

Chemical Combinatorics, Cascade Theory, and Polymer Science.

by

Manfred Gordon
University of Essex, Colchester, England

1. POLYMER DISTRIBUTIONS SIMULATED BY CASCADE PROCESSES ("STATISTICAL FORESTS").

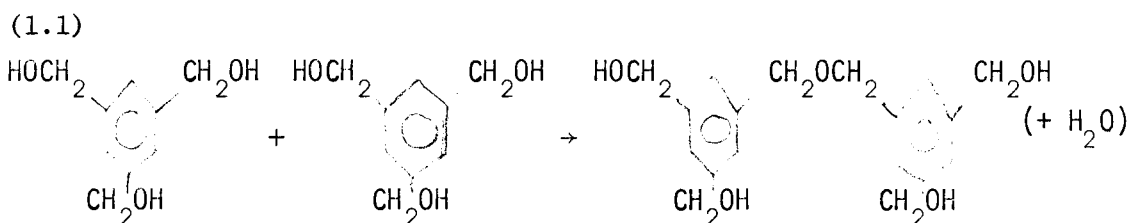
The representation of chemical molecules by graphs, in which the vertices represent atoms and the edges represent chemical bonds is well known (Harary 1967). Equilibrium distributions of chemical species may be represented as sets or 'ensembles' of graphs. The case of polymer distributions gives rise to important combinatorial problems, and the number of distinct graphs is infinite in such cases.

For concreteness, consider Figs. 1 and 2 which represent the simplest example viz., *random f-functional polycondensation*. In the extreme position of the equilibrium shown in Fig. 1, the ensemble or molecular system consists entirely of the *monomer* graph, which has one vertex of degree f (here $f = 3$) and f vertices of degree 1. The vertex of degree f is called a monomer unit, or simply a *node*. The vertices of degree 1 are unfortunately called functionalities by chemists. At position A in Fig. 1, we see the incipient process of chemical reaction, viz., a fusion of a pair of edges to produce a *dimer*. In Fig. 2, this process is well under way and has produced x -mer molecules of $x = 2, 3$, etc. For fixed f and x , two distinct graphs are called isomers (Harary 1967). (We

Lecture notes from a series of lectures delivered at the Summer Study Institute on Combinatorial Mathematics and Its Applications in the Natural Sciences, held on June 2-27, 1969, in Chapel Hill, and sponsored by the U. S. Air Force Office of Scientific Research under Air Force Grant No. AFOSR-68-1406 and the U. S. Army Research Office (Durham) under Grant No. DA-ARO-D-31-124-G910.

shall refer to the j -th x -mer as one of the possible isomer graphs in a suitable enumeration $j = 1, 2, \dots, J$ say.) Thus an x -mer has x nodes.

Edges incident on two nodes are termed *links*; those incident on a node and a functionality may be called *free edges*. Trees containing only vertices of degree f and unity are called $(1,f)$ -trees. The fraction α of the functionalities in Fig. 1 which have disappeared by fusion of free edges in Fig. 2 is called the degree of advancement of the polycondensation reaction by chemists. (In Fig. 1, $\alpha = 0$; in general $0 \leq \alpha \leq 1$). A chemical example of the process of random 3-functional polycondensation has as its first step:



Graphs with cycles (in addition to trees) occur in real polycondensation systems (see eq. 5.14) but are discounted in the simple model under discussion.

The power of the theory of branching ("cascade") processes (Harris, 1963) can be harnessed by constructing a *statistical forest* (s.f.) from the *molecular systems* (m.s.) as in Fig. 2. Between the m.s. and s.f. there subsists a $1 \rightarrow X$ correspondence of x -mer trees (*not* a one-one correspondence of trees), because the s.f. is constructed by choosing a node* at random in the m.s. to serve as a root. Thus, each of the X nodes of an x -mer tree has an equal chance of furnishing the root of a

* See footnote, p.18. Meanwhile equations happen to be simpler if roots are restricted to nodes.

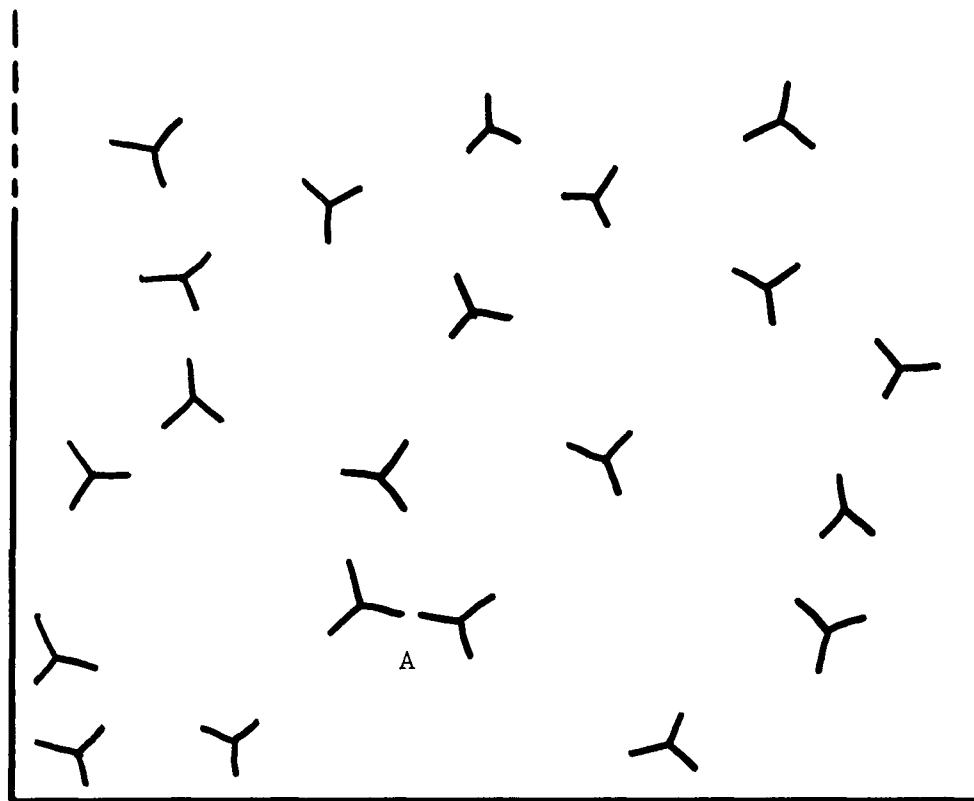


Figure 1

Part of infinite molecular system consisting entirely of 3-functional monomer graphs. At A, molecular collision favourable to production of dimer by fusion of two edges, see A, Fig. 2.

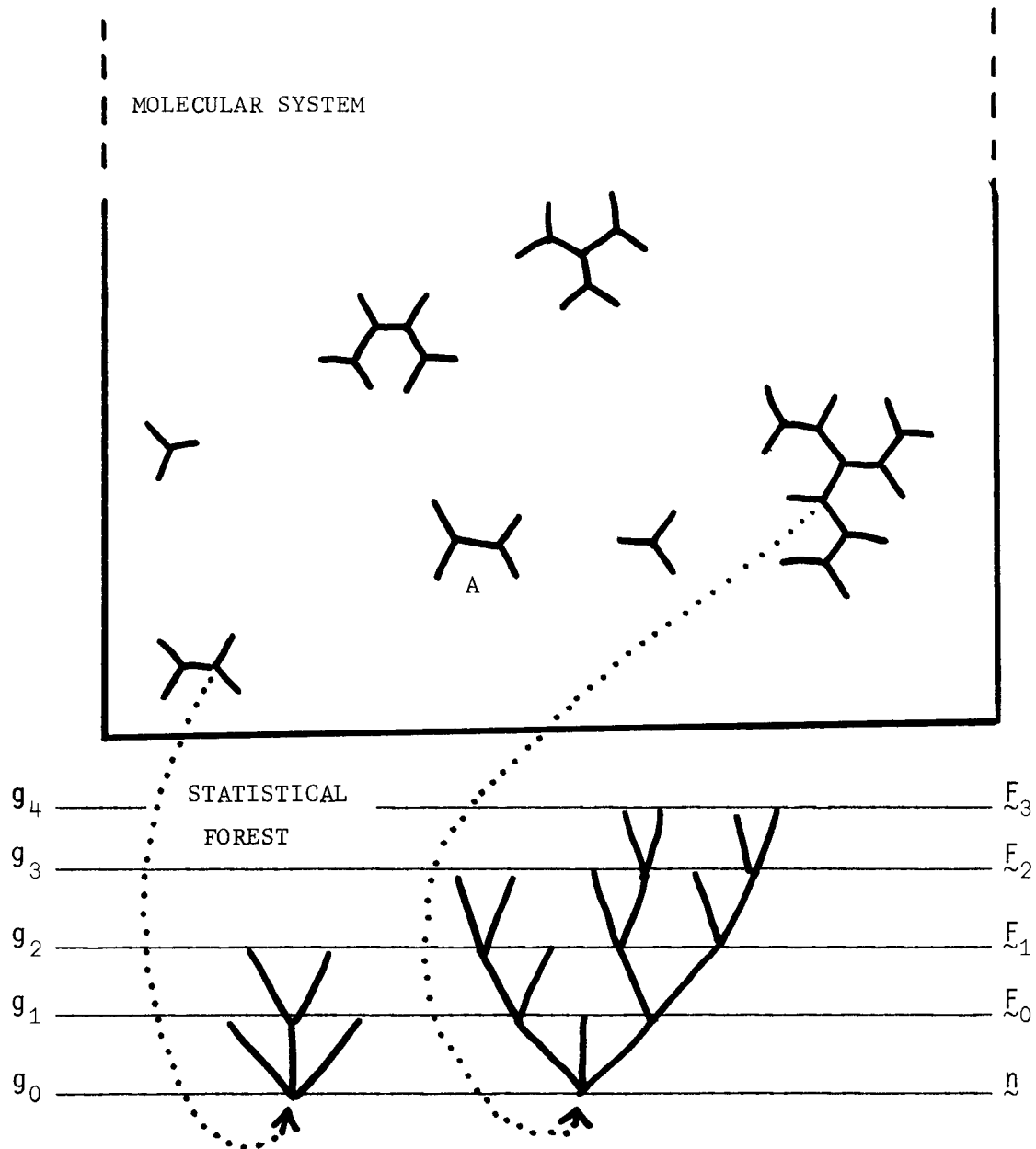


Figure 2

TOP: Molecular system of 3-functional polycondensation formed from Fig. 1 by raising the degree α of advancement of the reaction. A is a dimer; at top, two isometric tetramers occur. BOTTOM: Construction of the statistical forest of rooted trees is by random choice of nodes in the molecular system to serve as roots.

tree in the s.f. A statistical forest is defined by the pgf's $F_0(\theta), F_1(\theta), \dots$ for the number of offsprings of members of generations g_0, g_1, \dots . More generally, when nodes of distinct types may occur, an s.f. is defined by the vectorial quantities

$$(1.2) \quad \underline{n} = (n^1, n^2, \dots, n^S)$$

$$(1.3) \quad \underline{F}_n(\underline{\varrho}) = (F_{01}(\underline{\varrho}), F_{02}(\underline{\varrho}), \dots, F_{0S}(\underline{\varrho})) \quad n = 0, 1, \dots$$

$$(1.4) \quad \underline{\varrho} = (\varrho_1, \varrho_2, \dots, \varrho_S)$$

The n^i are the chances of a root belonging to type i (with $1 \leq i \leq S$). The coefficient of $\underline{\varrho}^{\underline{m}} \equiv \varrho_1^{m_1} \varrho_2^{m_2}, \dots, \varrho_S^{m_S}$ in $F_{jk}(\underline{\varrho})$ is the chance that a node of type k on g_j bears on g_{j+1} exactly m_1 offsprings (nodes) of type 1, m_2 of type 2, \dots , m_S of type S , (g_j means: generation j).

A master theorem (eq. 4.3-7) lays down necessary and sufficient conditions for an s.f. so defined to represent a *clone*, i.e., a forest generated by random choice of a root vertex within any fixed m.s. This theorem shows that the vector \underline{n} can be calculated from \underline{F}_0 , and all the \underline{F}_j ($j = 1, 2, \dots$) can be calculated (recursively) from \underline{F}_0 . This makes \underline{F}_0 , the so-called *link probability generating function* (lpgf), the central combinatorial tool of the whole subject.

1.1 CONNECTION BETWEEN RELEVANT ALGEBRA AND GRAPH-THEORY.

By repeated use of the cascade substitution (for the case $s = 1$), one obtains the weight fraction pgf, i.e., pgf of the chance of an x -mer tree in the s.f.:

$$(1.5) \quad W(\theta) = \sum_x \theta^x = \theta F_0(\theta F_1(\theta F_2(\theta F_3(\dots)))) .$$

Moreover, the master theorem shows that for $s=1$ types of nodes, $F_1 = F_2 = \dots$. For random f -functional polycondensation, it is readily apparent that

$$(1.6) \quad F_0 = (1 - \alpha + \alpha\theta)^f$$

$$(1.7) \quad F_j = (1 - \alpha + \alpha\theta)^{f-1} \quad j = 1, 2, \dots$$

The typographical trick of writing a product of n factors $a_1 a_2 \dots a_n$ so as to emphasize a_1 :

$$(1.8) \quad a_1 a_2 \dots a_n \equiv a_1 \begin{cases} a_2 \\ a_3 \\ \vdots \\ a_n \end{cases}$$

allows us to display the connection of the combinatorial formula (1.5) with graph theory. Substituting (1.6) and (1.7) in (1.5), for example, for the simple case $f=3$, we obtain Fig. 3.

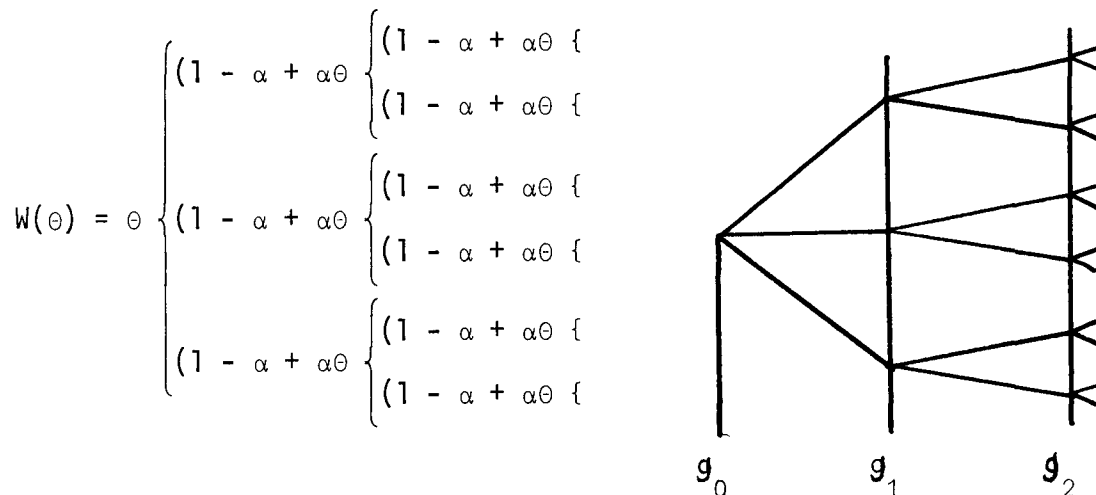


Figure 3

The relation between $W(\theta)$ and a lattice.

The θ -symbols now clearly stand in one-one correspondence to the nodes in the infinite lattice upon which the rooted trees of the s.f. can be arranged (Fig. 3).

Any single term in θ^x ($x = 1, 2, \dots$) in the expansion of $W(\theta)$ can only correspond to a singly-connected tree in Fig. 3 and, moreover, this tree must be rooted, in that one of the x θ -factors of such a term must be the one lying on g_0 . Moreover, every distinct ordered rooted x -mer tree on the lattice of Fig. 3 contributes an equal term $\alpha^{x-1}(1-\alpha)^{fx-2x+2}$ to the probability w_x of an x -mer tree in the s.f. This term corresponds to the probability α^{x-1} of finding the $x-1$ links times the independent probability $(1-\alpha)^{fx-2x+2}$ of finding the $fx-2x+2$ free edges which occur in any x -mer tree. The number of such equal terms, moreover, is the number of distinct node-rooted ordered $(1,f)$ -trees, i.e., the number $T_{mx}(f)$ of distinct node-rooted trees of x nodes which can be drawn on a lattice such as Fig. 3. It is shown by Lagrange expansion (Good 1963, Gordon and Scantlebury 1964, Butler, Gordon and Malcolm 1966) of (1.5), that

$$(1.9) \quad w_x = \frac{(fx-x)!f}{(x-1)!(fx-2x+2)!} \alpha^{x-1}(1-\alpha)^{fx-2x+2},$$

a classical result (Flory 1955) obtained by various ad hoc methods. It follows that

$$(1.10) \quad T_{mx}(f) = (fx-x)!f/(x-1)!(fx-2x+2)!$$

a result which can be established directly (cf., Good 1960). Examining the tree-distribution in finer detail than merely in terms of the number x of nodes in a tree, one shows that

$$(1.11) \quad w_{xj} = T_{mxj}(f) (1-\alpha)^{fx-2x+2} \alpha^{x-1} ,$$

where w_{xj} is the chance of a j -th x -mer $(1-f)$ -tree in the s.f., and $T_{mxj}(f)$ the number of distinct ordered rooted trees generated by the j -th x -mer $(1,f)$ -tree.

Eq. (1.10) and (1.11) can be written down (see Chapter 3) directly from statistical mechanics. This is because it has been established that any acyclic chemical structure of $M_{\nu n}$ of ν atoms and n links produces in its molecular partition function the combinatorial invariant T_M , i.e., the number of distinct ordered rooted trees generated by $M_{\nu n}$, which invariant corresponds to the whole of its so-called combinatorial entropy. The discovery of this factor in the special context of distributions of polymers (Gordon and Scantlebury 1964) is merely a historical accident. We return to the general chemical situation in Chapter 3

2. CRITICALITY. GELATION. SOL FRACTION. ELASTICITY.

Chemists refer to $E(x)$ as the weight-average degree of polymerization DP_w ; it is found from 1.5 as $W'(1)$ in the following steps:

$$(2.1) \quad W(\theta) = \theta F_0(u(\theta))$$

say, so that formally

$$(2.2) \quad u(\theta) = \theta F_1(\theta F_1(\dots)) .$$

Again, $u(\theta)$ can be found explicitly by Lagrange expansion, but it is more elegant to eliminate u by the formal assumption

$$(2.3) \quad u(\theta) = \theta F_1(u(\theta)) .$$

Differentiating (2.1) and (2.2) with respect to θ and eliminating u and u' , yields

$$(2.4) \quad DP_W = W'(1) = 1 + \frac{F_0'(1)}{1 - F_0'(1)} = E(x) .$$

Clearly, $u(\theta)$ is the pgf of the distribution of size x among all the proper subtrees within the trees of the s.f. If $W(\theta)$ itself exists, we may be confident that $u(\theta)$ exists and that (2.3) is justified. The function $W(\theta)$ has, in general, at least one classical singularity, known as the critical condition for survival of a family name in genetics (Galton 1889), or for a neutron cascade in a nuclear device, or for the so-called gel-point in polymer science. At this point, DP_W diverges because

$$(2.5) \quad F_1'(1) = 1 .$$

The generalization (Gordon 1962) of this condition for the singularity to the case of S different types of monomer units (trees with S differently labelled nodes) consists of the vanishing of the determinant

$$(2.6) \quad \left| \delta_i^j - F_{1i}^j \right| = 0 ,$$

where the general notation

$$(2.7) \quad F_{ni}^j \equiv (\partial/\partial\theta_j)F_{ni}$$

is employed.

The meaning of 2.4 is immediately apparent through the formal expansion

(2.8)

$$\begin{aligned}
 E(x) &= 1 + F'_0(1) + F'_0(1)F'_0(1) + F'_0(1)|F'_1(1)|^2 + \dots \\
 &\quad \downarrow \quad \downarrow \quad \downarrow \quad \downarrow \\
 &= E(g_0) + E(g_1) + E(g_2) + E(g_3) + \dots,
 \end{aligned}$$

where $E(g_i)$ denotes the expected number of nodes on generation g_i among the trees of the s.f.

If we substitute the pgf's of the simple 'random' case, eq. (1.6) and (1.7), in (2.4), we obtain immediately the classical formula (Flory 1953):

$$(2.9) \quad DP_w = (1 + \alpha)/(1 - \alpha(f-1)) = E(x)$$

and the critical value α_c of α at the gel point:

$$(2.10) \quad \alpha_c = 1/(f-1) .$$

For $\alpha > \alpha_c$, the s.f. splits formally into a set of trees of which a finite ('sol') fraction S is finite, while the remainder ('gel' fraction) is infinite, in the sense that there is no bound to the number n of generations g_n to which the nodes belong (Fig. 4). Experimentally (see, e.g., Charlesby 1955), the sol is found soluble in suitable solvent, while the single gel-molecule in the m.s. corresponding to the s.f. is found insoluble in all solvents - it is a molecule of microscopic dimensions.

A tree or subtree is finite if all its branches are finite. This leads to the recurrence relation for the 'extinction' probability v of a link (edge) which leads from a node on $g_n = 1, 2, \dots$ to g_{n+2} , viz.:

$$(2.11) \quad v = F_1(v) .$$

When solved explicitly for v , this equation provides the key to all interesting parameters in molecular systems consisting of sol and gel. In particular, the lpgf of the finite trees in the s.f. can be written

$$(2.12) \quad \hat{F}_0(\theta) = F_0(v\theta)/F_0(v).$$

This provides the lpgf for the sol fraction on its own, from which all interesting statistical parameters of that fraction can be calculated. The fraction of finite trees in the s.f. (i.e., the 'weight' fraction of sol in the m.s.) is given by

$$(2.13) \quad S = F_0(v).$$

The following note-worthy result follows from (2.12) for the case of random f -functional polycondensation (eq. (1.6) and (1.7)): the distribution $\hat{W}_x(\alpha)$ for the sol fraction alone passes, as the degree α of advancement of the chemical reaction increases, through the same family of functions (though in reverse order) as the whole system before the gel-point. The situation is graphically summed up by Fig. 5. The successive states of the system are characterised by the progress of single peak of a wave which spreads to an ever broader and lower form. At the gel point however, the wave 'splashes' over, progressively losing its substance to a delta-function (the gel fraction) rising up at $x = \infty$, while the remainder returns as the reflected sol-fraction wave. This reflected wave *sharpens* as it dies out (as $\alpha \rightarrow 1$), finally behaving as a delta function of vanishing magnitude at the origin: the last few tree-like molecules to join up with the infinite gel-tree are likely to be monomer trees.

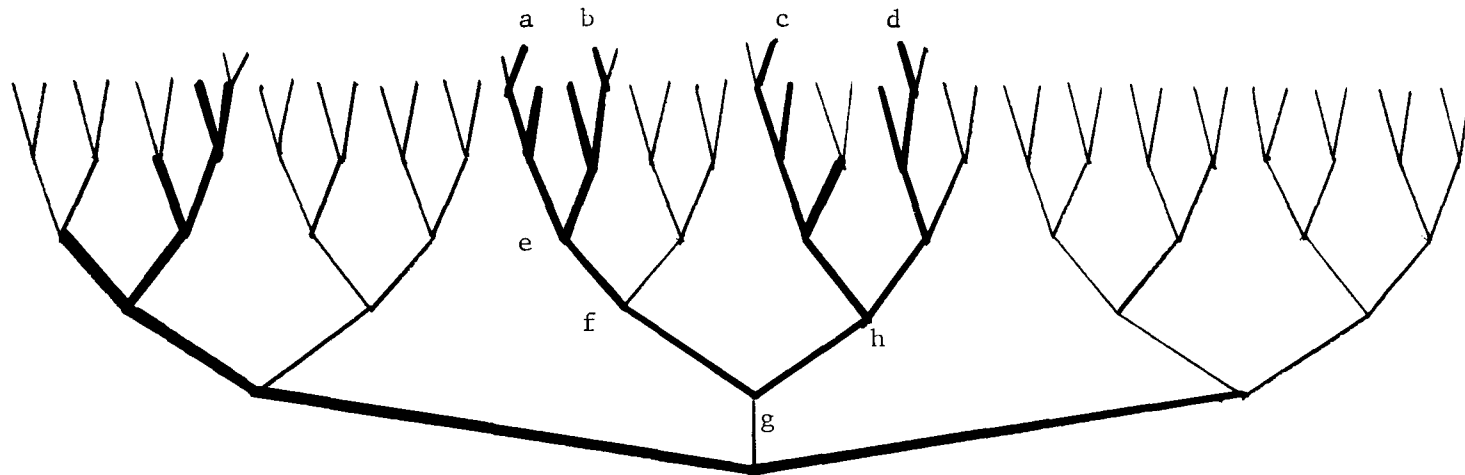


Figure 4

Gel molecule (heavy lines) as rooted tree, drawn on appropriate lattice (light lines). If branches a, b, c, and d can be continued to infinity, then efgh is the intersection (common part) of at least two infinite lines, say one through a and c, and one through b and d, while no proper part of efgh is the intersection of two infinite lines.

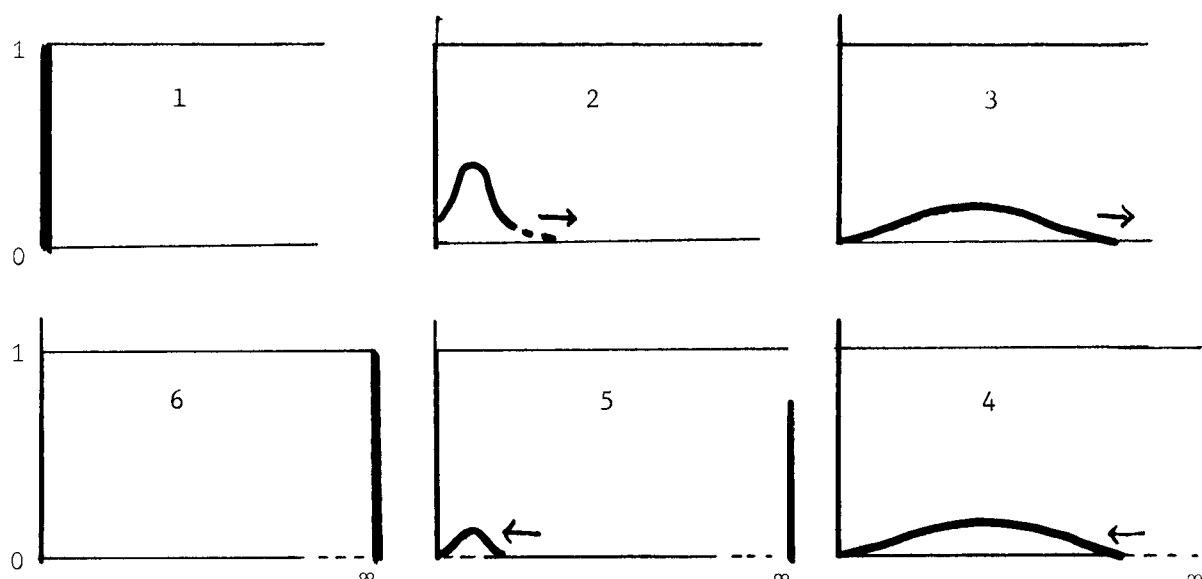
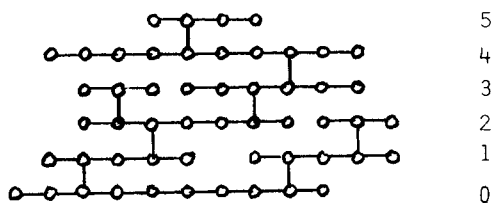


Figure 5

Schematic view of the return of the wave. The molecular size distribution W_X (ordinate) as a function of size X for increasing values of the parameter α . Case 1: $\alpha=0$; cases 1, 2, 3: $\alpha < \alpha_c$ (pre-gelation); cases 4, 5, 6: $\alpha > \alpha_c$ (post-gelation); case 6: $\alpha=1$. The gel fraction (cases 4, 5, 6) is a delta-function placed at infinity. The sol fraction distribution curves in 4 and 5 are scaled-down versions of W_X in 3 and 2 respectively.

Figure 6

Units providing the root and the nodes of the trees are here the primary chains, and *not* their constituent repeat units.



2.1. STATISTICS OF THE GEL AND RUBBER ELASTICITY.

Statistical parameters of the gel fraction (i.e., the infinite molecule) play important roles in theories of elasticity and swelling, etc. As α increases, formation of cycles in the gel molecule becomes important, but current theories neglect their effect and replace the ring-closing links by hypothetical links between pairs of tree-like gel molecules. Thus the gel 'network' is always pictured to consist of a *tree-like* graph with an infinite number of nodes. The elastic moduli are, according to statistical-mechanical theories (Dobson and Gordon 1965; Imai and Gordon 1969), proportional to the number N_e of distinct active network chains per monomer repeat unit. These monomer repeats may be small units like those discussed so far, or preexisting polymer chains which are being linked together in a vulcanization reaction (Fig. 6). In either case, the gel molecule can be represented as a rooted tree. In such a rooted tree, with an infinity of nodes, a set of infinite lines may be drawn each of which extends to infinity in both directions from some node of minimal generation number characteristic for a given line (Fig. 5). An active network chain (Dobson and Gordon 1965) is defined as a line segment which is an intersection of two such infinite lines, which must be minimal in the sense that no proper part of it is an intersection of two infinite lines. The number N_e of distinct active network chains per monomer repeat unit in the gel is readily calculated for most cases of interest in terms of the extinction probability v defined above. For instance, for random f -functional polycondensation,

$$(2.14) \quad N_e = f\alpha(1-v)^2(1-\beta)/2\alpha\gamma,$$

where

$$(2.15) \quad \beta \equiv (f-1)\alpha v / (1 - \alpha + \alpha v).$$

Experimental measurement of Young's modulus Y as a function of α generally reproduce quite well the results predicted from statistical theories of active network chains, especially qualitative features such as

$$(2.16) \quad \lim_{\alpha \rightarrow \alpha_c} dY/d\alpha = 0$$

which follows from (2.14), (2.15) and (2.11). For a recent review of rubber network theory, see Dusek and Prins (1969).

3. GRAPH THEORY AND COMBINATORIAL INVARIANTS OF MOLECULAR PARTITION FUNCTIONS.

A molecular graph or M-graph M_{vn} is singly connected, free of loops*, and has v vertices labelled with ('atomic') labels and n edges. Such an M-graph can be constructed in $n!$ 'synthetic routes' (not necessarily all distinct). Each route specifies a definite order in which the n edges are added one by one to the initial set of constituent atoms, i.e., the v trivial subgraphs of order 1. The k -th step ($1 \leq k \leq n$) in the synthesis may leave the connectivity unchanged, i.e., it may add an edge incident on two vertices belonging to a singly

* No edge is incident on the same vertex at both of its ends.

connected molecular sub-graph Y_k to produce another, X_k , with an additional edge (an 'intramolecular' step). Or it may link two unconnected molecular sub-graphs Y_k and Z_k into a singly connected X_k (an 'intermolecular' step).

Let the edge added in step k be incident in vertices v_1 and v_2 in X_k . Before the k -th edge is added, v_1 lies in Y_k and v_2 lies in Y_k (k -th step intramolecular) or in Z_k (k -th step intermolecular). Let $s(v_i; Y_k)$ denote the order of the equivalence class of vertex v_i under the automorphism group of Y_k , etc. Let $p(v_i, v_j; Y_k)$ denote the order of the equivalence class of the *ordered* pair (v_i, v_j) or vertices under the automorphism group of Y_k , etc.

The chemist's 'statistical equilibrium factor' of the k -th step is defined by

$$(3.1) \quad \kappa_k = p(v_1, v_2; Y_k)/p(v_1, v_2; X_k) \quad (\text{intramolecular})$$

$$(3.2) \quad \kappa_k = s(v_1; Y_k)s(v_2; Z_k)/p(v_1, v_2; X_k) \quad (\text{intermolecular}).$$

By multiplication over k and elementary group theory, the graph-theoretical result

$$(3.3) \quad |G_M|^{-1} = \prod_1^n \kappa_k$$

follows readily for all synthetic routes of all finite M -graphs. This theorem expresses the order of the automorphism group of a graph as a function of orders of equivalence classes occurring in all its sub-graphs. Remember that there are $n!$ different versions of eq. (3.3), in each of which the set of κ_k 's may be different.

The statistical mechanics of real molecules have combinatorial contributions from metrical geometry in addition to the graphtheoretical

result (3.3). G_M is the apex of the following hierarchy of four groups associated with any given molecule: $E_M \leq R_M \leq S_M \leq G_M$, where \leq means "is isomorphic with a proper or improper subgroup of". S_M is the symmetry group (point group) of the real state of the three-dimensional molecule, R_M its rotation group (comprising the operations of S_M devoid of mirror reflection), and E_M is the 'external rotation group' of the molecule as a whole (comprising the operations of R_M devoid of internal bond rotations). G_M , the group of automorphisms of the homology group of the molecular graph, can be explained to chemists as the point group of the molecule in its 'graph-like' (perfectly flexible) state.

Of these four groups, only the reciprocal $|R_M|^{-1}$ of the order of the rotation group features as a proportionality constant in gas or liquid partition functions Z_M , where it is usually introduced in *ad hoc* fashion as a product of 'symmetry numbers' for avoidance of overcounting configurations. It is more enlightening to factor it thus:

$$(3.4) \quad |R_M|^{-1} = g_m |G_M|^{-1} \propto Z_M,$$

where g_m is the index in G_M of a subgroup which is isomorphic with R_M . The factor $|G_M|^{-1}$ in (3.4) is then immediately traceable to eq. (3.3), and the factor g_m is explained below.

For a molecule in an ideal crystal, each atom (vertex) is distinctly labelled by its location, which insures that $g_m = |G_M|$. This simple remark has significance for the third law of thermodynamics.

We recall that chemical kinetics is linked through thermodynamics to statistical mechanics by the equations:

$$(3.5) \quad k_1/k_2 = K = \exp - \Delta F^0/kT = \prod Z_i^{n_i}.$$

Where the k_i are kinetic rate constants, K is an equilibrium constant, ΔF^0 the standard free energy change, and the n_i are stoichiometric coefficients (taken negatively for reactants). The Z_i are partition functions, with the ground states of the *atoms* taken as energy zero. We now apply (3.5) $(n+1)$ times to an $(n+1)$ -step synthesis of a real-state molecule from its atoms, so as to collect the combinatorial factor in its partition function Z as the product of κ_k -terms ($1 \leq k \leq n+1$) which occur as factors in the $n+1$ equilibrium constants. The first n steps are those described before, viz., the creation of the n edges in the molecular graph $M_{\nu n}$. All these n steps concern molecules only in their (possibly hypothetical) graph-like states. The final $(n+1)$ -th step consists in converting $M_{\nu n}$ from its graph-like into its real 3D state by determining bond lengths, angles (and free rotations, if any). This step has $\kappa_{k+1} = g_m$, the number of mappings of the M -graph $M_{\nu n}$ into the real-state molecule. Thus using (3.3) and (3.4), the combinatorial derivation of the factor $|R_M|^{-1}$ in Z is completed. The factorization of $|R_M|^{-1}$ in (3.4) is important. E.g., Z differs by a factor of 3^6 for two octanes: n -octane and 2,2,3,3-tetramethyl butane, rendering the former more stable and more abundant in equilibrium oil cracking mixtures. It may be shown that the factor 3^6 resides entirely in $|G_M|^{-1}$, while $g_m = 2^8$, the same for both octanes. Thus the difference in stability arises entirely from topology (connectivity) and not from 3D geometry, a feature obscured if we do not factor $|R_M|^{-1}$.

It can be shown that for any finite M -tree $M_{\nu n}$

$$(3.6) \quad |G_M|^{-1} = T_M \prod_1^{\nu} f_i / \underline{f}! \sum_1^{\nu} f_j ,$$

where T_M is the number of distinct rooted ordered trees corresponding to $M_{\nu n}$. The f_i ($i = 1, 2, \dots, \nu$) are the degrees of the vertices and $\underline{f}! = f_1! f_2! \dots f_\nu!$. In conjunction with (3.4) and (3.5), this suffices to develop in a few lines eq. (1.9) from perfectly general principles of statistical mechanics. It is possible to extend eq. (3.6) to all finite M-graphs (containing cycles) by a natural definition of ordered rooted *M-graphs* in general. Note that according to (3.4) and (3.6), there is an additivity of contributions of each equivalence class of atoms in an M-graph to the molecular partition function, since each equivalence class furnishes a set of distinct orderings which contributes to T_M when one of its member atoms is chosen as root of the M-tree.

4. COMBINATORICS OF A PAIR OF VERTICES IN A TREE. PATH-WEIGHTING FUNCTIONS. RADIATION SCATTERING.

Much important combinatorial theory arises from comparing the numbers T_{M1} and T_{M2} of distinct ordered rooted trees produced by a given M-tree when two distinct vertices* (here labelled 1 and 2) are chosen as the roots. It is shown that for any finite M-graph

$$(4.1) \quad T_{M1}/T_{M2} = f_1 s_1 / f_2 s_2 ,$$

where f_i and s_i are the degree and the order of the equivalence class of vertex i respectively. A simple summation yields for any equivalence

* Henceforth, in addition to nodes, vertices of degree unity are allowed as roots.

class E

$$(4.2) \quad T_{ME} = T_M f_E^S E / \sum_1^v f_j .$$

Although a simple group-theoretical proof can be given of (4.1) and (4.2), a probabilistic proof, given in Appendix A, is specially instructive. This is based on the master theorem concerning statistical forests (s.f., defined in eq. (1.2)-(1.4), also proved in the appendix). This master theorem lays down the necessary and sufficient condition for an s.f. to be a *clone*. A clone is defined as an s.f. in which the chance of an M-tree being found rooted on a vertex of an equivalence class E is proportional to s_E . (Nothing is stipulated about the relative chances of occurrence of different M-trees, however.) Thus, in a clone, each vertex of an M-tree has an equal chance of furnishing the root. The necessary and sufficient conditions are

$$(4.3) \quad n^j = n_{F_{0i}}^i (j) (1) / F'_{0(j)} (1) \quad (j = 1, 2, \dots, s)$$

$$(4.4) \quad F_{1j}(\varrho) = N_1 F'_{0j}(\varrho)$$

$$(4.5) \quad F_{2j}(\varrho) = N_2 F_{0j}^i(\varrho) F'_{1i}(1)$$

$$(4.6) \quad F_{nj}(\varrho) = N_n F_{0j}^i(\varrho) [F_{1i}(E_2(E_3 \dots (E_{n-1}(1))))]' .$$

Here N_1, N_2, \dots, N_n represent the normalizing factors, so that, e.g.,

$$(4.7) \quad N_n = \{F_{0j}^i(1) [F_{1i}(E_2(E_3 \dots (E_{n-1}(1))))]'\}^{-1} .$$

The following notation is used: $F_{ij}^k(\varrho)$ denotes $\partial F_{ij}(\varrho) / \partial \theta_k$, and $F'_{ij}(\varrho)$ denotes $\sum_k F_{ij}^k(\varrho)$. The summation convention applies so that any pair of indices, which comprises one subscript and one equal superscript,

must be summed over. A repeated index which is *not* to be summed over is put in brackets. The vector $\underline{1}$ has unit components:

$$\underline{1} = (1, 1, \dots).$$

The simple statistical forest of eq. (1.6) and (1.7) (random f-functional polycondensation), for which the number S of types of nodes is unity, satisfies eqs. (4.4)-(4.7) (and trivially (4.3)). In general, the clone concept covers chemical or genetic forests of trees in which the correlation between fertility patterns of vertices is absent or confined to the shortest range (chemical 'first-shell' substitution effect - see Chapter 5).

Pursuing the theme of chemically useful combinatorics arising from the consideration of two vertices of one M-tree at a time, we turn to 'trail-weighting functions'.

4.1 TRAIL-WEIGHTING FUNCTIONS.

A trail of size k is a connected linear (unbranched) progression of k edges, containing $k+1$ vertices. Clearly in a *tree*, the two terminal vertices alone fully determine the trail that links them. Much of the physical properties of molecular substances, but specially of polymers, can be tackled theoretically by expansions in successive terms which reflect properties of $1, 2, \dots$ vertices (atoms) in the structure at a time. Scattering of radiation, for instance, depends on phase differences between rays striking two atoms.

For polymer distributions in which every trail in an M-graph corresponds to a sub-chain obeying Gaussian end-to-end statistics in the real-state molecule, the mean square radius (variously weighted as a number, weight, etc., average) can be found easily by cascade combinatorics

(Dobson and Gordon, 1964; Butler, Gordon and Malcolm, 1966). Essentially the problem translates into finding the expected value $E(N_{xj}/x)$ for the distribution w_{xj} where N_{xj} is the sum of the numbers of edges in all distinct trails contained in the j -th x -mer. For the simple case of f -functional polycondensation, with the s.f. defined by eq. (1.6) and (1.7) and w_{xj} given by eq. (1.11), the z -average mean square radius (obtainable from light-scattering) is found on this assumption of Gaussian statistics to be given by the formula (Dobson and Gordon, 1964):

$$(4.8) \quad \begin{aligned} \overline{\langle R^2 \rangle}_z &= \frac{1}{2} b^2 E(N_{xj}/x)/E(x) \\ &= b^2 f\alpha/2(1 + \alpha)(1 - (f-1)\alpha) , \end{aligned}$$

where b^2 is the mean square distance between adjacent vertices (in 3D space). This is remarkably simple, considering $\overline{\langle R^2 \rangle}_z$ is an average over two complicated distributions: (a) the mean square distances of all atoms from the centre of gravity of a given molecule under the Gaussian random-flight type internal Brownian motions of each of its sub-chains, and (b) the distribution of molecules, i.e., of different isomers of different numbers x of vertices.

The method of Dobson and Gordon (1964) and Butler, Gordon and Malcolm (1966) may be generalized as follows (Burchard, Gordon and Kajiwara, to be published). Instead of each trail producing a weighting equal to its length (i.e., the number of its edges), it then contributes a weighting which is an arbitrary function of its length. Random f -functional polycondensation (eq. (1.6) and (1.7)) again provides the simplest example, though much more complicated cases occurring in practice can be dealt with analogously.

We construct a generating function by generalization of (1.5), turning the process of eq. (2.1) and (2.2) into an iteration procedure. Moreover, we raise the Θ -symbols to various powers ϕ_i according to the generation number g_i on which their corresponding nodes occur in the corresponding tree (see Fig. 3).

Thus let

$$(4.9) \quad u_0 = e^{\phi_0} (1 - \alpha + \alpha w_{\pi^1})^f$$

$$(4.10) \quad u_n = \Theta^{\phi_n} (1 - \alpha + \alpha u_{n+1})^{f-1} \quad n = 1, 2, \dots$$

For $\phi_i = 1$, ($i = 0, 1, \dots$), u_0 reverts to $W(\theta)$ of eq. (2.1), and then $u'_0(1) = DP_W$ (which means $E(x)$). But generally

$$(4.11) \quad u'_0(1) = \phi_0 + f\alpha \sum_0^{\infty} \phi_{n+1} (\alpha(f-1))^n$$

which can often be summed or approximated by a Laplace transform, or - if all else fails - we find $u'_0(1)$ by computer iteration from and 4.9-10. The usefulness of $u'_0(1)$ resides in the following equation:

$$(4.12) \quad u'_0(1) = \phi_0 + \sum_x \sum_j w_{xj} \sum_k N_{kxj} \phi_k / x,$$

where N_{kxj} is $2 \times$ number of distinct trails of length k in the j -th x -mer tree. The proof of 4.12 rests on a consideration of Figure 3, bearing in mind that the statistical forest is a clone, which implies that each end-vertex of every trail in the molecular system has an equal chance of furnishing a root in the s.f. (Fig. 2); eq. (4.1) is also relevant to the proof.

The structure factor or scattering function $P_z(I)$ (see Zimm, 1948) provides an example. Here

$$(4.13) \quad \phi_n = \exp - \chi^2 n/6 \quad n = 0, 1, \dots$$

with

$$(4.14) \quad \chi = b^2(4\pi)^2 \lambda^{-2} \sin^2(I/2)$$

and hence from the definition

$$P_Z(I)^{-1} \equiv +E(x)/u'_0(1)$$

eq. (2.9) and (4.11) give

$$(4.15) \quad P_Z(I) = \frac{1 + \frac{(f-1)\alpha}{1 - (f-1)\alpha} [1 - \exp - \chi^2/6]}{1 - \frac{\alpha}{1+\alpha} [1 - \exp - \chi^2/6]} .$$

The approximate formula derived by Zimm for the very simple case $f = 2$ of linear (unbranched) molecules:

$$(4.16) \quad P_Z(I)^{-1} = 1 + \frac{16}{3} \overline{R^2} \sin^2(I/2)$$

(cf., eq. (4.8)) gives slight inaccuracies at higher scattering angles for X-ray scattering ($\lambda \sim 1\text{\AA}$). Eq. (2.4) can also be obtained from

(4.11) using $\phi_k = 1$, and, using this result, (4.8) may be obtained from

(4.11) and (4.12) with $\phi_k = k$.

5. SUBSTITUTION EFFECTS AND FERTILITY CORRELATIONS.

Harris (1964) remarked that, in cases where the fertility pattern of a person depends on how many brothers he has, the Markovian nature of the cascade model might be restored by introducing different types of indi-

viduals corresponding to different fertilities. This device will be used here to solve such problems, which have very realistic counterparts in chemistry.

Deviations from randomness of statistical forests, due to local (not too long-range) correlations within each tree, can be computed with ease and profit. In chemistry, for instance, the distribution w_{xj}^* for *random* f -functional polycondensation (henceforth given a star) is often modified in practice by correlations of the following type. First, the probability of occurrence of the j -th x -mer is modified by a factor P_{xj} ; secondly each of the isomeric x -mers gets a weighting C_0^x where C_0 is adjusted for normalization:

$$(5.1) \quad w_{xj} = C_0^x w_{xj}^* P_{xj}; \quad \sum w_{xj} = 1 .$$

Our problem then resides in how to modify the pgf's F_0, F_1, \dots , etc., of the various generations so as to induce these weightings, while preserving the simulation of the real molecular system by the powerful cascade-type statistical forest (Fig. 2).

A trail of size k is defined as a connected linear (unbranched) progression of k links, containing the $k+1$ nodes (here all of degree f). In the simple case where the deviation from randomness of f -functional polycondensation arises from the so-called linear n -th shell substitution effect ($n = 0, 1, \dots$), P_{xj} takes the form of a trail weighting function; more specifically, $P_{xj} = N_n^{2r}$ where $r = r(x, j)$ is the number of distinct trails of length $n+1$ in the j -th x -mer. The *general* non-linear n -th shell substitution effect (of which the linear is a special case) introduces more parameters, and translates local environments into statistical weights in a more discriminating manner. Thus

the n -th shell general substitution effects are defined by

$$(5.2) \quad n = 0: \quad P_{0xj} = \phi^{x-1},$$

$$(5.2a) \quad n = 1 \quad P_{1xj} = \prod_{l=i}^f \phi(i)^{u_{i,xj}},$$

and

$$(5.3) \quad n = 2 \quad P_{2xj} = \prod_{k \leq \ell = 1} \phi(k, \ell)^{u_{k\ell, xj}}.$$

Here ϕ is a constant, the $\phi(i)$ and $\phi(k, \ell)$ are arbitrary functions, $u_{i, xj}$ is the number of nodes with i links in the j -th x -mer and $u_{k\ell, xj}$ the number of links in the j -th x -mer whose nodes bear $k+1$ and $\ell+1$ links respectively. (For definitions of link and node, see p. 1 and 2.)

The basic tricks for modifying the pgf's are as follows. Consider the weight fraction pgf (cf., eq. 1.5)

$$(5.4) \quad W(\theta) = \sum_x \sum_j w_{xj} \theta^x = \theta F_0(\theta F_1(\theta F_2(\dots))) .$$

Then

a) Simultaneous multiplication of all F_i ($i = 0, 1, \dots$) by C_0 causes w_{xj} to be multiplied by C_0^x .

b) The simultaneous transformations

$$(5.5) \quad F_i(\theta) \rightarrow F_i(C^{-1}\theta)$$

$$(5.6) \quad F_{i+1}(\theta) \rightarrow C F_{i+1}(\theta)$$

have, for arbitrary C and i , no effect on w_{xj} . This allows one to renormalize F_{i+1} , and then successively (with appropriate *different* constants C) F_i, F_{i-1}, \dots down to F_0 , after their coefficients have been given any desired weightings.

c) An individual who is one of a litter of i brothers is indexed i , so that the corresponding auxiliary variable of this type is written θ_i in the *vectorial* pgf's $F_n(\underline{\theta})$.

In this way, we can modify the lpgf. (eq. (1.5-7)) for the statistical forest corresponding to the random (uncorrelated) molecular system, so as to produce exactly the general superposed first and second shell substitution effects. Thus:

$$(5.7) \quad F_0(\underline{\theta}) = C_0 \sum_{i=0}^f \binom{f}{i} (\gamma^*)^i (1-\gamma^*)^{f-i} C_i^{-i} \phi(i) \theta_i^i ,$$

(5.8)

$$F_{1j}(\underline{\theta}) = C_0 C_j \sum_{i=0}^{f-1} \binom{f-1}{i} (\gamma^*)^i (1-\gamma^*)^{f-i-1} C_{i+1}^{-i} \phi(i+1) \phi(i, j-1) \theta_i^i \quad (1 \leq j \leq f),$$

$$(5.9) \quad F_{1,j+1}(\underline{\theta}) = F_{2j}(\underline{\theta}) = F_{3j}(\underline{\theta}) = F_{4j}(\underline{\theta}) = \dots \quad (1 \leq j \leq f-1),$$

and

$$(5.9a) \quad W(\underline{\theta}) = \theta F_0(\theta \wedge F_1(\theta \wedge F_2(\dots))) .$$

Here γ^* is a parameter used to generate different degrees $\alpha (= F'_0(1)/f)$ of conversion. For the 'random' case (without substitution effects), $C_i = \phi(i) = \phi(i, j) = 1$ for all i and j , and $\gamma^* = \alpha^*$. The constant C_0 and the f constants C_j are found by normalization:

$$(5.10) \quad F_0(\underline{1}) = F_{1j}(\underline{1}) = 1 \quad (0 \leq j \leq f) .$$

The following special case of pure second shell substitution effect with $f=3$ and:

$$(5.11) \quad \phi(i) = 1 \quad (0 \leq i \leq f) ,$$

$$(5.12) \quad \phi(j, k) = 1 \quad (0 \leq j \leq 2; 0 \leq k \leq 1)$$

except that

$$(5.13) \quad \phi(2,2) = 0$$

is instructive as a model for steric hindrance effects in chemical gels, and in its own right as an extension of a class of lattice-statistical problems from a linear lattice (Gordon and Hillier, 1963; McKenzie, 1962) to a branched one. Computer results will be shown on the statistics of the forest as a function of α . They indicate that the maximum attainable value of α (in the final gel molecule free from sol fraction) is $\approx 0.795 \pm 0.002$ for this model.

APPENDIX A

PROOF OF THE MASTER THEOREM (EQ. (4.3) - (4.7)) AND PROBABILISTIC PROOF OF EQ. (4.1).

A *replanting* is a change of rooting of an M-tree from an old root to a new root (Fig. 7). On replanting, the vertex which formed the old root on g_0 , appears on the same generation (g_n , say) on which the new root appeared in the old rooting. We shall merely say the path linking the roots is *inverted* by replanting; when the two roots belong to g_0 and g_n , we call the replanting a (g_0, g_n) -replanting.

Eq. (4.3) and (4.4) are proved by considering the statistical implications respectively of all (g_0, g_1) -replantings, eq. (4.5) of all (g_0, g_2) -replantings, and eq. (4.6) of all (g_0, g_n) -replantings involving at least one j-type. Thus eq. (4.3) merely expresses the fact that a j-type vertex on g_0 has the same average number $F'_{0j}(1)$ of offsprings (irrespective

of type) on g_1 as there are, on average, j -type offsprings on g_1 for vertices (irrespective of type) occurring as roots on g_0 .

Equation (4.6), which -suitably interpreted - contains (4.4) and (4.5) as special cases follows quite similarly. Thus we obtain the relation between F_0 and F_n as the statistical result of all possible (g_0, g_n) -replantings involving at least one j -type vertex. What happens to the set of offsprings of a j -type vertex when it is translated by path inversion from g_0 to g_n ? It carries with it, with one exception (Fig. 7) all its offsprings which move from g_1 to g_{n+1} . The only deletion is the *one* original offspring on g_1 (say an i -type) which lay on the path being inverted, which is carried from g_1 to g_{n-1} where it becomes the father of the translated j -type on g_n .

Eq. (4.6) can now be understood by inspection. It expresses the fact that pgf F_{nj} is essentially the same as F_{0j} , but the deletion of the i -type offspring must be effected by differentiating with respect to θ_i (which lowers the exponent of θ_i by unity, as required). Moreover, the same differentiation multiplies the probability of any given set of offsprings by the number of i -types in the set, which is required because each i -type offspring on g_1 provides a separate gate for an equal average number of paths from a j -type on g_0 to vertices of any type on g_n , which paths may thus translate the j -type up to g_n by path inversion. The average number N of paths which link an i -type on g_1 to any type of vertex on g_n is, by the usual methods of cascade theory, given by

$$(A1) \quad N = [F'_{1i}(F_2(F_3 \dots (F_{n-1}(1))))]$$

which is the remaining factor in eq. (4.6). As a last step in the

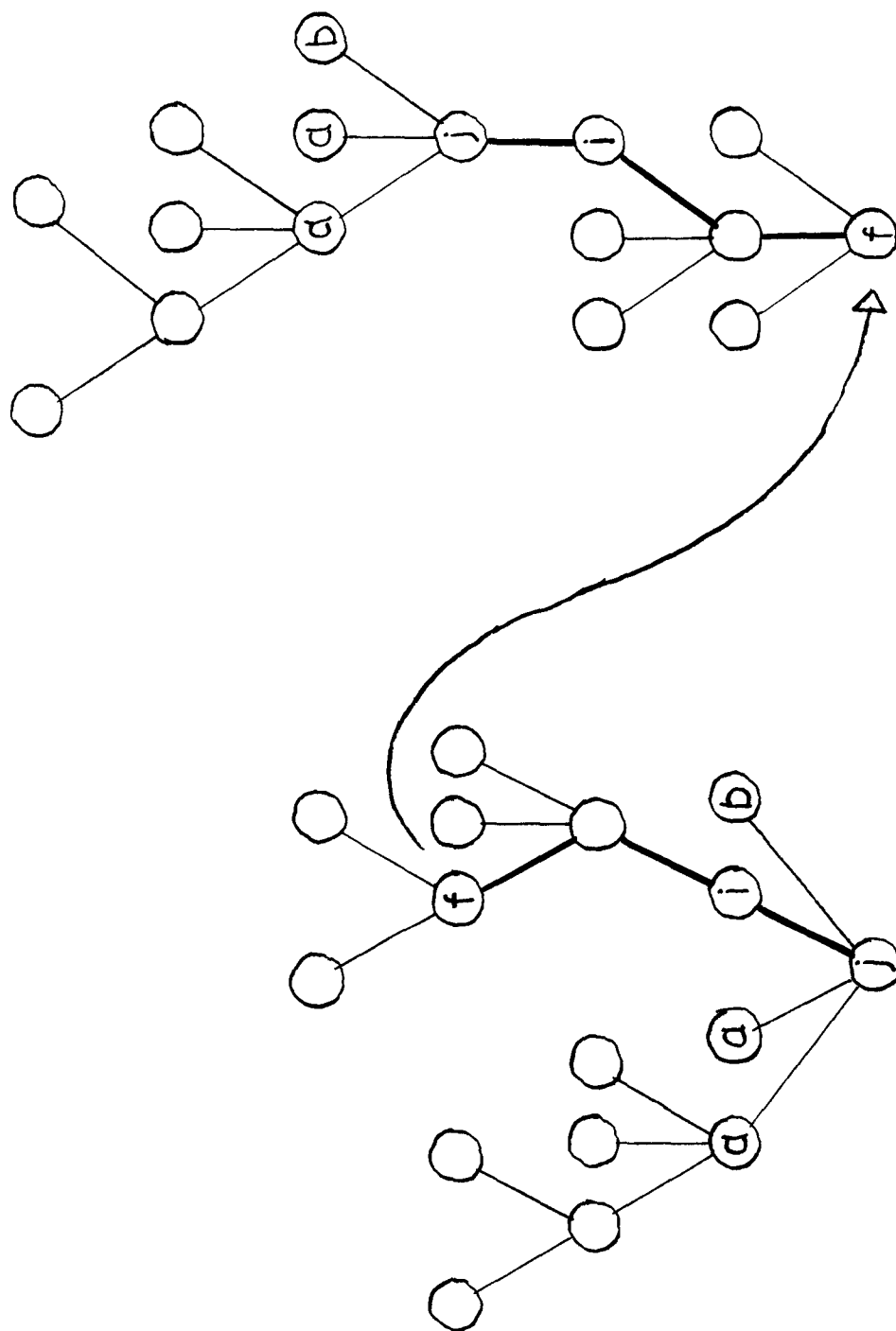


Figure 7

Re-rooting of an M-tree. The inverted path is shown as the thick line. For clarity, only the relevant atomic labels are shown.

derivation of (4.6), we note that we must sum with respect to i over all types, since the particular type i singled out for discussion was merely representative of any vertex on g_1 , all of which can feature on paths which are inverted on (g_0, g_n) -replantings.

A1. PROOF OF EQ. (4.1): GENERALIZATION OF THE NOTION OF STATISTICAL FOREST.

We now generalize the notion of different *types* of individuals in a statistical family tree, or vertices in an ordered rooted M-tree, by introducing auxiliary variables with three subscripts, thus:

$$(A2) \quad \varrho = [\theta_{111}, \dots, \theta_{ijk}, \dots] .$$

This enables us to classify individuals into types from three different points of view:

- I) A metrical classification denoted by subscript i , viz., the rank-label of a given individual defining his position in the litter (Fig. 8c). In this work, we thus generate for the first time statistical forests of *ordered* trees, using the subscript i to specify the ranking (from left to right) of an offspring within the litter to which he belongs;
- II) A topological classification denoted by subscript j , viz., the degree j of the vertex which the individual represents in the tree;
- III) A 'chemical' classification denoted by the subscript or atomic label k , which, as hitherto, specifies by convention some chemical type of atom in an M-tree.

The generalized statistical forest of ordered M-trees we contemplate is

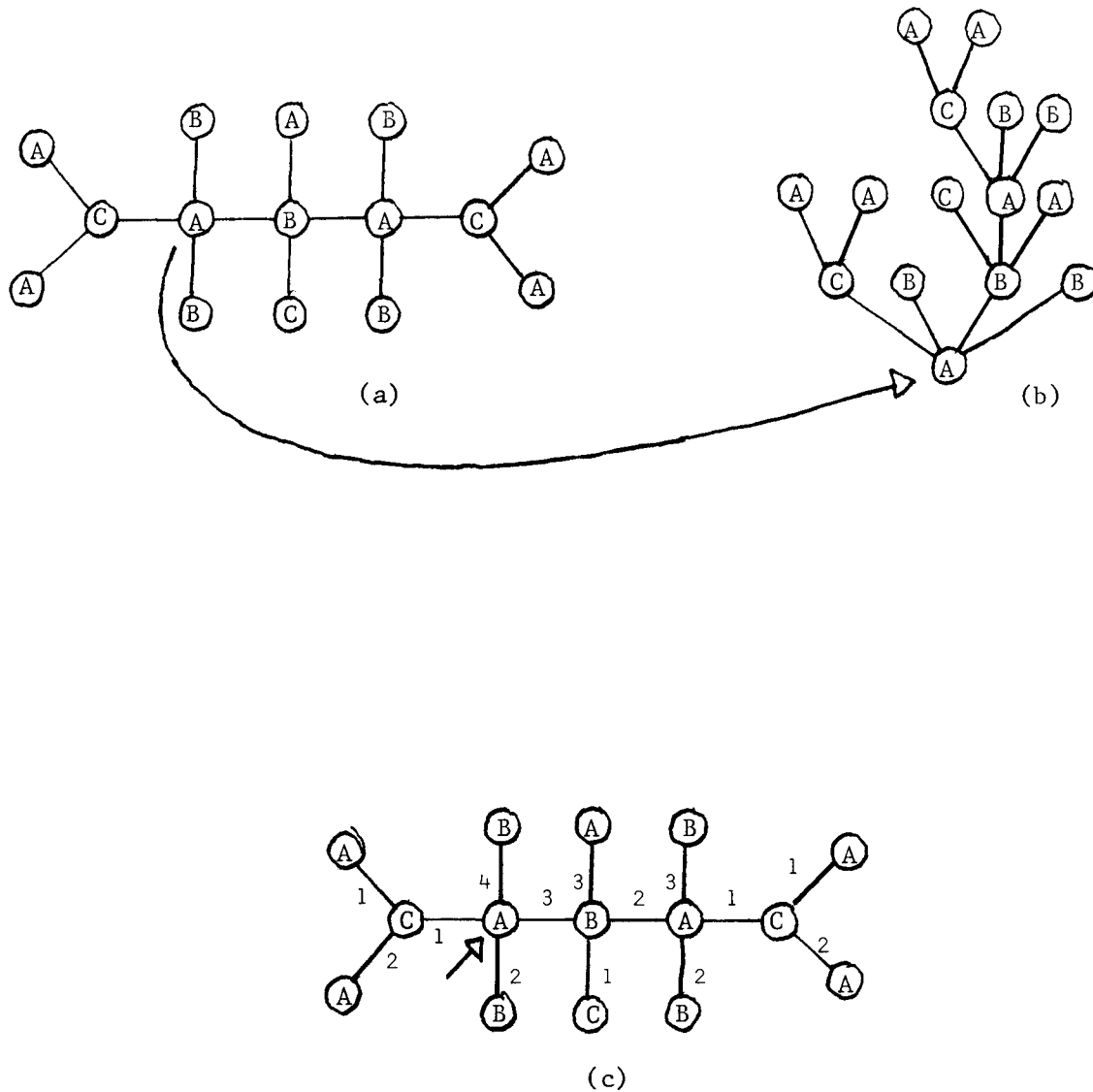


Figure 8

An *ordering* of an M-tree, such as (a), comprises the choice of a root-vertex, plus a positioning of the branches as in (b). The same information as in (b) is contained in (c), where the arrow shows the root-vertex, and the numerical rank-labels determine the positions from left to right of the edges sprouting upward from any vertex in (b).

thus specified by eq. (A3) and (A4):

$$(A3) \quad \underline{n} = [n^{11}, \dots, n^{jk}, \dots] .$$

Here n^{jk} is the chance that the root of an M-tree in the forest should have degree j and atomic label k , while superscript i is not required, because an individual on g_0 (forming the root of a tree) is necessarily devoid of siblings; and

$$(A4) \quad \underline{P}_n(\underline{\theta}) = \left[\sum_{\underline{m}} P_{n1, \underline{m}^{\underline{\theta}}}^{\underline{m}}, \sum_{\underline{m}} P_{n2, \underline{m}^{\underline{\theta}}}^{\underline{m}}, \dots, \sum_{\underline{m}} P_{ns, \underline{m}^{\underline{\theta}}}^{\underline{m}} \right] ,$$

where

$$(A5) \quad \underline{m} \equiv [m_{111}, \dots, m_{ijk}, \dots] .$$

The probability distributions P_{nr} are defined by (A4).

In the next section, we define a specific example of such a statistical forest, which is as random as can be, in the sense that wherever there arises a choice between various alternatives, each contingency carries equal probability. A forest of ordered rooted trees is a *thicket* if it embodies the following type of randomness. Every *distinct* ordering of any one given M-tree (Fig. 8) occurs with equal probability. There is no restriction, however, on the relative probabilities of orderings of distinct M-trees.

A2. THE THICKET T .

Thus consider the following example of eq. (A3):

$$(A6) \quad n^{fk} = 1/FK (= \text{const.}) ,$$

where henceforth

$$(A7) \quad 1 \leq f \leq F; \quad 1 \leq k \leq K .$$

This signifies that an individual on g_0 (a root) has equal chances of being found a vertex of any degree between 1 and F (an arbitrary constant positive integer) and bearing any atomic label between 1 and K (an arbitrary constant positive integer) inclusively. Next, let $F_{\sim 0}$ (cf., eq. (A4)) have F components, such that any individual of type (fk) has the same fertility pattern depending only on f , but independent of k :

$$(A8) \quad F_{0f} = (FK)^{-f} \prod_{i=1}^f \sum_{j=1}^F \sum_{k=1}^K \Theta_{ijk} .$$

Thus an individual of type (fk) on g_0 has, as by definition he must, always exactly f offsprings on g_1 , but there is no bias as regards the degrees j or the atomic labels k of these offsprings, each value in the ranges (A7) being of equal probability.

Finally, all individuals on generations g_1, g_2, \dots can produce litters of at most $F-1$ members, and their pgf's are taken to be independent of the generation to which they belong:

$$(A9) \quad F_{1f} = (FK)^{-f+1} \prod_{i=1}^{f-1} \sum_{j=1}^F \sum_{k=1}^K \Theta_{ijk}$$

$$(A10) \quad = F_{2f} = F_{3f} .$$

The statistical forest defined by (A8) - (A10) is a thicket, \mathcal{T} say, since every ordering of an M -tree occurs with equal chance: indeed, the probability of finding any ordering of any M -tree of order v is $(FK)^{-v}$. Any finite M -tree of $v \geq 2$ is included in \mathcal{T} with finite probability provided $F > v, K > v$.

A3. THE CLONE C.

The thicket T is *not* a clone. Eq. (4.4), (4.5) and (4.6), it is true, are verified* by the set (A8), (A9) and (A10), but eq. (4.3) is *not* satisfied by (A6). We are thus led to calculate the set of $F \times K$ components of n which, together with (A8) - (A10), do form a clone, which we denote by C . By substituting (A8) in (4.2), and using the normalization condition $\sum_f \sum_k n^{fk} = 1$ we find for all k :

$$(A11) \quad n^{fk} = f^{-1} \times \left(\sum_1^F p^{-1} \right)^{-1} (\alpha f^{-1}) .$$

Accordingly, eq. (A8) - (A10) can be combined with (A6) to form the thickened T , or with (A11) to form the closely related clone C .

Consider the relative frequencies of occurrence of one and the same finite M-tree when rooted on a vertex of degree f_1 belonging to equivalence class 1 of order s_1 , and when rooted on a vertex of degree f_2 belonging to equivalence class 2 of order s_2 . In the thicket T , the ratio of the two frequencies is merely the ratio, T_{M1}/T_{M2} , of distinct orderings belonging to these two rootings, since in a thicket every ordering occurs with equal frequency. In the clone C , however, comparison of (A6) and (A11) shows that the ratio of frequencies is modified to $T_{M1}f_2/T_{M2}f_1$. Moreover, because C is a clone, the same ratio of frequencies must be equal to s_1/s_2 so that eq. (4.1) is proved.

* Note that only *relative* values of rank labels are significant, so that, e.g., ${}^0_{156}{}^0_{284}{}^0_{379} = {}^0_{156}{}^0_{384}{}^0_{479} = {}^0_{256}{}^0_{384}{}^0_{479}$, etc., since $1 < 2 < 3$, $1 < 3 < 4$, $2 < 3 < 4$, etc.

REFERENCES CITED

1. D.S. Butler, M. Gordon and G.N. Malcolm, 1966, Proc. Roy. Soc. (London) A295, 29.
2. A. Charlesby, 1955, Proc. Roy. Soc. (London) A231, 521.
3. G.R. Dobson and M. Grodon, 1964, J. Chem. Phys. 41, 2389.
4. G.R. Dobson and M. Gordon, 1965, J. Chem. Phys. 43, 705, reprinted Rubb. Chem. Technol., 1966, 39, 1472.
5. K. Dusek^v and W. Prins, Adv. Polymer Sci., 1969, 6, 1.
6. P.J. Flory, 1953, Principles of polymer chemistry, Ithaca, Cornell University Press.
7. F. Galton, 1889, Natural Inheritance, London, McMillan.
8. M. Gordon, 1962, Proc. Roy. Soc. (London) 268, 240.
9. M. Gordon and I.A. Hillier, 1965, J. Chem. Phys., 38, 1376.
10. M. Gordon and G.R. Scantlebury, 1964, Trans. Faraday Soc., 60, 604.
11. I.J. Good, 1960, Proc. Camb. Phil. Soc., 56, 367.
12. I.J. Good, 1963, Proc. Roy. Soc. (London) A272, 54.
13. F. Harary, 1967, A Seminar on graph Theory, New York.
14. T.E. Harris, 1963, The theory of branching processes, Berlin, Springer.
15. S. Imai and M. Gordon, 1969, J. Chem. Phys., in press.
16. W. Kuhn, 1934, Kolloidz, 68, 2.
17. J.K. Mackenzie, J. Chem. Phys., 1962, 37, 723.
18. D.W. Matula, L.C.D. Groenweghe, and J.R. Van Wazer, 1964, J. Chem. Phys., 41, 3105.
19. B.J. Zimm, 1948, J. Chem. Phys., 16, 1093.

Other relevant references

1. M. Gordon and G.R. Scantlebury, 1967, J. Chem. Soc., B, 1.
2. M. Gordon and G.R. Scantlebury, 1968, J. Polymer Sci., C, No. 16, 3933.