Exact Solution of the Association Problem by a Matrix-Spinor Method with Applications to Statistical Mechanics

H. S. GREEN, *University of Adelaide, Adelaide, South Australia*

AND

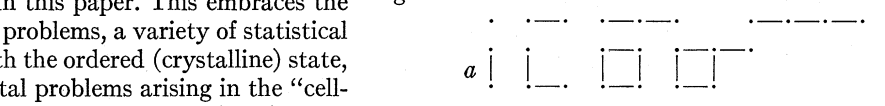
ROY LEIPNIK, *Michelson Laboratory Test Department, U. S. Naval Ordnance Test Station, China Lake, California*

1. INTRODUCTION

A UNIFIED treatment of a rather wide class of problems in statistical mechanics in terms of a generalized problem, which we call the association problem, is presented in this paper. This embraces the classical order-disorder problems, a variety of statistical problems concerned with the ordered (crystalline) state, and also the fundamental problems arising in the "cell-cluster" method applied to the disordered (fluid) state.

The object of statistical mechanical theory applied to these problems is to determine either the free energy as a function of temperature and densities, or the pressure as a function of temperature and activities. This normally involves a summation, or integration, over all the microscopic configurations of the system, a formidable task except for very simple systems, and one which usually cannot be carried out without some degree of approximation. There is a technical advantage in regarding a macroscopic system in extension as compounded of a large number of similar subsystems called cells, comparable in number with the particles of the system. In an ordered state, the number of particles of a particular kind in a cell is small (usually one) and the same for each cell. In a disordered state the number of particles of a particular kind may fluctuate within narrow limits. In any event the sum over all configurations of the large system can be broken down into a sum over configurations in each cell, and as the number of configurations in a cell is relatively small there is no difficulty in practice in performing the summation there.

The only difficulty arises in taking account of the correlation between configurations in different cells, and even this problem usually can be broken down into the computation of correlations in particular groups of cells. One is then left with the problem of assigning the correct statistical weight to each method of grouping the cells. This purely statistical problem is then the major task. It can be idealized to some extent by fixing attention on the mean centers of the cells, which form a space lattice, and picturing each group of cells as a geometrical figure with vertices at the lattice points. Looked at in this way, the statistical problem reduces to one of counting the number of ways in which a variety of oriented geometrical figures can be distributed over a lattice, in such a way that each lattice point is occupied by the vertex of one and only one figure. For example, on a square lattice in two dimensions, one has to determine the number of ways in which the figures



can be distributed so as to completely cover the lattice. The distance a is the distance between neighboring lattice points. The most convenient way of expressing the answer is by stating the generating function $F(c_0,c_1,c_2,\cdots,c_n,c_{1,1},c_{1n-1},c_{1n},c_{1,1,1},\cdots)$, constructed in such a way that the coefficients of products of various powers of the c's are the number of ways of distributing the corresponding numbers of the figures on the lattice. We propose as the main problem to be solved in this paper the evaluation of this generating function, when only a subclass of the entire class of c's is different from zero.

We have called this problem the association problem because it arises most naturally when we consider a system of particles, situated in the neighborhood of the lattice sites, and capable of forming associations by the pairing of electrons or ions localized in the individual particles. But the same problem has arisen in a variety of other connections.

An important example is the cell-cluster method of evaluating the partition function of a fluid. This originated in the lattice and cell models of the liquid state developed by Eyring and his associates[1] and Lennard-Jones and Devonshire.[2] The results were originally rather inaccurate owing to the neglect of multiple occupation of the cells and the neglect of correlations between configurations in neighboring cells. But, after Kirkwood[3] had shown the relation of the lattice-cell model to the exact statistical mechanical theory, a series of improvements were made by Mayer and Careri,[4] Wentorf, Buehler, Hirschfelder, and Cur-

[1] H. Eyring and J. Hirschfelder, J. Phys. Chem. **41**, 249 (1937); Hirschfelder, Stevenson, and Eyring, J. Chem. Phys. **5**, 896 (1937).
[2] J. E. Lennard-Jones and A. F. Devonshire, Proc. Roy. Soc. (London) **A163**, 53 (1937).
[3] J. G. Kirkwood, J. Chem. Phys. **18**, 380 (1950).
[4] J. E. Mayer and G. Careri, J. Chem. Phys. **20**, 1001 (1952).

tiss,[5] de Boer,[6] Barker,[7] and Green[8] which opened up the possibility of arbitrarily close approximation to the exact partition function. The only problem left unsolved in principle in this connection is the statistical one described above.

Another important example is the theory of solutions, which has also been the subject of a long series of improvements since the quasi-chemical method was devised independently by Bethe,[9] and by Guggenheim and his associates.[10] An important advance was made by Kikuchi,[11] which has been followed up by de Boer and his associates.[12] The degree of approximation which has been achieved, however, could be improved considerably by the exact solution of the combinatorial problem with which we are concerned. Another important application of the method is to order-disorder phenomena, where even the simple-appearing Ising problem in three dimensions awaits an exact solution. In two dimensions, the problem has been solved by Onsager,[13] and more expeditiously by Kaufman,[14] and Kac and Ward.[15]

We do not offer exact explicit solutions to all these problems, but show how they are reduced to the fundamental statistical problem enunciated, and present a unified formulation which we think is a useful step towards its solution. We further show how the solution can be reduced by the introduction of a matrix-spinor technique to the determination of the principal eigenvalue of a certain matrix, and extract this eigenvalue for various interesting special cases. We also solve explicitly and exactly by this method a two-dimensional association problem which, we believe, is comparable in difficulty to the Ising problem and has not been solved before. It is hoped that our general approach will lead to the solution of a variety of other outstanding problems of statistical physics.

2. REDUCTION OF THE ASSOCIATION TO A COMBINATORIAL PROBLEM

The reduction method described below is applicable to any system of atoms or molecules in thermodynamic equilibrium. We assume henceforth the validity of classical mechanics, though we believe the method could be adapted to take account of quantum mechanical corrections.

In an ordered state of the system, let L be the space lattice formed by the mean positions of the particles in question. Each lattice point is then also the mean center of a cell which can be constructed in such a way that each particle of the system is contained within exactly one cell.

In a disordered state of the system, the volume occupied by the system can be divided into cells such that the mean centers of the cells form a space lattice L and such that the volume of each cell is either equal to, or an integer multiple of, the mean particle (atomic or molecular) volume.

A particular configuration of the whole system can then be specified, in either case, by stating the cell configurations, that is, the kind, position, and orientation of each particle in each cell. In either case, we call L the association lattice of the system.

If C denotes the total configuration, let C_i denote the cell configuration in the ith cell. The energy of the particles in the ith cell is a function, say $E(C_i)$, of C_i. Similarly, let $E(C_i,C_j)$ denote the mutual energy of interaction between the particles in the ith and jth cells. The three-body energies between cells i, j, k might be denoted by $E(C_i,C_j,C_k)$. The energy of the whole system in the configuration C, denoted by $E(C)$, is given by

$$E(C)=\sum_i E(C_i)+\tfrac{1}{2}\sum_{i,j} E(C_i,C_j)+\cdots. \tag{1}$$

There is no difficulty in principle in taking account of three-body or many-body interactions, but for simplicity these are omitted, as they are usually negligible.

If z_{q,C_i} is the activity of the qth particle in the configuration C_i of the ith cell, let

$$Q(C_i)=\prod_q z_{q,C_i} \tag{2}$$

be the product of the activities of all particles in the configuration C_i of the ith cell. The theory of the grand partition function then gives the probability $P(C)$ of the configuration C as, in thermodynamic equilibrium,

$$P(C)=\exp(-\beta pV)\prod_i Q(C_i) \times\exp[-\beta E(C_i)]\prod_{i<j}\exp[-\beta E(C_i,C_j)], \tag{3}$$

where p is pressure, T absolute temperature, V volume, and $\beta=1/kT$. Thus

$$F_1=\exp(\beta pV)=\sum_C\prod_i A_1(C_i)\prod_{i<j} B_1(C_i,C_j), \tag{4}$$

where

$$A_1(C_i)=Q(C_i)\exp[-\beta E(C_i)], \qquad B_1(C_i,C_j)=\exp[-\beta E(C_i,C_j)]. \tag{5}$$

As is well known, the pressure, as a function of temperature and activities, determines all other thermodynamic functions of the system.

[5] Wentorf, Buehler, Hirschfelder, and Curtiss, J. Chem. Phys. **18**, 1484 (1950).
[6] J. de Boer, Proc. Roy. Soc. (London) **A215**, 4 (1952).
[7] J. A. Barker, Proc. Roy. Soc. (London) **A230**, 390 (1955).
[8] H. S. Green, J. Chem. Phys. **24**, 732 (1956).
[9] H. A. Bethe, Proc. Roy. Soc. (London) **A150**, 552 (1935).
[10] E. A. Guggenheim, Proc. Roy. Soc. (London) **A135**, 118 (1932).
[11] R. Kikuchi, Phys. Rev. **81**, 988 (1951).
[12] J. Hijmans and J. de Boer, Physica **21**, 471, 485, 499 (1955).
[13] L. Onsager, Phys. Rev. **65**, 117 (1944).
[14] B. Kaufman, Phys. Rev. **76**, 1232 (1949).
[15] M. Kac and J. C. Ward, Phys. Rev. **88**, 1332 (1952).

In practice it is often easier to compute, instead of F_1, the function

$$F_2=\exp[\beta(pV-\sum_{i<j}E_{i,j})], \tag{6}$$

where $E_{i,j}$ is the mean of the interaction energies $E(C_i,C_j)$ over all configurations C. For this purpose, it is helpful to introduce a number of auxiliary quantities, denoted by $E_j(C_i)$, $A_2(C_i)$, $B_2(C_i,C_j)$. As the notation suggests, $E_j(C_i)$ is the mean of the interaction energies $E(C_i,C_j)$ over all configurations C_j. Let

$$A_2(C_i)=Q(C_i)\exp[-\beta(E(C_i)+\sum_{j\neq i}E_j(C_i))], \tag{7}$$

$$B_2(C_i,C_j)=\exp[-\beta(E(C_i,C_j)-E_j(C_i)-E_i(C_j)+E_{i,j})],$$

and note that F_2 can now be expressed as

$$F_2=\sum_C\prod_i A_2(C_i)\prod_{i<j}B_2(C_i,C_j). \tag{8}$$

The same formal methods obviously can be applied to the evaluation of either (4) or (8).

If S is any set of cell numbers or lattice points, let C_S denote the configurations in these cells. Given any symmetric function $B(C_i,C_j)$ of pairs of cell configurations, we define, with Ursell,[16] functions $U(C_S)$ as follows: if T is any set of cell numbers, D_T the set of all partitions γ of T into disjoint sets of cell numbers, then

$$\prod_{\substack{i,j\epsilon T\\ i<j}}B(C_i,C_j)=\sum_{\gamma\epsilon D_T}\prod_{S\epsilon\gamma}U(C_S). \tag{9}$$

We make the convention that $U(C_S)=1$ whenever S is empty or contains just one cell number. If T is allowed to range over all sets of two or more cell numbers, then the functions $U(C_S)$ are completely determined by the above equation. In a less condensed notation, we write

$$U(C_S)=U(C_{i_1},C_{i_2},\cdots C_{i_m})\text{ when }S=\{i_1,i_2,\cdots i_m\}. \tag{10}$$

Then the foregoing equations become

$$\begin{aligned}B(C_i,C_j)&=U(C_i)U(C_j)+U(C_i,C_j),\\ B(C_i,C_j)B(C_i,C_k)B(C_j,C_k)&=U(C_i)U(C_j)U(C_k)\\ &\quad+U(C_i)U(C_j,C_k)+U(C_j)U(C_i,C_k)\\ &\quad+U(C_k)U(C_i,C_j)+U(C_i,C_j,C_k),\quad\text{etc.}\end{aligned} \tag{11}$$

We might call $U(C_S)$ the association coefficient for the cell cluster S, determined by the cell pair function $B(C_i,C_j)$. The association coefficients $U(C_S)$ can be defined directly in terms of the $B(C_i,C_j)$ by

$$U(C_S)=\sum_{\gamma\epsilon D_S}(-1)^{n(\gamma)-1}\times(n(\gamma)-1)!\sum_{T\epsilon\gamma}\prod_{\substack{i,j\epsilon T\\ i<j}}B(C_i,C_j), \tag{12}$$

where $n(\gamma)$ is the number of sets in the partition γ.

[16] H. D. Ursell, Proc. Camb. Phil. Soc. **23**, 685 (1927).

The importance of the $U(C_S)$ stems from the following: if S can be partitioned into subsets S_1, S_2 belonging to sets of cells separated by a large distance, then $U(C_S)$ is close to zero. (A chain of adjacent cells, however long, cannot be so partitioned.) Thus, approximations to F can be conveniently constructed by setting $U(C_S)=0$ for a set of appropriately chosen S.

The reduction of the expression

$$F=\sum_C\prod_i A(C_i)\prod_{i<j}B(C_i,C_j)=\sum_C\prod_i A(C_i)\sum_{\gamma\epsilon D_L}\prod_{S\epsilon\gamma}U(C_S) \tag{13}$$

to the generating function of a combinatorial problem depends on certain invariance properties of the $A(C_i)$ and $B(C_i,C_j)$. Suppose that C_i, C_j are configurations of the ith and jth cells, and τ is a translation, which sends i into $\tau(i)$ and j into $\tau(j)$ and also sends configurations C_i and C_j into identical configurations $C_{\tau(i)}{}^{(\tau)}$ and $C_{\tau(j)}{}^{(\tau)}$, respectively, of cells $\tau(i)$ and $\tau(j)$. We say that A and B are translation-invariant if

$$A(C_i)=A(C_{\tau(i)}{}^{(\tau)})\quad\text{and}\quad B(C_i,C_j)=B(C_{\tau(i)}{}^{(\tau)},\ C_{\tau(j)}{}^{(\tau)}) \tag{14}$$

for each translation, within the lattice. If B is translation-invariant, it follows from Eq. (14) that $U(C_S)$ is translation-invariant in the sense that

$$U(C_{\tau(S)}{}^{(\tau)})=U(C_S), \tag{15}$$

where $C_{\tau(S)}{}^{(\tau)}$ is the configuration of the cells $\tau(S)$ obtained by translation of the configuration C_S of cells S.

Because the energies $E(C_i)$, $E(C_i,C_j)$, and activities $Q(C_i)$ are translation-invariant in the above sense, it is clear that $A_1(C_i)$, $A_2(C_i)$, $B_1(C_iC_j)$, and $B_2(C_i,C_j)$ are also translation-invariant.

If $A(C_i)$ and $B(C_i,C_j)$ are translation-invariant, the evaluation of the quantity F can be given a combinatorial interpretation. If L_1 and L_2 are sets of lattice points such that $L_2=\tau(L_1)$ for some translation τ of L_1 within the lattice, we call L_1 and L_2 translation-equivalent; similarly, sets of cell numbers S_1 and S_2 are called translation-equivalent if they correspond to translation-equivalent sets of lattice points. The resulting equivalence classes, obtained by identifying equiyalent sets, can be regarded as oriented geometric figures capable of being attached to the lattice in a certain set of positions. If K is one of these oriented geometric objects, let L_K be the equivalence class of sets of cell numbers corresponding to K. Let $\kappa=(K_0,K_1,\cdots)$ be the set of such objects. For nonempty sets S, define coefficients b_S by

$$b_S=\sum_C U(C_S)\prod_{j\epsilon S}A(C_j). \tag{16}$$

More explicitly, if $S=\{i_1,\cdots i_m\}$, then

$$b_S=\sum_{C_{i_1}}\cdots\sum_{C_{i_m}}U(C_{i_1}\cdots C_{i_m})A(C_{i_1})\cdots A(C_{i_m}). \quad (17)$$

Clearly, $b_{S_1}=b_{S_2}$ whenever $S_1, S_2\epsilon L_K$, so we can write $b_S=b_K$ whenever $S\epsilon L_K$. We then have the identity (de Boer)

$$F=\sum_C\prod_i A(C_i)\prod_{i<j}B(C_i,C_j)$$
$$=\sum_{m_0,m_1,\cdots}N(m_0,m_1,\cdots)b_{K_0}{}^{m_0}b_{K_1}{}^{m_1}\cdots, \quad (18)$$

where $N(m_0,m_1,\cdots)$ is the number of ways of attaching m_0 copies of K_0, m_1 copies of $K_1,\cdots$ to the lattice L in such a way that each lattice point of L is occupied by one and only one vertex of a geometric object. (If n is the number of lattice points of L, and $n(K)$ the number of vertices in K, we have the constraint $\sum_i m_i n(K_i)=n$ in the foregoing summation.) This identity can be proved by induction on n with the help of the obvious relation

$$\sum_{C_1\cdots}\sum_{C_{n+1}}\prod_{i=1}^{n+1}A(C_i)\prod_{i<j}^{n+1}B(C_i,C_j)$$
$$=\sum_{C_1\cdots}\sum_{C_n}\prod_{i=1}^{n}A(C_i)\prod_{i<j}^{n}B(C_i,C_j)$$
$$\times\sum_{C_{n+1}}A(C_{n+1})\prod_k B(C_k,C_{n+1}), \quad (19)$$

which determines F in a lattice L' of $n+1$ points, obtained from a lattice L of n points by adjoining to L the result of applying a fundamental vector of the lattice L to one of its boundary points.

Thus we can approximate F by taking $b_K=0$ for a suitable class of geometric objects. In certain cases, $b_K=0$ is rigorously true for a class of geometric objects due to physical reasons. For example, in the Ising problem and other problems with only nearest neighbor interactions, $b_K=0$ for all objects not corresponding to chains of adjacent lattice points.

3. GENERATING FUNCTIONS FOR A COMBINATORIAL PROBLEM

Let L be a given one-, two-, or three-dimensional lattice. An oriented geometric object K consisting of linked vertices is said to be compatible with L if K can be attached to L, without reorientation, so that every vertex of K covers a lattice point of L. Two compatible objects K_1, K_2 are called vertex-equivalent if they can cover the same set of lattice points, regardless of the nature of the links in K_1, K_2. Let $\kappa=(K_0,\cdots K_p)$ be a given set of compatible objects. Let $N(m_0,m_1,\cdots m_p)$ be the number of ways of attaching m_0 copies of K_0, m_i copies of $K_i,\cdots$ to the lattice L so that every lattice point of L is covered by one and only one vertex of a geometric object. The combinatorial problem is to calculate $N(m_0,m_1,\cdots)$. The generating function for this combinatorial problem is

$$F(\mathbf{b})=\sum_{m_0,m_1,\cdots}N(m_0,m_1,\cdots)b_{K_0}{}^{m_0}b_{K_1}{}^{m_1}\cdots, \quad (20)$$

where b_k is given for each $K\epsilon\kappa$, $\mathbf{b}=(b_{K_0},b_{K_1},\cdots)$.

For statistical mechanical applications, it is $F(\mathbf{b})$ rather than $N(m_0,m_1,\cdots)$ itself that is wanted. However, the relation of $F(\mathbf{b})$ to $N(m_0,m_1,\cdots)$ makes possible the deduction of a generating function for $F(\mathbf{b})$. We assume that $b_{K_1}=b_{K_2}$ whenever K_1 and K_2 are vertex-equivalent. Let the lattice points of L be denoted by $x_1, x_2, \cdots x_n$ and let L_K be the set of ordered sets of subscripts of lattice points of L simultaneously covered by K. If K_1 and K_2 are vertex-equivalent, $L_{K_1}=L_{K_2}$.

The simplest generating function for $F(\mathbf{b})$ is now written down. Let

$$V(\mathbf{x},\mathbf{b})=\prod_{K\epsilon\kappa}\prod_{A\epsilon L_K}(1+b_K\prod_{i\epsilon A}x_i). \quad (21)$$

Then $F(\mathbf{b})$ is the coefficient of

$$\prod_{i=1}^{n}x_i=x_1\cdots x_n$$

in the multinomial expansion of $V(\mathbf{x},\mathbf{b})$ in powers of $x_1, x_2, \cdots x_n$. To prove this, fix $m_0, \cdots m_p$ and consider the coefficient of $b_{K_0}{}^{m_0}\cdots b_{K_p}{}^{m_p}\cdot x_1\cdots x_n$ in the expansion of $V(\mathbf{x},\mathbf{b})$ in powers of $b_{K_0}\cdots b_{K_p}$ and $x_1\cdots x_n$. This term can arise only from a product of m_0 factors of the form $(1+b_{K_0}\prod_{i\epsilon A}x_i)$ for $A\epsilon L_{K_0}$, m_1 factors of the form $(1+b_{K_1}\prod_{i\epsilon B}x_i)$ for $B\epsilon L_{K_1}$, etc., where $A, B, \cdots$ form a disjoint covering of the lattice L. Since each covering contributes one to the coefficient in question, we see after summation that

$$\sum_{m_0\cdots m_p}N(m_0\cdots m_p)b_{k_0}{}^{m_0}\cdots b_{k_p}{}^{m_p}x_1\cdots x_n=F(\mathbf{b})x_1\cdots x_n$$

is the term containing $x_1\cdots x_n$ in the expansion of $V(\mathbf{x},\mathbf{b})$ in powers of $x_1, x_2, \cdots x_n$.

More useful for our purposes is the exponential generating function

$$W(\mathbf{x},\mathbf{b})=\exp H(\mathbf{x},\mathbf{b}),$$

where

$$H(\mathbf{x},\mathbf{b})=\sum_{K\epsilon\kappa}\sum_{A\epsilon L_K}\prod_{i\epsilon A}x_i. \quad (22)$$

To prove that $F(\mathbf{b})$ is the coefficient of $x_1\cdots x_n$ in the expansion of $W(\mathbf{x},\mathbf{b})$, note that since $\exp(bu)=1+bu+[(bu)^2/2]+\cdots$, the terms in $\exp(b\prod_{i\epsilon A}x_i)$ in which no subscripts are repeated come already from $1+b\prod_{i\epsilon A}x_i$. In the calculation of the coefficient of $x_1\cdots x_n$ repeated subscripts may be disregarded so the coefficient of $x_1\cdots x_n$ in

$$\prod_{K\epsilon\kappa}\prod_{A\epsilon L_K}(1+b_K\prod_{i\epsilon A}x_i)=V(\mathbf{x},\mathbf{b})$$

is equal to the coefficient of $x_1 \cdots x_n$ in

$$\prod_{K \epsilon \kappa} \prod_{A \epsilon L_K} \exp(b_K \prod_{i \epsilon A} x_i) = W(\mathbf{x}, \mathbf{b}).$$

In the next section, a recurrence is derived from $W(\mathbf{x},\mathbf{b})$ for calculating $F(\mathbf{b})$.

4. REDUCTION TO A LINEAR RECURRENCE RELATION

The notational difficulties in deriving the desired recurrence relation are decreased by numbering the lattice points of L "systematically." Let $\mathbf{v}_1, \cdots, \mathbf{v}_r$ be the fundamental vectors of the lattice, so that every lattice point can be represented by a vector of the form $n_1\mathbf{v}_1 + \cdots + n_r\mathbf{v}_r$, where $n_1 \cdots n_r$ are integers. Not all the combinations $n_1\mathbf{v}_1 + \cdots + n_r\mathbf{v}_r$ represent lattice points, in general, since L is a finite lattice. Let $x_1, \cdots, x_n$ denote the lattice points. We say the lattice is numbered systematically if the result $\mathbf{v}_k(x_j)$ of translation of any interior point x_j of the lattice by the fundamental vector $\mathbf{v}_k$ is the lattice point $x_{j+t(k)}$, where $t(k)$ depends only on k. Thus the effect of any translation on a set of lattice points is to translate the subscripts by a certain integer, provided the translates still lie in L. Henceforth, systematic numbering is assumed. As in Sec. 3, let L_K be the set of ordered sets of lattice point numbers simultaneously covered by K, and let $S_K = [i_1(K), \cdots i_{q+1}(K)]$ be the member of L_K whose initial number $i_1(K)$ is the smallest. Clearly, K is specified up to vertex-equivalence, and therefore b_K is specified fully when S_K is given. K_0 denotes the figure consisting of a single vertex.

Let M_K be the set of p such that $\{i_1(K)+p, \cdots i_{q+1}(K)+p\} \epsilon L_K$. Note that M_K fixes the possible locations of K in L. The figure K is actually determined by the differences $\delta_1(K) = i_2(K) - i_1(K) \cdots \delta_q(K) = i_{q+1}(K) - i_1(K)$. It is convenient to define $\delta_0(K) = 0$. If $K \neq K_0$, we write $b_K = c_{\delta_1(K) \cdots \delta_q(K)}$, and we write $b_{K_0} = c_0$, $\mathbf{c} = (c_0, \cdots)$.

The exponent $H(\mathbf{x},\mathbf{b}) = H(\mathbf{x},\mathbf{c})$ can be rewritten as

$$H(\mathbf{x},\mathbf{c}) = \sum_{K \epsilon \kappa} c_{\delta_1(K) \cdots \delta_q(K)} \sum_{p \epsilon M_K} \prod_{j=0}^{q} x_{i_1(K)+\delta_j(K)+p}$$

$$= c_0 \sum_{i=1}^{n} x_i + \sum_{K_0 \neq K \epsilon \kappa} c_{\delta_1(K) \cdots \delta_q(K)} \sum_{p \epsilon M_K} \prod_{j=0}^{q} x_{i_1(K)+\delta_j(K)+p}. \tag{23}$$

Define the operator Y_j, $j = 1, \cdots n$ by

$$Y_j f(\mathbf{x}) = (\partial/\partial x_j) f(\mathbf{x}) \,|_{x_j = 0}. \tag{24}$$

In this notation, we have

$$F(\mathbf{b}) = F(\mathbf{c}) = \prod_{j=1}^{n} Y_j [\exp H(\mathbf{x},\mathbf{c})]. \tag{25}$$

Now let

$$F_r(x_{r+1}, \cdots x_n, \mathbf{c}) = \prod_{j=1}^{r} Y_j [\exp H(\mathbf{x},\mathbf{c})] \tag{26}$$

so that $F_n(\mathbf{c}) = F(\mathbf{c})$ and

$$F_r(x_{r+1}, \cdots x_n, \mathbf{c}) = Y_r F_{r-1}(x_r, \cdots x_n, \mathbf{c}). \tag{27}$$

To obtain a useful relation between F_r and F_{r-1}, we study the effect of applying Y_r (or $\partial/\partial x_r$) to $H(0, \cdots 0, x_r, \cdots x_n, \mathbf{c})$. Consider a typical term

$$\prod_{j=0}^{q} x_{i_1(K)+\delta_j(K)+p}$$

for $K \neq K_0$, $p \epsilon M_K$. If $i_1(K) + p < r$, then the result is zero, since at least one of the x_j's in the product has been set to zero. If $i_1(K) + p > r$, the product is independent of x_r, so differentiation yields zero. If $p = r - i_1(K) \epsilon M_K$, the result is

$$\prod_{j=1}^{q} x_{i_1(K)+\delta_j(K)+p} = \prod_{j=1}^{q} x_{r+\delta_j(K)}.$$

For $K = K_0$, the typical term is x_{p+1} and the effect is

$$\begin{cases} 1, & p+1 = r \\ 0, & p+1 \neq r. \end{cases}$$

Let κ_r be the set of $K \epsilon \kappa$ such that $r - i_1(K) \epsilon M_K$. From the above we have $Y_r H(0 \cdots 0, x_r \cdots x_n, \mathbf{c})$ equals

$$\sum_{K_0 \neq K \epsilon \kappa_r} c_{\delta_1(K) \cdots \delta_q(K)} x_{\delta_1(K)+r} \cdots x_{\delta_q(K)+r} + c_0 = G_r(x_{r+1}, \cdots x_{r+\mu_r}, \mathbf{c}), \tag{28}$$

where

$$\mu_r = \max_{K \epsilon \kappa_r} \delta_q(K).$$

Here G_r is linear in each x_j. In particular, for $r = 1$,

$$Y_1 H(x_1 \cdots x_n, \mathbf{c}) = G_1(x_2 \cdots x_{1+\mu_1}, \mathbf{c}). \tag{29}$$

Hence

$$F_1(x_2 \cdots x_n, \mathbf{c}) = Y_1 \exp H(\mathbf{x},\mathbf{c}) = \exp H(0, x_2 \cdots x_n, \mathbf{c}) \cdot G_1(x_2 \cdots x_{1+\mu_1}, \mathbf{c}). \tag{30}$$

Suppose now that for some $r \geqq 2$,

$$F_{r-1}(x_r \cdots x_n, \mathbf{c}) = \exp H(0 \cdots 0, x_r \cdots x_n, \mathbf{c}) \cdot R_{r-1}(x_r \cdots x_{\lambda_{r-1}}, \mathbf{c}), \tag{31}$$

where R_{r-1} is a polynomial in the x_j.

Then

$$\frac{\partial}{\partial x_r} F_{r-1}(x_r \cdots x_n, \mathbf{c}) = \exp H(0 \cdots 0, x_r \cdots x_n, \mathbf{c}) \times \left[R_{r-1}(x_r \cdots x_{\lambda_{r-1}}, \mathbf{c}) \frac{\partial}{\partial x_r} H(0 \cdots 0, x_r \cdots x_n, \mathbf{c}) + \frac{\partial}{\partial x_r} R_{r-1}(x_r \cdots x_{\lambda_{r-1}}, \mathbf{c}) \right], \tag{32}$$

and so

$$F_r(x_{r+1}\cdots x_n,\mathbf{c}) = Y_rF_{r-1}(x_r\cdots x_n,\mathbf{c}) = \exp H(0\cdots 0,x_{r+1}\cdots x_n,\mathbf{c}) \times[R_{r-1}(0,x_{r-1}\cdots x_{\lambda_{r-1}},\mathbf{c})G_r(x_{r+1}\cdots x_{r+\mu_r}) + Y_rR_{r-1}(x_r\cdots x_{\lambda_{r-1}},\mathbf{c})] = \exp H(0\cdots 0,x_{r+1}\cdots x_n,\mathbf{c}) \times R_r(x_{r+1}\cdots x_{\lambda_r},\mathbf{c}), \quad (33)$$

where

$$R_r(x_{r+1}\cdots x_{\lambda_r},\mathbf{c}) = R_{r-1}(0,x_{r+1}\cdots x_{\lambda_{r-1}},\mathbf{c})\cdot G_r(x_{r+1}\cdots x_{r+\mu_r}) + Y_rR_{r-1}(x_r\cdots x_{\lambda_{r-1}},\mathbf{c}) \quad (34)$$

is a polynomial in the x_j, and $\lambda_r=\max(\lambda_{r-1},r+\mu_r)$. Thus by induction we have the recurrence (34) for $2\leqq r\leqq n$. On comparison of (30) with (31), we have

$$R_1(x_2\cdots x_{\lambda_1},\mathbf{c})=G_1(x_2\cdots x_{1+\mu_1},\mathbf{c}), \quad (35)$$

so that $\lambda_1=\mu_1+1$, and thus $\lambda_r=\max(\mu_1+1,\mu_2+2,\cdots\mu_r+r)$ in the recurrence (34). For $r=n$, we have

$$F(\mathbf{c})=F_n(\mathbf{c})=\exp H(0\cdots 0,\mathbf{c})R_n(\mathbf{c})=R_n(\mathbf{c}). \quad (36)$$

Thus we can calculate $F(\mathbf{c})$ entirely from the recurrence (34) with the initial condition (35), where the coefficients G_r are given by (28).

These results can be given a more agreeable form by shifting the arguments. In the definition of G_r, make the shift $x_{r+1}\to x_1, \cdots x_{r+\mu_r}\to x_{\mu_r}$ to obtain

$$G_r(x_1\cdots x_{\mu_r},\mathbf{c}) = c_0+\sum_{K_0\neq K\,\epsilon\,\kappa_r} c_{\delta_1(K)\cdots\delta_q(K)}x_{\delta_1(K)}\cdots x_{\delta_q(K)}. \quad (37)$$

For some purposes, it is useful to write S_r for the set of subscripts of the form $\delta_m(K)$ for $m=1, 2, \cdots, q$, and $K\epsilon\kappa_r$. Thus, only the x_j for $j\epsilon S_r$ actually enter $G_r(x_1\cdots x_{\mu_r},\mathbf{c})$. Thus $S=S_1\cup S_2\cdots\cup S_n$ is the set of all subscripts j of variables x_j entering the shifted formulation. We let α denote the number of such "essential" variables.

In the recurrence (34), we make the shift $x_r\to x_s$, $x_{r+1}\to x_1$, $x_{r+2}\to x_2$, $\cdots$, $x_{\lambda_r}\to x_{\lambda_r-r}$ and let $\beta_r=\lambda_r-r=\max(\mu_1+1-r,\mu_2+2-r,\cdots\mu_r)$. We then obtain, if $s\geqq\beta_{r-1}$,

$$R_r(x_1,\cdots x_{\beta_r},\mathbf{c}) = G_r(x_1\cdots x_{\mu_r},\mathbf{c})R_{r-1}(0,x_1,\cdots x_{\beta_{r-1}-1},\mathbf{c}) + Y_sR_{r-1}(x_s,x_1,\cdots x_{\beta_{r-1}-1},\mathbf{c}),$$
$$R_1(x_1\cdots x_{\mu_1},\mathbf{c})=G_1(x_1\cdots x_{\mu_1},\mathbf{c}). \quad (38)$$

This is permissible since the bound variable x_s on the right-hand side is different from the free variables $x_1\cdots x_{\beta_{r-1}-1}$.

Another expression for $F_n(\mathbf{c})$ is obtained by applying a fixed cyclic shift to each of the $G_r(x_{r+1}\cdots x_{r+\mu_r},\mathbf{c})$, $r=1\cdots n$ rather than a variable shift. In (34), replace x_1 by $x_{n-t+1}\cdots$, x_{t+1} by x_1, x_{t+2} by $x_2\cdots$, x_n by x_{n-t} for $r=1\cdots n$, where t is fixed. We find for $r=2\cdots t-1$ the recursion

$$R_r(x_{n-t+r+1},x_{n-t+r+2},\cdots x_n,x_1,\cdots,\mathbf{c}) = G_r(x_{n-t+r+1},\cdots x_n,x_1,\cdots,\mathbf{c}) \times R_{r-1}(0,x_{n-t+r+1},\cdots x_n,x_1,\cdots,\mathbf{c}) + Y_{n-t+r}R_{r-1}(x_{n-t+r},\cdots x_n,x_1,\cdots,\mathbf{c}), \quad (39)$$

and for $r=t+1\cdots n$ the recursion

$$R_r(x_{r-t+1},x_{r-t+2},\cdots,\mathbf{c}) = G_r(x_{r-t+1},x_{r-t+2},\cdots,\mathbf{c})R_{r-1}(0,x_{r-t+1},x_{r-t+2},\cdots,\mathbf{c}) + Y_{r-t}R_{r-1}(x_{r-t},x_{r-t+1},\cdots,\mathbf{c}). \quad (40)$$

Clearly, $R_1=G_1$, and if $t>1$,

$$R_t(x_1,x_2,\cdots,\mathbf{c})=G_t(x_1,x_2,\cdots,\mathbf{c})R_{t-1}(0,x_1,x_2,\cdots,\mathbf{c}) + Y_nR_{t-1}(x_n,x_1,x_2,\cdots,\mathbf{c}). \quad (41)$$

As before, the result $R_n(\mathbf{c})$ of the recursion is equal to $F_n(\mathbf{c})$.

The dependence of $G_r(x_1\cdots x_{\mu_r},\mathbf{c})$ on r is due to the "corner" and "edge" effects limiting the possible locations of the geometric objects. It may be advantageous to neglect these effects, replacing the various G_r by a suitably chosen G. On the other hand, as will be seen from a later paper, the edge effects may supply an entering wedge for a recursive simplification of the problem (by diagonalization of certain corresponding matrices). Hence it may also be advantageous to consider the exact problem including edge effects, or even to introduce artificial edge effects, in order to attain a solution.

5. DIRECT SOLUTION OF THE RECURRENCE

In simple cases, the recurrence may be solved by direct methods. Cases so solved may have an intrinsic physical interest and also suggest the more general methods developed in the next section. We now discuss three examples of increasing difficulty, the linear lattice with nearest neighbor association, the linear lattice with nearest and next-nearest neighbor association, and the rectangular lattice with nearest neighbor association.

(a) Linear Lattice, Nearest Neighbors

The geometric figures are $\kappa=\{K_0,K_1\}$, where $K_0=\cdot$, $K_1=\cdot\!-\!\cdot$. Clearly, $\kappa_1=\cdots=\kappa_{n-1}=\kappa$, $\kappa_n=\{K_0\}$ for the linear lattice

Thus $G_k(x_1,\mathbf{c})=c_0+c_1x_1$, $k=1\cdots n-1$, $G_n(x_1,\mathbf{c})=c_0$. The recurrence is

$$R_r(x_1,\mathbf{c})=R_{r-1}(0,\mathbf{c})G_r(x_1,\mathbf{c})+R_{r-1}'(0,\mathbf{c}) = (c_0+c_1x_1)R_{r-1}(0,\mathbf{c})+R_{r-1}'(0,\mathbf{c}), r=2\cdots n-1,$$
$$R_n(\mathbf{c})=R_{n-1}(0,c)\cdot c_0+R_{n-1}'(0,\mathbf{c}). \quad (42)$$

Since each R_r is linear in x_1, we can write $R_r(x_1,\mathbf{c}) = d_0^{(r)}+d_1^{(r)}x_1$, so the recurrence becomes $d_0^{(r)}=c_0d_0^{(r-1)} + d_1^{(r-1)}$, $d_1^{(r)}=c_1d_0^{(r-1)}$. Elimination of $d_1^{(r)}$ yields the

second-order difference equation $d_0^{(r)}=c_0d_0^{(r-1)}+c_1d_0^{(r-2)}$ whose solution is $d_0^{(r)}=\alpha_1\lambda_1{}^r+\alpha_2\lambda_2{}^r$, where λ_1, λ_2, are the roots of $\lambda^2=c_0\lambda+c_1$. Hence $R_r(x_1,\mathbf{c})=\alpha_1\lambda_1{}^r+\alpha_2\lambda_2{}^r+c_1x_1(\alpha_1\lambda_1{}^{r-1}+\alpha_2\lambda_2{}^{r-1})$, where $\alpha_1\lambda_1+\alpha_2\lambda_2=c_0, \alpha_1+\alpha_2=1$, $r=1\cdots n-1$. Thus, finally

$$R_n(\mathbf{c})=c_0(\alpha_1\lambda_1{}^{n-1}+\alpha_2\lambda_2{}^{n-1})+c_1(\alpha_1\lambda_1{}^{n-2}+\alpha_2\lambda_2{}^{n-2})$$
$$=(\lambda_1{}^{n+1}-\lambda_2{}^{n+1})/(\lambda_1-\lambda_2). \tag{43}$$

A more elegant approach is obtained by writing

$$\mathbf{d}^{(r)}=\begin{pmatrix} d_0^{(r)} \\ d_1^{(r)} \end{pmatrix}$$

so that $R_r(x_1,\mathbf{c})=(1,x_1)\cdot\mathbf{d}^{(r)}$. The recurrence becomes $\mathbf{d}^{(r)}=M\mathbf{d}^{(r-1)}$, where

$$M=\begin{pmatrix} c_0 & 1 \\ c_1 & 0 \end{pmatrix},$$

with the result that

$$\mathbf{d}^{(r)}=M^{r-1}\mathbf{d}^{(1)}=M^{r-1}\begin{pmatrix} c_0 \\ c_1 \end{pmatrix}=M^r\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad r=1\cdots n-1.$$

Thus

$$R_r(x_1,\mathbf{c})=(1,x_1)M^r\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad r=1\cdots n-1,$$

$$R^n(\mathbf{c})=(c_0,0)M^{n-1}\begin{pmatrix} 1 \\ 0 \end{pmatrix}+(0,1)M^{n-1}\begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{44}$$
$$=(c_0,1)M^{n-1}\begin{pmatrix} 1 \\ 0 \end{pmatrix}=(1,0)M^n\begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

To recover the previous solution, it is sufficient to note that the characteristic equation of M is $\lambda^2=c_0\lambda+c_1$.

(b) Linear Lattice, Nearest and Next-Nearest Neighbors

Now $\kappa=\{K_0,K_1,K_2\}$, where $K_2=\cdot\!-\!\cdot\!-\!\cdot$, so $\kappa_1=\kappa_{n-2}=\kappa,\kappa_{n-1}=\{K_0,K_1\}$, $\kappa_n=\{K_0\}$. Thus $K_k(x_1,x_2,\mathbf{c})=c_0+c_1x_1+c_2x_2$, $k=1\cdots n-2$, $G_{n-1}(x_1,\mathbf{c})=c_0+c_1x_1$, $G_n(\mathbf{c})=c_0$. The recurrence is

$$R_r(x_1,x_2)=R_{r-1}(0,x_1)(c_0+c_1x_1+c_2x_2)$$
$$+\frac{\partial}{\partial x_2}R_{r-1}(x_2,x_1)\,|_{\,x_2=0}. \tag{45}$$

Obviously, $R_r(x_1,x_2)$ is linear in x_2. If we assume $R_{r-1}(x_1,x_2)$ is quadratic in x_1,

$$R_r(x_1,x_2)=(1,x_1,x_1{}^2,x_2,x_1x_2)\mathbf{d}^{(r)},$$

then

$$R_{r-1}(0,x_1)=(1,0,0,x_1,0)\mathbf{b}^{(r-1)},$$

$$\frac{\partial}{\partial x_2}R_{r-1}(x_2,x_1)\,|_{\,x_2=0}=(0,1,0,0,x_1)\mathbf{d}^{(r-1)},$$

and

$$R_r(x_1,x_2)=\begin{bmatrix} c_0+c_1x_1+c_2x_2 \\ 1 \\ 0 \\ c_0x_1+c_1x_1{}^2+c_2x_1x_2 \\ x_1 \end{bmatrix}\mathbf{d}^{(r-1)}$$
$$=(1,x_1,x_1{}^2,x_2,x_1x_2)M(c_0,c_1,c_2)\mathbf{d}^{(r-1)}$$
$$=(1,x_1,x_1{}^2,x_2,x_1,x_2)\mathbf{d}^{(r)}, \tag{46}$$

where

$$M(c_0,c_1,c_2)=\begin{bmatrix} c_0 & 1 & 0 & 0 & 0 \\ c_1 & 0 & 0 & c_0 & 0 \\ 0 & 0 & 0 & c_1 & 0 \\ c_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & c_2 & 0 \end{bmatrix}. \tag{47}$$

Hence

$$\mathbf{d}^{(r)}=[M(c_0,c_1c_2)]^{r-1}\mathbf{d}^{(1)}=[M(c_0,c_1c_2)^r]\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

for $r=1\cdots n-2$. From the last two steps of the recurrence for $r=n-1$, we find $\mathbf{d}^{(n-1)}=M(c_0,c_1,0)\mathbf{d}^{(n-2)}$, $\mathbf{d}^{(n)}=M(c_0,0,0)\mathbf{d}^{(n-1)}$, and thus

$$R_n(\mathbf{c})=(1\ 0\ 0\ 0\ 0)$$
$$\cdot[M(c_0,0,0)M(c_0,c_10)M(c_0,c_1,c_2)^{n-2}]\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

which can be simplified to

$$(1\ 0\ 0\ 0\ 0)[M(c_0,c_1,c_2)^n]\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

If the third row and column are deleted from the foregoing vectors and matrix, the resulting expression

$$(1\ 0\ 0\ 0)\begin{bmatrix} c_0 & 1 & 0 & 0 \\ c_1 & 0 & c_0 & 1 \\ c_2 & 0 & 0 & 0 \\ 0 & 0 & c_2 & 0 \end{bmatrix}^n\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{48}$$

is also equal to $R_n(\mathbf{c})$. The characteristic equation of the last matrix is $\lambda^4=c_0\lambda^3+c_1\lambda^2+c_0c_2\lambda+c_2{}^2$. If U is the diagonalizing matrix and Λ the diagonalized matrix so that $M(c_0,c_1,c_2)=U\Lambda U^{-1}$, then

$$R_n(\mathbf{c})=(1\ 0\ 0\ 0)U\Lambda^nU^{-1}\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

(c) Rectangular Lattice, Nearest Neighbors

Consider a p by q rectangular lattice, in which the lattice point in the ith row and jth column is denoted by $x_{(i-1)q+j}$. A unit horizontal translation shifts the

subscript by 1, a unit vertical translation by q. The figures $K_0=\cdot$, $K_1=\cdot\!-\!\cdot$, $K_2=\vdots$, and so

$$\begin{aligned} &G_{i,j}=G_{(i-1)q+j}=c_0+c_1x_1+c_qx_q, \quad i<p, \quad j<q,\\ &G_{p,j}=G_{(p-1)q+j}=c_0+c_1x_1, \quad j<q,\\ &G_{i,q}=G_{iq}=c_0+c_qx_q, \quad i<p,\\ &G_{p,q}=G_{pq}=c_0. \end{aligned} \tag{49}$$

For $r=1,\cdots q-1$, this is the same problem as (b), with c_2, x_2 replaced by c_q, x_q. If

$$N(c_0,c_1,c_q)=\begin{bmatrix} c_0 & 1 & 0 & 0\\ c_1 & 0 & c_0 & 1\\ c_q & 0 & 0 & 0\\ 0 & 0 & c_q & 0 \end{bmatrix}, \tag{50}$$

we thus have

$$R_r(x_1,x_q)=(1,x_1,x_q,x_1x_q)N(c_0,c_1,c_q)^{r-1}\begin{bmatrix}1\\0\\0\\0\end{bmatrix},$$

$r=1, \cdots q-1$.

Proceeding as in case (b), we find

$$R_q(x_1,x_q)=(1,x_1,x_q,x_1x_q)N(c_0,c_1,0)N(c_0,c_1c_q)^{q-1}\begin{bmatrix}1\\0\\0\\0\end{bmatrix},$$

thus completing the first row. By similar methods, we find for the end-of-row results

$$R_{(i-1)q}(x_1,x_q)=(1,x_1,x_q,x_1x_q)\cdot[N(c_0,0,c_q)N(c_0,c_1c_q)^{q-1}]^{i-1}\begin{bmatrix}1\\0\\0\\0\end{bmatrix}$$

for $i<p$. Completing the last row, we find

$$R_{pq}(\mathbf{c})=(1\ 0\ 0\ 0)[N(c_0,0,0)N(c_0,c_1,0)^{q-1}]\cdot[N(c_0,0,c_q)N(c_0,c_1,c_q)^{q-1}]^{p-1}\begin{bmatrix}1\\0\\0\\0\end{bmatrix}. \tag{51}$$

Once the form of matrix $N(c_0,c_1,c_q)$ without edge effects is known, the expression $R_{pq}(\mathbf{c})$ evidently can be written down by a simple prescription. The justification of such a prescription for solution in the general case is taken up in the next section.

In case (b), an extraneous dimension was originally present. In more complicated problems, the number of extraneous dimensions rises sharply. A desirable feature of a general solution method is the automatic omission of all extraneous dimensions.

6. SOLUTION BY A MATRIX-SPINOR METHOD

We first recall the definition of the direct, or Kronecker, product of matrices. If A and B are matrices, then $A\times B$ is the matrix formed by replacing each element b_{jk} of B by the matrix Ab_{jk}. The most important property of direct product used here is that $(A_1A_2)\times(B_1B_2)=(A_1\times B_1)(A_2\times B_2)$. Also, $\mathrm{tr}(A\times B)=\mathrm{tr}A\ \mathrm{tr}B$ for square matrices.

We now introduce a trio of singular matrices. Let

$$X=\begin{bmatrix}0&0\\1&0\end{bmatrix}, \quad D=\begin{bmatrix}0&1\\0&0\end{bmatrix}, \quad Z=\begin{bmatrix}1&0\\0&0\end{bmatrix}, \tag{52}$$

and check that $X^2=D^2=0$, $Z^2=Z$, $DX=Z$, $XD=I-Z$, $ZD=D$, $DZ=0$, $XZ=X$, $ZX=0$.

Now define the direct products

$$\begin{aligned} \mathbf{X}_p&=I_1\times\cdots\times I_{p-1}\times X_p\times I_{p+1}\times\cdots\times I_n\\ \mathbf{D}_p&=I_1\times\cdots\times I_{p-1}\times D_p\times I_{p+1}\times\cdots\times I_n\\ \mathbf{Z}_p&=I_1\times\cdots\times I_{p-1}\times Z_p\times I_{p+1}\times\cdots\times I_n \end{aligned} \tag{53}$$

where the subscripts on the right merely locate the factors, and I denotes the unit matrix. Clearly, $\mathbf{X}_p$, $\mathbf{D}_p$, $\mathbf{Z}_p$ are 2^n by 2^n matrices satisfying, for each p, the commutation rules for X, D, Z. Any matrix function of $\mathbf{X}_p$, $\mathbf{D}_p$, $\mathbf{Z}_p$ commutes with any matrix function of $\mathbf{X}_q$, $\mathbf{D}_q$, $\mathbf{Z}_q$ for $p\neq q$.

If $H(x_1\cdots x_n)$ is any polynomial in $x_1\cdots x_n$, then the replacement in $H(x_1\cdots x_n)$ of any nonempty subset of $\{x_1\cdots x_n\}$ by the corresponding subset of $\{\mathbf{X}_1\cdots\mathbf{X}_n\}$ produces a well-defined 2^n by 2^n matrix. The exponent of a matrix is defined in the usual way and is nonsingular. It is convenient to write $\mathbf{H}=H(\mathbf{X}_1\cdots\mathbf{X}_n)$ and $(T_j\mathbf{H}x_j)$ for $H(\mathbf{X}_1\cdots\mathbf{X}_{j-1},x_j,\mathbf{X}_{j+1}\cdots\mathbf{X}_n)$, etc.

We now establish the matrix identities

$$\begin{aligned} \mathbf{Z}_j\mathbf{D}_j\exp\mathbf{H}&=\left(\mathbf{D}_j+\mathbf{Z}_j\frac{\partial}{\partial x_j}\right)T_j\exp\mathbf{H}(x_j)\Big|_{x_j=0}\\ \mathbf{Z}_j\exp\mathbf{H}&=\mathbf{Z}_jT_j\exp\mathbf{H}(x_j)\,|_{x_j=0}\\ \mathbf{X}_j\exp\mathbf{H}&=\mathbf{X}_jT_j\exp\mathbf{H}(x_j)\,|_{x_j=0}. \end{aligned} \tag{54}$$

Let $H(x_1\cdots x_n)=\sum_{i=0}^{k}a_ix_j^i$, where $a_0\cdots a_k$ are polynomials in $x_1\cdots x_{j-1}$, $x_{j+1}\cdots x_n$. Thus

$$\frac{\partial}{\partial x_j}T_j(\exp\mathbf{H}x_j)\Big|_{x_j=0}=\mathbf{A}_1\exp\mathbf{A}_0$$

and $T_j\exp\mathbf{H}(0)=\exp\mathbf{A}_0$, where $\mathbf{A}_i=a_i(\mathbf{X}_1\cdots\mathbf{X}_{i-1},\mathbf{X}_{i+1}\cdots\mathbf{X}_n)$. On the other hand, since $\mathbf{X}_j^2=0$, we have $\mathbf{H}=\mathbf{A}_0+\mathbf{A}_1\mathbf{X}_j$, $\exp\mathbf{H}=\exp\mathbf{A}_0+\mathbf{A}_1\mathbf{X}_j\exp\mathbf{A}_0$, and therefore

$$\begin{aligned} \mathbf{Z}_j\mathbf{D}_j\exp\mathbf{H}&=\mathbf{D}_j\exp\mathbf{A}_0+\mathbf{Z}_j\mathbf{A}_1\exp\mathbf{A}_0\\ &=\mathbf{D}_j\mathbf{T}_j\exp\mathbf{H}(0)+\mathbf{Z}_j\frac{\partial}{\partial x_j}T_j\exp\mathbf{H}(x_j)\Big|_{x_j=0}\\ &=\left(\mathbf{D}_j+\mathbf{Z}_j\frac{\partial}{\partial x_j}\right)T_j\exp\mathbf{H}(x_j)\Big|_{x_j=0}, \end{aligned}$$

and

$$\begin{aligned} \mathbf{Z}_j\exp\mathbf{H}&=\mathbf{Z}_j\exp\mathbf{A}_0=\mathbf{Z}_jT_j\exp\mathbf{H}(x_j)\,|_{x_j=0},\\ \mathbf{X}_j\exp\mathbf{H}&=\mathbf{X}_j\exp\mathbf{A}_0=\mathbf{X}_jT_j\exp\mathbf{H}(x_j)\,|_{x_j=0}. \end{aligned}$$

Hence

$$(\prod_{j=1}^{n}\mathbf{Z}_j\mathbf{D}_j)\exp\mathbf{H}=\prod_{j=1}^{n}\left(\mathbf{D}_j+\mathbf{Z}_j\frac{\partial}{\partial x_j}\right)\cdot\exp H(x_1\cdots x_n)\Big|_{x_1=\cdots=x_n=0}. \quad (55)$$

We now prove that

$$F_n=\frac{\partial^n}{\partial x_1\cdots\partial x_n}\exp H(x_1\cdots x_n)\Big|_{x_1=\cdots=x_n=0}$$

can be written as

$$\operatorname{tr}(\prod_{j=1}^{n}\mathbf{Z}_j\mathbf{D}_j\exp\mathbf{H}). \quad (56)$$

First write

$$\prod_{j=1}^{n}\left(\mathbf{D}_j+\mathbf{Z}_j\frac{\partial}{\partial x_j}\right)$$

as

$$\sum_A(\prod_{i\in\Omega-A}\mathbf{D}_i)\prod_{j\in A}\left(\mathbf{Z}_j\frac{\partial}{\partial x_j}\right),$$

where $\Omega=\{1,2\cdots n\}$ and A ranges over all subsets of Ω. But

$$(\prod_{i\in\Omega-A}\mathbf{D}_i)\prod_{j\in A}\mathbf{Z}_j=W_1{}^{(A)}\times\cdots\times W_n{}^{(A)},$$

where

$$W_i{}^{(A)}=\begin{cases}D, & i\in\Omega-A\\ Z, & i\in A\end{cases}$$

in view of the definition of $\mathbf{D}_i$, $\mathbf{Z}_j$, and the main property of direct products. Thus

$$\operatorname{tr}(\prod_{i\in\Omega-A}\mathbf{D}_i\prod_{j\in A}\mathbf{Z}_j)=\prod_{i=1}^{n}\operatorname{tr}W_i{}^{(A)}=\begin{cases}1, & A=\Omega\\ 0, & A\neq\Omega\end{cases},$$

since $\operatorname{tr}D=0$, $\operatorname{tr}Z=1$. We see that

$$\operatorname{tr}\prod_{j=1}^{n}\left(\mathbf{D}_j+\mathbf{Z}_j\frac{\partial}{\partial x_j}\right)=\prod_{j=1}^{n}\frac{\partial}{\partial x_j}$$

and thus

$$\operatorname{tr}\prod_{j=1}^{n}\mathbf{Z}_j\mathbf{D}_j\exp\mathbf{H}=\prod_{j=1}^{n}\frac{\partial}{\partial x_j}\exp H\Big|_{x_1=\cdots=x_n=0}=F_n.$$

More convenient for our purposes is another expression for F_n. Let

$$\psi^{(0)}=\begin{pmatrix}1\\0\end{pmatrix},\quad \psi^{(1)}=\begin{pmatrix}0\\1\end{pmatrix},$$

and let

$$\Psi^{(0)}=\psi_1{}^{(0)}\times\cdots\times\psi_n{}^{(0)}$$

be the direct product of n copies of $\psi^{(0)}$,

$$\Psi^{(1)}=\psi_1{}^{(1)}\times\cdots\times\psi_n{}^{(1)}$$

be the direct product of n copies of $\psi^{(1)}$. (57)

If $\mathbf{M}=M_1\times\cdots\times M_n$, where $M_1\cdots M_n$ are 2 by 2 matrices, then

$$\Psi^{(0)\prime}\mathbf{M}\Psi^{(0)}=\psi^{(0)\prime}M_1\psi^{(0)}\times\cdots\times\psi^{(0)\prime}M_n\psi^{(0)}=\prod_{j=1}^{n}(\psi^{(0)\prime}M_j\psi^{(0)}),$$

since the direct product of scalars is a scalar. Hence

$$\Psi^{(0)\prime}\prod_{j=1}^{n}\left(\mathbf{D}_j+\mathbf{Z}_j\frac{\partial}{\partial x_j}\right)\Psi^{(0)}=\sum_A(\Psi^{(0)\prime}W_1{}^{(A)}\times\cdots\times W_n{}^{(A)}\Psi^{(0)})\prod_{j\in A}\frac{\partial}{\partial x_j}=\sum_A\left[(\prod_{i=1}^{n}\psi^{(0)\prime}W_i{}^{(A)}\psi^{(0)})\prod_{j\in A}\frac{\partial}{\partial x_j}\right]. \quad (58)$$

But $\psi^{(0)\prime}D\psi^{(0)}=0$, $\psi^{(0)\prime}Z\psi^{(0)}=1$, so that

$$\prod_{i=1}^{n}\psi^{(0)\prime}W_i{}^{(A)}\psi^{(0)}=\begin{cases}1, & A=\Omega\\ 0, & A\neq\Omega\end{cases},$$

and therefore

$$\Psi^{(0)\prime}\prod_{j=1}^{n}\left(\mathbf{D}_j+\mathbf{Z}_j\frac{\partial}{\partial x_j}\right)\Psi^{(0)}=\prod_{j=1}^{n}\frac{\partial}{\partial x_j}. \quad (59)$$

Hence we have

$$F_n=\Psi^{(0)\prime}\prod_{j=1}^{n}\mathbf{Z}_j\mathbf{D}_j\exp\mathbf{H}\Psi^{(0)}. \quad (60)$$

Since $\psi^{(0)\prime}ZD=\psi^{(1)\prime}$, it follows that

$$\Psi^{(0)\prime}\prod_{j=1}^{n}\mathbf{Z}_j\mathbf{D}_j=\Psi^{(1)\prime}$$

and $F_n=\Psi^{(1)\prime}\exp\mathbf{H}\Psi^{(0)}$.

We are now prepared to derive a general recurrence formula for $F_n(\mathbf{c})$. So far in this section, H has been an arbitrary polynomial. We specialize to the problem of interest by taking $H(x_1\cdots x_n)=H(x_1\cdots x_n,\mathbf{c})$, the polynomial for the combinatorial problem of Secs. 2 and 3.

Let $\mathbf{G}_r(\mathbf{c})$ be defined as the result $G_r(\mathbf{X}_{r+1}\cdots\mathbf{X}_{r+\mu_r},\mathbf{c})$ of replacing x_1 by $\mathbf{X}_1\cdots$, $x_{r+\mu_r}$ by $\mathbf{X}_{r+\mu_r}$ in $G_r(x_{r+1}\cdots x_{r+\mu_r},\mathbf{c})$ of (28), and let

$$\mathbf{F}_r(c)=\prod_{j=1}^{r}(\mathbf{Z}_j\mathbf{D}_j)\exp\mathbf{H},\quad r=1\cdots n. \quad (61)$$

Thus, from (54) we have

$$\mathbf{F}_1(c)=\mathbf{Z}_1\mathbf{D}_1\exp\mathbf{H}=\left(\mathbf{D}_1+\mathbf{Z}_1\frac{\partial}{\partial x_1}\right)T_1\exp\mathbf{H}(x_1)\Big|_{x_1=0}=\mathbf{D}_1\mathbf{Z}_1\exp\mathbf{H}+\mathbf{Z}_1\frac{\partial}{\partial x_1}T_1\exp\mathbf{H}(x_1)\Big|_{x_1=0}.$$

However, from (30) and the definition of $T_1\mathbf{H}(x_1)$ we find

$$\frac{\partial}{\partial x_1}T_1 \exp\mathbf{H}(x_1)\Big|_{x_1=0} = \exp\mathbf{H}(x_1)\Big|_{x_1=0}\cdot\mathbf{G}_1(\mathbf{c})$$

and thus $\mathbf{F}_1(\mathbf{c})=(\mathbf{D}_1+\mathbf{G}_1\mathbf{Z}_1)\mathbf{Z}_1 \exp\mathbf{H}$.

Now assume that

$$\mathbf{F}_{r-1}(\mathbf{c})=\mathbf{R}_{r-1}\mathbf{Z}_1\cdots\mathbf{Z}_{r-1}\exp\mathbf{H} \tag{62}$$

for some $r\geqq 2$, some matrix $\mathbf{R}_{r-1}$ commuting with $\mathbf{X}_r\cdots\mathbf{X}_n$, $\mathbf{D}_r\cdots\mathbf{D}_n$, $\mathbf{Z}_r\cdots\mathbf{Z}_n$, and hence with $\mathbf{G}_r\cdots\mathbf{G}_n$. Then

$$\begin{aligned}
&\mathbf{F}_r(\mathbf{c})\\
&=\mathbf{Z}_r\mathbf{D}_r\mathbf{R}_{r-1}\mathbf{Z}_1\cdots\mathbf{Z}_{r-1}\exp\mathbf{H}\\
&=\mathbf{Z}_r\mathbf{R}_{r-1}\mathbf{Z}_1\cdots\mathbf{Z}_{r-1}\mathbf{Z}_r\mathbf{D}_r\exp\mathbf{H}\\
&=\mathbf{Z}_r\mathbf{R}_{r-1}\mathbf{Z}_1\cdots\mathbf{Z}_{r-1}\left(\mathbf{D}_r+\mathbf{Z}_r\frac{\partial}{\partial x_r}\right)T_r\exp\mathbf{H}(x_r)\Big|_{x_r=0}\\
&=\mathbf{Z}_r\mathbf{R}_{r-1}\mathbf{Z}_1\cdots\mathbf{Z}_{r-1}(\mathbf{D}_r+\mathbf{G}_r)\mathbf{Z}_r\exp\mathbf{H}\\
&=(\mathbf{D}_r+\mathbf{G}_r\mathbf{Z}_r)\mathbf{R}_{r-1}\mathbf{Z}_1\cdots\mathbf{Z}_r\exp\mathbf{H}\\
&=\mathbf{R}_r\mathbf{Z}_1\cdots\mathbf{Z}_r\exp\mathbf{H},
\end{aligned} \tag{63}$$

where

$$\mathbf{R}_r=(\mathbf{D}_r+\mathbf{G}_r\mathbf{Z}_r)\mathbf{R}_{r-1}. \tag{64}$$

Note that $\mathbf{R}_r$ commutes with $\mathbf{X}_{r+1}\cdots\mathbf{X}_n$, $\mathbf{D}_{r+1}\cdots\mathbf{D}_n$, $\mathbf{Z}_{r+1}\cdots\mathbf{Z}_n$. By induction we have for $r=2, 3\cdots n$,

$$\begin{aligned}
\mathbf{R}_1&=\mathbf{D}_1+\mathbf{G}_1\mathbf{Z}_1\\
\mathbf{R}_r&=(\mathbf{D}_r+\mathbf{G}_r\mathbf{Z}_r)\mathbf{R}_{r-1}.
\end{aligned} \tag{65}$$

By definition,

$$\begin{aligned}
\mathbf{F}_n(\mathbf{c})&=\mathbf{R}_n\mathbf{Z}_1\cdots\mathbf{Z}_n\exp\mathbf{H}\\
&=\mathbf{R}_n\mathbf{Z}_1\cdots\mathbf{Z}_nT_1\cdots T_n\exp\mathbf{H}\,|_{x_1=\cdots=x_n=0}\\
&=\mathbf{R}_n\mathbf{Z}_1\cdots\mathbf{Z}_n=\mathbf{R}_n(Z_1\times\cdots\times Z_n).
\end{aligned}$$

Hence

$$\begin{aligned}
&F_n(\mathbf{c})=\Psi^{(0)\prime}\mathbf{F}_n(\mathbf{c})\Psi^{(0)}=\Psi^{(0)\prime}\mathbf{R}_n(Z_1\times\cdots\times Z_n)\Psi^{(0)}\\
&=\Psi^{(0)\prime}\mathbf{R}_n(Z_1\psi_1^{(0)}\times\cdots\times Z_n\psi_n^{(0)})=\Psi^{(0)\prime}\mathbf{R}_n\Psi^{(0)}.
\end{aligned} \tag{66}$$

Thus, it suffices to determine

$$\begin{aligned}
\mathbf{R}_n&=(\mathbf{D}_n+\mathbf{G}_n\mathbf{Z}_n)(\mathbf{D}_{n-1}+\mathbf{G}_{n-1}\mathbf{Z}_{n-1})\cdots(\mathbf{D}_1+\mathbf{G}_1\mathbf{Z}_1)\\
&\text{or}\quad \mathbf{R}_n=\mathbf{R}_n\Psi^{(0)}.
\end{aligned} \tag{67}$$

This formula has the important drawback of involving the calculation of the $3n$ matrices $\mathbf{D}_j$, $\mathbf{G}_j$, $\mathbf{Z}_j$, $j=1\cdots n$, the n matrices $\mathbf{D}_j+\mathbf{G}_j\mathbf{Z}_j$, $j=1\cdots n$, and finally their product, each of which is 2^n by 2^n. However, there are three methods of simplification of (62). The first is application of the shift technique which produced the convenient recursion (38) from the less convenient (34). The second is recursive diagonalization of the partial products obtained in the calculation of $\mathbf{R}_n$ or of $\mathbf{R}_n=\mathbf{R}_n\Psi^{(0)}$, which we will discuss in a later paper. The third is simply the combination of the first and second. In the remainder of this section we carry out the details of a matrix rendition of the shift technique. The first version reduces the problem to the product of n matrices, each 2^α by 2^α, of reasonably similar or even identical form, where α is the maximum number of distinct x's needed in the G_r.

If $1\leqq m\leqq n$, let ${}_m\Psi^{(0)}=\psi_1^{(0)}\times\cdots\times\psi_m^{(0)}$ and let ${}_mV(j)$ be the 2^m element column with 1 in the jth row and zero elsewhere. It is easy to verify that if $j-1=i_1+i_2\cdot 2+\cdots+i_m2^{m-1}$, then

$${}_mV(j)=\psi_1(i_1)\times\cdots\times\psi_m(i_m). \tag{68}$$

Thus the 2^m sets of choices of $i_1\cdots i_m$ from the set $\{0, 1\}$ give the 2^m columns ${}_mV(1)\cdots{}_mV(2^m)$.

For $1\leqq j\leqq m$, let

$${}_m\mathbf{X}_j=I_1\times\cdots\times X_j\times\cdots\times I_m. \tag{69}$$

Thus $\mathbf{X}_j={}_m\mathbf{X}_j\times(I_{m+1}\times\cdots\times I_n)$ for $1\leqq j\leqq m$. The subscript m signifies the reduction of the dimensionality of the direct product to that necessary to accomodate m independent nilpotents.

A straightforward direct-product induction argument shows that

$${}_m\mathbf{X}_{k_1}\cdots{}_m\mathbf{X}_{k_p}({}_m\Psi^{(0)})={}_mV\Big(\sum_{j=1}^{p}2^{k_j-1}+1\Big)\quad\text{for}$$

$$k_1\neq\cdots\neq k_p,\ 1\leqq k_1\cdots k_p\leqq m,\ 1\leqq p\leqq m. \tag{70}$$

Let ${}_m\mathbf{J}_{i,k}$ denote the result of interchanging the $2^{i-1}+1$ and $2^{k-1}+1$ columns of the 2^m by 2^m unit matrix. From (64) it follows that

$$\begin{aligned}
&{}_m\mathbf{J}_{k_1,k_2}\,{}_m\mathbf{X}_{k_1}\,{}_m\mathbf{X}_{k_2}\cdots{}_m\mathbf{X}_{k_p}\,{}_m\Psi^{(0)}\\
&\qquad={}_m\mathbf{X}_{k_1}\,{}_m\mathbf{X}_{k_2}\cdots{}_m\mathbf{X}_{k_p}\,{}_m\Psi^{(0)},\\
&{}_m\mathbf{J}_{k_1,k_2}\,{}_m\mathbf{X}_{k_2}\,{}_m\mathbf{X}_{k_3}\cdots{}_m\mathbf{X}_{k_p}\,{}_m\Psi^{(0)}\\
&\qquad={}_m\mathbf{X}_{k_1}\,{}_m\mathbf{X}_{k_3}\cdots{}_mX_{k_p}\,{}_m\Psi^{(0)},\\
&{}_m\mathbf{J}_{k_1,k_2}\,{}_m\mathbf{X}_{k_3}\cdots{}_m\mathbf{X}_{k_p}\,{}_m\Psi^{(0)}={}_m\mathbf{X}_{k_3}\cdots{}_m\mathbf{X}_{k_p}\,{}_m\Psi^{(0)},
\end{aligned} \tag{71}$$

for $k_1\neq k_2\cdots\neq k_p$, $1\leqq k_1\cdots k_p\leqq m$, $1\leqq p\leqq m$. Thus ${}_m\mathbf{J}_{k_1,k_2}$ has the effect of interchanging ${}_m\mathbf{X}_{k_1}$ and ${}_m\mathbf{X}_{k_2}$, *after application* to ${}_m\Psi^{(0)}$. The matrix ${}_m\mathbf{J}_{k_1,k_2}$ does not quite interchange ${}_m\mathbf{X}_{k_1}$ and ${}_m\mathbf{X}_{k_2}$ themselves.

We now "cut down" the dimensionality of the problem from 2^n to 2^α by considering the matrices ${}_\alpha\mathbf{X}_1$, ${}_\alpha\mathbf{X}_2\cdots{}_\alpha\mathbf{X}_\alpha$ only. In the remainder of this section, we omit the subscript α when it denotes the reduced matrices, columns, etc., (i.e., when α precedes the matrix, column, etc., to which it is a subscript).

We are now ready to derive the "shifted" form of the matrix equation matrix. Let $\mathbf{G}_r{}^\dagger$ be the result

$$G_r(\mathbf{X}_1\cdots\mathbf{X}_{\mu_r},\mathbf{c}) \tag{72}$$

of replacing x_1 by $\mathbf{X}_1\cdots$, x_{μ_r} by $\mathbf{X}_{\mu_r}$ in $G_r(x_1\cdots x_{\mu_r},\mathbf{c})$.

Let

$$\mathbf{C}_\alpha=\mathbf{J}_{\alpha,1}\cdot\mathbf{J}_{\alpha,2}\cdots\mathbf{J}_{\alpha,\alpha-1}. \tag{73}$$

A little manipulation shows that $\mathbf{C}_\alpha=(e_{jk})$, where

$$\begin{aligned} e_{jk}&=\delta(j,\tfrac{1}{2}k+\tfrac{1}{2}), \quad k \text{ odd},\\ e_{jk}&=\delta(j,2^{\alpha-1}+\tfrac{1}{2}k), \quad k \text{ even}. \end{aligned} \tag{74}$$

For example,

$$\mathbf{C}_1=\begin{bmatrix}1&0\\0&1\end{bmatrix},\quad \mathbf{C}_2=\begin{bmatrix}1&0&0&0\\0&0&1&0\\0&1&0&0\\0&0&0&1\end{bmatrix},$$

$$\mathbf{C}_3=\begin{bmatrix}1&0&0&0&0&0&0&0\\0&0&1&0&0&0&0&0\\0&0&0&0&1&0&0&0\\0&0&0&0&0&0&1&0\\0&1&0&0&0&0&0&0\\0&0&0&1&0&0&0&0\\0&0&0&0&0&1&0&0\\0&0&0&0&0&0&0&1\end{bmatrix}.$$

Note that the matrix $\mathbf{C}_\alpha$ can be partitioned for $\alpha\geqq 2$ into an upper matrix U_α and a lower matrix L_α, each of which is $2^{\alpha-1}$ by 2^α. These can be further partitioned into $2^{\alpha-2}$ (2 by 4) blocks, the off-diagonal blocks being zero blocks. The 2 by 4 diagonal blocks consist of

$$\begin{bmatrix}1&0&0&0\\0&0&1&0\end{bmatrix}=(1\ 0)\times I \text{ in } U_\alpha$$

and

$$\begin{bmatrix}0&1&0&0\\0&0&0&1\end{bmatrix}=(0\ 1)\times I \text{ in } L_\alpha.$$

Henceforth let $I_{(\alpha-1)}$ denote the $2^{\alpha-1}$ by $2^{\alpha-1}$ unit matrix. We see that $U_\alpha=(1\ 0)\times I_{(\alpha-1)}$, $L=(0\ 1)\times I_{(\alpha-1)}$.

For later purposes we want to know $\mathbf{D}_\alpha\mathbf{C}_\alpha$ and $\mathbf{Z}_\alpha\mathbf{C}_\alpha$. Since $\mathbf{D}_\alpha=I_{(\alpha-1)}\times D$, $\mathbf{Z}_\alpha=I_{(\alpha-1)}\times Z$, we find

$$\mathbf{D}_\alpha=\left(\begin{array}{c|c}0&\\ \vdots&I_{(\alpha-1)}\\0&\end{array}\right),\quad \mathbf{Z}_\alpha=\left(\begin{array}{c|c}I_{(\alpha-1)}&\\ &\vdots\\ &0\end{array}\right). \tag{75}$$

But

$$\mathbf{C}_\alpha=\begin{pmatrix}U_\alpha\\ \hline L_\alpha\end{pmatrix},$$

so that

$$\mathbf{D}_\alpha{}^\dagger=\mathbf{D}_\alpha\mathbf{C}_\alpha=\begin{pmatrix}L_\alpha\\ \hline 0\end{pmatrix}=\begin{pmatrix}0&1&0&0\\0&0&0&1\end{pmatrix}\times\begin{pmatrix}I_{(\alpha-2)}&\\ &0\end{pmatrix}$$

$$=(0\ 1)\times I_{(\alpha-2)}\times\begin{pmatrix}1\\0\end{pmatrix},$$

$$\mathbf{Z}_\alpha{}^\dagger=\mathbf{Z}_\alpha\mathbf{C}_\alpha=\begin{pmatrix}U_\alpha\\ \hline 0\end{pmatrix}=\begin{pmatrix}1&0&0&0\\0&0&1&0\end{pmatrix}\times\begin{pmatrix}I_{(\alpha-2)}&\\ &0\end{pmatrix}$$

$$=(1\ 0)\times I_{(\alpha-2)}\times\begin{pmatrix}1\\0\end{pmatrix}. \tag{76}$$

Now define recursively

$$\begin{aligned}\mathbf{R}_r{}^\dagger&=(\mathbf{D}_\alpha+\mathbf{G}_r{}^\dagger\mathbf{Z}_\alpha)\mathbf{C}_\alpha\mathbf{R}_{r-1}{}^\dagger\\&=(\mathbf{D}_\alpha{}^\dagger+\mathbf{G}_r{}^\dagger\mathbf{Z}_\alpha)\mathbf{R}_{r-1}{}^\dagger,\\ \mathbf{R}_1{}^\dagger&=(\mathbf{D}_\alpha+\mathbf{G}_1{}^\dagger\mathbf{Z}_\alpha)\mathbf{C}_\alpha=\mathbf{D}_\alpha{}^\dagger+\mathbf{G}_1{}^\dagger\mathbf{Z}_\alpha{}^\dagger.\end{aligned} \tag{77}$$

By the argument used to derive (66–67), modified in accordance with the form of (38), wherein the arguments in G_r are shifted through r, we find that

$$\begin{aligned}F_n(\mathbf{c})&=\Psi^{(0)\prime}\mathbf{R}_n{}^\dagger\Psi^{(0)}=\Psi^{(0)\prime}(\mathbf{D}_\alpha{}^\dagger+\mathbf{G}_n{}^\dagger\mathbf{Z}_\alpha{}^\dagger)\\&(\mathbf{D}_\alpha{}^\dagger+\mathbf{G}_{n-1}{}^\dagger\mathbf{Z}_\alpha{}^\dagger)\cdots(\mathbf{D}_\alpha{}^\dagger+\mathbf{G}_1{}^\dagger\mathbf{Z}_\alpha{}^\dagger)\Psi^{(0)}.\end{aligned} \tag{78}$$

(Note that $\mathbf{D}_\alpha{}^\dagger\Psi^{(0)}=0$.)

In formula (78) there are no extraneous dimensions!

If edge effects are absent or can be neglected, then $\mathbf{G}_1{}^\dagger, \mathbf{G}_2{}^\dagger, \cdots\mathbf{G}_n{}^\dagger$ can be replaced by a single matrix $\mathbf{G}^\dagger(\mathbf{X}_{i_1},\cdots\mathbf{X}_{i_\alpha},\mathbf{c})$, and so

$$F_n(\mathbf{c})=\Psi^{(0)\prime}(\mathbf{D}_\alpha{}^\dagger+\mathbf{G}^\dagger\mathbf{Z}_\alpha{}^\dagger)^n\Psi^{(0)}. \tag{79}$$

Hence, the problem reduces to that of diagonalizing $\mathbf{D}_\alpha{}^\dagger+\mathbf{G}^\dagger\mathbf{Z}_\alpha{}^\dagger$.

The matrix analogs of the fixed cyclic shift formulas (39)–(41) are less useful than (78), but in several examples are adaptable to recursive diagonalization. To save notational difficulties, we write down the matrix results for the shift through $t=1$ only. Let $\mathbf{G}_r{}^{(1)}$ be the result of replacing x_j by $\mathbf{X}_j$, $j=1\cdots r+\mu_r-1$, in $G_r(x_r,x_{r+1},\cdots x_{r+\mu_r-1},\mathbf{c})$, and recursively define

$$\begin{aligned}\mathbf{R}_r{}^{(1)}&=(\mathbf{D}_r+\mathbf{G}_r{}^{(1)}\mathbf{Z}_r)\mathbf{R}_{r-1}{}^{(1)}, \quad r=2\cdots n,\\ \mathbf{R}_1{}^{(1)}&=(\mathbf{D}_1+\mathbf{G}_1{}^{(1)}\mathbf{Z}_1).\end{aligned} \tag{80}$$

Then $F_n(\mathbf{c})=\Psi^{(0)\prime}\mathbf{R}_n{}^{(1)}\Psi^{(0)}$.

In general, we see that diagonalization of the product

$$\mathbf{R}_n{}^\dagger=(\mathbf{D}_\alpha{}^\dagger+\mathbf{G}_n{}^\dagger\mathbf{Z}_\alpha{}^\dagger)\cdots(\mathbf{D}_\alpha{}^\dagger+\mathbf{G}_1{}^\dagger\mathbf{Z}_\alpha{}^\dagger)$$

or of

$$\mathbf{R}_n=(\mathbf{D}_n+\mathbf{G}_n\mathbf{Z}_n)\cdots(\mathbf{D}_1+\mathbf{G}_1\mathbf{Z}_1)$$

or

$$\mathbf{R}_n{}^{(1)}=(\mathbf{D}_n+\mathbf{G}_n{}^{(1)}\mathbf{Z}_n)\cdots(\mathbf{D}_1+\mathbf{G}_1{}^{(1)}\mathbf{Z}_1)$$

is sufficient to solve the problem.

7. APPLICATIONS

In this section we briefly discuss a variety of applications of the foregoing method.

(a) Linear Chain with Next-Neighbor Interactions

Here $c_0\neq 0$, $c_1\neq 0$, $c_{12}\neq 0$, $c_{123}\neq 0$, $\cdots$. Let $c_{(0)}=c_0$, $c_{(r)}=c_{12\ldots r}$, so that, neglecting end effects,

$$\mathbf{G}=c_{(0)}+c_{(1)}\mathbf{X}_{(1)}+c_{(2)}\frac{\mathbf{X}_{(2)}{}^2}{2!}+\cdots+c_{(n)}\frac{\mathbf{X}_{(n)}{}^n}{n!},$$

where $\mathbf{X}_{(n)}=\mathbf{X}_1+\cdots+\mathbf{X}_m$. Inductively, we can show that $\mathbf{R}_r$ depends on $\mathbf{X}_{(1)}$, $\mathbf{X}_{(2)}{}^2$, $\cdots\mathbf{X}_{(n)}{}^n$ for each r.

Since

$$\frac{\partial}{\partial x_1}R_r(x_1,x_2,\cdots x_n)=\sum_{r=1}^{n}\frac{\partial}{\partial x_{(r)}}R_r(x_{(1)},\cdots,x_{(r)}),$$

where $x_{(m)}=x_1+\cdots+x_m$, and

$$R_r(x_{(1)},\cdots x_{(r)})\,|_{x_1=0}=R_r(0\cdots 0),$$

we can write $\mathbf{X}_{(1)}=\cdots=\mathbf{X}_{(n)}=\mathbf{X}$ and the fundamental recurrence becomes, in scalar form,

$$R_{r+1}(x)=\frac{\partial}{\partial x}R_r(x)+G(x)R_r(0),\quad R_1(x)=G(x),$$

where

$$G(x)=c_0+c_{(1)}x+\cdots+\frac{c_{(n)}x^n}{n!}.$$

Asymptotically, we assume $R_{r+1}(x)=\lambda R_r(x)$ and obtain

$$R_r(x)=R_r(0)e^{\lambda x}\left(1-\int_0^\infty e^{-\lambda x}G(x)dx+\int_x^\infty e^{-\lambda x}G(x)dx\right).$$

Since $R_r(x)$ is, like $G(x)$, a polynomial in x, we must have

$$1-\int_0^\infty e^{-\lambda x}G(x)dx=0.$$

But

$$\int_0^\infty e^{-\lambda x}G(x)dx=\sum_{m=0}^{n}\frac{c_{(m)}}{m!}\int_0^\infty x^m e^{-\lambda x}dx=\sum_{m=0}^{n}c_{(m)}\lambda^{-m-1}.$$

Hence λ is the (largest) root of $c_{(n)}+c_{(n-1)}\lambda+\cdots+c_{(0)}\lambda^n=\lambda^{n+1}$, a result already obtained (e.g., de Boer[12]) by other methods.

(b) Ising Problem

Here there are two configurations per cell, with next-neighbor interactions. $B(C_i,C_j)=1$ unless $j=i+t(k)$, where $t(k)$ is a translation between neighboring particles. Furthermore, if $j=i+t(k)$, the sum

$$\sum_{C_j}Q(C_j)\sum_{C_i}[B(C_i,C_j)-1]$$

vanishes on suitable choice of the interaction energies. Hence, $c_{i_1\cdots i_m}$ vanishes unless $(i_1\cdots i_m)$ forms a closed polygon with sides of nearest-neighbor distance. For if vertex i_j is connected only to vertex i_k, the Ursell function $U(C_{i_1}\cdots C_{i_m})$ is of the form

$$U(C_{i_1},C_{i_{j-1}},C_{i_{j+1}},\cdots C_{i_{k-1}},C_{i_{k+1}},\cdots C_{i_m})\cdot U(C_{i_j},C_{i_k}),$$

and thus

$$\sum_{C_{i_k}}Q(C_{i_k})U(C_{i_1}\cdots C_{i_m})=0.$$

The Ising problem then reduces to that of counting the closed polygons on the lattice of interest. Kac and Ward have solved the two-dimensional Ising problem from this viewpoint. Their determinantal equation can be obtained directly from the matrix formulation in Sec. 6, but as the reduction occupies some space we are obliged to postpone the demonstration of this.

(c) Problems Discussed in Sec. 5

The most general problem there considered contains two essential dimensions. Now

$$\mathbf{X}_1=\begin{bmatrix}0&0\\1&0\end{bmatrix}\times\begin{bmatrix}1&0\\0&1\end{bmatrix}=\begin{bmatrix}0&0&0&0\\1&0&0&0\\0&0&0&0\\0&0&1&0\end{bmatrix},$$

$$\mathbf{X}_2=\begin{bmatrix}1&0\\0&1\end{bmatrix}\times\begin{bmatrix}0&0\\1&0\end{bmatrix}\quad\begin{bmatrix}0&0&0&0\\0&0&0&0\\1&0&0&0\\0&1&0&0\end{bmatrix},$$

so that for an interior location,

$$\mathbf{G}_r{}^\dagger=c_0\mathbf{I}+c_1\mathbf{X}_1+c_q\mathbf{X}_2=\begin{bmatrix}c_0&0&0&0\\c_1&c_0&0&0\\c_q&0&c_0&0\\0&c_q&c_1&c_0\end{bmatrix},$$

$$\mathbf{Z}_\alpha{}^\dagger=\begin{bmatrix}1&0&0&0\\0&0&1&0\\0&0&0&0\\0&0&0&0\end{bmatrix},\quad\mathbf{D}_\alpha{}^\dagger=\begin{bmatrix}0&1&0&0\\0&0&0&1\\0&0&0&0\\0&0&0&0\end{bmatrix},$$

and thus

$$\mathbf{D}_\alpha{}^\dagger+\mathbf{G}_r{}^\dagger\mathbf{Z}_\alpha{}^\dagger=\begin{bmatrix}c_0&1&0&0\\c_1&0&c_0&1\\c_q&0&0&0\\0&0&c_q&0\end{bmatrix},$$

which is the matrix found in Sec. 5 for interior locations. The modifications for edge effects are in Sec. 5.

(d) Nearest Neighbors in Three Dimensions

Here we have

$$\mathbf{X}_1=\begin{bmatrix}0&0\\1&0\end{bmatrix}\times\begin{bmatrix}1&0\\0&1\end{bmatrix}\times\begin{bmatrix}1&0\\0&1\end{bmatrix}$$

$$\mathbf{X}_2=\begin{bmatrix}1&0\\0&1\end{bmatrix}\times\begin{bmatrix}0&0\\1&0\end{bmatrix}\times\begin{bmatrix}1&0\\0&1\end{bmatrix}$$

$$\mathbf{X}_3=\begin{bmatrix}1&0\\0&1\end{bmatrix}\times\begin{bmatrix}1&0\\0&1\end{bmatrix}\times\begin{bmatrix}0&0\\1&0\end{bmatrix},$$

so for an interior location

$$\mathbf{G}_r{}^\dagger=c_0\mathbf{I}+c_1\mathbf{X}_1+c_q\mathbf{X}_2+c_s\mathbf{X}_3$$

$$=\begin{bmatrix}c_0&&&&&&&\\c_1&c_0&&&&&&\\c_q&0&c_0&&&&&\\0&c_q&c_1&c_0&&&&\\c_s&0&0&0&c_0&&&\\0&c_s&0&0&c_1&c_0&&\\0&0&c_s&0&c_q&0&c_0&\\0&0&0&c_s&0&c_q&c_1&c_0\end{bmatrix}.$$

Also

$$\mathbf{Z}_\alpha{}^\dagger = \left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & & & & \\ 0 & 0 & 1 & 0 & & & & \\ \hline & & & & 1 & 0 & 0 & 0 \\ & & & & 0 & 0 & 1 & 0 \\ \hline & & & & & & & \end{array}\right],$$

$$\mathbf{D}_\alpha{}^\dagger = \left[\begin{array}{cccc|cccc} 0 & 1 & 0 & 0 & & & & \\ 0 & 0 & 0 & 1 & & & & \\ \hline & & & & 0 & 1 & 0 & 0 \\ & & & & 0 & 0 & 0 & 1 \\ \hline & & & & & & & \end{array}\right],$$

so that

$$\mathbf{D}_\alpha{}^\dagger + \mathbf{G}_r{}^\dagger \mathbf{Z}_\alpha{}^\dagger = \left[\begin{array}{cccc|cccc} c_0 & 1 & 0 & 0 & & & & \\ c_1 & 0 & c_0 & 1 & & & & \\ \hline c_q & 0 & 0 & 0 & c_0 & 1 & 0 & 0 \\ 0 & 0 & c_q & 0 & c_1 & 0 & c_0 & 1 \\ \hline c_s & 0 & 0 & 0 & & & & \\ 0 & 0 & c_s & 0 & & & & \\ \hline & & & & c_s & 0 & 0 & 0 \\ & & & & 0 & 0 & c_s & 0 \end{array}\right].$$

Modifications for side, edge, and corner effects can be determined in the manner of Secs. 5 and 6. The problem of obtaining the largest eigenvalue of matrices with this type of structure is under study.

CONCLUSION

We have shown that (1) a very general type of problem in statistical mechanics can be reduced to an associative combinatorial problem, and (2) this problem can again be reduced to the determination of the set of eigenvalues of a certain matrix.

A variety of problems already known to be soluble are easily handled by the automatic application of this technique, and at least one problem is removed from the unsolved category. We believe that the systematic exploitation of this method will provide many interesting new results.

The authors wish to thank the Test and Research Departments, Michelson Laboratory, U. S. Naval Ordnance Test Station, China Lake, California for cooperation and support.

REVIEWS OF MODERN PHYSICS VOLUME 32, NUMBER 1 JANUARY, 1960

Linearized Theory of Plasma Oscillations*

L. OSTER
Laboratory of Marine Physics, Yale University, New Haven, Connecticut

I. Hydrodynamic Treatment

INTRODUCTION

BASICALLY there are two ways of dealing with plasma problems: a microscopic gas-kinetic treatment using the Boltzmann equation together with Maxwell's equations of electrodynamics; or a macroscopic, hydrodynamic treatment using Euler's equation of motion together with the Maxwell equations. For various mathematical reasons it seems to be impossible to investigate the different general modes of oscillation using gas-kinetic methods without serious physical restrictions. To avoid excessively complicated mathematics in using the kinetic approach, it is necessary to make assumptions of such kind that it is more reasonable to use the hydrodynamic equations. For this reason we deal here only with the hydrodynamic equations together with Maxwell's equations. Questions concerning the range of validity of our treatment are deferred to Sec. II. The hydrodynamic treatment is always justified when there is a *stationary* distribution of velocities in the plasma which is not disturbed "essentially" by the collective oscillations.

Although one succeeds in this way in simplifying the procedure a great deal, the treatment of the unabridged hydrodynamic equations [except in a few cases such as the work of R. W. Larenz (1955)†] is further simplified. For this purpose, one supposes the plasma to be uniform and of infinite extent, and the oscillations to be small sinusoidal perturbations. These concepts are not very close to reality, and are unable to explain complicated processes such as the origin of cosmic radio-frequency radiation. Still, the linearized theory succeeds well in explaining the ionospheric observations, so one may hope that at least some idea is obtained of how and where to begin a later nonlinear approach.

All investigations of the linearized theory, until now, dealt with special cases—Langmuir oscillations, ionospheric theory, Alfvén's magnetohydrodynamics, etc.—

* Supported by the Office of Naval Research.

† References are given in alphabetical order in Bibliography.