text{critical}} - \beta} \right].$$

I refer the reader to their paper for their results. It may be that the situation depends on how one modifies the potential at short distances (to eliminate the ultraviolet singularities), and thus Kosterlitz and Thouless's result is one possible example which might occur.

VI. THE NON-LINEAR σ -MODEL

This section will show the equivalence of a non-linear σ -model and the sine-Gordon equation.²⁰ First consider the linear $O(2)$ σ -model with a linear symmetry breaking term:

$$\mathcal{L} = \frac{1}{2} (\partial\sigma)^2 + \frac{1}{2} (\partial\Pi)^2 - V(\sigma, \Pi), \quad (6.1)$$

and

$$V(\sigma, \Pi) = -2a\sigma + g(\sigma^2 + \Pi^2 - f_{\Pi}^2). \quad (6.2)$$

When $a=0$, $m_{\Pi}^2=0$ (Goldstone boson) and $m_{\sigma}^2=16gf_{\Pi}^2$ (the usual PCAC type relation of the σ -mass being proportional to f_{Π}). When $a \neq 0$ the minimum of V occurs at $\Pi=0$, $\sigma = (\text{sign } a)f_{\Pi} + \frac{a}{4f_{\Pi}^2} \left(\frac{1}{g}\right) + O\left(\frac{1}{g^2}\right)$ and the Π -field acquires a mass, $m_{\Pi}^2 = \frac{2|a|}{f_{\Pi}} + O\left(\frac{1}{g}\right)$. The non-linear σ -model with a linear symmetry breaking term is the limit of Eq. (6.2) as $g \rightarrow \infty$. This has the effect of requiring $\sigma^2 + \Pi^2 = f_{\Pi}^2$. In fact, if Eq. (6.1) was used in a functional integral the limit $g \rightarrow \infty$ would produce a functional delta-function, $\delta(\sigma^2 + \Pi^2 - f_{\Pi}^2)$. To enforce the relation $\sigma^2 + \Pi^2 = f_{\Pi}^2$ let $\sigma = -f_{\Pi} \cos\theta$ and $\Pi = -f_{\Pi} \sin\theta$ where θ is a new field. \mathcal{L} becomes

$$\mathcal{L} = \frac{1}{2} f_{\Pi}^2 (\partial\theta)^2 + \varepsilon \cdot f_{\Pi} \cos\theta. \quad (6.3)$$

which is the sine-Gordon lagrangian. Rescaling $\theta = \frac{1}{f_{\Pi}} \chi$ gives

$$\mathcal{L} = \frac{1}{2} (\partial \chi)^2 + 2af_{\Pi} \cos\left(\frac{1}{f_{\Pi}} \chi\right), \quad (6.4)$$

or $V(\chi) = -2af_{\Pi} \cos\left(\frac{1}{f_{\Pi}} \chi\right)$ from which one can translate the parameters of this model with the Coulomb gas parameters:

$$\lambda_0 = 2af_{\Pi}$$

$$\sqrt{4\pi\beta} q = \frac{1}{f_{\Pi}} \quad \text{or} \quad \beta q^2 = \frac{1}{4\pi f_{\Pi}^2}. \quad (6.5)$$

Since a phase transition takes place in the sine-Gordon field theory at $\beta q^2 \approx 2$ for λ_R small, a phase transition must occur in the non-linear $O(2)$ σ -model with small symmetry breaking term for $\frac{f_{\Pi}}{\Lambda} \approx 1/\sqrt{8\pi}$. This complements the results of E. Brézin and J. Zinn-Justin²¹ and W. A. Bardeen, B. W. Lee, and R. E. Shrock²² who find transitions in the $O(N)$ non-linear σ -model in $2+\epsilon$ dimensions for $N > 2$.

VII. RENORMALIZATION

S. Coleman¹³ has indicated that the only renormalization necessary in the two dimensional sine-Gordon model is for the self-energy tadpoles (such as in Figure 1b). These infinite contributions can be absorbed in λ_0 . One should use a renormalized activity λ_R and ignore the tadpole self-energies. S. Coleman's result is true for $\beta q^2 < 1$ ($\beta_C^2 < 4\pi$). For $\beta q^2 \geq 1$ the result is incorrect: although there are no divergent graphs in any finite order (in λ_R and β) of perturbation theory, there are divergences when graphs are summed. The reader has already seen an example of this: the connected vacuum bubbles of order λ_R^2 (the graphs of Figure 4 minus the first one). They sum to

$$\begin{aligned}
 \lim_{a_0 \rightarrow \infty} v \left(\frac{2\lambda_R}{a_0 \beta q^2} \right)^2 \int d^2 r \sum_{n=0}^{\infty} \left(\frac{2}{\beta} \ln \left| \frac{r}{a_0} \right| \right)^{2n} \frac{(-\beta q^2)^{2n}}{2(2n)!} - 2\lambda_R^2 v^2 \\
 = \lim_{a_0 \rightarrow \infty} \frac{2\lambda_R^2 v}{a_0^2 \beta q^2} \int_V d^2 r \left[\left(\frac{r}{a_0} \right)^{2\beta q^2} + \left(\frac{r}{a_0} \right)^{-2\beta q^2} \right] - 2\lambda_R^2 v^2 \\
 = 2\lambda_R^2 v \int_V d^2 r \left[r^{-2\beta q^2} - 1 \right], \tag{7.1}
 \end{aligned}$$

which converges for $\beta q^2 < 1$ and diverges for $\beta q^2 \geq 1$. Vacuum bubbles are not the only diagrams with divergences which cannot be absorbed in

λ_0 . Consider the contributions to the two-point function from the graphs of Figure 7. They sum to

$$S(x,y) = (2\lambda_R)^2 \int_V d^2z_1 \int_V d^2z_2 \ln \frac{|x-z_1|}{a_0} \ln \frac{|z_2-y|}{a_0} \left[\left(\frac{|z_1-z_2|}{a_0} \right)^{2\beta q^2} - \left(\frac{|z_1-z_2|}{a_0} \right)^{-2\beta q^2} - 1 \right]. \quad (7.2)$$

The parameter, a_0 , has been left in and naive Feynman rules have been used in calculating Eq. (7.2). This is because the Green's functions for the fields $\chi(x)$ are ill defined. The interesting and well defined operators are the $\phi_q(x) = e^{i\beta q \chi(x)}$. Eq. (7.2) converges for $\beta q^2 < 1$ and diverges for $\beta q^2 \geq 1$ [the same would be true if one calculated the Green's functions for the $\phi_q(x)$]. Note that any individual graph of Figure 4 or Figure 7 is convergent. Thus one has a situation ($\beta q^2 \geq 1$) where to any finite order (in βq^2) graphs have no ultraviolet divergences yet when perturbation theory is summed to all orders ultraviolet divergences appear, implying that non-perturbative renormalization methods are necessary. Of course, one can look at \mathcal{G} (Eq. [4.2]) directly to see that there are ultraviolet divergences for $\beta q^2 \geq 1$.

This section will consist of showing that to all orders in λ_R (as well as βq^2) \mathcal{G} is well defined when $\beta q^2 < 1$. I take this to be a proof that the 0(2) non-linear σ -model and the sine-Gordon field theory are renormalizable to all order for $\beta q^2 < 1$.²³ For $\beta q^2 \geq 1$ a cutoff must be introduced. Whether it may be removed by

wave function and coupling constant renormalizations is unknown. If additional interaction counter-terms must be added to the theory, then the sine-Gordon lagrangian would have to be modified and the equation of motion would be an inaccurate representation. An implication of this: the soliton-antisoliton doublets would not exist. I suspect that renormalization should be possible at least for $1 < \beta q^2 < 2$ since lattice methods¹⁷ have shown the existence of these doublets. It is still uncertain how to do this in the continuum field theory, although the equivalence of the non-linear σ -model and the sine-Gordon offers a possibility: since g has dimensions of $(\text{mass})^2$ and the linear σ -model (Eqs. [6.1] and [6.2]) is renormalizable, g acts as a cutoff for the sine-Gordon theory. If one can show the relevant quantities are g -independent (or g dependence can be absorbed into λ_0) for large g then this would provide the method of renormalization. The same is true in dimensions three and four where the equivalence between the two models still holds. Thus there is the possibility that using the linear σ -model one can renormalize the sine-Gordon lagrangian in three and four dimensions.

I will now present strong evidence that the partition function

$$\begin{aligned}
 \mathcal{Z}(\beta q^2, \lambda_R) = & \\
 & \sum_N \frac{\lambda_R^{2N}}{N!N!} \int d^2x_1 \cdots d^2x_N \int d^2y_1 \cdots d^2y_N \exp \left[2\beta q^2 \right. \\
 & \left. \times \left(\sum_{i < j} [\ln|x_i - x_j| + \ln|y_i - y_j|] - \sum_{i,j} \ln|x_i - y_j| \right) \right] \quad (7.3)
 \end{aligned}$$

converges for $\beta q^2 < 1$. λ_R , of course, is $e^{-\beta \text{self-energy}}$ absorbed in λ_j . The method is not intended to be rigorous: physical arguments are used to approximate \mathcal{Z} . Eq. (7.3) contains only neutral configurations because of the infrared singularity of the theory as discussed in Section IV. The arguments of this section also apply to the situation where neutrality is not required.

\mathcal{Z} acquires a big contribution whenever an x_i approaches a y_j and βq^2 is near 1. The nature of the singularity is governed by

$$\text{VI}(\beta q^2, \epsilon) = \int d^2x \int d^2y e^{-2\beta q^2 \ln|x-y|} \approx v \frac{\pi \epsilon^{2-2\beta q^2}}{1-\beta q^2} \quad (7.4)$$

$|x-y| < \epsilon$

Eq. (7.4) is the contribution to \mathcal{Z} when a plus and minus are within ϵ of each other (boundary effects being neglected). If ϵ is sufficiently small the plus-minus dipole will look like a neutral object and will interact very weakly with other charges and dipoles, even

when these objects approach the dipole. ϵ is just a small parameter. If a plus and minus are within ϵ of each other then one says a dipole is present in the system. Since there will be a mean density, ρ , for the plasma and in this dissociated phase charges are randomly located, the average distance between charges is roughly $\sqrt{\frac{1}{\rho}}$. One can take ϵ to be a fraction of this distance, say $\epsilon = 0.1\sqrt{\frac{1}{\rho}}$.

Consider the term in Z_N with N plus charges and N minus charges (i.e. Z_N)

$$Z_N = \frac{1}{N!N!} \int d^2x_1 \dots d^2x_N \int d^2y_1 \dots d^2y_N e^{-\beta U_N(x_i, y_j)} \quad (7.5)$$

and single out the contribution due to dipoles

$$Z_N = \frac{1}{N!N!} \left\{ \int d^2x_1 \dots d^2x_N \int d^2y_1 \dots d^2y_N e^{-\beta U_N} \right. \\ \left. |x_i - y_j| < \epsilon \text{ for some } i, j \right. \\ + \int d^2x_1 \dots d^2x_N \int d^2y_1 \dots d^2y_N e^{-\beta U_N} + \dots + \\ \left. |x_{i_1} - y_{j_1}| < \epsilon \right. \\ \left. |x_{i_2} - y_{j_2}| < \epsilon \text{ for some } i_1, j_1, i_2, j_2 \right.$$

$$\begin{aligned}
& + \int d^2x_1 \cdots d^2x_N \int d^2y_1 \cdots d^2y_N e^{-\beta U_N} \\
& \quad \left. \begin{array}{l} |x_{i_1} - y_{j_1}| < \epsilon \quad \text{for some ordering of} \\ |x_{i_2} - y_{j_2}| < \epsilon \quad \text{the } x\text{'s and } y\text{'s.} \\ \vdots \\ |x_{i_N} - y_{j_N}| < \epsilon \end{array} \right\} \\
& + \left. \int d^2x_1 \cdots d^2x_N \int d^2y_1 \cdots d^2y_N e^{-\beta U_N} \right\} \quad (7.6) \\
& \quad \text{no dipoles}
\end{aligned}$$

The first N terms in Eq. (7.6) have precisely $1, 2, \dots$, respectively N dipoles. The last term is the no dipole term where no x_i is within ϵ of any y_j . Because dipoles interact weakly they may be factored out of the summands of Eq. (7.6):

$$Z_N \approx \sum_{\ell=0}^N \frac{V^\ell [I(\beta q^2, \epsilon)]^\ell Z_{N-\ell}^{\text{No Dipole}}(\epsilon)}{\ell!}, \quad (7.7)$$

where

$$\begin{aligned}
Z_n^{\text{No Dipole}} = & \\
& \frac{1}{n!} \frac{1}{n!} \int d^2x_1 \cdots d^2x_n \int d^2y_1 \cdots d^2y_n e^{-\beta U_n(x_i, y_j)} \\
& \quad \text{no } x_i \text{ within } \epsilon \text{ of } y_j \quad (7.8)
\end{aligned}$$

The combinatorial factor is accounted for as follows: There are $\frac{[N(N-1)\dots(N-\ell+1)]^2}{\ell!}$ ways of pairing up ℓ pluses with ℓ minuses from a set of N pluses and N minuses. This number times $\frac{1}{N!} \frac{1}{N!}$ gives the $\frac{1}{\ell!(N-\ell)! \ell!(N-\ell)!}$ of Eqs. (7.7) and (7.8). Eqs. (7.3), (7.6), and (7.7) yield

$$Z \approx e^{VI(\beta q^2, \epsilon) \lambda_R^2} Z^{\text{No Dipole}}(\epsilon). \quad (7.9)$$

As long as $\beta q^2 < 1$, $I(\beta q^2, \epsilon)$ is small as one sees from Eq. (7.4). It remains to show $Z^{\text{No Dipole}}(\epsilon)$ does not diverge. Heuristically the reason for this is as follows: $Z^{\text{No Dipole}}$ vanishes if a plus approaches a plus or a minus approaches a minus and hence the charges must be evenly distributed when N becomes large. Consider a minus charge. Since the plasma is neutral it will see a charge distribution of $+1$. Neglecting boundary effects, one can lump this $+1$ charge distribution at some effective distance $r_{\text{eff}}(N, \epsilon)$. For ϵ small enough r_{eff} will be independent of ϵ and for N large enough it will be a slowly varying function of N . For simplicity take r_{eff} to be a constant as N goes to infinity. Then for large N

$$Z_N^{\text{No Dipole}} \sim \frac{1}{N!} \frac{1}{N!} \left(r_{\text{eff}}^{-2\beta q^2} \right)^{2N} v^{2N} \quad (7.10)$$

and this implies $Z^{\text{No Dipole}} = \sum_{n=0}^{\infty} \lambda_R^{2N} Z_N^{\text{No Dipole}}$ converges.

To make the above argument more precise let N be large. Break V into $2N$ square cells of volume $\frac{V}{2N}$. The length of the side of a

cell is $\sqrt{\frac{V}{2N}} = d(N)$. Approximate $Z_N^{\text{No Dipole}}$ by summing over all ways of placing the plus and minus charges into the $2N$ cells:

$$Z_N^{\text{No Dipole}} \approx \sum_C \frac{1}{N!N!} \left[\left(\sqrt{\frac{V}{2N}} \right)^2 \right]^{2N} e^{-8U(C)}. \quad (7.11)$$

$\left(\sqrt{\frac{V}{2N}} \right)^2$ is the area in which a charge is allowed to roam, C is

a placement of the pluses and minuses into cells, and $U(C)$ is the energy of such a configuration (calculated with the charges at the center of the cells). The minimum energy configuration by symmetry occurs when the plus and minus charges alternate as in Figure 8. To approximate the energy of this configuration, pick a charge. It has four nearest opposite charges (Figure 9a) contributing a factor of $(d^4)^{-2\beta q^2}$ and four nearest like charges (Figure 9b) contributing a factor of $[(\sqrt{2}d)^4]^{2\beta q^2}$. In the next row (Figure 9c) there are eight opposite charges and eight like charges giving a factor of $[(2d)^4]^{2\beta q^2} [(\sqrt{5}d)^8]^{-2\beta q^2} [(\sqrt{8}d)^4]^{2\beta q^2}$. To this order $e^{-8U(\text{single charge})} = [4 \left(\frac{1}{5}\right)^2 \left(\frac{8}{5}\right)^2]^{2\beta q^2}$. Taking into consideration all rows and neglecting edge effects

$$\begin{aligned} f(N) &= e^{-8U(\text{single charge})} \\ &\approx \prod_{n=0}^{\sqrt{N}} \prod_{m=1}^{\sqrt{N}} \left(\sqrt{\frac{V}{n^2+m^2}} \right)^{4(-1)^{m+n} 2\beta q^2} \\ &= \exp \left\{ \sum_{n=0}^{\sqrt{N}} \sum_{m=1}^{\sqrt{N}} (-1)^{n+m} 4\beta q^2 \ln(n^2+m^2) \right\}. \end{aligned} \quad (7.12)$$

Because the signs alternate f will be a slowly varying function of N (for example the third row multiplies the result of the first two by only $[(\frac{10}{9})^2(\frac{10}{14})^2(\frac{18}{14})^2]^{2\beta q^2} \approx (1.04)^{2\beta q^2}$). The total contribution

to $e^{-\beta U(C_{\min})}$ is $[f(N)^{2N}]^{1/2}$ ($2N$ for each particle and a $1/2$ for double counting). One can bound $Z_N^{\text{No Dipole}}$ by replacing $e^{-\beta U}$ by its maximum value $f(N)^N$. Since there are $(2N)!$ ways of putting the charges into the cells,

$$Z_N^{\text{No Dipole}} < \frac{(2N)!}{N!N!} [f(N)]^N \left(\frac{V}{2N}\right)^{2N}$$

$$\sim \frac{V^{2N} f(N)^N \sqrt{4\pi N} e^{-2N}}{N!N!}, \quad (7.13)$$

where Stirling's approximation has been used in the last step. Since $f(N)$ (Eq. [7.12]) is a slowly varying function of N , $Z_N^{\text{No Dipole}}$ is highly attenuate for N large and $\mathcal{Z}^{\text{No Dipole}}$ converges.

It has now been shown that $\mathcal{Z}(\lambda_R, \beta q^2)$ is finite for $\beta q^2 < 1$. \mathcal{Z} represents the sum of all vacuum bubble diagrams (in a finite volume limit). Since this sum converges this indicates that other functions (such as Green's functions of the relevant operators) will have no divergences. It is folklore that the vacuum bubbles represent the most ultraviolet divergent graphs.

The main result of this section is that the prescription of absorbing the self energies into the activity renormalizes the sine-Gordon to all orders when $\beta q^2 < 1$.

For the statistical analogue field theories discussed in this

paper, renormalization may be regarded as the removal of smearing functions. Consider the sine-Gordon. When \mathcal{Z} is smeared by an appropriate f (Eq. [2.19]) a well defined non-trivial partition function, $\mathcal{Z}(f)$ is obtained (well defined in the sense that no infinities occur and non-trivial in the sense that $\mathcal{Z}(f)$ is not 1 or $e^{2\lambda_0 V}$, which is the ideal gas grand partition function). $\mathcal{Z}(f)$ is, however, non-relativistic. One would like to take the limit $f(x) \rightarrow \delta^d(x)$ so as to recover Poincaré invariance. Doing this naively causes $\mathcal{Z}(f)$ to go to one because the $N \neq 0$ terms give zero due to the infinite self-energy. The way to avoid this problem is to let λ_0 , the bare activity, depend on f : $\lambda_0 \equiv \lambda_0(f)$. \mathcal{Z} becomes

$$\mathcal{Z}(f) = \frac{1}{N} \iint \mathcal{D}\chi e^{-\beta \int \frac{\nabla\chi \cdot \nabla\chi}{8\pi} + 2\lambda_0(f) \int_V \cos\beta q(\chi * f)} \quad (7.14)$$

Eq. (7.14) gives an upper bound on \mathcal{Z} of $e^{2\lambda_0(f)V}$ since the functional $\int_V \cos\beta q\chi * f \leq V$. \mathcal{Z} has a lower bound of 1 since the partition function is a sum of positive terms beginning with 1. $\lambda_0(f)$ is adjusted so that $\lim_{f \rightarrow \delta^d} \mathcal{Z}(f) \equiv \mathcal{Z}$ is non-trivial and any such limit should produce a good theory. The reader has seen one example of this, the sine-Gordon in two dimensions in the region $\beta q^2 < 1$. If one takes

$$\lambda_0(f) = C e^{-\beta q^2 \int f(r) \ln|r-r'| f(r') d^2r d^2r'} \quad (7.15)$$

where C is arbitrary (from $\lambda_R = \lambda_0 e^{-\beta \text{self-energy}}$ one identifies C with λ_R), then the limit $f(x) \rightarrow \delta^2(x)$ produces a well defined \mathcal{Z} . $\lambda_0(f)$ goes to infinity in such a way as to keep λ_R finite and prevent \mathcal{Z} from going to 1. In the region $\beta q^2 \geq 1$, the $\lambda_0(f)$, defined by Eq. (7.15), would cause $\mathcal{Z}(f)$ to become infinite as $f \rightarrow \delta^2$ and in this new region one must not let $\lambda_0(f)$ go to infinity as fast as in the $\beta q^2 < 1$ region. Possibly a sequence of $\lambda_0(f)$'s can be found which affects a cancellation between self-energy and interparticle interaction infinities.

Green's functions (and other relevant objects) must always be calculated using this limiting procedure. Consider

$$\begin{aligned} G(x,y) &= \langle e^{i\beta q\chi(x)} e^{-i\beta q\chi(y)} \rangle \\ &= \frac{1}{N} \iint \mathcal{D}\chi^a e^{-\beta \int \frac{\nabla\chi \cdot \nabla\chi}{8\pi} + 2\lambda_0 \int_V \cos\beta q\chi} \\ &\quad \times e^{i\beta q\chi(x)} e^{-i\beta q\chi(y)}. \end{aligned} \quad (7.16)$$

Even after smearing ($\int \cos \beta q \chi \star f$) and using $\lambda_0(f)$ of Eq. (7.15), this plus-minus Green's function is zero due to the ultraviolet infinite self-energy produced by $e^{i\beta q \chi(x)}$ and $e^{-i\beta q \chi(x)}$. The correct way of calculating is to replace Eq. (7.15) by

$$G(x,y) = \frac{1}{N} \iint \mathcal{D}\chi e^{-\beta \int \frac{\nabla\chi \cdot \nabla\chi}{8\pi} + 2\lambda_0(f) \int_V \cos \beta q(\chi \star f)} \times Z^2(f) e^{i\beta q(\chi \star f)(x)} e^{-i\beta q(\chi \star f)(y)}, \quad (7.17)$$

where $Z(f)$ is a wave function renormalization constant. From physical principles one knows $Z(f)$ must be proportional to $\exp\{-\beta q^2 \int f(r) \ln|r-r'| f(r') d^2r d^2r'\}$ since $G(x,y)$ has a physical interpretation: $G(x,y)$ is the partition function for a neutral Coulomb gas with a plus charge at x and a minus charge at y .

Non-renormalizability can be viewed as follows: As $f \rightarrow \delta^d$ dipoles, triatomic molecules, molecular rings, and other polyatomic structures will begin to form. One must introduce renormalized activities for each of these structures. If $f \rightarrow \delta^d$ is too singular a limit to take, an infinite number of polyatomic objects will form causing one to introduce an infinite number of renormalized activities.

This infinite set is reminiscent of what happens with polynomial field theories such as χ^6 in four dimensions. Here one is forced to introduce an infinite set of counterterms,

$$\sum_{n=1}^{\infty} (\delta g_{2n}) \chi^{2n}.$$

The δg_{2n} couplings correspond to the unrenormalized activities.

In the high temperature limit of the sine-Gordon $V(\chi) \approx 2\lambda_R$

$$\left[-1 + (4\pi\beta q^2) \frac{\chi^2}{2} - (4\pi\beta q^2)^2 \frac{\chi^4}{4!} + (4\pi\beta q^2)^3 \frac{\chi^6}{6!} \right]$$

which is a polynomial field theory with a χ^6 leading term. The high temperature sine-Gordon being similar to this polynomial potential implies the formation of polyatomic structures in the χ^6 theory. Thus one suspects that the cause for non-renormalizability for the statistical analogue field theories is the same as in non-renormalizable polynomial field theories.

VIII. TIDBITS

The correspondence between sine-Gordon and Coulomb gas indicates that many effects are completely missed in the naive treatment of the theory. Most important is the structure of the vacuum. When λ_R is very small perturbation theory is valid. The vacuum looks like a "vacuum" since few charges are present. When λ_R gets bigger, the vacuum is full of charges and the perturbation theory vacuum is a poor approximation to the real vacuum which contains many plus and minus ions. Also missed in perturbation theory is the phase transition at $\beta q''$ near 2. When $\beta q''$ is small the vacuum (and hence the entire theory) is radically different from when $\beta q''$ is large. The importance of the nature of the vacuum is neglected in most treatments of field theory. It will be a complicated vacuum structure which will lead to quark confinement, asymptotic freedom, and the hadron spectrum. A vacuum consisting of a gas of "quanta" would be compatible with asymptotic freedom and quark confinement. When two quarks are placed at small separation distance usually no vacuum quanta will be between them. The physical vacuum has little effect on these two quarks. Hence small distance behavior would be governed by bare vacuum and free interactions. At large separation distance the quanta would interpolate between the quarks and strange effects could occur. For example, a condensation into another phase might take place in the region between quarks especially if the gas is near a phase transition point. Such a condensation could provide a confining potential. In the two dimensional sine-Gordon in the

plasma phase (βq^2 small) one has asymptotic freedom. $G(x,y)$ (Eq. [7.16]) goes like $|x-y|^{-2\beta q^2}$ for small $|x-y|$. This is the same as in a free field theory (one would sum the diagrams of Figure 4 with the two vertices labelled by x and y). However, in this same phase there is charge screening for large distances, and hence the opposite effect one wants in confinement: largely separated charges have no interaction with each other, they merely interact individually with the vacuum. In the dipole phase, different effects arise. If one places a number of widely separated charges into the vacuum, the vacuum will immediately produce the opposite charges to form dipoles at the cost of βq per charge. On this physical basis one can write

$$\langle e^{i\beta q x(p_1)} \dots e^{i\beta q x(p_n)} \rangle \sim \lambda_H^n, \quad (8.1)$$

where $\int p_1(x) d^2x=1$, p_1 peaked about x_1 and all x_1 widely spaced. Because charges are immediately turned into dipoles the Green's function $G(x,y)$ (Eq. [7.16]) is constant for large $|x-y|$, and the two charges will not be confined.

Now consider the situation when fractional charges (say a $+1/2$ and a $-1/2$) are placed in the vacuum. These fractional charges cannot form dipoles because the vacuum quanta are integral. The fractional charges are not screened. The interaction between the $+1/2$ and the $-1/2$ will be essentially a logarithm mitigated by dipole effects. The dipole strength (which is determined by renormalization) will govern how much the dipoles influence the interaction

of the fractional charges. For weak dipoles, one expects the logarithmic potential to remain intact. One can write

$$\left\langle e^{i\beta \frac{q}{2} \chi(x)} e^{-i\beta \frac{q}{2} \chi(y)} \right\rangle \sim |x-y|^{-2\beta q^2(|x-y|)}, \quad (8.2)$$

with $q(r)$ an effective charge satisfying $q(r) \rightarrow 1/2$ for $r \rightarrow 0$ (asymptotic freedom) and $q(r)$ slowly varying for large r . This type of charge screening of the Green's function is reminiscent of a similar phenomenon¹¹ found in the massive Schwinger model. In general one has

$$\left\langle e^{i\beta(N+f)q\chi(x)} e^{-i\beta(N+f)q\chi(y)} \right\rangle \sim \lambda_R^{2N} |x-y|^{-2\beta f^2 q^2(|x-y|)}, \quad (8.3)$$

with $0 \leq f < 1$, $f q$ being the fractional excess charge. For $f=0$ the Green's functions are roughly constant. One can, of course, introduce triality operators

$$\phi_{q/3}(x) = e^{i\beta \frac{q}{3} \chi(x)} \quad \text{and} \quad \phi_{q/3}^*(x) = e^{-i\beta \frac{q}{3} \chi(x)} = \phi_{-q/3}(x). \quad \text{The Green's}$$

functions for these operators vanish unless the number of $\phi_{q/3}$'s minus the number of $\phi_{q/3}^*$'s is three times an integer. $\phi_{q/3} \phi_{q/3}^*$ configurations and $\phi_{q/3} \phi_{q/3} \phi_{q/3}^*$ configurations exist. This resembles the triality of the quark model. Of course, there also are n -ality

operators $\phi_{q/n}^{(2)} = e^{i\beta \frac{q}{n} \chi(x)}$ and $\phi_{q/n}^*(x) = \phi_{-q/n}(x)$. It is the nature of the vacuum that determines these unusual effects.

Fractional quanta have already been used as a possible quark confinement mechanism in four dimensions. The model in mind is the meron gas of Callan, Dashen, and Gross,²⁴ where the charge in the theory is not the usual charge but the topological charge, and the particles having fractional 1/2 charge are the merons. The use of fractional quanta is not as unnatural as one might think.

Another interesting effect is that the relevant operators are of the form $\phi_\rho = e^{i\beta \chi(\rho)}$ ($\chi[\rho] = \int \chi(x) \rho(x) d^2x$) since these operators produce charge distributions, ρ , in the Coulomb analogue model. These operators are precisely the ones Coleman used to show the equivalence between the sine-Gordon and the massive Thirring. Being use to perturbation theory one usually works with the bare vacuum and approximates the interacting fields by free fields in which case the interesting Green's functions are $\langle \chi(x_1) \dots \chi(x_n) \rangle$. The sine-Gordon theory shows that such simplistic vacuum expectation values may not be the interesting ones. In fact there is no reason why, in a particular theory, the relevant operators are not complicated functions of the fields. This conclusion may be applicable to gauge theories.

A third unusual effect is due to the infrared divergences in the sine-Gordon two dimensional theory. The normal Feynman rules are invalid. For example, one would conclude from $\mathcal{L}_1 = 2\lambda_0 \cos \beta q \chi$ that bubble diagrams of order λ_0^3 (such as in Figure 10) contribute

to the vacuum energy. This is incorrect since it violates charge neutrality. Renormalization procedures would be upset if non-neutrality is not maintained since the use of $\ln \frac{|x-y|}{a_0}$ for the propagator would make the theory depend on a_0 , which it should not. The modification of naive perturbation theory rules by infrared divergences may affect four dimensional theories such as the popular gauge theories. If such an effect occurs the usual Feynman rules are wrong and may upset the renormalization of infrared singularities.

Fourthly, it is curious that to every finite order (in βq^2) in perturbation theory there are no ultraviolet singularities; yet when all orders are summed an ultraviolet divergence arises when $\beta q^2 \geq 1$. The reader has already seen this in the bubble diagrams of Figure 4. They have no ultraviolet singularities to any finite order; yet when summed they are $\sim \lambda_R^2 \int r^{-2\beta q^2} d^2 r$ which diverges for $\beta q^2 \geq 1$.

One unanswered question still remains: since the sine-Gordon equations has soliton solutions, what do these solitons mean in the Coulomb gas context? Mandelstam¹⁶ has constructed the soliton operators. They are of the form

$$\begin{aligned} \psi_{1,2}(x,t) &= C_{1,2} \exp \left\{ -2\pi i \beta_M^{-1} \int_{-\infty}^x d\tau \Pi(\tau,t) \mp \frac{1}{2} \beta_M \phi(x,t) \right\} \\ \psi_{1,2}^*(x,t) &= C_{1,2}^* \exp \left\{ 2\pi i \beta_M^{-1} \int_{-\infty}^x d\tau \Pi(\tau,t) \pm \frac{1}{2} \beta_M \phi(x,t) \right\}. \end{aligned} \quad (8.4)$$

Using $\beta_H = \sqrt{4\pi\beta} q$, rescaling $\phi = \sqrt{\frac{\beta}{4\pi}} x$ and rotating to Euclidean space

$$\begin{aligned} \psi_{1,2}(x,t) &= C_{1,2} \exp\left\{ + \frac{1}{2q} \int_{-\infty}^x d\zeta \dot{\chi}(\zeta,t) \mp i\beta \frac{q}{2} x(x,t) \right\} \\ \psi_{1,2}^*(x,t) &= C_{1,2}^* \exp\left\{ - \frac{1}{2q} \int_{-\infty}^x d\zeta \dot{\chi}(\zeta,t) \pm i\beta \frac{q}{2} x(x,t) \right\}. \end{aligned} \quad (8.5)$$

The operator $e^{i\beta\vec{p}\cdot\nabla\chi(x)}$ produces a dipole of strength p since

$e^{i\beta\vec{p}\cdot\nabla\chi(x)} = \exp\left\{ i\beta \frac{p}{\alpha} \left[\chi\left(x + \frac{\vec{a}}{2}\right) - \chi\left(x - \frac{\vec{a}}{2}\right) \right] \right\}$ produces a charge $\frac{p}{\alpha}$ at $x + \frac{\vec{a}}{2}$ and a charge $-\frac{p}{\alpha}$ at $x - \frac{\vec{a}}{2}$ (where $\vec{p} = p\vec{a}$). $\psi_1^*(x)$ has the interpretation of producing a charge $+\frac{q}{2}$ at x and an imaginary dipole string. The dipoles point in the t -direction and have a strength density of $\frac{1}{2q\beta}$. The string is along the x -direction and ends at x . The imaginary dipole string makes little sense from the Coulomb point of view. Because of this complicated interpretation, the Coulomb gas analogy does not seem useful.

IX. SUMMARY

Here is a list of the main results:

A. (Section II) Certain field theories are equivalent to gases of interacting particles, in particular

1. The sine-Gordon corresponds to a neutral Coulomb gas.
2. The "massive" sine-Gordon and the massive Schwinger model at zero Coleman angle correspond to a gas of quanta interacting via Yukawa potentials.

B. (Section III) The Feynman diagrams for these theories have a statistical mechanical interpretation. The correspondence is outlined in Table 1.

C. (Section V) The vacuum of the two dimensional sine-Gordon undergoes a phase transition at βq^2 near 2. For $\beta q^2 < 2$ there is a plasma phase and for $\beta q^2 > 2$ there is a dipole gas phase. $\beta q^2 = 2$ is precisely the value Coleman¹³ finds a vacuum instability.

D. (Section VI) The two dimensional non-linear $O(2)$ σ -model with linear symmetry breaking term is equivalent to the sine-Gordon theory.

Results C and D imply

1. The σ -model undergoes a phase transition for $f_{\parallel} \approx \sqrt{\frac{1}{8\pi}}$.
2. The σ -model contains solitons and fermions.

- E. (Section VII) When $\beta q^2 < 1$ the sine-Gordon (massive Thirring and $O(2)$ σ -model) are renormalizable to all orders, i.e. they are well defined theories.²⁵
- F. (Section VIII) The dipole phase of the sine-Gordon completely shields integral charges but is unable to do so for fractional charges.
- G. The relevant operators for the sine-Gordon are not simply polynomials in the fields. They are $\phi_q(x) = e^{i\beta q\chi(x)}$ and have the simple physical interpretation of producing a charge, q , at x . They are the operators used by Coleman to prove the equivalence of the sine-Gordon and massive Thirring models.
- H. (Section VIII) Operators exhibiting the quark-like triality condition are $\phi_{q/3}(x)$. They have this property because of the infrared singular nature of the sine-Gordon field theory.
- I. (Section IV and VIII) Naive Feynman rules may be incorrect when infrared singularities occur. The sine-Gordon exhibits such a property.
- J. (Section VII and VIII) The sine-Gordon has no ultraviolet singularities to every finiter order of perturbation theory; yet when diagrams are summed an ultraviolet divergence appears.

K. Callan, Dashen, and Gross²⁶ have recently shown that the instanton approximation to 2-d charged scalar electrodynamics with massless fermions is equivalent to a neutral Coulomb gas. The instantons are Nielsen-Olesen vortices. The effect of the massless fermions is to raise the inverse temperature, βq^2 , from 0 to N , the number of fermions. Since the Coulomb interaction is mitigated at short distances their model has a natural renormalization. Their result together with Result I of this paper implies their model will possibly have (sine-Gordon) solitons. The existence of such solitons depends on how much the Coulomb force is modified and how good the instanton approximation is.

There is a good chance that a theory of strong interactions in four dimensions will have many of the properties exhibited by the two dimensional sine-Gordon theory. It is conceivable that the hadron vacuum has a complicated structure which must be treated using statistical mechanics. Strange effects can occur when such a vacuum has "lot's of quanta" in it. It would be able to support asymptotic freedom because at short distances the quanta are ineffective and at the same time it could provide confinement since at large distances the many body effects of such quanta can be unusual. It is conceivable that string like-structures or other types of extended objects could condense out of the vacuum when other quanta such as quarks are introduced. Further unusual effects created by strong interaction vacuum may provide for the triality condition now observed.

* Research sponsored by the National Science Foundation, Grant #PHY77-23512.

1. See, for example, R. Brout, Phys. Reports 10C, 1 (1974).
2. For a review see F. W. Wiegel, Phys. Reports 16C, 57 (1975).
3. S. Albeverio and R. Høegh-Krohn, Comm. Math. Phys. 30, 171 (1973); J. Fröhlich, Phys. Rev. Lett. 34, 833 (1975); J. Fröhlich, Renormalization Theory, Edited by G. Velo and A. S. Wightman, (D. Reidel Publishing Co., 1976).
4. J. Fröhlich and E. Seiler, Helvetica Physica Acta 49, 889 (1976); J. Fröhlich and Y. M. Park, Helvetica Physica Acta 50, 315 (1977).
5. J. Fröhlich, Comm. Math. Phys. 47, 233 (1976).
6. A. M. Polyakov, Nucl. Phys. B120, 429 (1977).
7. See, for example, J.D. Jackson, Classical Electrodynamics, (John Wiley and Sons, Inc., 1975) or L. D. Landau and E. M. Lifshitz, Statistical Physics, (Pergamon Press, 1970).
8. This is due to boson-fermion correspondence in two dimensions. For a summary see M. Bander, Phys. Rev. D13, 1566 (1976) and the references therein.
9. For two dimensions (which will be needed later) the propagator is $\frac{2}{\beta} \ln \frac{|r|}{a_0}$ where a_0 is an arbitrary scale.
10. See, for example, A. Isihara, Statistical Physics, (Academic Press, 1971).
11. The Schwinger model (both massive and massless) has many interesting properties. See J. Lowenstein and A. Swieca, Ann. Phys. 68, 172 (1971); A. Casher, J. Kogut, and L. Susskind, Phys. Rev. D10, 732 (1974);

- S. Coleman, R. Jackiw, and L. Susskind, *Ann. Phys.* 93, 267 (1975).
For a review of this last work see L. Susskind and J. Kogut, *Phys. Reports* 23C, 348 (1976).
12. There is an enormous amount of literature on the sine-Gordon theory. A good but limited set of references on the field theory aspects is contained in Fröhlich's Erice lectures (reference 3) while a good but again limited set of references to the classical and solitons aspects is contained in reference 18.
 13. S. Coleman, *Phys. Rev.* D11, 2088 (1975).
 14. M. B. Halpern and P. Senjanovic, *Phys. Rev.* D15, 3629 (1977).
 15. E. H. Hauge and P. C. Hemmer, *Physica Norvegica* 5, 209 (1971);
J. M. Kosterlitz and D. J. Thouless, *J. Phys.* C6, 1181 (1973);
J. M. Kosterlitz, *J. Phys.* C7, 1046 (1974); C. Deutsch and M. Lavard, *Phys. Rev.* A9, 2598 (1974).
 16. S. Mandelstam, *Phys. Rev.* D11, 3026 (1975).
 17. A. Luther, *Phys. Rev.* B14, 2153 (1976). Mass ratio's were first obtained using semiclassical techniques. See R. F. Dashen, B. Hasslacher, and A. Neveu, *Phys. Rev.* D11, 3424 (1975). A review of this work is presented by R. Rajaraman, *Phys. Reports* 21C, 227 (1975).
 18. A. C. Scott, F. Y. F. Chu, and D. W. McLaughlin, *Proc. IEEE*, Vol. 61, 1443 (1973).
 19. I would like to thank K. Bardakci for this point.
 20. I thank K. Bardakci for showing me this.
 21. E. Brézin and J. Zinn-Justin, *Phys. Rev. Lett.* 36, 691 (1976);
Phys. Rev. B14, 3110 (1976).

22. W. A. Bardeen, B. W. Lee, R. E. Shrock, Phys. Rev. D14, 985 (1976).
23. For similar results on the convergence of this theory see J. Fröhlich's work in references 3, 4, and 5.
24. C. G. Callan, R. Dashen, and D. J. Gross, Phys. Lett. 66B, 375 (1977).
25. J. Fröhlich has proved that in such a case ϕ is a rigorously defined Wightman field. See references 3 and 4.
26. C. G. Callan, R. Dashen, and D. J. Gross, Phys. Rev. D16, 2526 (1977).

Table 1: FEYNMAN GRAPH CORRESPONDENCE

| <u>Statistical Mechanics</u> | | <u>Field Theory</u> |
|---------------------------------|---|--|
| particles | ↔ | vertices |
| interactions | ↔ | propagators |
| λ_0, β | ↔ | coupling constants |
| Z_N | ↔ | N^{th} order diagrams |
| cluster expansion | ↔ | expansion in $2\lambda_R \int_V (\cos \beta q \chi - 1)$ |
| correlation functions | ↔ | Green's functions |
| a charge, q , produced at x | ↔ | the operator, $\phi_q(x) = e^{i\beta q \chi(x)}$ |

Table 2: NOTATION TRANSLATION

| <u>Reference</u> | <u>Inverse Temperature</u> | <u>Activity</u> | <u>Dimensional Factor</u> |
|------------------------------------|--|--|-------------------------------|
| This Paper | β | λ_o, λ_R | a_o |
| J. Fröhlich (Ref. 5) | $\beta_F = \beta$ | $Z_F = 8\lambda_R \pi \beta q^2$ $\epsilon_F = \sqrt{4\pi} q$ | — |
| Coleman (Ref. 13) | $\beta_C = \sqrt{4\pi\beta} q$ | $\alpha_o = 8\pi\lambda_o \beta q^2$ $\alpha = 8\pi\lambda_R \beta q^2$ | $\frac{1}{\sqrt{c} \mu}$ |
| Mandelstam (Ref 16) | $\beta_M = \sqrt{4\pi\beta} q$ | $\mu_M^2 = 8\pi\lambda_o \beta q^2$ | $\frac{1}{c\mu}$ |
| Kosterlitz & Thouless (Ref. 15) | $\beta_{K.T.} = \beta$ | $\lambda_{K.T.}$ | r_o |
| Hauge & Henner (Ref. 15) | $\beta_{H.H.} = \beta$ | — | $L_o = 1$ |
| Luther (Ref. 17) | $\theta_L = \frac{1}{2} \beta q^2$ $\mu_L = \pi(1 - \frac{1}{2} \beta q^2)$ $\gamma_L = 4\pi\beta q^2(1 - \frac{1}{2} \beta q^2)^{-1}$ | — | — |

NOTES TO TABLE 2

As previously shown, a correct definition of the energy makes the theory independent of a_0 even though the potential is $2\ln \frac{r}{a_0}$. Varying the dimensional factor redefines the self-energy (and hence the activity) at the cost of redefining interparticle energies. This is done in such a way that the total energy remains unchanged.

The massive Thirring model, $\mathcal{L} = \bar{\psi}(i\gamma - M)\psi - 1/2 g(\bar{\psi}\gamma^\mu\psi)(\bar{\psi}\gamma_\mu\psi)$ is known^{13,16} to be equivalent to the sine-Gordon: $g = \left(\frac{1}{8q} - 1\right)\pi$ and $2\lambda_0 \cos \sqrt{4\pi\beta} qx$ corresponds to $-M\bar{\psi}\psi$. M , a cutoff dependent quantity (like λ_0) depends on how one renormalizes and thus may be chosen arbitrarily.

The $O(2)$ non-linear σ -model with linear symmetry breaking term, $\mathcal{L} = \lim_{g \rightarrow \infty} \left\{ 1/2(\partial_\mu \sigma)(\partial^\mu \sigma) + 1/2(\partial_\mu \pi)(\partial^\mu \pi) + 2a_0 + g(\sigma^2 + \pi^2 - f_\pi^2) \right\}$, is also equivalent to the sine-Gordon:

$$f_\pi = \frac{1}{\sqrt{4\pi\beta} q} \quad \text{and} \quad a = \lambda_0 \sqrt{\pi\beta} q.$$

$\lambda_{K.T.} = e^{-2\beta\mu_{K.T.}}$. Kosterlitz and Thouless interpret $\mu_{K.T.}$ as the energy required to create a plus-minus pair at a distance of their cut-off, r_0 .

Hauge and Hemmer deal with a partition function so that the activity never appears.

In Luther's paper the coupling constant corresponding to the activity comes in via the inverse lattice spacing.

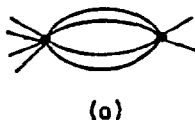


FIGURE 1. SOME FEYNMAN GRAPH COMBINATORIAL FACTORS. In (a) there are four lines connecting the two vertices. According to rule (e) there is a factor of $\frac{1}{4!}$. (b) is a three-fold self energy tadpole. According to rule (f) there is a factor of $\frac{1}{3!} \frac{1}{2^3}$.



FIGURE 2. THE EFFECT OF SELF-ENERGY TADPOLES. The bare vertex is replaced by a sum of terms, each one with an additional tadpole attached.

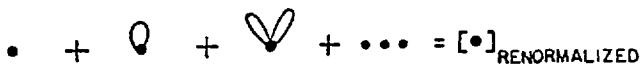


FIGURE 3. THE DIAGRAMS OF ORDER λ_0 .

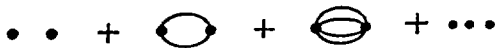


FIGURE 4. THE DIAGRAMS OF ORDER λ_R^2 . The self-energy tadpoles are dropped according to the modified (b) and (f) rules.

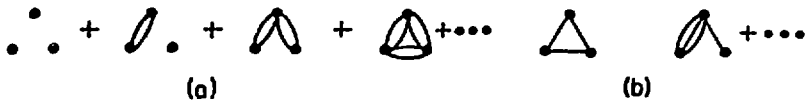


FIGURE 5. DIAGRAMS OF ORDER λ_R^3 . Diagrams of figure (a) have an even number of propagators between vertices, whereas (b) have only an odd number.

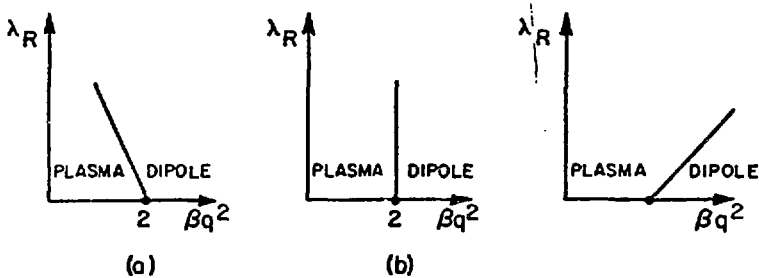


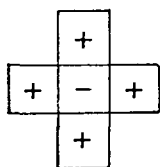
FIGURE 6. THREE POSSIBLE PHASE DIAGRAMS.



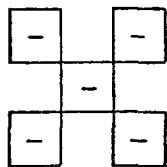
FIGURE 7. CONTRIBUTIONS TO THE TWO POINT FUNCTION. Although individual graphs are not divergent, the sum of these graphs gives a divergent contribution to the two-point function when $\beta q^2 \geq 1$.

| | | | | | |
|---|---|---|---|---|---|
| | | + | - | + | |
| | + | - | + | - | |
| + | - | + | - | + | - |
| - | + | - | + | - | + |
| | - | + | - | + | |
| | | - | | | |

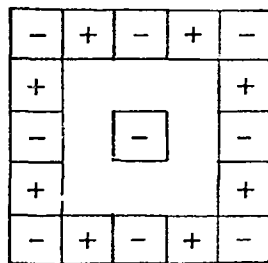
FIGURE 8. MINIMUM ENERGY CONFIGURATION.



(a)



(b)



(c)

FIGURE 9. LEADING CONTRIBUTIONS TO THE ENERGY OF A MINUS CHARGE.



FIGURE 10. AN INFRARED-FORBIDDEN DIAGRAM.

CHAPTER V

TOPOLOGICAL SYMMETRY BREAKDOWN
AND QUARK CONFINEMENT

I. INTRODUCTION

In a recent paper¹, 't Hooft has stressed the importance of the center of the group in non-abelian gauge theories. This has exposed new concepts and created new possibilities for quark confinement. The center of $SU(N)$ is Z_N , isomorphic to the set of integers, $0, 1, 2, \dots, N - 1$, under the operation of addition modulo N . Z_N , with its finite number of elements and unusual modulo addition property, is a feature distinguishing non-abelian theories from abelian ones. It may be the crucial factor explaining why non-abelian theories confine though abelian theories do not. Instantons also differentiate $SU(N)$ gauge theories. However, dilute instanton gases do not confine and dense instanton gases are hard to handle. The runaway scale and density problem has made instanton calculations virtually impossible to do. There are now conflicting views^{2,3} on their relevance to physical processes, problems, and confinement. 't Hooft Z_N type excitations, on the other hand, have only been discussed in a formal manner. The calculations remain to be done and it is unknown whether they will encounter similar difficulties. Thus, this is an important area of research. This is what we will be discussing in this paper. In particular, we will show how some calculations, such as Wilson loop integrals, can be done in a manner similar to instanton ones. We will also discuss many physically interesting ideas although no computations will be done to support them.

Unlike 't Hooft, who used a Hamiltonian approach, we shall use a Euclidean formulation. We find that the properties of Z_N excitations are particularly simple from this point of view,

especially when considering Wilson loops, where the effect of the excitations is expressed in terms of linking numbers.

Although 't Hooft had suggested attacking the problem by going to a lattice and many people have begun considering Z_N lattice gauge theories⁴, we shall work in the continuum.

't Hooft discussed, in detail, but in a formal way, the nature of confinement in $2 + 1$ dimensions. Here is a review of how it works. One starts with an $SU(N)$ gauge theory. Using a symmetry breaking Higgs potential, $SU(N)$ is broken down to Z_N . Topological solitons can then occur. These are not so different from Nielsen-Olesen vortices⁵ except that the non-zero gauge potentials are proportional to λ^8 (for example) in the $SU(3)$ case and more Higgs fields are involved. Away from a Nielsen-Olesen vortex, field configurations look like a gauge transformation, $U(x)$, with $U(x) = \exp(i\phi)$, so that $A_\mu = -\frac{i}{e} \left[\partial_\mu U(x) \right] U^{-1}(x)$ and $\phi(x) = U(x)F$ (see Sec. III for notation). Likewise, 't Hooft vortices are approximately singular gauge transformations, $U_{x_0}(x)$ (x_0 is the location of the soliton and x is the point where the gauge transformation is applied). When $U_{x_0}(x)$ is written as $\exp \left[i \sum_{i=1}^8 \alpha_i(x) \frac{\lambda_i}{2} \right]$, that is, the fundamental matrix representation is used, $U_{x_0}(x)$ has the property that when x encircles x_0 and returns to x , $U_{x_0}(x)$ is an element of the center of the group. Such a transformation is multivalued (and globally ill-defined), however when written in the adjoint representation (for example), it becomes single-valued. In 't Hooft's model, gauge and Higgs fields are in the adjoint representation so that $U_{x_0}(x)$ is well-defined.

If $\langle \phi_i(x) \rangle = F_i$ are a set of vacuum

expectation values which minimize the Higgs potential, then the classical configuration for a 't Hooft Z_N soliton at x_0 is $A_\nu = -\frac{i}{g} \left[\partial_\nu U_{x_0}(x) \right] U_{x_0}^{-1}(x)$ and $\phi_i(x) = U_{x_0}(x) F_i$ far from the vortex. Of course, nearby the field configuration is non-trivial just as in the Nielsen-Olesen case.

In $2+1$ dimensions the above solitons are particles with extended structure and non-zero form factors. They are stable for topological reasons and carry a topological conserved charge. This charge is governed by the group, Z_N , which means that charge is conserved modulo N so that, in principle, N charges may annihilate. This Z_N group is completely different from the one associated with the center of $SU(N)$ and should not be confused with it. There are now two Z_N 's. The one associated with the soliton charge will be called topological Z_N . 't Hooft argues that it may be possible that topological Z_N is spontaneously broken, a phenomenon we call topological symmetry breakdown. This is an interesting phase in which quark confinement occurs. The argument is as follows: If topological Z_N is broken then there are N different kinds of vacuums characterized by their topological Z_N numbers (compare this to spontaneous symmetry breakdown of a $U(1)$ symmetry by a Higgs potential, where, instead, there is a continuum (a circle) of vacuums defined by the direction in which the Higgs field points). In general, at any instant in time, the physical vacuum will look like a collection of domains each characterized by its Z_N value. Separating these domains will be Bloch walls. They carry an energy per unit length and may be associated with a new quanta in the theory: closed strings.

When particles in the fundamental representation are introduced, 't Hooft argues that they will be confined. A quark and an antiquark will have a Bloch-wall-like string between them. This will provide a linear confining potential. For $SU(3)$ three strings may join so that baryons can consist of three confined quarks. This is 't Hooft's $2 + 1$ dimensional quark confinement scheme. He derived it using simple, physical, intuitive arguments.

Several questions are generated. First is how does one extend these notions to $3 + 1$ dimensions. Following the same line of reasoning there will be volumes of Z_N vacua (instead of areas) and closed surfaces separating them (instead of closed strings). Clearly closed surfaces are unable to interpolate between quarks and create a linear potential as in one lower dimension. For this reason 't Hooft conjectured that confinement in $3 + 1$ is different. Instead of proposing a confinement scheme, he settled for an operator algebra which allowed the determination of the different phases of the theory. One of the important results of our paper will be the extension of the $2 + 1$ dimensional scheme to $3 + 1$ dimensions. The reason we are able to do this is that in a Euclidean formulation of the $2 + 1$ dimensional model we find a slightly different picture of the confinement, which has a straightforward generalization to one higher dimension.

A second question is how does one do calculations. 't Hooft has used formal powerful arguments, but it remains an open problem as to how to do computations. Of particular interest is the coefficient in front of the linear potential and its companion, the slope parameter. More generally, how does one calculate in a

theory with topological symmetry breakdown? We are able to supply some of the answers. We treat the Nielsen-Olesen case in Sec. III and the 't Hooft model in Sec. IV. To do these calculations requires a new calculation method. We develop it in Sec. II. The problem is equivalent to treating a gas of closed loops. Similar problems arise in lattice field theories. Examples are the three-dimensional $O(2)$ classical Heisenberg model and the four-dimensional abelian lattice gauge theory discussed by Banks, Myerson, and Kogut⁶. These authors must deal with a gas of monopole loops. Our gas consists of Nielsen-Olesen or 't Hooft vortex loops. We treat such a system by developing a field theory to describe it. Very simple arguments then tell us about the quark-antiquark and three quark potentials. Although our main interest is in topological symmetry breakdown and Wilson loop calculations, Sec. II discusses several field theory phenomena in this new picture. These concepts are enumerated and briefly discussed.

A third question is what are the essential ingredients in 't Hooft's $2 + 1$ confinement scheme? Certainly topological symmetry breakdown is one of them but are there others? We find the answer is yes. Topological symmetry breakdown leads only to a logarithmic quark-antiquark potential unless another ingredient is also present. It is the monopole. The 't Hooft model has monopoles in it. They play an instrumental role in the quark confinement. Before topological symmetry breakdown, the monopoles are bound together in monopole-antimonopole pairs. These dipoles have little effect. Topological symmetry breakdown liberates these monopoles. They, in turn, confine charges in a

manner not so different from Polyakov⁷ and Mandelstam⁸. In Sec. VII, we relate these ideas to Mandelstam's quark confinement scheme. The Z_N confinement in $3 + 1$ is practically the same as Mandelstam's. We consider this to be an important result: two seemingly different confinement schemes are, in fact, the same.

The remaining open problems and questions (and there are several) are presented at the end.

II. CLOSED LOOP GAS AS A FIELD THEORY

This section will relate a gas of closed continuous loops to a relativistic field theory⁹. Connections between statistical mechanics and field theory often prove useful¹⁰. We find this to be the case here and will use it to extract results in a physical and almost intuitive manner.

We shall proceed in steps. Note, first, however, that a continuous curve when broken into N segments, resembles a polymer with vertices acting as atoms and line segments acting as bonds (Fig. 1). Sometimes this analogy is useful. Consider an open macromolecule (or polymer) which goes from x_0 to x_N via x_1, x_2, \dots, x_{N-1} (Fig. 1b). To enforce the condition that the curve be continuous, we demand that the i^{th} atom be near its two neighbors. This can be done by requiring two neighboring atoms to be a distance ϵ from each other, that is, the bonds have a fixed length, ϵ . The total length is $N\epsilon$. Allowing curves of different lengths means summing over N in the partition function. To make the model more physical, assume the atoms have a chemical potential, μ , and interact with a "strength" g to an external potential, $V(x)$. The grand partition function for

a macromolecule with ends fixed at x_0 and $x_f \equiv x_N$ is

$$Z(x_0, x_f) = \sum_{N=1}^{\infty} \exp(-\beta\mu N) Z_N(x_0, x_f),$$

$$Z_N(x_0, x_f) = \left[\prod_{i=1}^{N-1} \int d^3x_i \right] \left[\prod_{i=1}^N \frac{\delta(|x_i - x_{i-1}| - \epsilon)}{4\pi\epsilon^2} \right] \quad (2.1)$$

$$\times \exp \left[-\beta \sum_{i=1}^N gV(x_i) \right].$$

Z_N , the partition function for a macromolecule with $N + 1$ atoms and N bonds, is a summation (integration) over all positions of intermediate atoms weighted by Boltzmann factors, $\exp \left[-\beta gV(x_i) \right]$, with the constraint (the delta functions) that bonds have length, ϵ . The factor, $\frac{1}{4\pi\epsilon^2}$, normalizes this so that given the position of the i^{th} atom, the integration over the location of the $(i+1)^{\text{th}}$ atom is unity. Modification of this factor can always be absorbed in μ . As long as x_f and x_0 are far apart and ϵ is small, the sum over N effectively begins with the enormous value, $N_0 = \frac{|x_f - x_0|}{\epsilon}$. It does no harm to start the sum at $N = 1$.

Eventually, we will take ϵ to zero so as to recover continuous curves from segmented N -step ones. In this limit, Z_N resembles a Feynman path integral. Path integrals have been widely used to account for the statistical properties of macromolecules¹¹. We shall review this. As is common in statistical

mechanics, one must coarse-grain: write N as nm with both n and m large, that is, break up the macromolecule into m units of n atoms.

Consider the situation where ϵ is small, n is large, but $\sqrt{n}\epsilon$ is also small. Then

$$\prod_{i=v_n}^{(v+1)n-1} \frac{1}{4\pi\epsilon^2} \delta(|x_i - x_{i-1}| - \epsilon) \rightarrow \left(\frac{3}{2\pi n\epsilon^2}\right)^{3/2} \\ \times \exp\left[-\frac{3}{2n\epsilon^2}(x_{(v+1)n} - x_{vn-1})^2\right] \quad (2.2)$$

In Eq. (2.2) $i = vn$ to $(v+1)n - 1$ are the atoms in the v^{th} unit. Equation (2.2) is true because the left hand side represents a random walk between x_{vn-1} and $x_{(v+1)n}$ which by the Central Limit Theorem approaches a Gaussian. If $\sqrt{n}\epsilon$ is small compared to the distance over which $V(x)$ varies appreciably, then $V(x)$ may be treated as a constant in each unit. We obtain

$$Z = \sum_{N=1}^{\infty} \left[\prod_{v=1}^{m-1} d^3x_v \right] \left[\left(\frac{3}{2\pi n\epsilon^2}\right)^{\frac{3m}{2}} \exp\left\{-\frac{3}{2} \sum_{v=1}^m \frac{(x_v - x_{v-1})^2}{n\epsilon^2}\right\} \right] \\ \times \exp\left[-\beta \sum_{v=1}^m n V(x_v) - \beta \sum_{v=1}^m n\mu\right] \quad (2.3)$$

The x_i 's have been relabelled so that x_v is the average value of x in the v^{th} cell. As ϵ goes to zero, Eq. (2.3) approaches

a Feynman integral. Using the variables

$$s \equiv \frac{n\epsilon^2 v}{6} , \quad (2.4)$$

$$\tau \equiv \frac{\epsilon^2 N}{6} ,$$

and the "bare" mass and coupling defined by

$$m_0^2 = \frac{6\beta\mu}{\epsilon^2} , \quad (2.5)$$

$$g_0 = \frac{6g}{\epsilon^2} ,$$

Equation (2.3)' becomes

$$Z = \frac{6}{\epsilon^2} \int_0^\infty dt \iint_{\substack{x(0) = x_0 \\ x(\tau) = x_f}} \exp \left[- \int_0^\tau \left(\frac{\dot{x}^2(s)}{4} + m_0^2 + \beta g_0 V(x(s)) \right) ds \right] . \quad (2.6)$$

The sum over N has become an integral over τ .

Remarks: (a) A bond-vector field interaction of the form $\sum_i q(x_i - x_{i-1})_\mu H^\mu(\bar{x}_i)$ can also be considered ($\bar{x}_i \equiv \frac{x_i + x_{i-1}}{2}$ is the average value of x along the bond). The effect is to add a term $-\beta q \int_0^\tau \dot{x}_\mu(s) H^\mu(x(s)) ds$ to the action in Eq. (2.6).

(b) Equation (2.6) resembles the Green's functions which arise in a particle dynamics representation of a field theory¹².

(c) The factor $\frac{6}{\epsilon^2}$ can be removed by normalizing Z appropriately (a sort of wave-function renormalization). In any case, physical quantities (averages) do not depend on the overall normalization of Z .

Summarizing, the Green's functions, $G(x_f, x_0)$, which appear in a particle dynamics representation of a field theory, correspond to the grand partition functions of polymers with ends at x_f and x_0 . The sum is over all continuous paths of arbitrary shape and length. The mass squared is proportional to the chemical potential which must be appropriately scaled to obtain a continuum limit. Other inputs are (a) that $n \rightarrow$ large so that products of delta functions approach Gaussians, (b) that $\sqrt{n} \epsilon \rightarrow 0$ so that $\sum_{x_1 \in \text{unit}} V(x_1) \rightarrow nV(\bar{x})$, and (c) that $\epsilon \rightarrow 0$ so that segmented curves approach smooth ones.

Representations such as Eq. (2.6) are well known to chemical physicists¹¹. Equation (2.6) is not a new result. We have rederived it as a warm-up for the next step: polymers in bulk. It is not hard to believe that a gas of polymers might be describable as a Euclidean field theory, and we shall show this. We have known about this correspondence for some time and have thought about using it to obtain field theory results as in reference 10. Until the present application, this analogy did not seem fruitful because a gas of closed loops is a complicated statistical system of which little is known. Fortunately, we shall not use the statistical

mechanics side of this analogy in an essential way.

Now consider closed polymers, obtained by setting $x_f = x_0 \equiv x$ in Eq. (2.6). Allowing loops to be located anywhere necessitates integrating over x . There is, however, an overcounting problem. For a closed polymer of N atoms, it is impossible to differentiate which atom was the starting point, that is, the N different starting points cannot be distinguished. For each configuration that begins at x_0 , traces an N -fold segmented path, and returns to x_0 , there is one which begins at x_1 , traces out the self-same path, and returns to x_1 . Thus configurations are overcounted by a factor of N . The partition function for a closed polymer of arbitrary length and location is

$$z = \int d^3x \sum_{N=1}^{\infty} \frac{1}{N} \exp[-\beta\mu N] Z_N(x, x), \quad (2.7)$$

where Z_N is given in Eq. (2.1). Proceeding as before, we obtain

$$z = \int_0^{\infty} \frac{d\tau}{\tau} \int d^3x \int_{\substack{x(0)=x \\ x(\tau)=x}} \mathcal{H} x \exp \left[- \int_0^{\tau} \left(\frac{\dot{x}^2}{4} + m_0^2 + \beta g_0 V(x) \right) ds \right]. \quad (2.8)$$

Finally, the grand partition function for the gas of loops is

$$\begin{aligned} \hat{\mathcal{Z}} &= \sum_{M=0}^{\infty} \frac{1}{M!} z^M = \exp z \\ &= \int \exp \left\{ - \text{tr} \ln \left[p^2 + m_0^2 + \beta g_0 V \right] \right\} \end{aligned} \quad (2.9)$$

(2.9) Continued on next page.

$$= \mathcal{N} \int \int \mathcal{H} \phi \exp \left\{ - \int \left[\partial_{\mu} \phi \partial^{\mu} \phi(x) + m_0^2 \phi^2(x) + \beta g_0 V(x) \phi^2(x) \right] d^3x \right\} .$$

\mathcal{N} is an overall (infinite) normalization constant. We have used

$$\int_0^{\infty} \frac{d\tau}{\tau} \left[\exp(-a\tau) - \exp(-b\tau) \right] = \int_a^b dz \int_0^{\infty} \exp(-z\tau) d\tau = -\ln \frac{a}{b} .$$

The ill-defined innocuous extra piece, $\exp \left[- \int d^3x \int_0^{\infty} \frac{d\tau}{\tau} \exp(-b\tau) \right]$

is \mathcal{N} . The operator, $i\partial_{\mu}$, is p_{μ} and ϕ is a scalar field.

Remarks: (a) As usual, the infinity in \mathcal{N} is harmless since it divides out when calculating expectation values.

(b) For oriented curves (that is, curves with a direction for which curves of different direction are distinguishable) a functional integral over a complex (charged) field is obtained. The orientation direction is identified with the flow of charge. In general, a gas of T different types of macromolecules leads to a T -component field.

(c) With oriented curves and a bond interaction, the action, $A = \int \left[|(\partial_{\mu} - ieA_{\mu})\phi|^2 + m_0^2 \phi^* \phi \right] d^3x$ can be obtained.

(d) Interactions between atoms (and/or bonds) can be introduced using auxiliary fields. Suppose the interactions are of the form $\sum_{\substack{\text{all pairs} \\ \text{of atoms}}} g^2 V(x_i, x_j)$. Define

$$\int G(x, y) V(y, z) d^3y = \delta^3(x - y) . \quad \text{Then}$$

$$\begin{aligned}
Z &= \frac{1}{\int \mathcal{A}_\chi \exp \frac{\beta}{2} \int \chi(x) G(x, y) \chi(y) d^3x d^3y} \times \\
&\times \int \mathcal{A}_\chi \int \mathcal{A}_\phi \exp \left\{ - \int \left[\partial_\mu \phi \partial_\mu \phi + m_0^2 \phi^2 + \beta g_0 \chi(x) \phi \right] d^3x \right\} \\
&\times \exp \left\{ \frac{\beta}{2} \int \chi(x) G(x, y) \chi(y) d^3x d^3y \right\}. \quad (2.10)
\end{aligned}$$

This is verified by first doing the ϕ integration to yield a gas of closed loops and then by doing the χ integration.

(e) In particular, a ϕ^4 theory corresponds to a repulsive interatomic delta function potential.

(f) Three dimensional scalar QED corresponds to a bond-bond interaction of the form

$$\frac{e^2}{4\pi\beta} \sum_{\text{pairs of } \vec{b}, \vec{b}'} \frac{\vec{b} \cdot \vec{b}'}{r_{bb'}}. \quad (2.11)$$

bonds, b and b'

In Eq. (2.11) $\vec{b} = \vec{x}_{i+1} - \vec{x}_i$ is the bond vector between the i^{th} and $(i+1)^{\text{th}}$ atoms, $r_{bb'}$ is the distance between the two bonds, and the sum is over all pairs of bonds. The interaction is attractive for antiparallel bonds and repulsive for parallel ones.

(g) The Lagrangian, $\mathcal{L} = -\frac{1}{2}(\partial_\mu \phi)^2 - \frac{1}{2}m_0^2 \phi^2 - \beta g_0 \chi \phi^2 - \frac{\beta}{2}(\partial_\mu \chi)^2 - \frac{\beta m_\chi^2}{2} \chi^2$, corresponds to attractive interatomic Yukawa potentials.

When $m_x = 0$, Coulomb interactions are obtained.

(h) The method works in any dimension, of course.

(i) Our method can be applied to obtain local field theories for solitons^{9, 13}. As reference 13 points out, perturbative expansion about vacuum expectation values misses soliton solutions. These extra solitons carry topological charges which are conserved. They generate closed loops in the appropriate dimension, via a macromolecule analogy yield partition functions similar to $Z_N(x, x)$ of Eq. (2.1) and Z of Eq. (2.7), and therefore result in a field theory. Of course, the solitons interact with the original fields. This can be taken into consideration using auxiliary fields and Lagrange multipliers (as done in reference 13). The macromolecule technique would replace the left-hand-side of Eq. (3.3) of reference 13 by

$$\sum_{M=0}^{\infty} \frac{1}{M!} Z^M, \quad (2.12)$$

with

$$Z = \int_0^{\infty} \frac{d\tau}{\tau} \int dx \int_{\substack{x(0)=x \\ x(\tau)=x}} \delta^4 x \exp \left[- \int_0^{\tau} \left(\frac{\dot{x}^2(s)}{4} + m_0^2 + iQ_{\mu}(x(s))\dot{x}^{\mu}(s) \right) ds \right]. \quad (2.13)$$

This would provide an alternative derivation to the one of Appendix A of reference 13. Although the normalization of delta functions in Eq. (2.1) is unknown and could generate an unknown

mass parameter, m_0^2 , our intuitive feeling is that $m_0^2 = 0$, although we cannot prove it. If it were non-zero, the measure factors in going to a sum over trajectories, would determine it in terms of the parameters in the original Lagrangian. One might think that m_0^2 could be the missing factor cancelling self-energy infinities found in reference 13 (Sec. IV), but we believe this is not the case. The infinities occur because near the soliton the Higgs field must go to zero. Expanding about non-zero Higgs field vacuum expectation values cannot deal with such a constraint. Regardless, the mass, m_0 , is geometrical in nature and has nothing to do with the soliton's physical mass.

(j) The chemical potential per atom and m_0^2 are proportional. For $m_0^2 < 0$, the chemical potential is negative and loops will populate the vacuum indefinitely. When a symmetry breaking potential is used, repulsive delta function potentials between atoms stabilize the proliferation of loops. Therefore, spontaneous symmetry breakdown looks like a dense gas of loops from the symmetric vacuum point of view (the "spaghetti vacuum").

(k) Renormalization infinities occur because the macromolecule potentials corresponding to interacting relativistic field theories are singular. For example, a ϕ^4 theory corresponds to repulsive delta function potentials between atoms. This extremely short-ranged singular potential consequently ruins the coarse-graining procedure used in going from Eq. (2.1) to (2.6). It will no longer be true that these "random walks" approach Gaussian distributions. A ϕ^4 theory is not so different from the

self-avoiding random walk problem^{11, 14}. The non-Gaussian nature of this process is well known¹⁴. This will lead to wave function and mass renormalization as perturbation theory tries to approximate non-Gaussian processes by Gaussian ones. The singular nature of potentials can also cause other problems. For attractive potentials, the interatomic forces might be too strong ("non-renormalizability") and cause macromolecules to collapse into "balls of wire". Perturbation theory is insensitive to the sign of g and therefore such effects manifest themselves for repulsive potentials, also. The higher the dimension of space-time, the more singular the forces. This is why fewer renormalizable theories occur in higher dimensions.

III WILSON LOOPS IN THE PRESENCE OF TOPOLOGICAL VORTICES

This section will calculate the Wilson loop in the presence of a gas of Nielsen-Olesen vortices⁵. The next section will treat 't Hooft Z_3 vortices. These calculations are similar to instanton calculations, where multiple instanton configurations generate a gas. Various computational devices have put instanton calculations on a solid foundation⁷. Statistical mechanics and physical intuition determine their properties quite easily. This is how Callan, Dashen, and Gross² are able to determine the magnetic properties of a dense BPST¹⁵ instanton gas. Contrast this with a vortex gas. The vortices may vary in number, position, and the way they are imbedded. They may twist in the most unruly manner. These complications lead one to think that such a system is too difficult to deal with, however, the methods of Section II make the problem tractable. We are able to do Wilson loop calculations. In fact, gases of vortices are as easy to

handle as gases of instantons. We will show how calculations involving topological spontaneous symmetry breakdown are done.

Take the Euclidean space version of the 2 + 1 dimensional Nielsen-Olesen model⁵. This is scalar QED with a Higgs potential. The Lagrangian is

$$\mathcal{L} = \frac{1}{4} F_{\mu\nu} F^{\mu\nu} + |(\partial_{\mu} - ieA_{\mu})\phi|^2 + \lambda (\phi^*\phi - F^2)^2, \quad (3.1)$$

and has vortex-like solutions along the third axis of the form

$$\phi(x) = f(\rho)\exp(i\phi),$$

$$A_{\phi}(x) = a(\rho), \quad (3.2)$$

$$A_z(x) = A_{\rho}(x) = 0.$$

Cylindrical coordinates, ρ, z, ϕ , have been used. Graphs of $f(\rho) = |\phi(\rho)|$ and $H_z(\rho)$ are given in Fig. 1 of reference 5.

The important properties of the solution are

- i) (ϕ_1, ϕ_2) points radially outward from the vortex.
- ii) $|\phi(\rho)|$ vanishes at the vortex (at $\rho = 0$) and goes to F far from it ($\rho = \infty$).

iii) $a(\rho) \rightarrow \frac{1}{e\rho}$ far from the vortex. This means that the vortex contains a tube of magnetic flux. The total flux is

$$\int dx dy H_z(x, y) = \int_0^{2\pi} A_{\phi}(\rho) \rho d\phi = \frac{2\pi}{e}.$$

Property (i) is topological in nature and makes the soliton topologically stable.

Property (ii) has important implications (short-ranged ones and long-ranged ones). First, near the vortex where $|\phi(\rho)| = 0$, the photon is effectively massless, whereas far away it has mass because the Higgs field has a vacuum expectation value. This is one way of understanding why flux is channelled into tubes. There is a tubular "mass confinement" bag. Also $|\phi(\rho = 0)| = 0$ indicates symmetry restoration in the vortex, a point which we shall discuss in detail later. Finally, we should mention that the Higgs field prevents the vortex from collapsing to zero size and gives it a finite mass. Secondly, the Higgs field at infinity must take on vacuum expectation values and be covariantly constant. The latter means that $A_\mu(\rho)$ is determined in terms of the phase, χ , of the Higgs field. A_μ is pure gauge and $\chi(x) - \chi(y) = e \int_x^y A_\mu dx^\mu$. Whenever y loops around a circle and returns to x the phase, χ , must be an integral multiple of 2π . This causes the flux to be quantized. We shall return to this point in Sec. VIII.

Property (iii) contains the physics: the vortices are quantized tubes of magnetic flux. This is the key physical characteristic.

The vortex soliton has a topological number, which can be seen in two ways: using gauge potentials or using Higgs fields¹⁶. Of course, the two are interrelated. In terms of A_μ , consider $\exp\left[ie \int_x^y A \cdot dx\right]$. Such a pure phase takes values on the unit circle in the complex plane. Fix x and move y around the vortex

as in Fig. 2. In regions where A_μ is pure gauge and the Higgs field is covariantly constant, $\exp \left[ie \int_x^y A \cdot dx \right]$ must return to 1 when y returns to x . This forms a map from a circle in Euclidean space to the unit complex circle. These maps are characterized by winding numbers, $\pi_1(S^1) = \mathbb{Z}$, the set of integers. Winding numbers count the number of times the image curve loops around the unit circle. Regions of space with $\pi_1 = 1$ or $\pi_1 = -1$ contain a vortex or an antivortex.

In terms of the Higgs field the topology is as follows: far away $|\phi|$ must be F , that is, $\phi(x)$ must take on values which minimize the potential, $\lambda(\phi^*\phi - F^2)^2$; they form a circle of minima. Again consider moving y around a closed curve (Fig. 2). Then $\phi(y)$ forms a map from a circle in Euclidean space to a circle (of minima). Again these maps are characterized by $\pi_1(S^1) = \mathbb{Z}$.

Because winding number is invariant under continuous deformations (homotopies), winding number can neither be created nor destroyed unless an (unallowed?) discontinuity occurs in field configuration. This provides the topological stability of the vortices.

In three dimensional space-time, orient the vortex lines. The orientation indicates the direction of the flux and the flow of topological charge. In this way antivortices act like particles travelling backward in time. Of course, vortices need not be straight lines. In general, they twist and curve in an arbitrary manner. Since they may never end, they form closed loops. If they have a positive mass, these loops will be *trall*, will act

like neutral objects, and will have few physical consequences. In contrast, when their mass is negative, they fill up the vacuum and their effects can be dramatic.

Having reviewed the properties of Nielsen-Olesen vortices, let us put a Wilson loop (with an arbitrary charge, q) in the system (see Fig. 3a):

$$\langle \exp[iq \oint_C A \cdot dy] \rangle . \quad (3.3)$$

Consider a single vortex. What is its effect? Using Stoke's theorem (or simply physical intuition), one sees that the Wilson loop measures the linking number¹⁷ of the two curves. If the vortex links n times the contribution is $\exp(i \frac{2\pi}{e} qn)$. Examples with $n = +1$, -1 , and $+2$ are shown in Figs. 3b, 3c, and 3d. For several vortices the contribution is $\exp[i \frac{2\pi q}{e} (n_+ - n_-)]$, where n_+ and n_- are the number of positive and negative linkages. Idealize the situation to the case where vortices are thin (trace out curves rather than tubes) and are not mutually interacting. This can be rectified if the form of the interactions is known. For simplicity set the mass equal to zero. In the end, we will restore a non-zero mass. The linking number of two curves can be expressed as an integral¹⁷:

$$n = - \frac{1}{4\pi} \int_{C_1} d\vec{x} \cdot \int_C \frac{(\vec{x} - \vec{y}) \times d\vec{y}}{|\vec{x} - \vec{y}|^3} . \quad (3.4)$$

Here, n is the number of times C_1 and C link. For several curves the total linking number with C is

$$n = \sum_i -\frac{1}{4\pi} \int_{C_i} d\vec{x}_i \cdot \int_C \frac{(\vec{x}_i - \vec{y}) \times d\vec{y}}{|\vec{x}_i - \vec{y}|^3} = \sum_i \int_{C_i} d\vec{x}_i \cdot \vec{B}(\vec{x}_i), \quad (3.5)$$

where

$$\vec{B}(\vec{x}) = -\frac{1}{4\pi} \int_C \frac{(\vec{x} - \vec{y}) \times d\vec{y}}{|\vec{x} - \vec{y}|^3}. \quad (3.6)$$

Note that $B(x)$ has the same form as the magnetic field produced by a current flowing in C . Take C to be the Wilson loop and think of the C_i as vortex loops. Using the method of Sec. II, the calculation of (3.3) in a gas of idealized Nielsen-Olesen vortices is

$$\langle \exp[iq \oint A \cdot dy] \rangle =$$

$$\sum_{M=0}^{\infty} \frac{1}{M!} \prod_{i=1}^M \left[\int_0^{\infty} \frac{d\tau_i}{\tau_i} \int_{d\omega_i} \int \mathcal{B}_x^i \right] \times$$

$$x^i(0) = x_i$$

$$x^i(\tau_i) = x_i$$

$$\times \exp \left\{ - \int_0^{\tau_i} \left(\frac{(\dot{x}^i)^2}{4} + \frac{2\pi i q}{e} B^\mu(x^i) x_\mu^i \right) ds^i \right\} \quad (3.7)$$

(3.7) continued on next page

$$\times \frac{1}{[\text{same term with } B_\mu = 0]} \left\{ \int \int \psi \exp \left\{ - \int \left(a_\mu - \frac{i2\pi q}{e} B_\mu \right) \psi \right\}^2 \right\} \\ = \frac{\int \int \psi \exp \left\{ - \int \left(a_\mu - \frac{i2\pi q}{e} B_\mu \right) \psi \right\}^2}{[\text{same term with } B_\mu = 0]}$$

If the vortices interact and have a mass Eq. (3.7) would be replaced by

$$\frac{\int \int \psi \exp \left\{ - \int \left(a_\mu - \frac{i2\pi q}{e} B_\mu \right) \psi \right\}^2 + m^2 \psi^* \psi + \mathcal{V}(\psi^* \psi) \right\}}{[\text{same term with } B_\mu = 0]} \quad (3.8)$$

$\int \int (\psi^* \psi)$ might be, for example, $\int d^3x \int d^3y \psi^*(x) \psi(x) \frac{\exp[-\mu|x-y|]}{4\pi|x-y|} \times \psi^*(y) \psi(y)$. This would correspond to a Yukawa interaction between points on the vortex. In any case \mathcal{V} must be a functional of $\psi^*(x)\psi(x)$ only. Given the interactions and mass of vortices, the exact form of Eq. (3.8) can be determined. In principle, m^2 and $\mathcal{V}(\psi^*\psi)$ may be approximately deduced using semiclassical methods or perhaps by examining the Lagrangians of reference 13 in more detail.

Of course ψ is the vortex field. In fact, inserting $\psi^*(x)\psi(y)$ in the integrand of Eq. (3.8) (with $B_\mu = 0$ so that the Wilson loop is absent) and returning to a macromolecule description, one obtains a gas of closed vortices in the presence of one open one which starts at x and ends at y . Hence, $\psi^*(x)$ produces the vortex endpoint at x and $\psi(y)$ destroys it at y .

Regardless of the detailed nature of $\mathcal{V}(\psi^*\psi)$, we can evaluate Eq. (3.8) semiclassically for various cases. i) When

$m^2 > 0$, $\psi = 0$ is expected to be the vacuum. This is also the solution to the equations of motion for a non-zero B_μ and the Wilson loop to this approximation is 1. Vortices do not contribute to the Wilson loop as expected. This is because for $m^2 > 0$ the vortices are small loops and rarely link with the Wilson loop.

ii) When $m^2 < 0$, there is topological symmetry breakdown in topological charge. The vacuum fills up with vortices. Presumably $\mathcal{L}'(\psi^*\psi)$ contains repulsive forces which eventually stabilize the proliferation of loops (an example is $\mathcal{L}'(\psi^*\psi) = g\psi^*\psi\psi^*\psi$ which corresponds to repulsive delta function forces between points on vortices). ψ then acquires a vacuum expectation value, $\langle\psi\rangle = \psi_0$. Because the (denominator) Lagrangian of Eq. (3.8) is a function of $\psi^*\psi(x)$, $\psi_0 \exp(i\theta)$ for $0 < \theta < 2\pi$ are also action minima. Consider a long rectangular Wilson loop of width, r , and length, t (Fig. 3a). Trying $\psi = \psi_0$ as a trial solution yields

$$\langle \exp [i q \oint A \cdot dy] \rangle \sim \exp \left[- \left(\frac{2\pi q}{e} \right)^2 \int d^3x B^2(x) \right] . \quad (3.9)$$

$B(x)$ is given in Eq. (3.6) and is the magnetic field created by two parallel wires with opposite current flowing through them. The evaluation of Eq. (3.9) is a problem of undergraduate electromagnetism¹⁸: $\int d^3x B^2 \sim \left(\frac{1}{4\pi} \right)^2 2t \ln(r/r_0)$. The constant r_0 should be of the order of the vortex width, since thick vortices can partially intersect a Wilson loop and our idealized approximation breaks down. For $\frac{q}{e}$ greater than 1, we can find screening type solutions which better minimize the action. Let m be the

nearest integer to $\frac{q}{e}$ and set $\Delta q = m - \frac{q}{e}$. Let

$$\psi(x) = \exp \left[i\chi(x) \right] \psi_0 ,$$

$$\chi(x) \equiv m \left\{ \tan^{-1} \left| \frac{x - x_1}{y - y_1} \right| - \tan^{-1} \left| \frac{x - x_2}{y - y_2} \right| \right\} , \quad (3.10)$$

where (x_1, y_1) and (x_2, y_2) are the (x, y) coordinates of the two lines comprising the Wilson loop. In Eq. (3.10) m must be integer-valued so that $\chi(x)$ is single-valued when circling around (x_1, y_1) or (x_2, y_2) . We obtain

$$\langle \exp \left[i q \oint_{\Lambda} \mathbf{A} \cdot d\mathbf{y} \right] \rangle \sim \exp \left[- \frac{(\Delta q)^2}{2} \tau \ln \frac{r}{r_0} \right] . \quad (3.11)$$

We expect Eq. (3.10) to be approximately the correct solution for arbitrary $\mathcal{V}(\psi^*\psi)$. Only an attractive singular potential could cause the vortices to form neutral bound objects and ruin the picture. Our result is obtained in an almost model independent manner.

The Wilson loop test is sensitive only to the excess charge. This is the omnipresent periodic (in q) screening effect which occurs when the potential is due to topological configurations.

Equation (3.11) indicates a logarithmic potential between charges because of topological symmetry breakdown. Before ψ had acquired a vacuum expectation value, there was no logarithmic potential due to the photon because the Higgs mechanism had given the photon a mass. Spontaneous symmetry breakdown has restored

a two dimensional Coulomb-like force.

In addition to this topological non-perturbative $\ln r$ potential, we expect the perturbative $\ln r$ potential of the unbroken $U(1)$ gauge theory to be present also: Recall property (ii) of the Nielsen-Olesen vortex, that ϕ , the charged scalar field, must vanish at the vortex. Along the vortex the photon is "massless" in contrast to outside the vortex where it has mass. When topological symmetry breakdown occurs the vacuum is filled with vortices until repulsive forces take over. Presumably this occurs when they begin to overlap. This means that $\langle \phi(x) \rangle$ must be zero since vortices occupy all of space-time. The photon must be massless, since $\langle \phi(x) \rangle \neq 0$ was the factor contributing to its mass. The original $U(1)$ charge symmetry is restored and another perturbative $\ln r$ potential due to the photon is expected. The sequel is depicted in Fig. 4.

SECTION IV Z_N VORTICES

We will now consider 't Hooft vortices. We will proceed in a manner similar to Sec III: first reviewing the soliton solutions and their properties and then performing the Wilson loop calculation.

The topology¹⁶ of an $SU(N)$ vortex is most easily discussed for $N = 2$. For this case, a possible Lagrangian is

$$\begin{aligned} \mathcal{L} = & \frac{1}{4} F_{\mu\nu}^2 + \frac{1}{2} |(\partial_\mu - igA_\mu^a L^a)\phi_{(1)}|^2 + \frac{1}{2} |(\partial_\mu - igA_\mu^a L^a)\phi_{(2)}|^2 \\ & + V(\phi_{(1)}, \phi_{(2)}), \end{aligned} \quad (4.1)$$

where

$$V(\phi_{(1)}, \phi_{(2)}) = \lambda_1(\phi_{(1)}^2 - F_{(1)}^2)^2 + \lambda_2(\phi_{(2)}^2 - F_{(2)}^2)^2 + \lambda_3(\phi_{(1)} \cdot \phi_{(2)})^2. \quad (4.2)$$

$\phi_{(1)}, \phi_{(2)}$ are two $SU(2)$ triplet Higgs fields, L^a are the 3×3 $SU(2)$ matrix generators, and $\lambda_1, \lambda_2, \lambda_3, F_{(1)},$ and $F_{(2)}$ are constants. Other Higgs potentials will also work.

Because all fields are in the adjoint representation, the symmetry of this model is $SU(2)/Z_2$ or $O(3)$. The Higgs potential breaks this symmetry completely, so that all three gauge fields acquire mass. The remaining bosonic excitations are also massive. From this point of view, there can be no long range forces.

As in the Nielsen-Olesen case, there are two types of topological numbers, one related to Higgs fields, the other connected with gauge fields. The solitons have both types.

Consider a static vortex at $x = y = 0$ (Fig. 2). Go far from it and circulate around it. The Higgs fields must take on vacuum expectation values. In going around the circle, these values trace out a curve in the minima, M , of the Higgs potential. Such curves are characterized by $\pi_1(M)$. Are there curves in M not deformable to a point for the V in Eq. (4.2)? M is the set of pairs of three dimensional vectors, $v_{(1)} \equiv \phi_{(1)}/F_{(1)}$ and $v_{(2)} \equiv \phi_{(2)}/F_{(2)}$ satisfying $v_{(1)} \cdot v_{(1)} = v_{(2)} \cdot v_{(2)} = 1, v_{(1)} \cdot v_{(2)} = 0$. If we append

the vector $v_{(3)} = v_{(1)} \times v_{(2)}$, an orthonormal frame is obtained. Therefore, M is the set of orientations of this frame. Fix a reference frame $v_{(1)}^0 = (1, 0, 0)$, $v_{(2)}^0 = (0, 1, 0)$, $v_{(3)}^0 = (0, 0, 1)$. An arbitrary frame is determined by an $O(3)$ rotation of this reference frame. Therefore, M is the set of orthogonal transformations. Characterize the rotation ala Schiff¹⁹ by a vector in the direction of the rotation whose magnitude is the angle of rotation. M becomes isomorphic to the solid three-dimensional sphere with antipodal points identified. This space has curves which cannot be deformed to a point (see Fig. 5). However, a path which goes twice along the route of Fig. 5 can be deformed to a point (see Fig. 6). Physically this is demonstrated in M.T.W.²⁰. The fundamental group, $\pi_1(M)$, is Z_2 , and this characterizes the vortex. The vortices carry a topological charge conserved modulo two. Again do not confuse this Z_2 with the center of $SU(2)$. This is another Z_2 , topological in nature. Typical non-trivial Higgs configurations are shown in Fig. 7.

The topological number associated with the gauge field is similar to the Nielsen-Olesen case. Let $A_\mu = A_\mu^a L^a$, where L^a are the 3×3 "angular momentum" matrices.

$\left| P \exp \left\{ ig \int_x^y A_\mu \cdot dx^\mu \right\} \right|_{\alpha\beta}$ is the path ordered product from x to y . As y goes around the circle and back to the starting point, x , (Fig. 2), this matrix traces a curve in $O(3)$. Because we are far away, in the region where the Higgs fields are covariantly constant, this matrix is precisely the $O(3)$ rotation that takes the Higgs frame at x to the Higgs frame at y . Again a closed

curved in $O(3)$ is obtained which begins and ends at the identity element. Again Z_2 characterizes the topological charge.

Letting $A_\mu = \frac{\tau^a}{2} A_\mu^a$ with $\frac{\tau^a}{2}$, the Pauli matrices:

$$\left[\nu \exp \left(i g \oint A_\mu \cdot dx^\mu \right) \right]_{\alpha\beta} = -\delta_{\alpha\beta}, \quad (4.3)$$

for a gauge field vortex. The line integral is around a circle containing the vortex.

The important local properties of the vortex are

(i) $\phi_{(1)}$ and $\phi_{(2)}$ become parallel at the vortex (Fig. 7c) so that the frame becomes ill-defined²¹.

(ii) For point-like vortices and A_μ^k not varying rapidly in $SU(2)$ space (as the vortex is approached, we expect A_μ^k to go to zero rapidly but do not expect the color direction to vary rapidly), there is a delta-function-like contribution to the flux in the direction of $SU(2)$ space parallel to $\phi_{(1)}$ and $\phi_{(2)}$:

$$\hat{\phi}_{(1)}^k \cdot F_{12}^k(x, y) = \hat{\phi}_{(2)}^k \cdot F_{12}^k(x, y) \sim \frac{2\pi}{g} \delta(x) \delta(y). \quad (4.4)$$

For $SU(N)$ we expect the following to be true:

(a) The Higgs potential is chosen to break the symmetry completely. The Higgs fields are in the adjoint representation.

For SU(3) two Higgs fields are sufficient²². Let $\phi_{(i)}^0$ be the vacuum expectation value of the i^{th} Higgs field. Let $v_{(i)} = L^{\ell} \phi_{(i)}^0 / F_{(i)}$, where L^{ℓ} are the Lie algebra adjoint matrices of SU(N). The symmetry is completely broken if for any η^{ℓ} , $[\eta^{\ell} \cdot L^{\ell}, v_{(i)}] \neq 0$ for some i .

(b) Vortices are characterized by $\pi_1(\text{SU}(N)/Z_N) \simeq Z_N$.

(c) Define $v_{(i)}(x) = \hat{\phi}_{(i)}^{\ell}(x) L^{\ell}$. The "hat" over $\phi_{(i)}$ indicates that it is normalized to 1, so that

$$\hat{\phi}_{(i)}(x) = \phi_{(i)}(x) / \left[\phi_{(i)}^{\ell}(x) \phi_{(i)}^{\ell}(x) \right]^{1/2}.$$

Then at the vortex, there is at least one η^{ℓ} such that

$$[\eta^{\ell} \cdot L^{\ell}, v_{(i)}(x)] = 0 \text{ for all } i, \text{ where } x \text{ is the location of}$$

the vortex. Let \mathcal{J} be the set of vectors, η , such that

$$[\eta^{\ell} \cdot L^{\ell}, v_{(i)}(x)] = 0 \text{ for all } i. \text{ Then } a\eta_1 + b\eta_2 \text{ is in } \mathcal{J}$$

and $[\eta_1, \eta_2]$ is in \mathcal{J} , if η_1 and η_2 are in \mathcal{J} . The

latter is true because $[[\eta_1, \eta_2], v_{(i)}(x)] = -[[\eta_2, v_{(i)}(x)], \eta_1]$

- $[[v_{(i)}(x), \eta_1], \eta_2] = 0$. The set of matrices $\eta^{\ell} \cdot L^{\ell}$ for

$\eta \in \mathcal{J}$ forms a Lie subalgebra which generates a subgroup, H, of SU(N).

(d) We expect that only one η^{ℓ} occurs, so that \mathcal{J} is one dimensional and $H \simeq U(1)$ (in SU(3), for example, the two Higgs

fields might point in the $\frac{\tau^2}{2}$ and $\frac{\tau^3}{2}$ directions so that only one

η occurs and $\eta^{\ell} = \delta^{\ell 8}$). Furthermore $\eta^{\ell} \cdot \frac{\lambda^{\ell}}{2}$ has N-1

eigenvectors with eigenvalue $\frac{1}{N}$ and one eigenvector of eigenvalue

$\frac{N-1}{N}$ when $\eta^{\ell} \cdot \eta^{\ell} = \frac{2(N-1)}{N}$ and $\frac{\lambda^{\ell}}{2}$ are the fundamental matrix

representation normalized so that $\text{tr} \frac{\lambda^{\ell}}{2} \frac{\lambda^m}{2} = \frac{\delta^{\ell m}}{2}$. For SU(3) we

believe the set of η^k satisfying this condition is of the form,
 $\eta \cdot \frac{\lambda}{2} = \exp(i\alpha \cdot \frac{\lambda}{2}) \frac{2}{\sqrt{3}} \frac{\lambda^8}{2} \exp(-i\alpha \cdot \frac{\lambda}{2})$, for eight vectors, α .

(e) There is a delta-function-like contribution to the flux of the form

$$\eta^k \cdot F_{12}^k(x) - \frac{2\pi}{g} \delta(x) \delta(y), \quad (4.5)$$

when η^k is normalized so that $\eta^k \cdot \eta^k = \frac{N}{2(N-1)}$.

(f) When more than one η^k occurs, the flux might rapidly fluctuate along the vortex in different color directions. Such fluctuations may be the color zitterbewegung phenomenon²³ associated with particles carrying an internal symmetry. There is the speculative possibility that the particles associated with such trajectories are non-abelian ones and the fields related to them form a representation of H . This is highly conjectural and probably impossible to prove. This is a quantum mechanical effect. In this way a dual gauge group may be generated²⁴. See also reference 21.

(g) That the flux must be in the subgroup, H , is reasonable physically. At the vortex gluonic excitations associated with H are massless, because the symmetry, H , is restored. Outside, these same fields are massive. This is a sort of tubular bag-like mass confinement mechanism of field strengths. They are restricted to the massless regions of space, i.e. the vortices. The important point is vortices carry tubes of magnetic flux in the group, H .

For the rest of this paper, we will restrict ourselves to the relevant case of $SU(3)$. Let us repeat the calculation of the last section for $SU(3)$. Idealize to zero width vortices. The topological charge is conserved mod. 3. Thus, there are two non-trivial types of vortices. One is characterized by $\text{tr} \left[P \exp \left(ig \oint A \cdot dx \right) \right] = 3 \exp \left(\frac{2\pi i}{3} \right)$, the other by $\text{tr} \left[P \exp \left(ig \oint A \cdot dx \right) \right] = 3 \exp \left(-\frac{2\pi i}{3} \right)$ (we are now dotting A_μ^k into the fundamental representation, $\frac{\lambda^k}{2}$, the 3×3 matrices). The path in the path ordered product is to be taken around the vortex. The 3's in the above equations are trace factors and get replaced by N for $SU(N)$. Again, vortices will trace out particle trajectories in three dimensional Euclidean space. If we assign orientations, then oppositely oriented vortices carry opposite units of flux and may be regarded as antiparticles. The vacuum will be a gas of them. If they have a positive mass they will be small and sparsely located. If they have a negative mass they will fill up the vacuum. The calculation of the Wilson loop in the presence of these vortices proceeds as in Sec. III, except that $q = \frac{2}{3}$. The result is

$$\langle \text{tr} \left[P \exp \left(ig \oint A \cdot dx \right) \right] \rangle^{-1} = \frac{\int \int \mathcal{P} \psi \exp \left\{ - \int d^3x \left[\left| \left(\partial_\mu - \frac{2\pi i}{3} B_\mu \right) \psi \right|^2 + m^2 \psi^* \psi + \mathcal{U}(\psi^* \psi) \right] \right\}}{\left[\text{same as numerator with } B_\mu = 0 \right]} \quad (4.6)$$

The function, B_μ , is given in Eq. (3.6). We have allowed for a mass and for interactions. In principle, these are determined from the original Lagrangian (Eq. (4.1) for $SU(2)$). Again semiclassical methods and/or local field theory soliton methods^{13, 21} should be helpful in this respect. The only input in Eq. (4.6) is the commutation relations of vortex fields and Wilson loops which in Euclidean formulation become linking numbers. The factor of 3 is due to color.

The solutions and conclusions are the same as in the Nielsen-Olesen case. There is no long-range potential unless topological symmetry breakdown takes place, in which case the potential is a logarithm. We conclude that topological symmetry breakdown of Z_N vortex loops is not enough to give a linearly confining potential.

The above calculation considered only closed loops and did not allow for the possibility that three flux tubes could annihilate. Charge conservation, being modulo three, permits such events. Whether it actually happens is a question which can only be answered by finding the effective soliton Lagrangian. It may be that these events occur with zero probability. This question must be answered by doing the proper analysis of the original Lagrangian (the $SU(3)$ analog of Eq. (4.1)). No arbitrariness is involved. Let us redo the Wilson loop calculation making the ad hoc assumption that three vortices can annihilate. Configurations such as Fig. 8a as well as more complicated ones (Fig. 8c) are allowed. We shall call these configurations triplets. As indicated by 't' Hooft they may be generated by adding a term, $\lambda_0(\psi^3 + \psi^{\#3})$, to the Lagrangian of Eq. (4.6)²⁵:

$$\mathcal{L} = |\partial_{\mu}\psi|^2 + m^2\psi^*\psi + \mathcal{V}(\psi^*\psi) + \frac{\lambda_0}{3!}(\psi^3 + \psi^{*3}) . \quad (4.7)$$

This is seen by doing perturbation theory in λ_0 : zero'th order in λ_0 is equivalent to our previous gas of interacting loops. Second order (there are no first order terms) in λ_0 produces the configurations of Fig. 8a. Higher order terms yield a gas of those Fig. 8c as well as more complicated ones. Consider the second order term from the macromolecule point of view. Neglect interactions (i.e. set $\mathcal{V}(\psi^*\psi) = 0$) for simplicity.

$$\begin{aligned} Z &= \eta \iint \mathcal{D}\psi \mathcal{D}\psi^* \exp \left[- \int (\partial_{\mu}\psi \partial^{\mu}\psi^* + m_0^2\psi^*\psi + \frac{\lambda_0}{3!}(\psi^3 + \psi^{*3})) \right] = \\ &= \int_0^{\infty} d\tau_0 \left[1 + \frac{\lambda_0}{3!} \int d^3x_0 \int d^3x_f \int_0^{\infty} d\tau_1 \int_0^{\infty} d\tau_2 \int_0^{\infty} d\tau_3 \right. \\ &\quad \times \iint_{\mathcal{K}^3} \mathcal{K}^3 x(1) \quad \iint_{\mathcal{K}^3} \mathcal{K}^3 x(2) \quad \iint_{\mathcal{K}^3} \mathcal{K}^3 x(3) \\ &\quad x(1)(0) = x_0 \quad x(2)(0) = x_0 \quad x(3)(0) = x_0 \\ &\quad x(1)(\tau_1) = x_f \quad x(2)(\tau_2) = x_f \quad x(3)(\tau_3) = x_f \\ &\quad \left. \times \exp \left\{ - \int_0^{\tau_1} \left[\frac{\dot{x}(1)^2}{4} + m_0^2 \right] - \int_0^{\tau_2} \left[\frac{\dot{x}(2)^2}{4} + m_0^2 \right] - \int_0^{\tau_3} \left[\frac{\dot{x}(3)^2}{4} + m_0^2 \right] \right\} \right. \\ &\quad \left. + o(\lambda_0^4) \right] \quad (4.8) \end{aligned}$$

Equation (4.8) continued on next page.

$$z = z_0 \left[1 + \frac{\lambda_0^2}{3! \epsilon^3} \left(\frac{1}{8}\right)^3 \int d^3 x_0 \int d^3 x_f z_p + 0 (\lambda_0^4) \right], \quad (4.8)$$

where

$$z_p = \epsilon^9 \sum_{N_1=1}^{\infty} \sum_{N_2=1}^{\infty} \sum_{N_3=1}^{\infty} \left[\prod_{i_1=1}^{N_1-1} \int d^3 x_{i_1}^{(1)} \right] \left[\prod_{i_1=1}^{N_1} \frac{\delta(|x_{i_1}^{(1)} - x_{i_1-1}^{(1)}| - \epsilon)}{4\pi\epsilon^2} \right] \\ \times \left[\prod_{i_2=1}^{N_2-1} \int d^3 x_{i_2}^{(2)} \right] \left[\prod_{i_2=1}^{N_2} \frac{\delta(|x_{i_2}^{(2)} - x_{i_2-1}^{(2)}| - \epsilon)}{4\pi\epsilon^2} \right] \quad (4.9)$$

$$\times \left[\prod_{i_3=1}^{N_3-1} \int d^3 x_{i_3}^{(3)} \right] \left[\prod_{i_3=1}^{N_3} \frac{\delta(|x_{i_3}^{(3)} - x_{i_3-1}^{(3)}| - \epsilon)}{4\pi\epsilon^2} \right] \\ \exp \left\{ - \sum_{i_1=1}^{N_1} \beta\mu - \sum_{i_2=1}^{N_2} \beta\mu - \sum_{i_3=1}^{N_3} \beta\mu \right\}.$$

Comments:

(a) z_0 is the previously discussed grand partition function for a gas of non-interacting 't Hooft vortex loops. z_p is the partition function for three macromolecules which all begin at x_0 and all end at x_f (Fig. 8b). The endpoints, x_0 and x_f , are

arbitrarily located in three dimensional Euclidean space, hence the integrals, $\int d^3x_o \int d^3x_f$. There is also an arbitrary number of atoms, N_1 , N_2 , and N_3 , for each macromolecule; $x_s^{(j)}$ is the position of the s^{th} atom of the j^{th} macromolecule. The chemical potential is $\mu = \frac{\epsilon^2 m_o^2}{\beta}$, where β is the inverse temperature and ϵ is the length of a bond. The latter is the cutoff parameter in our segmented line approach to a continuous curve.

(b) The three macromolecules in Eq.(4.9) are 't Hooft vortices. Reintroducing $\mathcal{A}(\psi^* \psi)$, they undergo the same monomer or bond interactions as loops. Each carries $\frac{2\pi}{g}$ units of flux from x_o to x_f .

(c) Because \mathcal{Z}_o multiplies Z_p , the second term, $\mathcal{Z}_o Z_p$, is a system of an arbitrary number of closed loops and one triplet configuration. Let

$$\lambda_p = \frac{\lambda_o^2}{3! \epsilon^3} \left(\frac{1}{6}\right)^3 \equiv (\lambda_R^m)^2 \quad (4.10)$$

be a renormalized triplet activity (the subscript p stands for pair since there is a pair of vertices in Figs. 8a and 8b). Higher order terms in λ_p will generate multiple triplet configurations and will lead to $\mathcal{Z}_o \left(\sum_{M=0}^{\infty} \frac{1}{M!} \lambda_p^M Z_p^M \right)$, a grand partition function for a system of closed $M=0$ loops and triplets. The combinatorial factors in Feynman rules precisely give the $\frac{1}{M!}$ factor necessary for a grand partition function (vacuum bubbles exponentiate).

Let us redo the Wilson loop integral test, allowing for triplets and more complicated configurations. Both loops and triplets

lead to phase factors dependent on linking number. Figure 9 illustrates some possibilities.

Trouble arises in trying to repeat the calculation of Sec. III. Equation (3.4) is no longer valid. It works only for closed loops and not for those of Fig. 8. Fortunately, Eqs. (3.4) and (3.5) are not the only ones for linking number. Suppose

B_μ is replaced by $B_\mu + \partial_\mu \chi$, where χ is an arbitrary smooth function. Then, according to Eq. (3.5), $n = \sum_i \int_{C_i} B_\mu(x) dx^\mu \rightarrow$

$$\sum_i \int_{C_i} (B_\mu + \partial_\mu \chi) dx^\mu = \sum_i \int_{C_i} B_\mu(x) dx^\mu. \text{ The effect of } \partial_\mu \chi$$

disappears because of Stokes theorem; there is a kind of "gauge invariance" in defining B_μ . Does this arbitrariness affect the conclusions of Sec. III? The answer is no as seen from Eq. (3.8).

A change in B_μ by $\partial_\mu \chi$ can be absorbed in the solution for ψ by multiplying $\psi(x)$ by $\exp \left[i \frac{2\pi q}{e} \chi(x) \right]$. This leads to the same action and the same $\ln r$ behavior of the potential. The

Lagrangian of Eq. (3.8) has a global U(1) invariance which allows a redefinition of B_μ to be absorbed in a redefinition of ψ .

Previously, the 't Hooft vortex Lagrangian (without the $\lambda_0(\psi^3 + \psi^{*3})$ term) had this U(1) invariance also. The charges $\frac{2}{3}\pi$ and $-\frac{2}{3}\pi$ were absolutely conserved. Such a system looked like ordinary charge. Only when ψ^3 type terms are added can one "see" charge conserved modulo three. This is why the conclusions of the 't Hooft model were similar to the Nielsen-Olesen case. Now that triplets are present we expect different conclusions.

We can use the "gauge invariance" to define a B_μ which works. Let S be any two dimensional surface which spans the Wilson loop; the boundary of S is the Wilson loop. For simplicity take S to be a surface of minimal area. Let

$$B_\mu(x) = \int_S \delta^3(x-y) dS^\mu(y). \quad (4.11)$$

Here dS^μ is the surface element directed normal to the surface (the sign of the normal is determined by the orientation of the Wilson loop). As a clarifying example, take the Wilson loop in Fig.9a:

$$B_\mu(x) = -\delta_{\mu 3} \delta(z) \theta(x) \theta(L-x) \theta(y) \theta(L-y). \quad (4.12)$$

If a curve, $x(\tau)$, pierces the surface then $\int d\tau \dot{x}_\mu(\tau) B^\mu(x(\tau))$ is plus or minus one depending on the piercing direction. Roughly speaking, $B_\mu(x)$ [of Eq. (4.11)] points in the direction normal to the surface, acts like a delta function in this direction, and vanishes away from the surface. This $B_\mu(x)$ is obtainable from the old one by a gauge transformation. The old $B_\mu(x)$ [of Eq. (3.6)] is in the "Lorentz gauge", $\partial^\mu B_\mu = 0$. The new one is in a "surface axial gauge". Both B_μ 's have the same curl. The new B_μ , however, can handle triplets. This follows from the above discussion when one does perturbation theory in λ_0 and returns to the macromolecule analog gas.

We proceed as in Sec. III. The result is

$$\langle \text{tr} \left\{ P \exp \left(i g \oint_A \cdot dl \right) \right\} \rangle = \frac{\int \psi^\dagger \psi \psi^* \exp \left\{ - \int \left[\left| \left(\partial_\mu - i \frac{2\pi}{3} B_\mu \right) \psi \right|^2 + m_0^2 \psi^\dagger \psi + \mathcal{L}(\psi^3, \psi^*, \psi^* \psi) \right] \right\}}{\left[\text{Same term as numerator with } B_\mu = 0 \right]} \quad (4.13)$$

Because of the singular nature of B_μ , the action in the numerator is infinite unless ψ vanishes on S . The solutions are as follows:

(a) When $m_0^2 > 0$ and $\langle \psi \rangle = 0$, $\psi = 0$ is the solution and there is no confinement due to 't Hooft vortices.

(b) When $m_0^2 < 0$ and $\langle \psi \rangle \neq 0$, the equations of motion must be solved with the constraint that ψ vanishes on S . This means that ψ will have non vacuum expectation values near S , the action will go like the area of S , and the potential will grow with r . Spontaneous symmetry breakdown with triplets present gives linear confinement.

We now make some observations:

(i) The effect of choosing a general gauge for B_μ is as follows: Suppose the B_μ of Eq. (4.11) is replaced by $B_\mu + \partial_\mu \chi$. When expanding Eq. (4.13), the macromolecules in a triplet, going from x_0 to x_f , get multiplied by $\exp \left[3i\chi(x_f) - 3i\chi(x_0) \right]$. This transformation is innocuous only if $\chi(x_0)$ and $\chi(x_f)$ are multiples of $\frac{2\pi}{3}$. Because there is a gas of triplets, whose position may be anywhere, general gauges

are not allowed. The arbitrariness in the triplet's location constrains $\chi(x)$. A gauge transformation causes triplets to be multiplied by unwanted phase factors and ruins the Wilson loop calculation. A singular surface gauge must be chosen.

(ii) What happens if a non minimal surface is chosen? Does the action go as the area of this surface? Consistency demands that the physics be independent of S . Suppose another non minimal surface, S' , is chosen. The minimal surface, S , and the non minimal surface, S' , form a closed surface (Fig. 10). Let V be the enclosed volume. Redefine ψ by

$$\psi(x) \longrightarrow \psi(x) \exp \left[i \frac{2\pi}{3} \chi_V(x) \right], \quad (4.14)$$

where $\chi_V(x)$ is the characteristic function for V , that is, $\chi_V(x) = 1$ if x is in V and $\chi_V(x) = 0$ if x is outside V . Plugging into Eq. (4.13) with B_μ given by Eq. (4.11) one sees that

$$\left(\partial_\mu \psi - i \frac{2\pi}{3} B_\mu^{S'} \psi \right) \rightarrow \left(\partial_\mu \psi - i \frac{2\pi}{3} B_\mu^S \psi \right). \quad (4.15)$$

The surface may be moved around by doing a step function gauge transformation where steps occur in multiples of $\frac{2\pi}{3}$. According to (i), such a gauge transformation is allowed. In Eq. (4.15)

$B_\mu^{S'}$ is the B_μ of Eq. (4.11) for the surface, S' , and B_μ^S is the B_μ for the surface S . Thus, for a non minimal surface, the solution is the one for the minimal surface multiplied by a step function phase factor and leads to the same action. The Wilson

loop action again goes as the minimal spanning surface area.

(iii) The solution is periodic in $q = 2\pi n$: when a higher dimensional representation is used so that the effective charge is $q = 2\pi n$, "screening" occurs and the action no longer goes like the area. The Lagrangian

$$\mathcal{L} = |(\partial_\mu - iqB_\mu)\psi|^2 + m^2\psi^*\psi + \mathcal{V}(\psi^*\psi, \psi^3, \psi^{*3}), \quad (4.16)$$

for $q = 2\pi n$ has trivial solutions where the phase of ψ jumps by $2\pi n$ across the surface, S . The singularity in taking the derivative of this phase cancels the singularity in B_μ . Since the phase of ψ is defined modulo 2π , there is no mismatch of phases in ψ when going around the line of a Wilson loop; there are no global difficulties with this solution. For $q = 2\pi n \pm \frac{2\pi}{3}$, a similar procedure yields solutions whose action is the same as $q = \pm \frac{2\pi}{3}$. Quarks will be confined but gluons will be screened. Going back to the original macromolecule partition function, one sees trivially that the Wilson loop is 1 for $q = 2\pi n$. However, it is non-trivial that classical solutions reproduce this phenomenon since saddle point is an approximation. This gives confidence to our methods.

(iv) The solution is virtually independent of $\mathcal{V}(\psi^*\psi, \psi^3, \psi^{*3})$ although a strong potential between vortices may cause them to form dipole-like objects and ruin confinement, an unlikely possibility we feel. Thus confinement is a general phenomenon (almost)

independent of the forces between vortices. However, m^2 being less than (or equal to) zero must be the reason that $\langle\psi\rangle \neq 0$ because of our reliance on the analogy with a gas of loops. By m we mean the physical effective mass which includes the energy per unit length as well as entropy embedding effects. There are other ways in which $m^2 > 0$ but $\langle\psi\rangle \neq 0$ (see the potential of Fig. 11). These potentials result in a liquid-gas type phase transition discussed by Langer²⁶ and Coleman²⁷. Bubbles of true vacuum form rather than a dense spaghetti vacuum. The phase transition occurs via barrier penetration instead of vacuum instability.

(v) The difference between having and not having triplets is the difference between having and not having a Goldstone phenomenon with a broken invariance. Without triplets the Lagrangian of Eq. (4.7) (with $\lambda_0 = 0$) has a global $U(1)$ symmetry, because, as we have noted, topological charge, unable to be created or destroyed, behaves like ordinary charge. Ordinary charge is associated with a $U(1)$ invariance. When spontaneous symmetry breakdown takes place, a Goldstone boson occurs. We conjecture that this massless particle creates long-ranged forces which ruin confinement and lead to only a logarithmic potential, although we cannot explicitly demonstrate this. Contrast this to when triplets occur. Charges in threes are created and destroyed; the symmetry is Z , a discrete group. In this case, no Goldstone boson occurs to disrupt the linear confinement.

Proving triplets exist will be difficult. We know no obvious way to use semiclassical methods to calculate ϵ_0 . It is also easy to overlook such configurations using Bardakci's and Samuel's local field theory formulation^{13, 21} because of the powers of ϵ in eq. (4.9).

(vi) We conjecture that when $\langle \psi \rangle \neq 0$ the symmetry subgroup, H (which is probably $U(1)$), is restored. Along the vortices the Higgs fields are ineffective in breaking H , the gauge fields associated with H are massless, and the symmetry is present. At the vortex the symmetries associated with the generators commuting with "parallel" Higgs fields are not broken. When $m^2 = 0$, the vacuum is filled with vortices, so that, virtually, in every square centimeter of space the symmetry is restored. We conclude that topological symmetry breaking will restore at least a $U(1)$ subgroup of the original color group. Hence, in addition to the linear confinement there will be a logarithmic potential due to these gluons. This logarithmic potential will be analytic in g and calculable via perturbation theory, whereas the topologically generated linear potential is non-analytic in g .

(vii) What do our solutions look like in a Hamiltonian approach? In particular, how do our ideas relate to 't Hooft's and are there any differences?

Let us first reproduce his result that $\text{tr} P[\exp(i\oint_C A \cdot d\ell)]$ creates a region of vacuum with $\langle \psi(x) \rangle = \psi_0 \exp\left(\frac{2\pi i}{3}\right)$ for x inside C (throughout our discussion ψ_0 is the vacuum expectation value of ψ and C is a closed loop contained in a time slice of

three dimensional Euclidean space). In a Euclidean formulation this is seen as follows: consider the propagator of two Wilson loops

$$\langle \text{tr} \left[P \exp \left(- i g \oint_C A \cdot d\ell \right) \right] \text{tr} \left[P \exp \left(i g \oint_{C'} A \cdot d\ell \right) \right] \right\rangle \quad (4.17)$$

$$= \langle \text{tr} \left[P \exp \left(- i g \oint_{C'} A \cdot d\ell \right) \right] \right\rangle \langle \text{tr} \left[P \exp \left(i g \oint_C A \cdot d\ell \right) \right] \right\rangle$$

where C' occurs at a much later time than C (see Fig. 12). To evaluate $\langle \dots \rangle$ (4.17), resort to the "gas of loops" analogy used to calculate a single Wilson loop [See Eq. (3.8)]. Choose B_μ [Eq. (4.11)] to be a sum of two terms, one resulting from using the minimal surface of C and one resulting from the minimal surface of C' . On each of these two surfaces ψ must vanish. For t large, the solution to the equations of motion is approximately the sum of the solutions of each Wilson loop. When the vacuum expectation value of each is subtracted off as in Eq. (4.17), the contribution cancels. There are, however, other surfaces which span a pair of Wilson loops, which are not the union of two surfaces, one for C and one for C' (see Fig. 12). They look like "hour glasses". They occur when $\psi(x) \sim \psi_0 \exp\left(\frac{2\pi i}{\sigma}\right)$ inside the "hour glass", in which case, the singularity in B_μ on C and C' is cancelled and reappears on the "hour glass" surface. The contributions from these do not cancel in Eq. (4.17). The new ψ must vanish on this new surface so that the action goes like the hour glass's surface area.

Classically, the hour glass will try to be small-necked. This, of course, is the instability of a classical closed string. Quantum mechanically, there is a sum over all surfaces, each weighted by its surface area (the Nambu action). Equation (4.17) will result in the propagator of two closed strings. These new solutions yield a ψ with the extra phase factor, $\exp\left(\frac{2\pi i}{3}\right)$, inside. This coincides with 't Hooft's conclusion; the physical interpretation of this process is that $\text{tr}\left[\text{P exp}\left(ig\int_C A \cdot d\lambda\right)\right]$ produces a region of $\psi_0 \exp\left(\frac{2\pi i}{3}\right)$ vacuum which propagates until $\text{tr}\left[\text{P exp}\left(-ig\int_C A \cdot d\lambda\right)\right]$ destroys it. The Wilson loop operator does, indeed, produce regions of topological Z_3 vacuum.

Now let us deal with a quark-antiquark system or equivalently put a Wilson loop in the system. Choose a non-minimal surface, S' , as the surface in the surface axial gauge. By looking at a time slice, we can relate our solutions to a Hamiltonian picture. Figure 13 shows, in this time slice, the curve C' , which is the intersection of S' with this slice, and the curve C , which is the intersection of the minimal surface, S , with this slice. C' and C enclose a region, R . The classical solution has $\psi(x) = \psi_0$ far from R , $\psi(x) = \psi_0 \exp\left(\frac{2\pi i}{3}\right)$ well inside R , and $\psi(x) = 0$ on C . In this sense we reproduce 't Hooft's conclusion; across the line, C , ψ undergoes a continuous phase change. Unlike 't Hooft, we have the extra constraint that $\psi = 0$ on C . This constraint is extremely important.

There is a possible misconception concerning how one might calculate the force between quarks. The following, although it leads to a linear potential, is incorrect; it gives the wrong coefficient in front of r . As 't Hooft noted, there are Bloch walls separating regions of topological Z_3 vacua (Fig. 14a). Their energy per unit length can be calculated by going to one lower dimension ($1+1$) and looking for static solutions, that is, solutions to the equations of motion which depend on x but not on y or t . They look like one dimensional solitons which go to ψ_0 as $x \rightarrow -\infty$ and go to $\psi_0 \exp\left(\frac{2\pi i}{3}\right)$ as $x \rightarrow \infty$ (Fig. ~~14b~~). An incorrect conclusion is that these selfsame solitons form strings between quarks so that the potential is $m_s r$ (m_s is the soliton's mass). As demonstrated the correct solution requires $\psi(x)$ to vanish between the quarks. It is most likely that the Bloch wall solitons do not satisfy this²⁸; ψ can undergo a phase change without going to zero.

What should be done is to find solutions, independent of y and t but x dependent, under the constraint that at $x = 0$, $\psi = 0$ (Fig. 14c). They will have an energy, m_c . The potential between quark and antiquark will be $m_c r$. This m_c differs from m_s since the equations are solved using different constraints and boundary conditions. Although this distinction may seem unimportant, it is crucial in extending our ideas to one higher dimension.

By using a spatially dependent chemical potential, it is seen that $\langle \psi^*(x)\psi(x) \rangle$ is proportional to the density of monomers at x .

Topological spontaneous symmetry breakdown give $\langle \psi^*(x)\psi(x) \rangle$ a classical value indicating that the vacuum is filled with \mathbb{Z}_3 vortices. The quark antiquark solutions have $\langle \psi^*(x)\psi(x) \rangle = 0$ along a string between them. Hence, this region contains no vortices. We would say that the confining string obtains its energy not from being a Bloch wall soliton but by expelling topological vortex vacuum. The essential physical property is not that ψ undergoes a phase change but that it vanish on C .

We conclude that, technically speaking, it is not a Bloch wall soliton which ties the quarks together. The potential energy goes as the distance, because equations of motion are solved with constraints. As pointed out in the introduction, there are no soliton-type vortices to bind quarks in four dimensions; the Bloch wall solitons are surfaces and cannot serve as confining structures. The crucial point, however, is that Wilson loops will still induce constraints that lead to confinement. The naive extension of the 't Hooft $2 + 1$ dimensional model to one higher dimension will yield a similar type of confinement.

(viii) The confinement criterion for baryons is different from mesons²³. What replaces the Wilson loop is (see Fig. 15a)

$$t_{\alpha\beta\lambda} t_{\alpha'\beta'\lambda'} \left[P \exp \left(ig \int_{C_1}^b A \cdot dl \right) \right]_{\alpha\alpha'} \left[P \exp \left(ig \int_{C_2}^b A \cdot d\ell \right) \right]_{\beta\beta'}$$

$$\times \left[P \exp \left(ig \int_{C_3}^b A \cdot d\ell \right) \right]_{\lambda\lambda'}. \quad (4.18)$$

Take C_1 far from C_2 and C_3 . If the action decreases exponentially with area then confinement in baryons occurs. Figure 15a resembles the "dual" of a triplet. Similarly, an arbitrary gauge for B_μ is not possible. One, therefore, might think that even without triplets topological symmetry breakdown would confine quarks in baryons. This is not true, though: First choose a gauge which is singular on three surfaces such as in Fig. 15b.

A gauge transformation, $B_\mu \rightarrow B_\mu + \partial_\mu \chi$, can be performed as long as $\chi(a) = \frac{2\pi}{3} n$ and $\chi(b) = \frac{2\pi}{3} (a$ and b are the endpoints in Fig. 15). Hence, a gauge can be chosen for which B_μ is smoothed out. If a and b are far apart, it can be made to look like the Lorentz gauge in the region far from both a and b . This will lead to only a logarithmic potential. Triplets are needed for baryon confinement also. Concerning the restrictive effect on gauge choice, the difference between baryons and triplets is that the latter form a gas. Endpoints of triplets can be located anywhere; given any space-time point there is a configuration in the statistical ensemble with a triplet vertex there. This forces χ to be a step function in units of $\frac{2\pi}{3}$ everywhere. On the other hand a baryon constrains χ at only two points, a and b .

In the presence of triplets and topological symmetry breakdown baryons will be confined. A gauge for B_μ singular on surfaces (Fig. 15b) must be chosen. One must solve the ψ equations of motion with $\psi = 0$ constraints on surfaces to determine interquark forces.

Assuming constituent quarks are far apart and that the energy goes as the area of the singular surface, we can see how a static baryon looks. In dual string models, there were several speculations: (a) three quarks at the ends of three strings with the other three string endpoints joined (Fig. 16a), (b) one quark in the middle of a single string with the other two quarks at the endpoints (Fig. 16b), and (c) quarks in a triangular string configuration (Fig. 16c). Case (c) cannot occur in our formalism. Case (b) is a special case of (a) when one of the three string has zero length. When does (a) occur and when does (b) occur? As an example of what happens, constrain the third quark to be equidistant from the other two (Fig. 16a). The energy of this configuration is

$$E(d) = m_c \left(2 \sqrt{l^2 + d^2} + y - d \right). \quad (4.19)$$

The notation is as in Fig. 16a and m_c is a constant. The point d is determined by setting $\frac{\partial E}{\partial d} = 0$, for which we find that for $d < \frac{l}{\sqrt{3}}$ the third quark sits in the middle [case (b)], whereas for $d > \frac{l}{\sqrt{3}}$ three strings form [case (a)]. This is reasonable physically: as the third quark moves farther away, energetically it becomes favorable to pull new string out, rather than stretch two strings.

Because of the different baryon string picture, baryon Regge trajectories might behave differently from meson ones. It may be that at low energies the third quark sits in the middle. This would give similar Regge trajectories and slopes. At higher

energies third quark excitations might form causing baryon trajectories to become different from meson ones.

Let us summarize the key points when topological symmetry breakdown in the Hooft vortex charge takes place:

- I. A logarithmic potential is obtained in the absence of triplets and a linear potential is obtained in their presence.
- II. The potential is triality dependent. Representations with integral hypercharge are screened. Representations with fractional hypercharge are confined. The potential is periodic in $1/3$ charge.
- III. Original color symmetries are at least partially restored.

V MONOPOLES

In this section, we shall show how monopoles arise in the Hooft Σ_N model. The important conclusion will be that, in the presence of triplets, the phase transition from $\langle \psi \rangle = 0$ to $\langle \psi \rangle \neq 0$ is a transition from an ensemble of monopole - antimonopole pairs (magnetic dipoles) to a liberated plasma of monopoles and antimonopoles. This phase transition might be compared to that of a two-dimensional Coulomb gas^{10, 29} except that the interaction between monopoles gets changed from a linear one to a non-confining one, most likely a Yukawa, $\frac{1}{r} \exp(-\mu r)$, potential (the expected Coulomb-like $\frac{1}{r}$ potential probably gets screened due to plasma effects^{7, 8}).

Two points need clarifying: In $3+1$ dimensions for an abelian theory, we know what a monopole is: We compute $\int_S \mathbf{B} \cdot d\mathbf{A}$

over a closed surface, S . If $\frac{2\pi}{e}$ results, there is a monopole of strength $\frac{2\pi}{e}$. First, what is a monopole in $2 + 1$ dimensions? Solitons in $3 + 1$ dimensions are instantons in $2 + 1$ dimensional Euclidean space. The monopoles in the 't Hooft model will be instantons (like Polyakov's, for example⁷). Secondly, what is a monopole in an $SU(3)$ gauge theory? Without Higgs fields, we don't know how to precisely define one, but with Higgs fields we can. A classic example is the 't Hooft - Polyakov monopole³⁰ in the Georgi - Glashow model. Recall that 't Hooft defined a gauge invariant $F_{\mu\nu}$ which, in the presence of the "hedgehog" solution, behaved as a monopole field. The Higgs fields were an essential ingredient in $F_{\mu\nu}$. Likewise, we can define an $F_{\mu\nu}$, but only when $H = U(1)$ ³¹ (H , described in Sec. IV, is the subgroup in which the flux points). When this happens, there is a vector, η [Eq. (4.5)], constructed out of Higgs fields which indicates the color direction of vortex flux. This vector transforms in the adjoint representation under gauge transformations. Hence $\eta \cdot F_{\mu\nu}$ will be gauge invariant and can be used to measure the flux³². Normalize η so that $\eta \cdot \eta = \frac{N}{2(N-1)} = \frac{3}{4}$ for $SU(3)$. Definition: There is a monopole of N Dirac units if

$$\int_S B \cdot dA = \frac{2\pi N}{g}, \quad (5.1)$$

where S is a closed surface and $B^i = \frac{1}{2} \epsilon^{ijk} \eta^k F_{jk}^i$.

With this definition triplet configurations are monopole-antimonopole instanton pairs. Perform the measurements, $\int_S B \cdot dA$,

over a sphere enclosing one end of a triplet (Fig. 17a). Since each vortex acts like a delta function of flux, we will measure three contributions of flux of $\frac{2\pi}{g}$ units. The total flux emanating from the endpoint is $\frac{6\pi}{g}$. Triplets are monopole - antimonopole pairs. Each has three Dirac units. Ordinary vortex loops are not monopoles because $\frac{2\pi}{g}$ units enters at one point but exits at another (Fig. 17b).

Without Higgs fields it is difficult to know what constitutes a monopole. It is important to know whether it can be defined in a pure SU(3) gauge theory. Aesthetically, one would like to do away with the Higgs fields. They are only being used as a crutch.

One might try the following. Take a sphere, S . Break it into small regions, R_i . Each region, R_i , has a closed boundary, C_i . Make measurements $\left\{ \text{tr } P \exp \left(ig \oint_{C_i} A \cdot d\ell \right) \right\}$. Define $\frac{1}{g}$ times the phase to be the "flux". Add up all the fluxes to obtain the total flux. Several problems arise: first, θ units of flux is indistinguishable from $2\pi n + \theta$ units. In treating triplets, one might conclude that the first two vortices contribute $\frac{2\pi}{g}$ while the last contributes $\frac{-4\pi}{g}$ so that the total is zero. This procedure would give an incorrect result. Secondly, it is unclear what is being measured because there is no Stoke's theorem for non-abelian gauge theories³³. We are not really measuring the total flux. Because non-abelian flux is not additive like abelian flux, this procedure will almost always yield a non-zero result. Although gauge invariant, it is useless. Another attempt chooses a vector, $V^a(x)$, appropriately normalized. If

$$\frac{1}{2} \int_S V^l \epsilon_{ijk} F_{ij} dS^k \neq 0, \quad (5.2)$$

one might say there is a monopole. For an arbitrary V^l , this is not gauge invariant so that Eq. (5.2) is also meaningless.

Fortunately the Z_N 't Hooft model has Higgs fields and we are able to circumvent these difficulties. We shall return to this point in Sec. VIII.

When $\langle \psi \rangle = 0$ and $m^2 > 0$, the three strings of a monopole carry energy per unit length (these strings are the analogs of superconductor vortices). This means that a monopole must always be paired very closely to its antimonopole partner. Detection of monopoles is difficult unless the sphere, S , in Eq. (5.1) is miniscule. Dipoles will also have little effect on confinement as the calculations of Sec. IV demonstrated. A dipole's three strings can link with a Wilson loop only when it is near the quark trajectory. This produces only a perimeter effect and a mass renormalization. Next consider what happens when $m^2 < 0$ and topological spontaneous symmetry breakdown takes place. The chemical potential per atom (in the macromolecule analogy) is negative and vortices of large size are favored. One can see that the monopole-antimonopole constituents in a dipole are liberated by considering a triplet with endpoints far apart. Focus on one of the monopoles. There are three flux strings emanating from it. Unlike the previous case, these strings do not head directly for the partner antimonopole. Instead they move in arbitrary directions.

The partition function (or the Feynman path integral in a particle dynamics description) sums over all these directions. Consider measuring the flux, $\int_A \mathbf{B} \cdot d\mathbf{S}$, flowing through a small disk, A , of a sphere enclosing the monopole (see Fig. 17a). In doing this measurement only a fraction of the time will a string pierce A . This averaging effect spreads the flux. Symmetry demands we measure $\frac{6\pi}{8} \frac{A}{4\pi r^2}$, where A is the area of the disk and r is the radius of the sphere. In short, averaging over all vortex paths quantum mechanically spreads out the flux so that a "normal" monopole with a radial magnetic field is observed. Since the action of the gauge fields is $-\frac{1}{4} \int F_{\mu\nu}^2 d^3x$, we expect³⁴ that these monopoles will interact very much like ordinary ones, with Coulomb-like potentials. In a semiclassical approximation to the Lagrangian of Eq. (4.7), the monopole's activity is $\frac{1}{\sqrt{3!}} \lambda_0 \langle \psi \rangle^3$. As it should, the monopole's density goes to zero when the vacuum expectation value of ψ goes to zero. This picture of confinement in the $2+1$ Z_N model is similar to Polyakov's instanton one⁷.

VI FROM $2+1$ TO $3+1$

This section extends the ideas in $2+1$ dimensions to $3+1$ dimensions, relates the $2+1$ model to the $3+1$ one, and indicates that proof of confinement in $2+1$ is probably sufficient to prove confinement in $3+1$. This means that the $2+1$ Z_N model is more than just a toy laboratory. It is important to calculate the vortex properties and find the effective Lagrangian.

Once these are known, we will probably know whether gauge theories in $3 + 1$ dimensions confine via the Z_N mechanism.

Just as instantons in $1 + 1$ dimensions are solitons in $2 + 1$, the Z_N solitons in $2 + 1$ are strings in $3 + 1$. The vortex solution, which is x and y dependent in Euclidean $1 + 1$, becomes t independent in $2 + 1$ and t and z independent in $3 + 1$. It respectively looks like a point, a line, and a sheet. The latter two manifest themselves as loops and closed surfaces and their associated quanta are particles and closed strings. The ideas in Euclidean $2 + 1$ are relevant for $3 + 1$ dimensions because a time slice of $3 + 1$ looks like Euclidean $2 + 1$. The solitons in $2 + 1$ which were particles tracing out trajectories now become strings tracing out surfaces. Hence, in the physical world the topological objects are closed strings manifesting themselves in Euclidean space as " Z_N surface solitons."

The key defining property is linking number. In $2 + 1$ dimensions two non-intersecting loops can link and linking number is well-defined for oriented curves. In $3 + 1$ dimensions a closed surface and a loop can link. Again, for oriented surfaces and oriented loops the linking number is well-defined. Figure 18 illustrates using "time lapse photography" how a sphere³⁵ and a loop can link. Because of this, a Z_N topology characterizes "surface solitons". It works just like it does in one lower dimension. An idealized, that is infinitely thin, surface soliton, S , satisfies

$$\left[\text{tr } P \exp \left(i g \oint_C \mathbf{A} \cdot d\mathbf{l} \right) \right] = 3 \exp \left(\frac{2\pi i}{3} \right), \quad (6.1)$$

whenever C links with S . Far from the surface where potentials are pure gauge, Eq. (6.1) is valid. Near the surface of a smeared or physical surface soliton, Eq. (6.1) is incorrect. We shall always use "idealized" solitons. Whenever a loop, C , links with S , there is a map from C into the gauge group (as discussed in Sec. IV). These maps are characterized by $\prod_1(SU(3)/Z_3) \cong Z_3$. Eq. (6.1) says this map is a non-trivial element of \prod_1 . Similarly, the topology can be discussed in terms of Higgs fields. Far from the surface, the Higgs fields take on values in the minimum, M , of the potential. Moving along C traces a closed curve in M . Again, these are characterized by $\prod_1(M) = Z_3$. As long as C and S are kept away from each other, there is no way to unlink them. Likewise, as one moves C , the topological element $\prod_1(M)$ or $\prod_1(SU(3)/Z_3)$ cannot jump since such continuous movements are homotopies. The topology is virtually the same as in one lower dimension.

The topological surfaces will have properties similar to the $2 + 1$ dimensional case:

(i) On the surface the Higgs fields become "aligned" so that at least one generator commutes with them. We expect only one generator, $\eta^{\ell} \lambda^{\ell}$.

(ii) For "idealized" surfaces, there is a delta function contribution to the flux in the η direction. Let $e^{(1)}(x)$ and $e^{(2)}(x)$ be two orthonormal tangent vectors to the surface at x . Then

$$\eta^\ell(x) e_\alpha^{(1)}(x) e_\beta^{(2)}(x) \frac{1}{2} F_{\mu\nu}^\ell(x) \epsilon_{\alpha\beta\mu\nu} = \frac{2\pi}{g} \delta_1^2(x). \quad (6.2)$$

The vector, η^ℓ , is normalized so that $\sum_\ell \eta^\ell \eta^\ell = \frac{3}{4}$ and $\delta_1^2(x)$ is a delta function in the variables perpendicular to the surface.

More accurately,

$$F_{\mu\nu}^\ell(x) = \frac{2\pi}{g} \oint_{\substack{\text{over} \\ \text{surface} \\ \text{soliton}}} \frac{4}{3} \eta^\ell(y) \delta^4(x-y) \frac{1}{2} \epsilon_{\alpha\beta\mu\nu} dS_{\alpha\beta}. \quad (6.3)$$

(iii) The mass of the soliton in $2+1$ is (roughly) the energy per unit length of the string. The engineering dimensions of the parameters that give the mass in $2+1$ have dimensions of mass squared in $3+1$; couplings acquire different dimensions in different dimensions. For example, the mass of the Nielsen-Olesen vortex due to the Higgs potential goes like $\langle \phi \rangle^2$.⁵ In $2+1$, $\langle \phi \rangle^2$ has dimensions of mass, whereas in $3+1$, it has dimensions of mass squared.

Linking number formulas in four dimensions also exist.

Reference 17 gives

$$n = \frac{1}{2\pi^2} \int_{x \in S} \oint_{y \in C} dy_\beta \frac{1}{2} \epsilon_{\alpha\beta\mu\nu} \frac{(y-x)_\alpha}{|y-x|^4}$$

$$\begin{aligned}
 &= \frac{1}{2\pi^2} \int_0^1 d\sigma \int_0^1 d\tau \int_0^1 ds \frac{1}{2} \epsilon_{\alpha\beta\mu\nu} \frac{[y(s)-x(\sigma,\tau)]_\alpha}{|y(s)-x(\sigma,\tau)|^4} \frac{dy_\beta(s)}{ds} \frac{dx_\mu(\sigma,\tau)}{d\sigma} \frac{dx_\nu(\sigma,\tau)}{d\tau} \\
 &= \oint_S \frac{1}{2} B_{\mu\nu}(x) dS_{\mu\nu}(x) , \tag{6.4}
 \end{aligned}$$

where

$$B_{\mu\nu}(x) = \frac{1}{2\pi^2} \oint_C \epsilon_{\alpha\beta\mu\nu} \frac{(y-x)_\alpha}{|y-x|^4} dy_\beta . \tag{6.5}$$

In Eq.(6.4) S is a closed oriented surface, $dS_{\mu\nu}$ is the surface element, C is a closed oriented curve, dy_β is its line element, and $\frac{1}{2\pi^2}$ is the volume of the three dimensional unit sphere. In the second form, the curve, C , is parametrized by s so that $y(s)$ is the location at "time", s , $y(0) = y(1)$, and $\frac{dy_\beta}{ds}$ is the "velocity". The variables σ and τ parametrize the surface in the same manner as in dual string theory. They take on values in the unit square, and $x(\sigma,\tau)$ is the location of the surface at that (σ,τ) value. Because the surface is closed, $x(0,\lambda) = x(1,\lambda) = x(\lambda,0) = x(\lambda,1) =$ a constant for $0 \leq \lambda \leq 1$, that is, the boundary of the square is mapped to the same point. The linking number is an integer.

If one were doing static electromagnetism in $4 + 1$ dimensions, the $B_{\mu\nu}(x)$ of Eq. (6.5) would be the magnetic field at x produced by a unit current flowing through C . It has a "vector field" gauge invariance:

$$\bar{B}_{\mu\nu}(x) + B_{\mu\nu}(x) + \partial_\mu \chi_\nu(x) - \partial_\nu \chi_\mu(x). \quad (6.6)$$

Under the transformation of Eq. (6.6), Eq. (6.4) is unchanged. This means that there are many other permissible forms for $B_{\mu\nu}$.

There is also a gauge invariance for the gauge since $\chi_\mu \rightarrow \chi_\mu + \partial_\mu \chi$ leaves Eq. (6.6) invariant.

The vacuum will be a gas of Z_N closed surfaces. This will be an interesting statistical mechanics ensemble. Such a gas might yield a field theory for strings for which the methods of sections II, III, and IV could be mimicked. Even in the absence of interactions between points on the surface, where a free field theory of closed strings is expected, constructing such a field theory will be difficult³⁶. We are, therefore, unable to calculate the Wilson loop. Such a calculation involves summing over the gas of surfaces, weighting each surface, S , by the factor

$$\exp \left[\frac{2\pi i}{3} \int_S \frac{1}{2} B_{\mu\nu}(y) dS_{\mu\nu}(y) \right], \quad (6.7)$$

with $B_{\mu\nu}$ given by Eq. (6.5) for C being the Wilson loop. Instead, we make the following observations and conjectures: Assume that the soliton in $2+1$ dimensions has a positive mass so that in $3+1$ the surfaces carry a positive surface energy. The topological sector generates a theory of strings. Such closed strings do not carry quantum numbers. They occur as space-time events: a point sudden appears, sketches into a ringlet, and shrinks away. These ringlets are restricted to be small and to last for brief

durations because of the large surface action generated. An exception to this might be the tachyons in the dual string model. These two low lying states, as minute closed strings, resemble particles of negative mass squared. They trace out long trajectories which because there is no conserved quantum number can begin or end in the vacuum. As long thin cylinders, they have little effect on Wilson loops except for possibly a mass renormalization; such thin tubular configurations are unlikely to link. All in all, such a system does not produce confinement.

Because of the Z_3 structure of the topology there may exist other types of surfaces. These are the analogs of the $2 + 1$ triplets. Figure 19 shows a "triplet surface" imbedded in the slice, $z = 0$. The temporal evolution of this configuration is shown in Fig. 20. Time slices sometimes yield triplets and sometimes ringlets. These string configurations are very different from interacting dual strings. The latter interact by breaking or by joining ends as well as a "four point" interaction where two strings touch in the middle and exchange string halves.

These interactions occur at a specific location at a particular time. Figure 20 shows that triplet surfaces look like an open string circumscribed by a closed one or more accurately three open strings joined at the ends; the three are in constant interaction. Thus, this string theory is unlike anything previously considered in dual models.

Triplet surfaces contain, of course, monopoles. The intersection of the three surfaces is a curve which is to be identified with the monopole's trajectory. Like closed soliton surfaces, triplet surfaces must also be assigned an orientation. This

induces an orientation on the boundary, that is, for the monopole loop³⁷. Hence, as should be the case, the monopole trajectories are oriented and closed, indicating a conserved quantum number (monopole charge). Since three surfaces span the monopole loop, the system at a particular time looks like a monopole-antimonopole pair joined by three vortices (Fig. 8). Because these three vortices carry a positive energy per unit length, the monopole-antimonopole pair are inexorably bound by a linear potential. Triplet surfaces, comprised of the monopole loop and three spanning surfaces, must be small and hence monopole vacuum loops are rare events. These "neutral" objects have few physical effects. As in one lower dimension, there is no confinement. The interesting case of topological symmetry breakdown, where confinement is expected, will be discussed shortly. We first must show how linking number can be defined for triplet surfaces.

Equation (6.4) no longer works for triplet surfaces. This is the higher dimensional analog of the problem discussed in Sec. IV. The resolution is similar: a singular "surface axial gauge" for $B_{\mu\nu}$ must be used. Let S_C be any surface which spans the Wilson loop, C. Then

$$B_{\mu\nu}^{S_C}(x) = \int_{S_C} \frac{1}{2} \epsilon_{\mu\nu\alpha\beta} \delta^4(x-y) dS_{\alpha\beta}(y), \quad (6.8)$$

works. This $B_{\mu\nu}^{S_C}$, when substituted into Eq. (6.7) and integrated over the triplet surface, S, yields the correct phase factor. The surfaces S_C and S should not be confused; the former is any

surface whose boundary is the Wilson loop while the latter is the triplet surface. It is not hard to find a gauge transformation which moves S_C : Let S_C^1 and S_C^2 be two Wilson loop spanning surfaces. They form a closed surface. Let V be any volume, i.e., a three-dimensional manifold, whose boundary is their union. For $y \in V$, let $n_\mu(y)$ be the vector orthonormal to V at y in four space. Then

$$\begin{aligned} \chi_\mu(x) &= \int_{y \in V} n_\mu(y) \delta^4(x-y) d^3y \\ &= \int_{y \in V} \frac{1}{3!} \epsilon_{\mu\alpha\beta\gamma} \delta^4(x-y) d^{\alpha\beta\gamma}(y), \quad (6.9) \end{aligned}$$

when used in Equation (6.6) affects a gauge transformation from $B_{\mu\nu}^{S_C^1}$ to $B_{\mu\nu}^{S_C^2}$. Actually, we have not defined how the sign of χ_μ is to be chosen. One choice gauges $B_{\mu\nu}^{S_C^1}$ into $B_{\mu\nu}^{S_C^2}$; the other gauges $B_{\mu\nu}^{S_C^2}$ into $B_{\mu\nu}^{S_C^1}$. The fact that V is not unique (there are many volumes, V , whose boundary is $S_C^1 \cup S_C^2$) reflects the fact that the gauge transformation is not unique ("a gauge invariance for the gauge", as previously discussed).

What would happen if a general gauge was chosen in Eq. (6.7)?

Take the $B_{\mu\nu}$ of Eq. (6.8) and perform the gauge transformation

$B_{\mu\nu} \rightarrow B_{\mu\nu} + \partial_\mu \chi_\nu - \partial_\nu \chi_\mu$. Then, each of the three surfaces comprising a triplet surface would contribute an extra factor

$$\int_S (\partial_\mu X_\nu - \partial_\nu X_\mu) \mathcal{L}_{\mu\nu}$$

$$= \oint_{\partial S} X_\mu dx^\mu = \oint_{C_M} X_\mu dx^\mu . \quad (6.10)$$

The boundary of each surface is C_M , the monopole loop. Triplet surfaces would get multiplied by the unwanted phases

$$\exp \left(3i \oint_{C_M} X_\mu dx^\mu \right) . \quad (6.11)$$

This is innocuous if $\oint_{C_M} X_\mu dx^\mu = \frac{2\pi}{3} n$. For a gas of triplet surfaces, we must require that

$$\oint_C X_\mu dx^\mu = \frac{2\pi}{3} n , \quad (6.12)$$

for all loops, C . By shrinking C to a point in a plane, P , Eq. (6.12) implies that $\partial_\mu X_\nu - \partial_\nu X_\mu$ must be singular on the plane, P^\perp , perpendicular to P . Hence gauge transformations can only move S^C around and Eq. (6.8) is the most general form for $B_{\mu\nu}$; one cannot smooth $B_{\mu\nu}$ out.

The interesting case is when topological symmetry breakdown occurs, that is, when the vortices in $2+1$ dimensions have a negative (or perhaps zero) mass squared. Then, topological surfaces are expected to have a negative surface action density and will populate the four-dimensional world. This implies that the closed topological strings will have a negative Regge slope, a thought that, at first,

seems preposterous because of the infinite number of vacuons.

However, the situation is not as bad as it appears and even has a simple physical interpretation. First of all, a total collapse is not expected. The same repulsive vortex forces which stabilize the proliferation of closed loops in $2 + 1$ dimensions will be present in $3 + 1$ dimensions and will stabilize the vacuum³⁸. Secondly, triplet surfaces will become monopoles with their magnetic flux spread out. Consider, for example, a large monopole loop with its three topological Z_3 surfaces. Unlike the case, $m^2 > 0$, where the surfaces are the ones of minimal surface area, the surfaces can be anywhere. The quantum mechanical sum over all possibilities will make the flux evenly distributed rather than being focused in tubes. We conjecture that a negative slope parameter for these types of strings results in "fields". By this, we mean a string theory gets transformed into a field theory. There is another example of this phenomenon which should clarify what we mean, namely Wilson's lattice gauge theory³⁹. Keep the lattice spacing finite. In the strong coupling limit the electric field is focused into tubes. States of two quarks connected by an electric flux tube or a torus of flux are permissible. The theory has strings with positive slope parameters. As the coupling constant is lowered, the electric flux begins to spread out more and more until a phase transition occurs where it spreads out uniformly in the usual Coulomb-like manner. At the phase transition, spontaneous symmetry breakdown of electric strings has occurred. The effective surface action density (including entropy contributions) has become zero. We conjecture that the phase transition

from $\langle \psi \rangle = 0$ to $\langle \psi \rangle \neq 0$ is the same phenomenon except with "dual" magnetic fields (see Table 1). This is what we mean by strings of negative Regge slope being metamorphosed into fields. When this happens topological surfaces become magnetic fluctuations and monopoles bound in dipoles become liberated. These monopoles will now be instrumental in confining quarks. If $m^2 < 0$ for 't Hooft vortices and triplets exist in $2 + 1$ dimensions, then in $3 + 1$ dimensions monopoles, previously confined in monopole-antimonopole pairs by magnetic flux tubes, get liberated resulting in a monopole plasma.

We conjecture that the phenomena exhibited in $2 + 1$ dimensions will be present in $3 + 1$. Already discussed are magnetic fluctuations due to close surface solitons and monopoles due to surface triplets. Also guaranteed is that the non-perturbative potential will be "periodic in charge" because of Eqs. (6.4) and (6.7). This screening phenomenon means that integral hypercharge multiplets will not be confined. Any approximation scheme to a Wilson loop calculation should be able to reproduce this phenomenon. Next, we expect that topological symmetry breakdown will restore the symmetry associated with H . The reason is the same as in one lower dimension: the Z_N surfaces will fill up the vacuum until overlap repulsive forces take over. Since H is restored on these surfaces, H will be restored virtually everywhere. Finally, several arguments show why confinement in $3 + 1$ dimensions will occur:

(i) The vacuum is a gas of monopoles. Roughly speaking, such a system confines because of a "dual Meissner effect"^{8,40}.

Just as a gas of current loops confines monopoles in a superconductor, a gas of magnetic current loops confines charges.

(ii) Although there is now no interpolating soliton as there was in $2 + 1$ dimensions, confinement is still expected. Physically, surface solitons are expelled from a region (the surface, S_C) between the quarks. This creates a tube of "abnormal vacuum" which carries energy. A linear potential results.

(iii) The Wilson loop calculation involves a system with an area constraint (the "surface axial gauge" constraint of Eq. 6.8). If a non-trivial situation exists, the action must go as the area.

(iv) A time slice of $3 + 1$ looks like $2 + 1$ Euclidean space. Confinement in $2 + 1$ might indicate confinement in $3 + 1$. Putting a quark loop in a time slice appears to reduce the calculation to one lower dimension. Let us define the term, dimensional reduction, as when a Wilson loop in such a time slice of $3 + 1$ can be calculated in $2 + 1$ as a static situation. Dimensional reduction does not always occur. Consider, for example, the lower dimensional analog: a comparison of the $1 + 1$ instanton gas of Nielsen-Olesen vortices to the Nielsen-Olesen model in $2 + 1$ with vortex topological symmetry breakdown. The former looks approximately like a time slice of the latter. However, Sec. III obtained a logarithmic potential for the $2 + 1$ model, whereas a linear potential occurs in $1 + 1$ dimensions. The reason for this is clear. When the Wilson loop is placed in a time slice, the P_μ of Eq. (3.6) is non-zero for all times. The Wilson loop affects the vortex gas throughout the $2 + 1$ dimensional world and not just at one time.

Dimensional reduction does not happen. If, however, B_μ were zero outside the time slice it would happen. This is the case for the 't Hooft model with triplets. The B_μ of Eq. (4.12) has its support in the time slice containing the Wilson loop. This is why linear confinement occurs: The calculation dimensionally reduces to the $1 + 1$ instanton calculation, which is known to confine.

Returning to the physical world with triplet surfaces absent, the $B_{\mu\nu}$ of Eq. (6.5) is non-zero for all times. For this situation, the calculation does not dimensionally reduce. If $B_{\mu\nu}$ were forced to be in a gauge with the support of $B_{\mu\nu}$ in a time slice, then topological symmetry breakdown would yield the logarithmic potential obtained in $2 + 1$. But this is not the case. Our guess is that the action will go like $\int P_{\mu\nu}^2 d^4x$ so that closed surfaces yield only a $1/r$ potential. When there are triplet surfaces the situation is completely different; $B_{\mu\nu}$ is forced to be in a "surface axial gauge" such as Eq. (6.8), the calculation dimensionally reduces, and a linearly confining potential (the same one as in Sec. IV) is obtained. With triplet surfaces present, a time slice of the real world does indeed look like the $2 + 1$ Z_N model with triplets and confinement.

VII RELATION TO MANDELSTAM'S SCHEME

There is a similarity between Mandelstam's confinement scheme⁸ and topological symmetry breakdown with triplet surfaces. We shall touch upon the common points and differences. Here is a quick review of his quark confinement:

(a) The Coulomb gauge contains a term $E_L^\alpha \cdot E_L^\alpha$ in the Hamiltonian where

$$E_L^\alpha = \nabla \left[\nabla^2 \delta^{\alpha\gamma} - g f^{\alpha\beta\gamma} A_i^\beta \nabla_i \right]^{-1} \rho^\gamma \quad (7.1)$$

The inverse operator, $\left[\nabla^2 \delta^{\alpha\gamma} - g f^{c\beta\gamma} A_i^\beta \nabla_i \right]^{-1}$, akin to $|ax-b|^{-1}$ in a one-dimensional quantum system, will produce infinities unless the vacuum is suitably chosen.

(b) In the $A_z = 0$ axial gauge, this problem becomes equivalent to whether local (x, y) dependent (but z -independent) gauge transformations annihilate the vacuum. The bare vacuum fails to do this and makes a poor starting point to perturb around.

(c) A vacuum comprised of monopoles has the right properties: the Dirac tubes (in $A_z = 0$ gauge, the Dirac strings become tubes for finite-sized smeared-out monopoles) produce random gauge rotations. Such a state will be a singlet under local (x, y) dependent gauge transformations. It is a candidate for the vacuum state.

(d) Such a monopole gas confines in not so different a way from the Polyakov model⁷.

In short, Mandelstam's gas results in (i) a restoration of $SU(3)$ gauge symmetry and (ii) quark confinement.

Likewise, we have shown similar results when topological symmetry breakdown occurs in the presence of triplets: a monopole vacuum is generated and at least a global $U(1)$ color symmetry is restored. We have been unable to prove the restoration of the complete $SU(3)$ gauge invariance. Like Mandelstam's vacuum, confinement is a consequence. The ideas of Sections IV, V, and VI neatly jell with Mandelstam's.

We have given support to Mandelstam's work in showing how

monopoles naturally arise in an $SU(3)$ gauge theory. The type of monopoles Mandelstam has been using (Wu-Yang ones⁴¹) are probably different from those in an $SU(2)$ gauge theory. For $SU(3)$ our monopoles carry $\frac{6n}{g}$ units of flux. We have also elucidated on the dynamics of the system; in particular, how negative Regge slopes transform magnetic flux tubes into magnetic fields. We have not shown why topological symmetry breakdown takes place. If Mandelstam is correct then he has given us the reason: such a breakdown occurs because of the $|ax-b|^{-1}$ problem. It is a matter of symmetry, monopoles, and disorder. In essence, the difference between no confinement and confinement is the difference between an ordered system with a broken symmetry and a disordered system with monopoles acting as the symmetry restoring agent.

VIII OPEN QUESTIONS

A. How does the phase diagram look for a 't Hooft $SU(N)$ gauge theory? In particular, how many phases are there? In Figure 4, we have drawn three phases, corresponding to (a) a normal boson potential, (b) a Higgs boson potential, and (c) a Higgs soliton potential. 't Hooft conjectured that phases (a) and (c) are the same. He argues that the soliton's mass squared must be proportional to the Higgs mass squared, the proportionality sign being negative since when the Higgs bosons are tachyonic the solitons have a positive mass. Analyticity implies that the solitons are tachyons

when the bosons have a positive mass squared. We feel this argument needs further justification since phase transitions induce non-analyticities. This remains an open question. In general, it is important to determine the phase diagram so that the confinement phase (if it exists) and the coupling constant values which yield this phase can be found⁴².

B. Can a string field theory be constructed to describe the statistical ensemble of topological surfaces? As previously discussed, with such a construction the calculations of Sections II, III, and IV could be performed in $3 + 1$ dimensions.

C. What effects do instantons have? Callan, Dashen, and Gross have shown that instantons create a paramagnetic vacuum which tends to expel electric fields². This will surely affect the dynamics of topological surfaces. Instantons might aid in (or even cause) the topological symmetry breaking. Roughly speaking, the surfaces should couple to instantons because monopoles and associated magnetic fields are long-ranged. They can contribute to

$$\int \Gamma_{\mu\nu} \tilde{F}^{\mu\nu} d^4x, \quad (8.1)$$

in contrast to short-ranged field configurations which cannot due to the fact that Eq. (8.1) can be written as a surface integral.

D. Are Higgs bosons necessary? As indicated in

Sections III and IV, topological vortices can be characterized solely in terms of gauge potentials by using path ordered products. This suggests that, perhaps, they exist independently of the Higgs fields. We feel this may be the case for an $SU(N)$ gauge theory but certainly not for a $U(1)$ theory. The Higgs bosons in the $2 + 1$ dimensional Nielsen-Olesen model serve two important purposes:

(i) they smooth out the short distance singularities which would otherwise occur.

(ii) they ensure that $\exp\left(i e \int_x^y A \cdot d\ell\right)$ returns to 1 when y loops around the vortex and returns to x .

Purpose (i) is not as important as purpose (ii) for the existence of vortices. Purpose (i) is a short distance phenomenon which might be cured quantum mechanically or through renormalization.

In the absence of Higgs fields, there are still finite energy vortex configurations (obtained by using the classical values of $A_\mu(x)$ when the Higgs fields were present) although they are not energy minima. In contrast purpose (ii) is essential. With Higgs bosons, the Higgs fields must take on values in the minima of the potential and be covariantly constant far from the vortex. This means that $\exp\left(i e \int_x^y A \cdot d\ell\right)$ is precisely

the rotation (in this case, phase) which transforms the classical Higgs field at x to its value at y .

As one goes around the loop and returns to x , this guarantees that $i \oint A \cdot d\ell = \frac{2\pi n}{e}$. This means that magnetic flux gets quantized. Without Higgs fields, configurations such as $A_\phi(x) = a(\rho)/\rho$, with $a(\rho) = 0$ at the vortex and $a(\rho) = \frac{c}{e}$ far away, have finite energy. Here, c is an arbitrary constant. This means that flux is not quantized and $\int A \cdot d\ell = \frac{2\pi c}{e}$ is arbitrary. The solitons lose their identity as tubes of conserved quantized flux. Instability occurs because tubes of $\frac{2\pi}{e}$ units can dissolve into many smaller tubes, say n tubes of $\frac{2\pi}{en}$ units. Thus purpose (ii) is the essential stabilizing effect of Higgs bosons in the Nielsen-Olesen $U(1)$ model.

In the non-abelian case, purpose (i) still functions but it is possible that property (ii), which is now modified to

$$\left[P \exp(i g \oint_C A \cdot d\ell) \right]_{\alpha\beta} = \quad (8.2)$$

[a center of the group element] $_{\alpha\beta}$,

for loops, C , which encircle the vortex, holds even in the absence of Higgs fields. This is because Yang-Mills theories are self-interacting. These self-interactions possibly act as a replacement for the Higgs

boson - gauge field interactions. If a flux tube, at some instance, points in the z direction, thus contributing to $r^{\ell} B_z^{\ell}$, two of the three gauge potentials might act like the two Higgs fields while the third is the flux generating gauge potential. This point needs further clarification, but Yang-Mills theories offer the aesthetically pleasing possibility of eliminating Higgs fields without ruining any of the physical results discussed in this paper.

If the Higgs fields are eliminated, the problem of defining a monopole returns. Previously, the vector, r^{ℓ} , constructed out of Higgs fields, was used. We now must manufacture an η^{ℓ} using gauge potentials. There are many ways of doing this: at each point, x , attach a closed loop, C_x . Define

$$\left[\exp \left(\eta^{\ell}(x) \frac{\lambda^{\ell}}{2} \right) \right]_{\alpha\beta} = \left[P \exp \left(\oint_{C_x} A \cdot d\vec{t} \right) \right]_{\alpha\beta}. \quad (8.3)$$

In Eq.(8.3) the path ordering starts at x , proceeds along C_x , and ends at x . Beginning at any other starting point is not possible. The vector, $\eta^{\ell}(x)$, transforms in the octet representation under gauge transformations. Using the magnetic field, $B_i(x) = \frac{1}{2} \epsilon_{ijk} F_j^{\ell}(x) F_k^{\ell}(x)$, one can "test" for monopoles by integrating $\int B \cdot dS$ over closed surfaces. Of course, this method generates

an infinite number of magnetic fields, all of which are gauge invariant. Only a prudent choice of the C_x will yield a $B_i(x)$ with the desired properties, that is, that this magnetic field be the one contained in vortex flux tubes and be the one associated with the monopoles contained in triplet surfaces.

E. Is there a non-abelian dual group? Implicit in our discussion is the dualities between magnetic fields and electric fields, monopoles and charges, etc. Table 1 illustrates some of these. When there are Higgs bosons and a unique η^2 , i.e. $H = U(1)$, the symmetry group of the surfaces (and monopoles) is either $U(1)$ or Z_3 depending on whether triplets exist. If the Higgs fields can be eliminated will there be a non-abelian dual group? This question will be hard to answer because of the quantum mechanical effects inherently associated with an internal symmetry group, namely, color zitterbewegung²³. This goes beyond our discussions which have been classical in character.

IX SUMMARY

To prove confinement the following must be shown:

- (a) that the 't Hooft Z_N solitons in $2 + 1$ dimensions have a phase with $m^2 < 0$ and that there are repulsive forces which stabilize the vacuum.
- (b) that $\lambda_0(\psi^3 + \psi^{*3})$ (or similar) terms exist, that is, triplet are present.
- (c) that a time slice of the four dimensional world is described by the physics of the $2 + 1$ dimensional model.

When the above are satisfied, we expect

- (i) confinement
- (ii) restoration of some or all of the original color symmetries.

Without (b) topological symmetry breakdown gives only a $\ln r$ potential in $2 + 1$ and (most likely) a $\frac{1}{r}$ potential in $3 + 1$. Showing (a) and (b) is the next calculational step. These problems can be approached by using:

- (i) Semi-classical methods⁴³. These can determine how the soliton's mass depends on the parameters in the Higgs potential.
- (ii) A field theory for solitons^{13,21}. This should be useful in determining the forces between vortices and whether the vacuum stabilizes for $m^2 < 0$.
- (iii) Mandelstam's operator methods⁸. These should be applied to Z_N type monopoles. These methods might be useful in determining vacuum instabilities and hence why the soliton's mass is negative.
- (iv) Halpern's dual field strength formulation⁴⁴. This approach might exhibit the topological solitons and their properties directly. It might also be helpful in determining whether vacuum instabilities exist.

Most difficult will be showing (b), that triplet configurations exist, although there is no a priori reason (in the sense of a conservation law) why they shouldn't.

Additional problems particular to four dimensions are:

- (a) Whether the closed soliton strings have negative Regge slopes and whether this makes sense as this paper suggests.
- (b) Whether triplet surfaces occur. A new type of dual string model is needed.
- (c) Whether the Higgs potential can be done away with. Are there singular but stable topological solitons in pure Yang-Mills theories?
- (d) If (c) holds, there are no mass scales. How does dimensional transmutation⁴⁵ come about?

We wish to emphasize the following points:

- (a) The non-abelian (and to a lesser extent non-perturbative) nature of the confinement. The selfsame method can at most yield a $\ln r$ potential for a U(1) theory. This is because monopoles, i.e. triplets, cannot occur in an abelian theory and they are essential in the linear confinement.
- (b) The potential between quarks is expected to have (i) a linear piece due to triplet surfaces, (ii) most likely a $1/r$ non-perturbative large r piece due to closed surfaces, and (iii) a (possibly screened or antiscreened) perturbative piece which dominates the short distance physics.

Let us conclude by summarizing what we have shown:

- (i) In $2 + 1$ dimensions we have exhibited how to do calculations when topological symmetry breakdown occurs by using a macromolecule analogy. We believe this will form the prototype of future topological symmetry breakdown calculations.
- (ii) We have extended 't Hooft's confinement scheme in $2 + 1$ to $3 + 1$. Previously, only an operator algebra yielding different phases was obtained.
- (iii) We have discussed the dynamics of quark confinement, namely, how monopoles naturally arise in a non-abelian gauge theory; how negative Regge slopes make sense, change Z_N strings into magnetic field fluctuations, generate the usual $1/r$ radial monopole magnetic field, and liberate the monopoles (of the bound monopole-antimonopole pairs) which are so instrumental to quark confinement.
- (iv) We have connected the physics of the Z_N models to Mandelstam's confinement scheme.

In short, we believe we are at the verge of proving confinement in non-abelian gauge theories.

REFERENCES

1. G. 't Hooft, Nucl. Phys. B138, 1 (1978).
2. C. G. Callan, R. F. Dashen, and D. J. Gross, A Theory of Hadronic Structure, Princeton Preprint (June, 1978).
3. E. Witten, Instantons, The Quark Model, and the $1/N$ Expansion, Harvard Preprint HUTP-78 (Oct., 1978).
4. T. Yoneya, Nucl. Phys. B144, 195 (1978).
A. Casher, Self Dual $Z(N)$ Gauge Theories, DESY Preprint (1978).
5. H. B. Nielsen and P. Olesen, Nucl. Phys. B61, 45 (1973).
6. T. Banks, R. Myerson, and J. Kogut, Nucl. Phys. B129, 493 (1977).
7. A. M. Polyakov, Nucl. Phys. B121, 429 (1977).
8. S. Mandelstam, Confinement in Non-Abelian Gauge Theories, talk given at the Washington Meeting of the American Physical Society, (April, 1977).
9. M. Stone and P. R. Thomas have used such an analogy in Phys. Rev. Letts. 41, 351 (1973). We prefer to rederive their result in a lattice independent way.
10. S. Samuel, Phys. Rev. D18, 1916 (1978).
11. Two reviews are K. F. Freed, Adv. Chem. Phys. 22, 1 (1972) and F. W. Wiegel, Phys. Reports 16c, 59 (1975). References to earlier work can be found in these two.
12. M. B. Halpern and P. Senjanovic, Phys. Rev. D15, 1655 (1977).
M. B. Halpern and W. Siegel, ibid. D16, 2486 (1977).
M. B. Halpern, A. Jevicki, and P. Senjanovic, ibid. D16, 2476 (1977).
13. K. Bardakci and S. Samuel, Phys. Rev. D18, 2849 (1978).

14. R. J. Rubin, *J. Chem. Phys.* 20, 1940 (1952); *J. Math Phys.* 8, 576 (1967). S. F. Edwards, *Proc. Phys. Soc.* 85, 613 (1965); 88, 265 (1966). C. Domb, *Adv. Chem. Phys.* 15, 229 (1969).
15. A. A. Belavin, A. M. Polyakov, A. S. Schwartz, and Yu. S. Tyupkin, *Phys. Lett.* 59B, 85 (1975).
16. This review of vortex topological properties is similar to that of P. Goddard and D. I. Olive who elucidate the topological properties of monopoles in New Developments in the Theory of Magnetic Monopoles, CERN preprint TH. 2445 (1978).
17. H. Flanders, Differential Forms (Academic, New York, 1963). See p. 74.
18. E. M. Purcell, Electricity and Magnetism (McGraw-Hill, New York, 1965). See chapter 6 section 1.
19. L. I. Schiff, Quantum Mechanics, third edition (McGraw-Hill, New York, 1968). Sec. 27, p. 196.
20. C. W. Misner, K. S. Thorne, J. A. Wheeler, Gravitation, (W. H. Freeman, 1973). See p. 1149.
21. K. Bardakci, Field Theory for Solitons, II, Berkeley Preprint (Dec. 1978).
22. We thank Joe Polchinski for this point. Two Higgs fields suffice for all $SU(N)$, $N \geq 2$.
23. S. Samuel, Color Zitterbewegung, to be published in Nuclear Physics.
24. Of interest in this context is P. Goddard, J. Nuyts, and D. Olive, *Nucl. Phys.* B125, 1, (1977).

25. Many other terms can be added which also generate triplets. They are non-local interaction terms. They are permitted, of course, since there is no reason to expect the effective soliton Lagrangian to be a local field theory.
26. J. S. Langer, *Annals of Physics* 41, 108 (1967); *Annals of Physics* 54, 258 (1969); *Proc. IUPAP Conf. on Statistical Mechanics, Amsterdam* (1973). For a review see Wiegand, reference 11.
27. S. Coleman, *Phys. Rev.* D15, 2929 (1977).
28. For $SU(2)$ the Bloch wall solitons do satisfy $\psi(x) = 0$ on C , because, ψ is a real field. ψ must vanish because the boundary conditions are $\psi(x) \rightarrow \psi_0$ for $x \rightarrow -\infty$ and $\psi(x) \rightarrow -\psi_0$ for $x \rightarrow +\infty$. In this respect $SU(2)$ is a singular case.
29. J. M. Kosterlitz and D. J. Thouless, *J. Phys.* C6, 1181 (1973).
30. A. M. Polyakov, *JETP Lett.* 20, 194 (1974). G. 't Hooft, *Nucl. Phys.* B79, 276 (1974). For a review see reference 16.
31. For groups, H , larger than $U(1)$, it becomes just as problematic to define a monopole as in a non-abelian gauge theory without Higgs fields.
32. The analog of $\eta \cdot F_{\mu\nu}$ for the model of reference 30 is $\hat{\phi} \cdot F_{\mu\nu}$. This is not the 't Hooft $F_{\mu\nu}$ but it works equally well as noted by S. Coleman in Classical Lumps and Their Quantum Descendants, *Proceedings of the 1975 International School of Physics "Ettore Majorana"* and emphasized in reference 16.

33. See, however, N. H. Christ, Phys. Rev. Lett. 32, 355, and reference 16.
34. This is true in an "abelized" gauge where the η vector's color direction is fixed, that is, spatially constant.
35. Besides spheres there will be tori and other oriented closed surfaces. The most general oriented surface is topologically equivalent to a sphere with n handles (an object of genus, n). These are allowed; the other two dimensional surfaces, the non-orientable ones such as the projective plane, are not allowed because linking number cannot be defined.
36. M. Kaku and K. Kikkawa have constructed a field theory for strings. Unfortunately, it is non-covariant and hence not useful from our point of view. See Phys. Rev. D10, 1110 (1974).
37. The definition of orientation for a manifold can be found in M. Spivak, Calculus on Manifolds (Benjamin, New York, 1965) page 119. The way in which an oriented manifold induces an orientation for its boundary is on this same page.
38. K. Bardakci has explicitly shown this in a string model, private communication.
39. K. G. Wilson, Phys. Rev. D10, 2445 (1974).
40. S. Mandelstam, Phys. Reports 23C, 245 (1976); G. 't Hooft, in High Energy Physics: Proceedings of the E. P. S. International Conference, Palermo, June 1975, ed. A. Zichichi, p. 1225.
41. T. T. Wu and C. N. Yang, in Properties of Matter Under Unusual Conditions, ed. H. Mark and S. Fernach (Interscience, New York, 1969), p. 349.

42. Of interest in this connection is T. Banks and E. Rabinovici, Finite Temperature Abelian Lattice Gauge Theories, IAS Preprint (Jan. 1979).
43. J. Goldstone and R. Jackiw, Phys. Rev. D11, 1486 (1975).
R. Dashen, B. Hasslacher, and A. Neveu, Phys. Rev. D10, 4115 (1974);
10, 4130 (1974); 10, 4138 (1974); 11, 3424 (1975). For reviews
see R. Jackiw, Reviews of Mod. Phys. 49, 681 (1977) and
R. Rajaraman, Phys. Rep. 21C, 227 (1975).
44. M. B. Halpern, Field-Strength and Dual Variable Formulations of Gauge Theory, LBL Preprint-7980 (Sept. 1979).
45. S. Coleman and E. Weinberg, Phys. Rev. D7, 1888 (1973).
C. Thorn has presented a pedagogical example of dimensional transmutation in the two-dimensional quantum mechanics problem with a potential, $g\delta^2(x)$. Talk given at Lawrence Berkeley Laboratory. See also C. Thorn, Quark Confinement in The Infinite Momentum Frame, MIT preprint (Sept. 1978).

| | |
|--|--|
| Wilson Lattice Gauge Theory | SU(3) Gauge Theory |
| Electric Fields | Magnetic Fields |
| Quarks | Monopoles |
| Closed Electric String ↔ Electric Fluctuations | Closed Magnetic String ↔ Magnetic Fluctuations |
| Positive Slope ↔ Zero or Negative Slope | Positive Slope ↔ Zero or Negative Slope |

Table 1

- Figure 1. The Macromolecule Approximation. (a) A continuous curve between x_0 and x_f , (b) the macromolecule approximation of (a), (c) a closed curve, and (d) its macromolecule approximation.
- Figure 2. A Vortex. This is a cross section of the vortex in the 1-2 plane. The 3rd axis is out of the paper.
- Figure 3. Linking Numbers. (a) The Wilson loop. Its width and length are r and t . Figures (b), (c), and (d) show a vortex linking with this Wilson loop with linking numbers +1, -1, and +2 respectively.
- Figure 4. The Phases of the 2 + 1 Dimensional Nielsen-Olesen Model. (a) The symmetric phase. (b) The Higgs phase (c) The Topological Symmetry broken phase.
- Figure 5. A Non-trivial Path. Here is a solid sphere with antipodal points identified. This solid is topologically equivalent to $O(3)$. A closed path is said to be trivial if, via continuous deformations, it can be shrunk to a point. The path shown here begins at A , makes its way through the sphere, and ends at A' . The path is closed since A' is

the same point as A . Because when A is moved A' must move so as to be opposite A , it is impossible to bring A to A' so as to shrink the path to a point. This path is non-trivial.

Figure 6. The Product of Two Non-Trivial Paths Equals a Trivial One.

Multiplying two paths, P_1 and P_2 , which begin and end at the same point is defined to be the path, $P_3 = P_2 \cdot P_1$, formed by first traversing P_1 and then traversing P_2 . (a) Here is a path which begins at A , goes to A' via P , where it "reappears" at A . It then goes to A' via Q , hence back to the starting point, A . This path is the product of two "Figure 5" paths. (b) We deform the curve a bit. The new path again starts at A , but instead goes to a point, B , nearby A' . It "reappears" at B' , whence it goes to A' via Q . This closes the path since A' is identified with A . (c) Move the point B (and hence the point, B') around the sphere until it comes to A . (d) Shrink the two loops to the points A and A' . Since we have continuously deformed this path to a point, it is trivial.

Figure 7. Non-Trivial Higgs Configurations. Figures (a) and (b) show the behavior of two non-trivial Higgs configurations far from the vortex. They carry the topological Z_2 charge. Figure (c) shows what happens near the vortex: the Higgs fields become almost parallel but still rotate when going around the vortex. At the center they become exactly parallel.

Figure 8. Triplets. Figure (a) is the simplest triplet and Fig. (b) is its macromolecule approximation. Figure (c) is a more complicated structure.

Figure 9. Linking Number of a Triplet With a Wilson Loop. (a) The Wilson loop. It has dimensions $L \times L$ and sits in the x - y plane. The z -axis comes out of the paper. The nearby triplet does not link with the Wilson loop so $n = 0$. (b) A linking configuration yielding the phase factor $\exp\left(\frac{2\pi i}{3}\right)$ and (c) a linking configuration yielding the phase factor $\exp\left(\frac{-2\pi i}{3}\right)$.

Figure 10. Spanning Surfaces. The dark line is the Wilson loop. The non minimal surface, S' , is the "cup-like" surface below the loop. The minimal surface, S , forms a "lid". Together they enclosed the volume, V .

Figure 11. A "Bad" Topological Symmetry Breaking Potential.

Figure 12. The Propagator of Two Wilson Loops, C and C' . There are two types of spanning surfaces giving contributions; the ones where each Wilson loop annihilates into the vacuum and the "hour-glass" ones.

Figure 13. A Time Slice of a Wilson Loop. At an instant in time, the non minimal gauge surface, S' , and the minimal surface, S , will look like the curves C' and C . The classical solution has the phase factor $\exp\left(\frac{2\pi i}{3}\right)$ inside the region, R , enclosed by C and C' .

Figure 14. Bloch Wall Solitons. (a) A region of $\exp\left(\frac{2\pi i}{3}\right)$ vacuum surrounded by normal vacuum. The physical ground state will contain such domains of Z_3 vacua. The objects separating them are Bloch walls. Their excitations are associated with closed strings. (b) The mass per unit length of the Bloch wall is calculated by considering a straight one at $x = 0$. Then the problem is similar to finding a one-dimensional soliton. (c) The correct constraint for calculating the confining potential between quarks.

Figure 15. Quark Trajectories in a Baryon. (a) A baryon consisting of three quarks is produced at a and destroyed at b. In between the three quarks travel along paths, C_1 , C_2 , and C_3 . (b) A singular gauge surface. It consists of three disks, each of which is bounded by a quark trajectory and the line from a to b.

Figure 16. The Shape of Baryons. (a) Quarks at the ends of Three strings. A time slice of Fig. 15b would yield this configuration. (b) A quark in the middle of a string. (c) A triangular quark and string configuration. The numbers 1, 2, and 3 label the quarks. The dark lines are the strings.

Figure 17. Monopoles. (a) A sphere of radius, r , surrounding the endpoint (monopole) of a triplet. The total magnetic flux emanating is $\frac{6\pi}{g}$. The region, A, is a disk on the sphere. (b) For a vortex the total flux emanating is zero.

Figure 18. Linking of a Sphere and a Loop in Four Dimensions. These three figures show a temporal sequence in which a sphere and a loop link. Each $t = n$ value represents a time slice. In general, a time slice of a loop and a sphere yields respectively two points and a loop. Exceptions to this occur when the loop or sphere are contained within a single time slice, in which case they respectively look like a loop (Fig. (c) at $t = 3$) and a sphere (Fig. (b) at $t = 3$). Figure (a) shows the generic case: a pair of "particles" and a small "closed string" are produced out of the vacuum. One of the particles shoots through the loop, which subsequently shrinks and disappears. The particles then annihilate. In Fig. (b) the sphere is contained in the $t = 3$ slice. Again, a pair of particles is produced and the two separate. One of them is instantaneously surrounded by the sphere, which subsequently vanishes. The two then annihilate. In Fig. (c) the loop is contained in the $t = 3$ slice. A closed string is produced. It expands; then with the sudden appearance of the loop, it links. The loop instantly disappears and the closed string shrinks and vanishes.

Figure 19. A Triplet Surface. This surface is contained in the $z = 0$ slice of space-time, hence the z direction is not shown. The object is like a smaller bubble stuck to a bigger one.

Figure 20. The Temporal Sequence of Fig. 19.



(a)



(b)



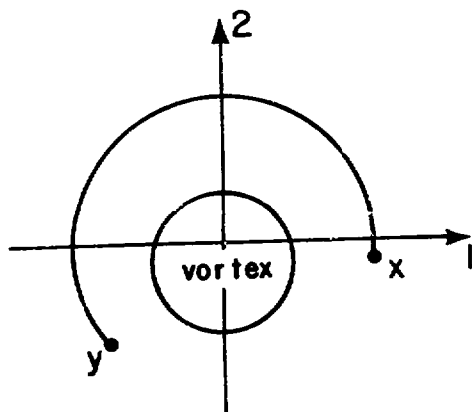
(c)



(d)

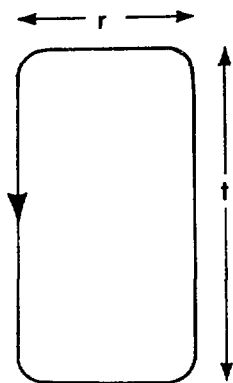
Figure 1

XBL 791-215

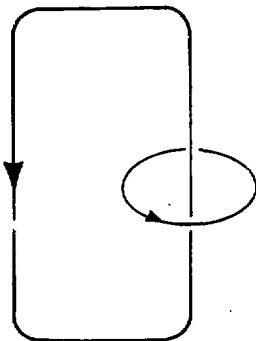


XBL 791-216

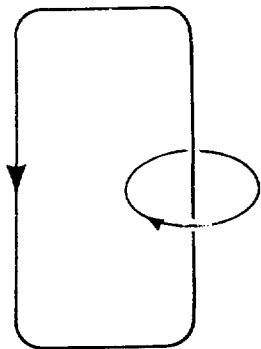
Figure 2



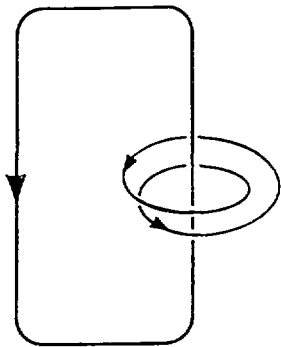
(a)



(b)



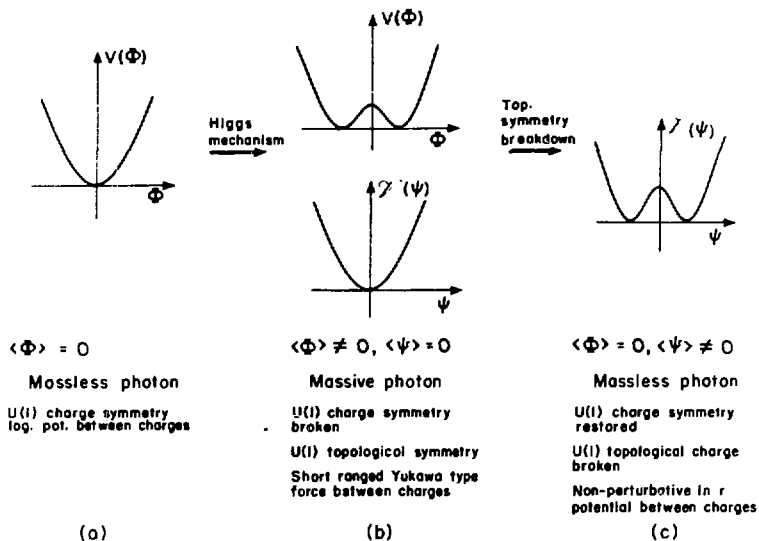
(c)



(d)

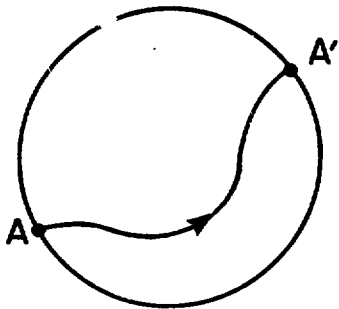
XBL 791-217

Figure 3



XBL 791-218

Figure 4



XBL 791-219

Figure 5

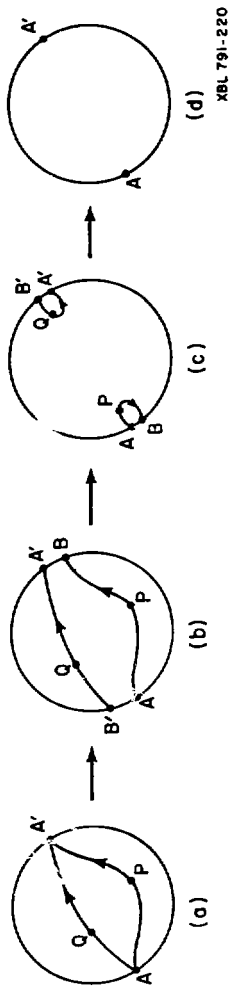
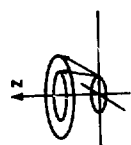
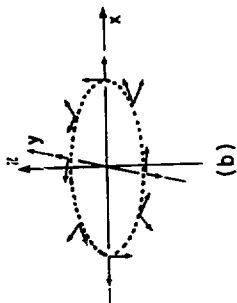


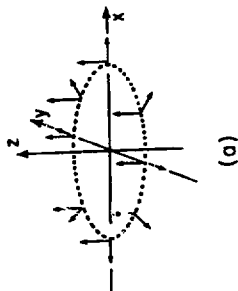
Figure 6



(c)
XBL 791-221

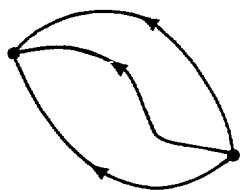


(b)

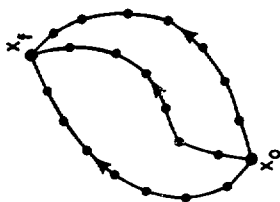


(a)

Figure 7



(a)



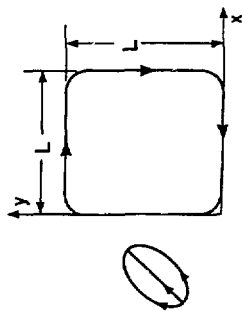
(b)



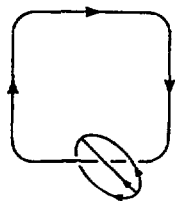
(c)

XBL 791-222

Figure 8



(a)

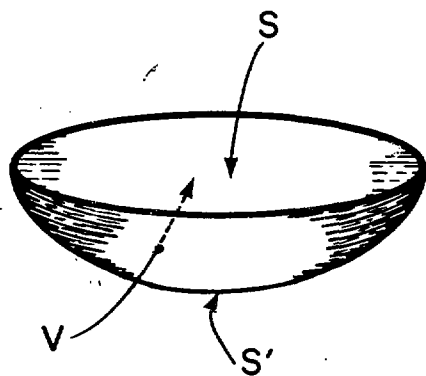


$\exp(2\pi i/3)$
(b)

$\exp(4\pi i/3) = \exp(-2\pi i/3)$
(c)

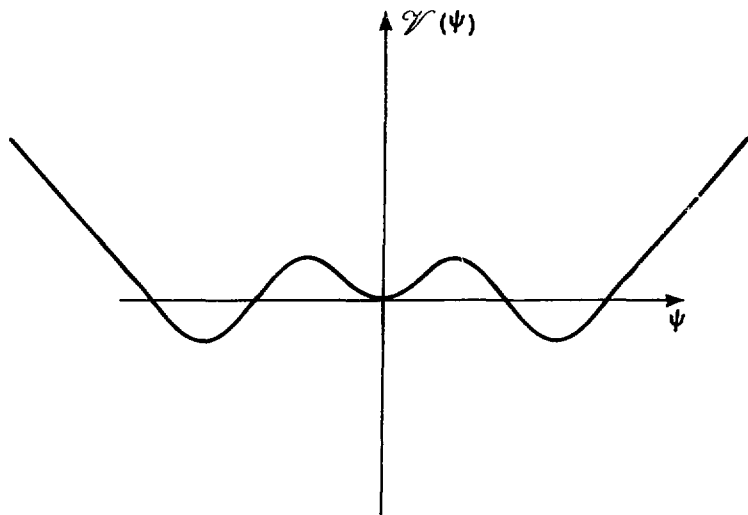
XBL 791-223

Figure 9



XBL 791-224

Figure 10



XBL 791-225

Figure 11

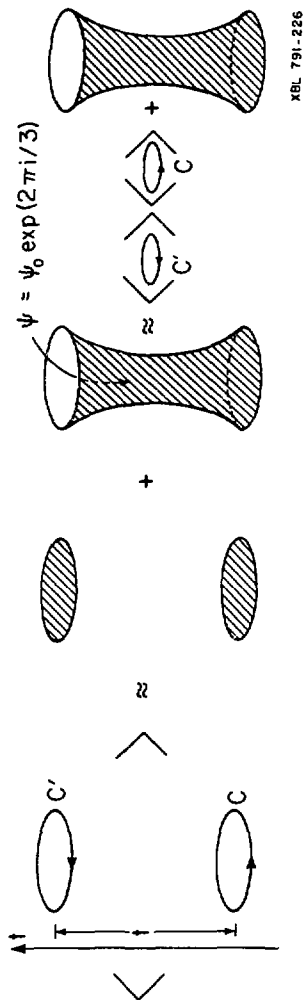
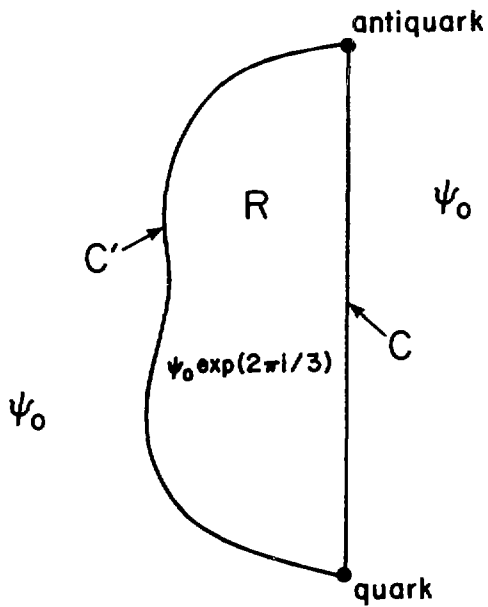
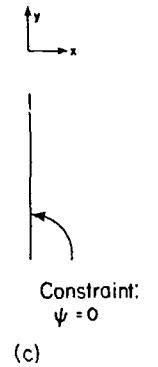
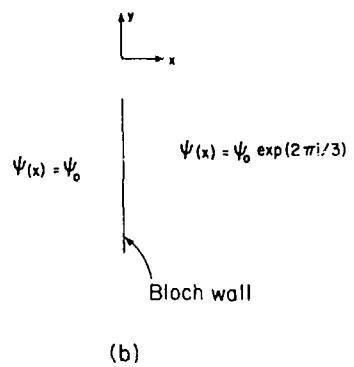
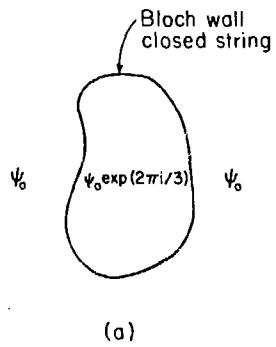


Figure 12



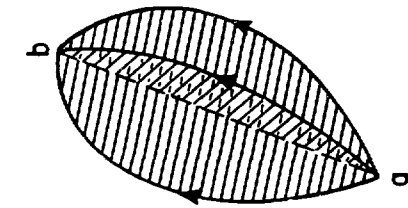
XBL 79I-227

Figure 13



XBL 791-228

Figure 14



XBL 791 - 233

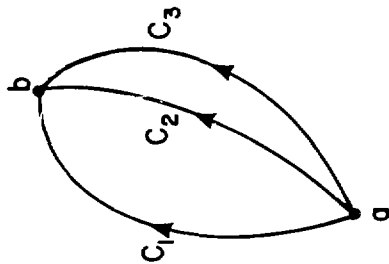
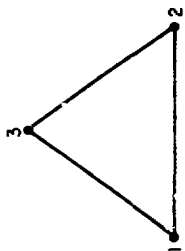
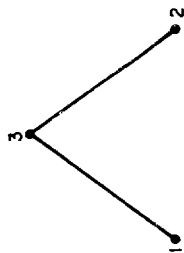
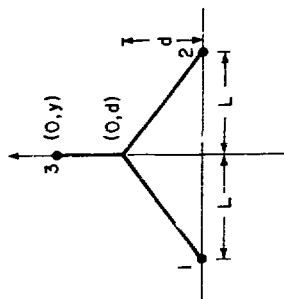
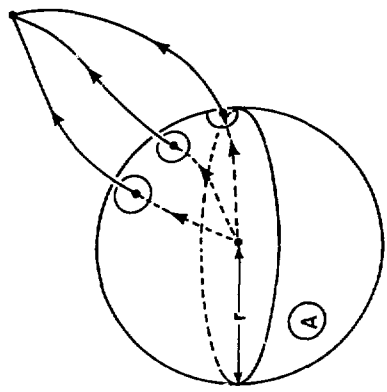


Figure 15

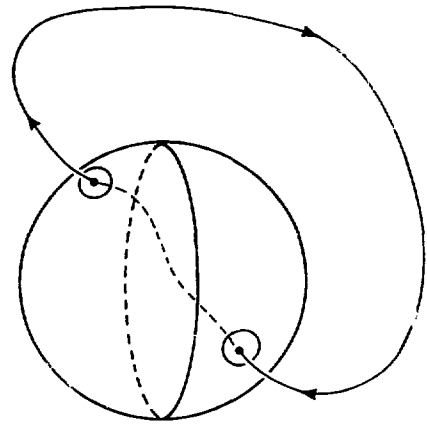


XBL 791-229

Figure 16



(a)



(b)

Figure 17



(a)



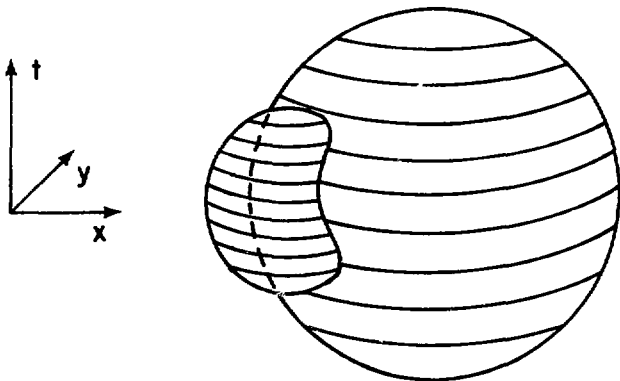
(b)



(c)

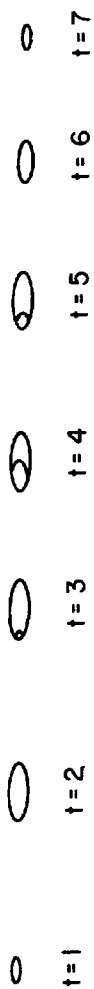
$t = 1$ $t = 2$ $t = 3$ $t = 4$ $t = 5$

Figure 18



XBL 791.. 232

Figure 19



XBL 791-234

Figure 20