Polymer Models and Generalized Potts–Kasteleyn Models

P. Whittle[1]

The well-established relation between Potts models with v spin values and random-cluster models (with intracluster bonding favored over intercluster bonding by a factor v) is explored, but with the random-cluster model replaced by a much generalized polymer model, implying a corresponding generalization of the Potts model. The analysis is carried out in terms a given defined function $R(\rho)$, an entropy/free-energy density for the polymer model in the case $v = 1$, expressed as a function of the density ρ of units. The aim of the analysis is to determine the analog $R_v(\rho)$ of $R(\rho)$ for general nonnegative v in terms of $R(\rho)$, and thence to determine the critical value of density ρ_{vg} at which gelation occurs. This critical value is independent of v up to a value v_P, the Potts-critical value. What is principally required of $R(\rho)$ is that it should show a certain given concave/convex behavior, although differentiability and another regularizing condition are required for complete conclusions. Under these conditions the unique evaluation of $R_v(\rho)$ in terms of $R(\rho)$ is given in a form known to hold for integral v but not previously extended. The analysis is carried out in terms of the Legendre transforms of these functions, in terms of which the phenomena of criticality (gelation) and Potts criticality appear very transparently and v_P is easily determined. The value of v_P is 2 under mild conditions on R. Special interest attaches to the function $R_0(\rho)$, which is shown to be the greatest concave minorant of $R(\rho)$. The naturalness of the approach is demonstrated by explicit treatment of the first-shell model.

1. POLYMER MODELS; CRITICALITY

The assumptions and conclusions of the paper are summarized at the end of this introductory section.

---

[1] Statistical Laboratory, University of Cambridge, Cambridge, England.

There is a useful and interesting class of polymer models[3,4,7-13] for which the metric aspects of space are neglected, in that the assembly of units is regarded simply as a random graph. Units correspond to nodes of this graph and bonds between units to arcs between nodes. In this equivalence a polymer (i.e., a polymer molecule) then corresponds to a component of the graph. Usually one considers symmetric statistics on the complete graph, in that a bond can in principle be formed between any pair of units and the probability of a configuration is invariant under permutation of the units. One can regard such models as mean-field versions of an explicitly spatial model.

In such models it is useful to include a parameter v which allows intrapolymer bonds (i.e., a new arc in an existing component) to be formed v times as easily as interpolymer bonds (i.e., a new arc between existing components). In the extreme case $v = 0$ intrapolymer bonding is forbidden, and all polymers must be trees. The assumption of large v would reflect the view that units within the same polymer molecule are contiguous in some sense, and so form new bonds more easily than would units in different molecules.

This additional parameter corresponds exactly to the parameter q of the Potts model, as was revealed when Kasteleyn and Fortuin[5] established an equivalence between the Potts model and a simple random-cluster model. In ignorance of this work the author[13] established this same equivalence, but working from the other direction, working from a polymer model which is a considerable generalization of the random-cluster models considered in the Potts–Kasteleyn literature. This paper essentially explores this generalization; we comment further on the mutual implications between the two problems in Sections 2 and 3.

Consider then the statistics of N units. Despite the fact that the model has no metric structure, it is natural to introduce a parameter V which can be interpreted as 'volume,' in that the bonding rate between two given units is of order V^{-1}. In going to the thermodynamic limit one will then let N and V tend to infinity in constant ratio,

$$\rho = N/V$$

where ρ is then to be interpreted as the spatial density of units.

Suppose the probability of a configuration $\mathscr{C}$ of these N identifiable units in a volume V is proportional to $Q(\mathscr{C} \,|\, N, V)$, where the proportionality factor is such that $Q(N, V) = \sum_{\mathscr{C}} Q(\mathscr{C} \,|\, N, V)$ can be identified with the partition function of the system. Define

$$A(N, V) = \frac{V^N}{N!} Q(N, V)$$

If interactions are 'local' in that the thermodynamic limit indeed exists, then one will expect $\log Q$, and so $\log A$, to be extensive in that they are $O(V)$ for large V and prescribed ρ. More specifically, that

$$A(N, V) = e^{VR(\rho) + o(V)} \tag{1.1}$$

for large V, so that $V^{-1} \log A(N, V)$ has a limit $R(\rho)$ in the thermodynamic limit. In more familiar notation one has $R = \rho - \rho \log \rho - \beta F$, where F is the free energy per unit volume and β is proportional to inverse temperature. However, we shall be taking density of units ρ rather than temperature as the parameter of interest, whose variation exhibits the phase transitions of the system, and so work rather in terms of R.

As an example, the version taken by the author[12] of the so-called 'first-shell' model corresponds (in the case $v = 1$) to the choice

$$Q(\mathscr{C} \mid N, V) = \left[\prod_a \prod_b (2\kappa V)^{-s_{ab}} / s_{ab}! \right] \left(\prod_j H_j^{n_j} \right) \tag{1.2}$$

where s_{ab} is the number of bonds from node a to node b and n_j is the number of units with j bonds (nodes of degree j). The configuration $\mathscr{C}$ is determined by the s_{ab}. The terms in κ and the H_j then represent a Gibbs factor $e^{-\beta \mathscr{E}}$, where $\mathscr{E}$ is the potential energy of the configuration. The terms $(V^{s_{ab}} s_{ab}!)^{-1}$ are essentially combinatorial, reflecting the decreasing probability of bonding between a given pair of units with increasing volume and the indistinguishability of distinct bonds between this pair of units. (Although the probability that multiple bonds exist between two prescribed units, conditional on the event that they are bonded, tends to zero in the thermodynamic limit.)

It was shown by the author[11,12] that, for this case,

$$Q(N, V) = \left(\frac{\kappa V}{2\pi} \right)^{1/2} \int_{-\infty}^{\infty} H(\xi)^N e^{-\kappa V \xi^2 / 2} \, d\xi \tag{1.3}$$

where

$$H(\xi) = \sum_{j=0}^{\infty} H_j \xi^j / j! $$

If $H(\xi)$ is such that the integral (1.3) is convergent for all positive N, V, then it follows that, for this model,

$$R(\rho) = \rho - \rho \log \rho + \sup_{\xi} \left[\rho \log H(\xi) - \tfrac{1}{2} \kappa \xi^2 \right] \tag{1.4}$$

However, there is no need to argue in terms as specific as this. Let us assume merely that $R(\rho)$ belongs to the class $\mathscr{R}$ of functions which, as ρ

increases from zero, are first strictly concave with infinite positive derivative at $\rho = 0$, and then become strictly convex at some density ρ_g, to remain so thereafter. We illustrate the behavior in Fig. 1. The function (1.4) behaves in this fashion (see Section 8); we see that the initial concavity comes from the entropy term $-\rho \log \rho$ and the later convexity from the final term in (1.4), proportional to the free energy.

The value ρ_g is then the unique value of ρ at which the derivative $R'(\rho)$ is minimal. It can be said to mark a critical point, at least in the non-degenerate case when it is neither zero nor infinite. To see this, consider two replicas of the model which are independent but communicating, in that the probability of configurations $\mathscr{C}_1$ and $\mathscr{C}_2$ in the two replicas is proportional to $Q(\mathscr{C}_1 | N_1, V) Q(\mathscr{C}_2 | N_2, V)$ subject only to $N_1 + N_2 = 2N$, where N_i is the number of units in configuration $\mathscr{C}_i$. The joint distribution of the N_i is then obtained by summing this expression both over configurations consistent with the N_i and over permutations of the units, and so is given by

$$P(N_1, N_2) \propto A(N_1, V) A(N_2, V) \qquad (N_1 + N_2 = N)$$

The most probable value of the vector of densities $\rho_i = N_i/V$ in the thermodynamic limit is then that maximizing $R(\rho_1) + R(\rho_2)$ subject to $\frac{1}{2}(\rho_1 + \rho_2) = \rho$.
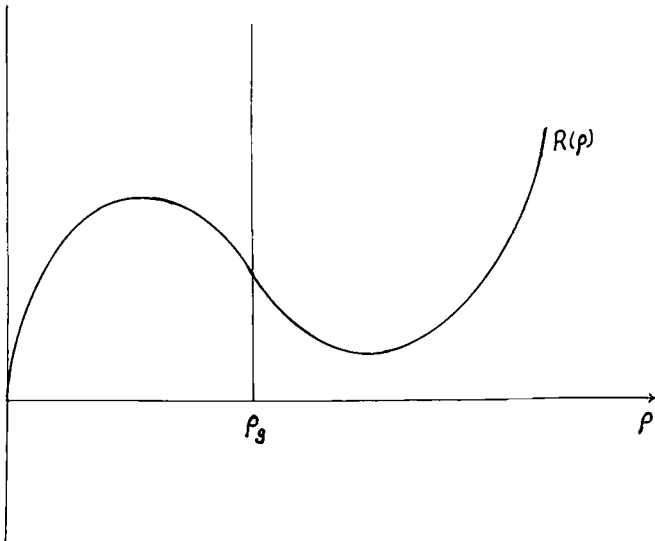

Fig. 1. The graph of $R(\rho)$ (an entropy/free-energy density for the polymer model) as a function of the density ρ of units. It is initially concave, with infinite positive gradient at zero density, and then changes to being convex from ρ_g. The value ρ_g then marks the point of minimal gradient, and locates the gel point.

Proposition 1. Suppose that $R(\rho) \in \mathscr{R}$ and $R'(\rho)$ is minimal at ρ_g. If $\rho < \rho_g$, then the configuration of equidistribution between two replicas $\rho_1 = \rho_2 = \rho$ is at least metastable, in that $R(\rho - \delta) + R(\rho + \delta)$ is locally maximal with respect to δ at $\delta = 0$. If $\rho > \rho_g$, then it is not.

That is, matter tends to lump in one or the other of the replicas if ρ exceeds ρ_g, but will resist at least local perturbations from equidistribution if ρ is less than ρ_g. In the polymer context this transition marks passage from the sol state to the gel state; the subscript g in ρ_g is intended to indicate 'gel point.'

Proof. If $0 \leqslant \rho - \delta < \rho + \delta \leqslant \rho_g$, then

$$R(\rho - \delta) + R(\rho + \delta) < 2R(\rho)$$

by strict concavity. If $\rho_g \leqslant \rho - \delta < \rho + \delta$, then

$$R(\rho - \delta) + R(\rho + \delta) > 2R(\rho)$$

by strict convexity. The proposition then follows. ∎

One can now ask whether the assertion of metastability can be strengthened to that of global stability, in that $R(\rho_1) + R(\rho_2)$ has its global maximum, subject to prescription of mean density ρ, at $\rho_1 = \rho_2 = \rho$. We shall see in Section 7 that, under additional conditions, this strengthening is only just valid, in that equipartition gives the global maximum for any ρ less than ρ_g, but that this assertion does not hold if one considers more than two replicas. Indeed, this statement holds even if we allow the number of replicas to be nonintegral, an extension which turns out to be necessary, natural, and unique.

Consider indeed the case of an integral number ν of replicas, independent but communicating. If ρ_i is the density in the ith replica, then the most probable distribution of matter subject to prescription of the average density

$$\nu^{-1} \sum_i \rho_i = \rho \tag{1.5}$$

is that which maximizes $\sum_i R(\rho_i)$ subject to this constraint. Let us indeed define

$$R_\nu(\rho) = \sup \nu^{-1} \sum_i R(\rho_i)$$

where the supremum is over the distribution $\{\rho_i\}$ subject to condition (1.5). Let us refer to values of density which are less than or greater than ρ_g as *subcritical* and *supercritical*, respectively.

Proposition 2. Suppose that $R \in \mathscr{R}$. Then in the most probable distribution over v replicas at most one replica is supercritical and all subcritical replicas have the same density. Further,

$$R_v(\rho) = \sup v^{-1}[(v-1)R(\rho_1) + R(\rho_2)] \tag{1.6}$$

where the supremum is over ρ_1 and ρ_2 and subject to

$$v^{-1}[(v-1)\rho_1 + \rho_2] = \rho \tag{1.7}$$

The maximizing ρ_1 is necessarily subcritical, and, if the maximizing ρ_2 is also subcritical, then necessarily $\rho_1 = \rho_2 = \rho$.

The first assertion follows from the concave/convex nature of R by standard arguments; the others follow immediately from it. In polymer terms, the assertion is that the sol fraction has the same density in all replicas, and that the gel fraction, if there is one, is concentrated in a single replica.

It will later transpire that, for any prescribed positive ρ less than ρ_g, the inequality $R_v(\rho) > R(\rho)$ holds for all large enough v. That is, if one puts sufficiently many replicas into communication at any prescribed positive average density of units, then fluctuations will bring one of the replicas to criticality and the ensuing gelation is stable. This is Potts criticality, remarked in ref. 14 for polymers.

If we view the v replicas as v adjacent compartments in physical space, then we might term this a 'compartmental' model. However, such a view raises fundamental points, which we discuss briefly in Section 9.

However, the point well recognized in the Potts–Kasteleyn literature, and which we shall discuss in the next section, is that the modification of $R(\rho)$ to $R_v(\rho)$ can be brought about, not only by the consideration of v replicas of the original system, but by the introduction of a term in the distribution which favors the creation of a new bond *within* a polymer over one *between* polymers by a factor v. This modification is meaningful and interesting in itself, and is one not restricted to integral v.

A preliminary aim of the paper is, first, to obtain an evaluation of $R_v(\rho)$ in terms of $R(\rho)$ for all nonnegative v. It is shown in Proposition 8 that the determination (1.6) for integral v generalizes simply, although the analysis is best pursued in terms of the Legendre transform of R_v as a function of ρ. Qualitative features of the function are listed in Proposition 6 and depicted in Fig. 2. One striking and significant conclusion is that $R_0(\rho)$ is the greatest concave minorant of $R(\rho)$.

A more fundamental aim is evaluation of the critical point as a function ρ_{vg} of v. This evaluation is implied by the evaluation of $R_v(\rho)$. It turns out that $\rho_{vg} = \rho_g$ for v less than or equal to a critical value of v which we

shall term the *Potts critical value* and denote by v_P. For $v > v_P$ the critical point ρ_{vg} decreases with increasing v.

A strengthening of assumptions yields the evaluation (7.1) of v_P. Under rather weak assumptions this reduces to the evaluation $v_P = 2$ familiar from the literature, but there are interesting processes for which $v_P > 2$.

The basic assumption is that R should belong to $\mathscr{R}$, where $\mathscr{R}$ is the set of scalar functions of the nonnegative variable ρ which are first concave and then convex and have infinite positive derivative at the origin. A stronger assumption is that R should belong to $\mathscr{R}_c$, the set of those members of $\mathscr{R}$ which have continuous first derivative. For the strongest conclusions we require that R belong to $\mathscr{R}^*$, the set of those members of $\mathscr{R}_c$ for which the function $D(\gamma)$ defined in (7.2), (7.3) is increasing. The intuitive content of this additional condition is at least partly exhibited by the geometric interpretation given after (7.3).

Finally, it is shown in Section 8 that all these conditions hold for the first-shell model, and some general implications of the analysis are briefly notes in Section 9.

2. INTRA/INTERPOLYMER DIFFERENTIATION AND POTTS–KASTELEYN EFFECTS

The usual random graph model adopted by mathematicians is a rather degenerate one from the physical point of view, as indeed is the 'random-cluster' model which arises in connection with the usual Potts model. It allows single bonds to form between distinct units (arcs between distinct nodes) independently, with a probability a/N of occurrence for some a. The analog in our formulation would be to set $H_i \equiv 1$ in the single-shell model (1.2), so that there is no sensitivity to degree at a node, and the N^2 random variables s_{ab} (the number of arcs directed from node a to node b) are independent Poisson variables with common expectation $1/2\kappa V$. The choice of Poisson variables rather than 0/1 variables is much more natural mathematically, which probably means also physically. It implies the possibility of multiple bonding and self-bonding, although with a relative probability tending to zero for any given unit or unit pair in the thermodynamic limit. In this limit the two models show the same behavior. We shall refer to it as the 'Poisson model'; it is in a certain sense the zeroth-order model, in that others are obtained by superimposing additional interactions upon it. For the Poisson model $H(\xi) = e^\xi$ and the evaluation (1.4) of $R(\rho)$ becomes

$$R(\rho) = \rho - \rho \log \rho + \rho^2/(2\kappa) \tag{2.1}$$

The relation between the parameters in the two versions is $a = \rho/\kappa$.

Suppose the possible distinguishable polymers (graph components) are labeled $r = 1, 2,...$, with the rth type referred to as an *r-mer*. Then a generalization of model (1.2) (the maximal generalization in a certain sense, for symmetric bonding models on the complete graph) is to assume that

$$Q(\mathscr{C} \mid N, V) = \prod_r (\Gamma_r V^{-b_r})^{m_r} \tag{2.2}$$

where Γ_r is a constant independent of N and V, b_r is the number of bonds in an r-mer, and m_r is the number of r-mers in configuration $\mathscr{C}$. Since the configuration is one on N units, one must have

$$\sum_r m_r n_r = N \tag{2.3}$$

where n_r is the number of units in an r-mer. The configuration $\mathscr{C}$ is determined by prescription of the m_r and prescription of the places of identified units. Distribution (1.2) is indeed a special case of (2.2).

Define now the generating function

$$\tilde{Q}(\theta, V) = \sum_N \theta^N A(N, V) = \sum_N (\theta V)^N \frac{Q(N, V)}{N!}$$

which is the grand partition function. It is possible (and, indeed, usual in this context) that this series diverges for all nonzero θ. This fact is not disturbing: relations between such generating functions give the concise and natural expression of relations between the coefficients in their expansions in powers of θ. However, this lack of convergence seemingly makes it necessary to form hypotheses in terms of $A(N, V)$ rather than of the generating function $\tilde{Q}$ itself.

The following result is proved in ref. 13; it is of such fundamental importance that we quote it as a proposition.

Proposition 3. For model (2.2)

$$\tilde{Q}(\theta, V) = \exp \left(\sum_r \Gamma_r \theta^{n_r} V^{n_r - b_r} \right) \tag{2.4}$$

Suppose now that model (2.2) is modified to

$$Q(\mathscr{C} \mid N, V) = Q_v(\mathscr{C} \mid N, V) = v^{B + C - N} \prod_r (\Gamma_r V^{-b_r})^{m_r} \tag{2.5}$$

where $B = \sum_r m_r b_r$ is the number of number of bonds in the configuration and C is the number of components. Then $B + C - N$ is the cyclomatic number of the graph—the number of bonds which are in excess in that it is the maximal number of bonds which could be removed while still leaving the polymers (components) connected. (Recall that $b_r \geqslant n_r - 1$, with equality in just those cases for which the r-mer is a tree.)

The effect of increasing v is then to favor such excess bonding—to favor the formation of intrapolymer bonds relative to interpolymer bonds. If we denote the grand partition $\tilde{Q}$ for model (2.5) by $\tilde{Q}_v$ (so that $\tilde{Q}_1 = \tilde{Q}$), then we have the following important identity.

Proposition 4.

$$\tilde{Q}_v(\theta, V) = \tilde{Q}(\theta, V/v)^v \qquad (2.6)$$

Proof. Model (2.5) is again of the form (2.2), but with $\Gamma_{vr} = v^{b_r + 1 - n_r} \Gamma_r$. Making this substitution in evaluation (2.4), we deduce just the relation (2.6). ∎

Identity (2.6) has a fascinating implication. Suppose that v is integral. Then relation (2.6) implies that the statistics for prescribed N, V, and v are identical with those for the case when the N units are distributed over v independent but communicating replicas of volume V/v in which the statistics are those for the case $v = 1$. That is, we return to the compartmental models of Section 1. The same holds for the case of nonintegral v, except that one must then accept (and handle mathematically) the situation of a nonintegral number of replicas.

Relation (2.6) is of course that established in a special case by Kasteleyn and Fortuin[5] and which revealed the equivalence between Potts and random-cluster models. As Wu[15] observes, this equivalence had been well known to mathematicians. However, relation (2.6) is in some respects a very considerable generalization of those to be found in the Potts literature. It is more special than the identities reviewed in ref. 1, for example, in that, as we shall see in the next section, it corresponds to a Potts model on the complete graph, a mean-field model. However, it corresponds to a considerable generalization of the random-cluster model, and so of the interaction pattern permitted in the Potts model. Indeed, it constitutes the most general model on the graph which is symmetric in the vertices and allows components to be statistically independent in an open system (i.e., in the grand canonical ensemble). One can of course add the features of incomplete symmetry and of color to the graph, generalizations which we shall forswear in this paper.

We assume that if the logarithm of the partition function has the extensive form (1.1) in the case $v = 1$, then it continues to do so in the case of general fixed v,

$$A_v(N, V) = e^{V R_v(\rho) + o(V)} \tag{2.7}$$

for some function $R_v(\rho)$. This is a hypothesis which certainly places no additional constraint upon the function R. We shall see in Section 6 that if R is in $\mathscr{R}$, then it is 'infinitely divisible' in that a function R_v exists which satisfies (2.6) and (2.7) for any $v \geqslant 0$. Indeed, this determination of R_v in terms of R is unique if R belongs to $\mathscr{R}_c$ and the additional condition is imposed that R_v be increasing in v, as would be the case if it were generated from the distribution (2.5). Whether $A_v(N, V)$ *necessarily* has the form (2.7) has yet to be demonstrated. It certainly does for positive integral v, as follows from the calculations at the end of Section 1. In the particular case when $A_v(N, V)$ is derived from expression (2.5) it follows both that $A_v(N, V)$ is increasing in v and that assertion (2.7) holds for $v = 0$ [see Eqs. (9.2) and (9.3)]. It thus follows that $\log A_v(N, V)$ is of order V in the thermodynamic limit for any nonnegative v, which is part way to the conclusion that $V^{-1} \log A_v(N, V)$ has a limit.

Propositions 2 and 4 imply the evaluation (1.6) of R_v in terms of $R_1 = R$ in the case of integral v. Our task is now to extend this evaluation to the case of nonintegral v and to deduce the effect of variation of the parameter v on phase transitions and the like.

3. RELATION WITH THE POTTS–KASTELEYN MODEL

The Potts model is an Ising model which allows q values of spin at each site. If one considers a mean-field model in which sites interact only with sites of the same spin value, and if N_i out of N sites have the ith spin value, then the joint probability distribution of the N_i will be

$$P(\{N_i\}) \propto \prod_{i=1}^{q} \frac{\exp(q N_i^2 / 2\kappa V)}{N_i!} \qquad \left(\sum_i N_i = N \right) \tag{3.1}$$

Here κ is an interaction constant and V is a variable measuring the size of the system, proportional to N itself. We see then that this is exactly the situation of a distribution of N units over q independent but communicating replicas of a system. Indeed, as we see from comparison of (3.1) with (2.1), the system would be one with Poisson statistics, the statistics of (1.2) in the special case $H_j \equiv 1$. We know from Proposition 4 that the statistics of such a replicated system is identical with that in which the Poisson distribution of bonds is weighted by a factor $q^{B + C - N}$, where $B + C - N$ is the

cyclomatic number of the graph constituted by the whole configuration. The parameter q is then to be identified with our parameter v.

The random-cluster model implied in formula (3.1) is then the very simplest one, the Poisson model. By replacing this by our general polymer model we are replacing expression (3.1) by

$$P(\{N_i\}) \propto \prod_i A(N_i, V/q)$$

where A can take a rather general form—roughly, one for which (1.1) holds with R in $\mathscr{R}$. The other point is that we now consider nonintegral values v of q.

The connection of Potts models with polymer models has been discussed by Lubensky and Isaacson.[6] They make the point that the range $0 < v < 1$ is of particular interest (so that excess bonding is discouraged) and deduce some qualitative equivalences. Blöte et al.[2] consider the case of nonintegral v computationally. The present paper pursues these questions analytically for R in the general class $\mathscr{R}$ and, more explicitly, for the first-shell model.

As far as polymer models are concerned, the author has followed a line over a sequence of publications[10–14] which is continued here. The first two papers[10,11] set the Flory–Stockmayer models on a firm basis as Markov processes. These concerned the case $v = 0$, when all polymers were trees. However, the natural extension of this version to general v was perceived at the end of ref. 11 and partially achieved in ref. 12. A very much simpler and more powerful analysis was then found in ref. 13 and set out as part of a systematic exposition in ref. 14.

One should perhaps comment generally on the status of the Flory–Stockmayer theory, formulated initially in refs. 3, 8, and 9. As indicated above, this is a mean-field theory, and as such has been compared in ref. 7 with nearest-neighbor bonding models on a d-dimensional lattice, the so-called 'percolation' models of polymerization. A systematic difference in critical exponents is remarked for $d = 2$, which is, however, believed to vanish for d greater than some quite moderate value. The original Flory–Stockmayer model (which is the one considered in ref. 7) was one of irreversible aggregation, parametrized by the 'conversion factor' (the proportion of possible bonds which have been realized, corresponding to the 'probability of arc formation' in the simple mathematical models of random graphs) and in which cycle formation was impossible (implying restriction to the case $v = 0$). However, the model allows relaxation on all these points, as demonstrated in refs. 10–14. One can allow reversible dynamics and consider the consequent stochastic equilibrium; the model is then best

parametrized by the density of units ρ. One can allow the binding energy of a polymer to depend upon its configuration in almost arbitrary fashion, and there is no hindrance to cycle formation.

4. THE FUNCTIONAL EQUATION FOR $R_v(\rho)$

Note an implication of identity (2.6): that

$$\tilde{Q}_v(\theta, V) = \tilde{Q}_{v_1}(\theta, v_1 V/v)\, \tilde{Q}_{v_2}(\theta, v_2 V/v) \tag{4.1}$$

where $v = v_1 + v_2$. This implies in turn a corresponding relation for $R_v(\rho)$, as shown in the next proposition.

Proposition 5. If $v = v_1 + v_2$, then

$$R_v(\rho) = \sup[(v_1/v)\, R_{v_1}(\rho_1) + (v_2/v)\, R_{v_2}(\rho_2)] \tag{4.2}$$

where the supremum is over ρ_1 and ρ_2 subject to

$$(v_1/v)\, \rho_1 + (v_2/v)\, \rho_2 = \rho \tag{4.3}$$

Proof. It follows from identity (4.1) that

$$A_v(N, V) = \sum_{N_1, N_2} A_{v_1}(N_1, v_1 V/v)\, A_{v_2}(N_2, v_2 V/v)$$

where the summation is subject to $N_1 + N_2 = N$. Substituting the asymptotic form (2.7) with $\rho_i = (v N_i)/(v_i V)$ and equating the dominant terms on each side of the last equation, we deduce relations (4.2), (4.3). ∎

5. A CONJECTURED SOLUTION

We shall demonstrate in the next section that Eq. (4.2) is solved by

$$R_v(\rho) = \text{ext}\; v^{-1}[(v-1)\, R(\rho_1) + R(\rho_2)] \tag{5.1}$$

where ext denotes the taking of an extremum with respect to ρ_1 and ρ_2 subject to

$$\rho_1 < \rho_g, \qquad (v-1)\, \rho_1 + \rho_2 = v\rho \tag{5.2}$$

and the extremum is a supremum or an infimum according as $v > 1$ or $v < 1$. Of course, v takes values only in the range $v \geqslant 0$; if we speak of 'all v' we mean all v in this range.

We know already from (1.6) that formula (5.1) holds for v a positive integer. There are simple plausibility arguments which extend this result. However, these simple arguments become complicated when one tries to convert them into proofs, and the less direct approach of the next section turns out to be the economical and revealing one.

The course of the function $R_v(\rho)$ for varying ρ and v is illustrated in Fig. 2 and described more explicitly as follows.

Proposition 6. Suppose that $R \in \mathscr{R}_c$. Then the following assertions hold for the function R_v as specified by (5.1).

(i)   It belongs to $\mathscr{R}$ for all v.

(ii) It is nondecreasing in v.

(iii) It is the unique nondecreasing solution of the functional equation (4.2).

(iv) R_0 is the greatest concave minorant to R.

(v)   The Potts critical value v_P is greater than 1. If $v < v_P$, then $R_v(\rho) = R(\rho)$ for $\rho \leqslant \rho_g$ and it continues convexly with continuous first derivative for $\rho > \rho_g$.

(vi) The critical value of density ρ_{vg} equals ρ_g for $v \leqslant v_P$ and decreases continuously with increasing v thereafter. For $v > v_P$ the function
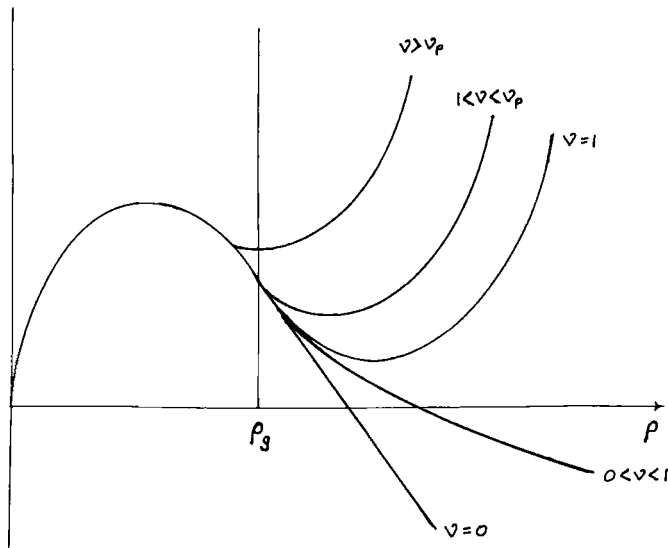

Fig. 2.   The graph of $R_v(\rho)$ for varying v.

$R_\nu(\rho)$ equals $R(\rho)$ for $\rho \leqslant \rho_{\nu g}$; as a function of ρ it shows a discontinuous increase in first derivative at $\rho_{\nu g}$ and then continues convexly.

We shall prove the various assertions in Section 6. In Section 7 we shall obtain some results on the determination of the Potts critical value ν_P and shall in fact find that $\nu_P = 2$ under rather general conditions. The case of two replicas that we used in Section 1 to determine ρ_g is then exactly on the verge of Potts criticality in most cases.

Assertion (iii) does not of course prove that expression (5.1) can be identified with the actual $R_\nu(\rho)$. There is the question of uniqueness to be settled, and we can easily see that some kind of condition is needed to ensure this (in fact, that R should have a continuous first differential). Suppose that we are given the function $R_\mu(\rho)$ for some integer μ exceeding ν_P. We know then that $R_\mu(\rho)$ and $R(\rho)$ are related by relation (5.1), with ν set equal to μ. However, we could never completely determine the function R in terms of R_μ from this relation. Suppose the maximizing value of ρ_2 in (1.4) for $\nu = \mu$ and $\rho = \rho_{\mu g}$ is ρ^*, necessarily a supercritical value. Then R_μ does not depend upon the value of $R(\rho)$ for $\rho_{\mu g} < \rho < \rho^*$; consequently, the form of $R(\rho)$ in this interval cannot be determined from the prescription of R_μ (apart from the fact that it must be consistent with the prescribed value of $\rho_{\nu g}$). So, should $R(\rho)$ itself show such behavior, then the value of $R_\nu(\rho)$ for nonintegral ν would not be completely determinable.

One assertion that we can conveniently and usefully prove now is assertion (iv): that R_0 [as given by (5.1)] is the greatest concave minorant to R. This is in itself an interesting observation, whose significance will be revealed in Section 9.

The following geometrical proof can easily be formalized. Suppose that ρ is given a prescribed value $\bar\rho$. We can write relation (5.2) in the form

$$\nu^{-1}(\rho_2 - \rho_1) = \bar\rho - \rho_1 \tag{5.3}$$

where we know that ρ_1 is subcritical. We have drawn the situation for a value of ν less than 1 in Fig. 3. One can say that the aim is to choose ρ_1 and ρ_2 consistent with (5.3), i.e., such that the length of the line segment AB is ν times that of AC, in such a way as to bring the point C as low as possible. In the limit as ν tends to zero, then, $\rho_2 - \rho_1$ must also tend to zero and the line ABC becomes tangent to the graph of $R(\rho)$. If $\bar\rho$ is less than ρ_g, then the height of C is minimized when the line is tangent at $\bar\rho$, so that $R_0(\bar\rho) = R(\bar\rho)$. If $\bar\rho$ is greater than ρ_g, then the height of C is minimized when the line is tangent to the graph at ρ_g. Thus the graph of $R_0(\rho)$ for $\rho > \rho_g$ is given by this fixed tangent line, and R_0 is indeed just the greatest concave minorant to R.
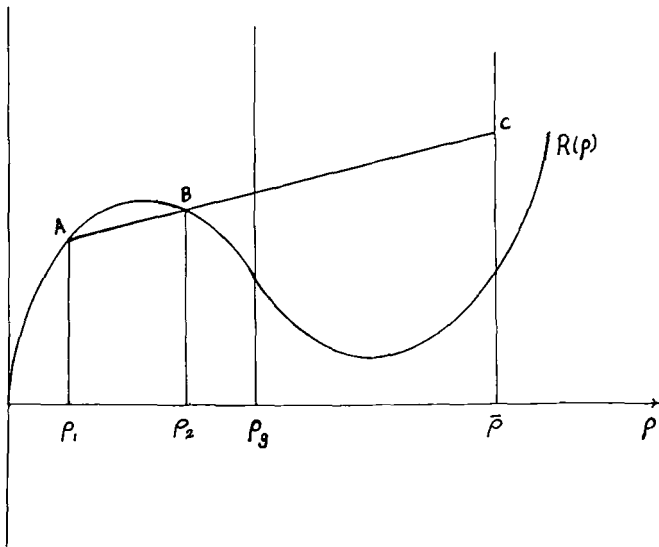
Fig. 3. A demonstration that $R_0(\rho)$, as determined by formula (5.1), is indeed the greatest concave minorant to $R(\rho)$.

6. Deduction of the solution and its properties

We shall find that the key to the whole treatment is the use of the Legendre transform (or contact transformation). The Legendre transform of $R(\rho)$ is defined as

$$\bar{R}(\gamma) = \underset{\rho}{\text{stat}} \ [R(\rho) - \gamma\rho] \qquad (6.1)$$

where by 'stat' we mean the evaluation of the bracket at a stationary point with respect to ρ; the conjugate variable γ is then necessarily the gradient of R at the stationary point. $\bar{R}(\gamma)$ is then in general a multivalued function of γ, in that there are several stationary points for a given value of γ. The function thus has several branches, which we may write as $\bar{R}_i(\gamma)(i = 1, 2,...)$, each generated as the tangent to the graph of R is rolled along a segment of R which is concave or convex, and so on which there is a unique tangent with a given slope γ. As this tangent line rolls over the graph of $R(\rho)$, the value (or values) of $\bar{R}(\gamma)$ is read off as the intercept on the vertical axis by a tangent of slope γ; see Fig. 4.

So, in our particular case the function R is first concave and then convex, and so $\bar{R}$ has two segments, generated as the tangent rolls over the
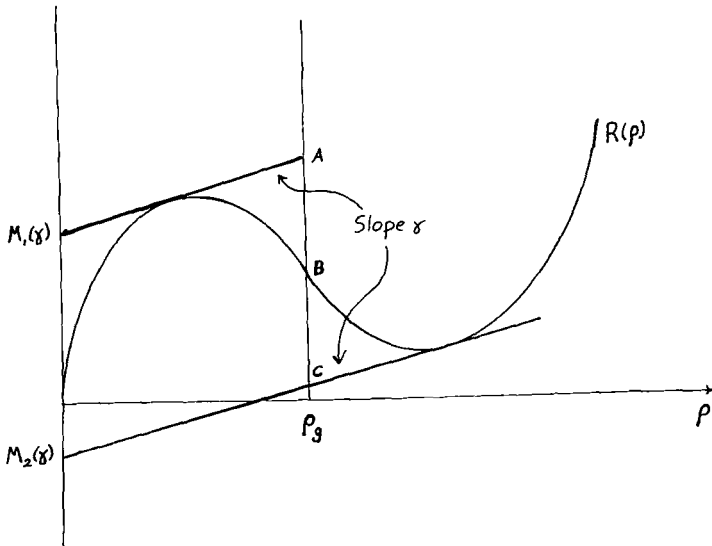
Fig. 4. The generation of the two branches $M_1(\gamma)$ and $M_2(\gamma)$ of the Legendre transform $\bar{R}(\gamma)$ of $R(\rho)$ by the rolling of a tangent of variable slope γ over the graph of $R(\rho)$. Also an indication of the additional property required later: that the ratio of the length of the line segment BC to that of AB should not decrease with increasing γ.
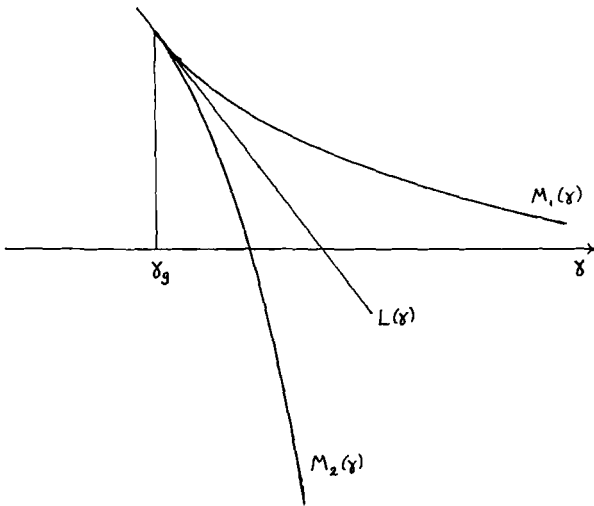

Fig. 5. The two branches, $M_1(\gamma)$ and $M_2(\gamma)$, of the Legendre transform $\bar{R}(\gamma)$ of $R(\rho)$. The line $L(\gamma)$ is their common tangent at the cusp point at which they meet.

concave section or over the convex section, respectively; see Figs. 4 and 5. These two branches of $\bar{R}(\gamma)$ are central to the ensuing argument, so we shall give them their own notation: $M_1(\gamma)$ and $M_2(\gamma)$, respectively. They are respectively convex and concave, and meet in a cusp at the critical value γ_g, the minimal slope $R'(\rho_g)$ of R, and both branches will be tangent at this point to a line $L(\gamma)$ with slope $-\rho_g$; see Fig. 5. At least this is the case if there are no discontinuities in the first derivative of $R(\rho)$. If there is such a discontinuity anywhere, then there will be a break in the corresponding branch of $\bar{R}(\gamma)$. In particular, if there is an upward jump in first derivative at ρ_g, from γ_g^- to γ_g^+, say, then there will be a gap in the M_2 branch; see Fig. 6.

The inverse of transform (6.1) is simply

$$R(\rho) = \operatorname*{stat}_{\gamma} \left[\bar{R}(\gamma) + \gamma\rho \right] \tag{6.2}$$

This inversion is single-valued: there is only one branch of $\bar{R}(\gamma)$ on which its derivative takes a prescribed value $-\rho$. In our particular case, as one rolls a tangent line with slope $-\rho$ decreasing from 0 to $-\infty$ over the graph of $\bar{R}(\gamma)$, one rolls first over the graph of M_1 (so generating the concave part of R by the intercept on the vertical axis) and then onto the graph of M_2 (so generating the convex part). In general, the case when the permitted range of ρ is limited needs special consideration. However, the point does not arise in our case: the restriction to the permitted set $\rho \geqslant 0$ is automati-
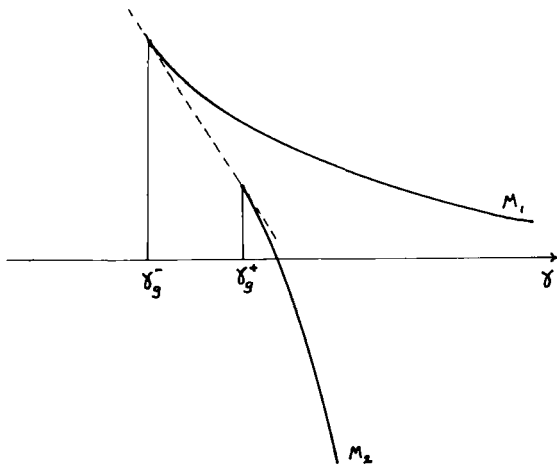

Fig. 6.   The two branches of $\bar{R}(\gamma)$ in the case that $R(\rho)$ has a discontinuity in derivative at the gelpoint.

cally achieved, thanks to the assumption that $R(\rho)$ has infinite derivative at the origin. This has the implication that, in the most probable distribution of matter over replicas, no replica is ever empty.

We come now to the assertion which is the key one as far as this paper is concerned.

Proposition 7. Suppose that functions A, B, and C of a variable σ are related by

$$A(\sigma) = \sup[B(\sigma_1) + C(\sigma_2)] \tag{6.3}$$

where the supremum is with respect to σ_1 and σ_2 subject to

$$\sigma_1 + \sigma_2 = \sigma \tag{6.4}$$

Then one can write

$$A(\sigma) = \sup_{j,k} \operatorname*{stat}_{\theta} \left[\bar{B}_j(\theta) + \bar{C}_k(\theta) + \theta\sigma \right] \tag{6.5}$$

Proof. Suppose we weaken the taking of a supremum in (6.3) to the taking of a stationary point, subject still to condition (6.4). Then one finds easily that $\operatorname{stat}[B(\sigma_1) + C(\sigma_2)]$ has, as a function of σ, the Legendre transform $\bar{B}(\gamma) + \bar{C}(\gamma)$. This has many branches, $\bar{A}_{jk}(\gamma) = \bar{B}_j(\gamma) + \bar{C}_k(\gamma)$. One now finds that the inversion (6.2) of the $\bar{A}$ thus determined is in general nonunique, because there will in general be many branches $\bar{A}_{jk}(\gamma)$ on which there is a point of prescribed gradient $-\sigma$. This nonuniqueness simply reflects the fact that, when we weakened the taking of a supremum in (6.3) to the taking of a stationary point, then we replaced one function of σ by a whole class. We must maximize over this class, which is exactly what evaluation (6.5) achieves. ∎

We are now in a position to deduce our principal conclusion.

Proposition 8. Suppose that $R \in \mathscr{R}_c$. Then $R_\nu(\rho)$ is uniquely determined for $\nu \geq 0$. Its Legendre transform $\bar{R}_\nu(\gamma)$ has just two branches: the upper branch given by $M_1(\gamma)$ and the lower by

$$M_{\nu 2}(\gamma) = \nu^{-1}[(\nu - 1) M_1(\gamma) + M_2(\gamma)] \tag{6.6}$$

The inverse Legendre transform of this determination of $\bar{R}_\nu(\gamma)$ is just the evaluation (5.1) asserted for $R_\nu(\rho)$.

Proof. Define

$$S_\nu(\sigma) = \nu R_\nu(\sigma/\nu)$$

In the case of integral v this amounts to working in terms of total probabilities and unit numbers for the set of v replicas rather than per replica. In terms of S the functional equation becomes

$$S_{v_1 + v_2}(\sigma) = \sup[S_{v_1}(\sigma_1) + S_{v_2}(\sigma_2)] \qquad (6.7)$$

where the supremum is with respect to σ_1 and σ_2 subject to (6.4). The assertion of the proposition would imply that $\bar{S}_v$ has the branches vM_1 and $(v-1)M_1 + M_2$. If this were the case, then, according to Eq. (6.5), $\bar{S}_{v_1 + v_2}$ has possible branches vM_1, $(v-1)M_1 + M_2$, and $(v-2)M_1 + 2M_2$, where $v = v_1 + v_2$. But this third branch is irrelevant. It competes with the second branch, in that both show the same range of gradients $(-\rho_g$ to $-\infty)$, but the third lies below the second, and so can never yield the maximizing option in (6.5). We thus have the two branches, consistent with the proposed solution, and confirm that the proposed solution satisfies the functional equation (6.7). Correspondingly, the solution R_v corresponding to the proposed evaluation of $\bar{R}_v$ solves (4.2). That the form asserted for $\bar{R}_v(\gamma)$ implies the form (5.1) asserted for $R_v(\rho)$ follows again by an application of Proposition 7.

Uniqueness can be proved if we can show that expression (5.1) provides the only possible solution of the functional equation (4.2) for $v = 1/2$. This is because it then gives only possible solution for $v = 2^{-n}$ for any positive integral n, and so for $v = m2^{-n}$ for any positive integral m, n. That is, the asserted and actual solutions agree on a set of v dense on the half-line. They then agree everywhere, since both are monotonic in v.

We shall prove the uniqueness of the solution for $v = 1/2$ by analysis of a particular case which plainly generalizes. Suppose that $\bar{R}_{1/2}$ consists of three branches $F_j(\gamma)$ $(j = 1, 2, 3)$. If numbered in order of decreasing gradient, they must then have gradients in contiguous intervals $[0, -\rho_1]$, $[-\rho_1, -\rho_2]$, and $[-\rho_2, -\infty]$ with $0 < \rho_1 < \rho_2 < \infty$. The branches M_1 and M_2 of $\bar{R}$ must then be obtained by considering the nine possible branches $\frac{1}{2}(F_j + F_k)$ and taking the options which yield maximal R, as in (6.5). The branches chosen must be such as to make $\bar{R}$ continuous and with continuous derivative (i.e., such that R has continuous derivative and that there is no range of ρ for which it is undetermined). As the gradient $-\rho$ decreases from zero, $M_1 = F_1$ is initially the only option which has numerically small enough gradient. The only way to continue this consistent with continuity of both function and derivative is along either F_2 or $\frac{1}{2}(F_1 + F_2)$. We know from the argument above that the second option is the maximizing one, so that $M_2 = \frac{1}{2}(F_1 + F_2)$. For agreement of this continuation with $\bar{R}$ we must then also have $\rho_1 = \rho_g$. But this evaluation of M_2 has minimal gradient $-\frac{1}{2}(\rho_1 + \rho_2)$, whereas the actual M_2 shows

indefinitely large negative gradients and there are no further cusps or branches. We must thus have $\rho_2 = +\infty$ and the branch F_3 absent, and so have the determinations

$$F_1 = M_1, \qquad F_2 = 2M_2 - M_1$$

which is exactly the solution asserted for $v = 1/2$. ∎

Note the necessity of the continuity of the first derivative of R for uniqueness of the determination of R_v. Had there been a discontinuity in derivative at ρ_g, say, then the branches M_1 and M_2 would have failed to meet, as in Fig. 6, and the argument which established the uniqueness of $\bar{R}_{1/2}$ would have failed.

In Fig. 7 we sketch the two branches of $\bar{R}(\gamma)$ for varying v and see that many of the conclusions of Proposition 6 follow immediately. The lower branch M_{v2} is increasing in v, which implies the nondecreasing character of R_v. The fact that one inverts by rolling the tangent line first under M_1 and then over M_2 (as ρ increases from 0 to $+\infty$) implies that R_v belongs to $\mathscr{R}$: first a concave section [equal initially at least to $R(\rho)$] and then a convex section.
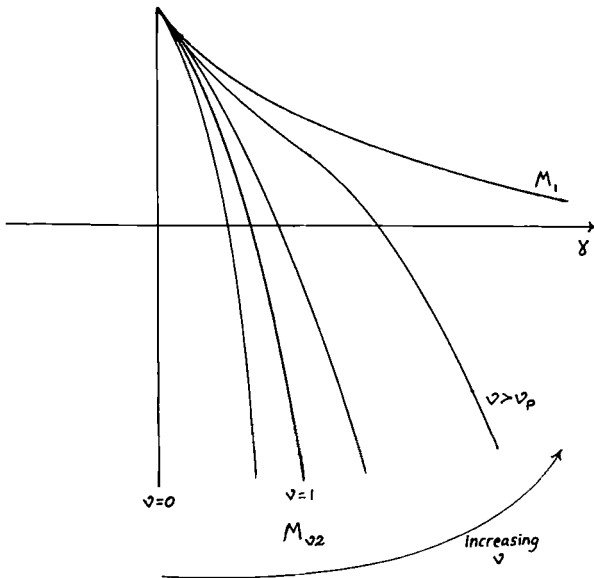

Fig. 7.   The graph of the branch $M_{v2}(\gamma)$ for varying v. The development of any convexity in this branch as v increases signals the onset of Potts criticality.
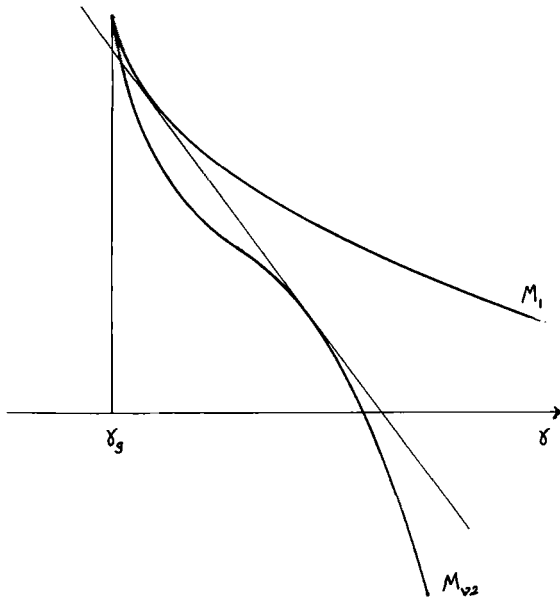
Fig. 8. A more explicit picture of a Potts supercritical case in terms of the Legendre transform.

As v tends to zero, then the lower branch tends to a vertical line, M_{02}. The transform $\bar{R}_0$ then has as Legendre inverse exactly the greatest concave minorant to $R(\rho)$, consistently with Proposition 6(iv).

However, the striking point is the transparent emergence and characterization of Potts criticality. As v increases from unity, the lower branch $M_{v2}(\gamma)$ remains concave to begin with, but then as more of M_1 is admixed, begins to develop a local convexity, presumably initially at γ_g. This marks the onset of Potts criticality. In Fig. 8 we have drawn a case somewhat beyond this point. As the tangent is rolled leftward under M_1 it meets M_{v2} (at slope $-\rho_{vg}$) before it has rolled as far as the cusp point. Transference to the lower branch takes place at this point already [since to stay on the upper branch would yield a smaller value of $R_v(\rho)$]. At this transference there is a jump in γ, and so in the derivative of $R_v(\rho)$.

7. EVALUATION OF THE CRITICAL VALUE OF v

One expects that when Potts criticality first occurs then it manifests itself at the gel point ρ_g if we consider the behavior of $R_v(\rho)$ and at the cusp coordinate γ_g if we consider the behavior of $\bar{R}_v(\gamma)$. That is, when v first exceeds v_p then this is manifested by the fact that $R_v(\rho)$ exceeds $R(\rho)$ in the

range $0 \leqslant \rho \leqslant \rho_g$ first at ρ_g and by the fact that the graph of $M_{v2}(\gamma)$ crosses the common tangent at the cusp first at a γ value arbitrarily close to γ_g. The first assertion is true under a condition no stronger than that R should lie in $\mathscr{R}_c$. The second implies that the most probable configuration [i.e., the extremizing values of ρ_1 and ρ_2 in (5.1)] at ρ_g should change continuously as v increases through v_P, and stronger conditions seem to be needed to ensure its validity.

Potts criticality first occurs when the graph of $M_{v2}(\gamma)$ meets a line which is tangent to the graph of M_1 at some value $\gamma_1 > \gamma_g$, meeting it at some value $\gamma > \gamma_1$. That is, when the inequality

$$M_{v2}(\gamma) < M_1(\gamma_1) + M_1'(\gamma_1)(\gamma - \gamma_1)$$

is first violated for some $\gamma > \gamma_1 > \gamma_g$.

Proposition 9. Suppose that $R \in \mathscr{R}_c$. Then ρ_{vg} decreases from ρ_g as v increases from v_P.

Proof. The assertion is evident geometrically, although the equivalent manipulative proof is easily supplied. Suppose that $v > v_P$, so that we can draw a common tangent to the graphs of M_1 and M_{v2} as in Fig. 8. The slope of this tangent line is just $-\rho_{vg}$. As v increases, then so does M_{v2}, and the gradient of the common tangent must evidently increase. That is, ρ_{vg} decreases. Furthermore, the tangent line to M_1 which will first be crossed by M_{v2} as v increases is that for $\gamma_1 = \gamma_g$, with slope $-\rho_g$. Hence ρ_{vg} decreases from ρ_g. ∎

This last-mentioned tangent line is the common tangent to the graphs of M_1 and M_2 at γ_g, illustrated in Fig. 5. It is the graph of the linear function $L(\gamma) = \alpha - \rho_g(\gamma - \gamma_g)$, where α is the common value of $M_j(\gamma_g)$ $(j = 1, 2)$ and $-\rho_g$ is the common value of $M_j'(\gamma_g)$. Define the function

$$D(\gamma) = \frac{L(\gamma) - M_2(\gamma)}{M_1(\gamma) - L(\gamma)}$$

This is nonnegative, and defined in the range $\gamma \geqslant \gamma_g$. As just asserted, Potts criticality manifests itself when $M_{v2}(\gamma)$ exceeds $L(\gamma)$ for some γ. We see from the form (6.6) of $M_{v2}(\gamma)$ that this is exactly when v first exceeds $1 + D(\gamma)$ for some γ. That is,

$$v_P = 1 + \inf_{\gamma} D(\gamma)$$

and the breakdown occurs at the infimizing value of γ. We thus immediately derive the following result.

Proposition 10. Suppose that $D(\gamma)$ is nondecreasing in γ. Then Potts criticality manifests itself first at γ_g, and

$$\nu_P = 1 + \lim_{\gamma \downarrow \gamma_g} D(\gamma) \tag{7.1}$$

Now, we can also express $D(\gamma)$ as

$$D(\gamma) = \frac{R(\rho_g) - R(\rho_2) - \gamma(\rho_g - \rho_2)}{R(\rho_1) + \gamma(\rho_g - \rho_1) - R(\rho_g)} \tag{7.2}$$

where ρ_1 and ρ_2 are the lower and upper roots of

$$R'(\rho) = \gamma \tag{7.3}$$

Expressed graphically, the requirement is that the ratio of the length of the segment BC to that of AB in Fig. 4 should not decrease as the common gradient γ of the two tangents to the graph of R increases from its minimal value γ_g. At the end of Section 1 we defined $\mathscr{R}^*$: the set of functions R which both belong to $\mathscr{R}_c$ and obey this latter condition. Part of the assertions of the last two propositions can then be translated into the following assertion in terms of $R(\rho)$.

Proposition 11. Suppose that $R \in \mathscr{R}^*$. Then ν_P is determined by (7.1), with $D(\gamma)$ determined by (7.2), (7.3), and the solution is continuous in that, as ρ increases from ρ_g at this value of ν, the values of ρ_1 and ρ_2 in solution (5.1) are just those yielded by (7.3) as γ increases from γ_g.

The condition that R should belong to $\mathscr{R}^*$ is a perfectly realistic and verifiable one. We shall see in the next section that it is satisfied for all first-shell models for which the integral (1.3) converges.

The evaluation (7.1) can be made much more specific under mild differentiability conditions on R.

Proposition 12. Suppose that $R \in \mathscr{R}^*$. If R also has a continuous nonzero third derivative in the neighborhood of ρ_g, then $\nu_P = 2$.

Proof. Denote the first and third derivatives of R at ρ_g by R' and R'''. Necessarily $R'' = 0$ and $R''' > 0$. Choose $\gamma = R' + \varepsilon$ for small, positive ε. Then (7.3) has the roots

$$\rho_1, \rho_2 = \rho_g \mp (6\varepsilon/R''')^{1/2} + o(\sqrt{\varepsilon})$$

whence it follows that $D(\gamma_g + \varepsilon) = 1 + o(1)$. Letting ε tend to zero, we deduce the assertion. ∎

The hypotheses of the proposition hold for the Poisson model, so implying the well-known result that $v_P = 2$ for this case. However, we now see how little dependent this conclusion is on the Poisson character of the model.

One can find examples for which v_P can take any value exceeding unity. Suppose that we do have a model for which $v_P = 2$, so that ρ_{v_g} decreases from ρ_g as v increases from 2. Take R_α as the R for a new process. It follows then that the critical value of v for the new process is $2/\alpha$ if $\alpha \leqslant 2$, unity otherwise.

However, there are less contrived examples. Suppose that R has a continuous first derivative at ρ_g but that the second derivative is discontinuous there, taking values R''_- and R''_+ at values of ρ immediately less and greater than ρ_g. These values must be respectively negative and positive. Evaluation (7.1) then reduces to

$$v_P = 1 - (R''_- / R''_+) \tag{7.4}$$

For an example of such a case, consider the 'cosh' model, a first-shell model for which

$$H(\xi) = \cosh(\xi) = \tfrac{1}{2}(e^\xi + e^{-\xi})$$

This has previously been considered by the author[12]; it can be regarded as a Poisson model with the added constraint that units may form only an even number of bonds. The critical density ρ_g has the same value κ as it does for the Poisson model. However, the model shows an additional degree of discontinuity: the maximizing value of ξ in (1.4) is zero for $\rho \leqslant \kappa$; the stationarity condition

$$\rho H'(\xi)/H(\xi) = \kappa\xi \tag{7.5}$$

has a nonzero solution first for $\rho > \kappa$. That is, there is essentially no bonding at all below criticality. Otherwise expressed, the sol phase consists essentially of monomers; only when large (gel) molecules are stable do any molecules form at all. Then R has a continuous first derivative but a discontinuous second derivative at criticality. One finds for the general first-shell model that

$$R'' = -\rho^{-1} + \kappa^{-1} \left[\frac{H'}{H}\right]^2 \left[1 + \xi\left(\frac{H'}{H} - \frac{H''}{H'}\right)\right]^{-1}$$

where H and its derivatives are evaluated at the appropriate root ξ of (7.5). For the cosh model this leads to the evaluations

$$R''_- = -\kappa^{-1}, \qquad R''_+ = \tfrac{1}{2}\kappa^{-1}$$

whence we deduce from (7.4) that $v_P = 3$. That is, the increased difficulty of bonding does not postpone the onset of gelation, but does postpone the onset of Potts criticality.

8. THE PARTICULAR CASE OF THE FIRST-SHELL MODEL

For the first-shell model (1.2) one readily finds that ρ and $R'(\rho)$ are related to the maximizing value of ξ in (1.4) by

$$\rho = \kappa\xi H/H', \qquad \gamma = R'(\rho) = \log(H/\rho) = \log(H'/\kappa\xi)$$

where H and its derivatives have the argument ξ. Suppose that branching is possible, so that $H_j > 0$ for some $j > 2$. There is then a gel point ρ_g which corresponds to the value of ξ minimizing H'/ξ, and so determined by

$$1 - \xi H''/H' = 0$$

Since H'/ξ is strictly convex on the positive half-axis, this equation has a unique solution there.

Regularity requirements imply one further condition. If expressions (1.3) or (1.4) are to be finite, then $\log H(\xi)$ must grow at a less than quadratic rate as ξ goes to $+\infty$. This ensures that configurations with an excessive number of bonds are improbable. One could also frame the condition by requiring that the turning points of the function $\rho \log H(\xi) - \frac{1}{2}\kappa\xi^2$ of ξ should be maxima, whatever ρ. This implies the condition

$$1 + \xi\left(\frac{H'}{H} - \frac{H''}{H'}\right) > 0 \tag{8.1}$$

Proposition 13. For the first-shell model the Legendre transform $\bar{R}(\gamma)$ of $R(\rho)$ has two branches; these are given by

$$M_j(\gamma) = \underset{\xi}{\text{ext}}\left[e^{-\gamma}H(\xi) - \frac{1}{2}\kappa\xi^2\right] \qquad (j = 1, 2) \tag{8.2}$$

where the extremum is a local maximum for $j = 1$ and a local minimum for $j = 2$.

Proof. From the expression (1.4) for $R(\rho)$ we derive

$$\bar{R}(\gamma) = \underset{\rho}{\text{stat}}\{\rho - \rho \log \rho + \underset{\xi}{\text{sup}}[\rho \log H(\xi) - \frac{1}{2}\kappa\xi^2] - \rho\gamma\} \tag{8.3}$$

The stationarity conditions with respect to ρ and ξ are

$$-\log \rho + \log H - \gamma = 0 \qquad (8.4)$$

$$\rho H'/H - \kappa \xi = 0 \qquad (8.5)$$

Substituting for ρ from relation (8.4) into relations (8.3) and (8.5), we deduce that $\bar{R}(\gamma)$ is given by the bracketed expression in (8.2) with the stationarity condition

$$e^{-\gamma} H'(\xi) = \kappa \xi \qquad (8.6)$$

Expression (8.2) thus follows, with the convex and concave segments of $\bar{R}(\gamma)$ presumably being given by the maximizing and minimizing options, respectively. However, while the maximizing option certainly yields a convex function of γ, we must appeal to condition (8.1) to establish that the minimizing option yields a concave function. One finds that the second derivative of the function defined by (8.2) is

$$M_{\gamma\gamma} = \left(\frac{\kappa \xi H}{H'}\right)\left(1 - \xi \frac{H''}{H'}\right)^{-1} \left[1 + \xi\left(\frac{H'}{H} - \frac{H''}{H'}\right)\right]$$

Inequality (8.1) implies that this expression has the sign of $(1 - \xi H''/H')$, whence the convexity/concavity of the two branches follows. ∎

Proposition 14. For the first-shell model $R(\rho)$ belongs to $\mathscr{R}^*$.

Proof. That it has the concavity/convexity and continuous-differential properties required of a function in $\mathscr{R}_c$ follows from the form of its Legendre transform $\bar{R}(\gamma)$. We require then only to establish that $D(\gamma)$ is nondecreasing. We can write

$$D(\gamma) = \frac{L - M_2}{M_1 - L} = -\frac{\phi(\xi_2)}{\phi(\xi_1)}$$

where we have used the relation (8.6) to express the quotient as a function of ξ instead of γ. The functions M_1 and M_2 will then have the same expression as functions of ξ, M_1 being the evaluation of that function for $\xi < \xi_g$ and M_2 for $\xi > \xi_g$. The values ξ_1 and ξ_2 are the lesser and greater roots of (8.6) for given γ.

We have then

$$\log D(\gamma) = \log[-\phi(\xi_2)/\phi(\xi_1)] = \int_{\xi_1}^{\xi_2} [\phi'(\xi)/|\phi(\xi)|] \, d\xi \qquad (8.7)$$

with

$$\phi(\xi) = \text{const} + (\kappa \xi H/H') - \tfrac{1}{2}\kappa \xi^2 + \rho_g \log(H'/\xi)$$

where the constant is such that this expression changes sign at ξ_g. One verifies directly that the differential of this last expression can be written

$$\phi'(\xi) = \frac{1}{\xi}\left(1 - \frac{\xi H''}{H'}\right)\left(\frac{\kappa \xi H}{H'} - \rho_g\right) = \xi^{-1}\left(\frac{1 - \xi H''}{H'}\right)(\rho - \rho_g) \geqslant 0$$

Here ρ is the value corresponding to ξ, whence it is evident that both brackets in the final expression change sign at ξ_g and the whole expression remains nonnegative. It follows than that the integrand in (8.7) is nonnegative, and hence (since ξ_1 and ξ_2 respectively decrease and increase as γ increases) that $D(\gamma)$ is an increasing function of γ. ∎

9. FURTHER IMPLICATIONS

The case $\nu = 0$ has an interesting and significant character. The density ρ_g is indeed a critical density for all ν, in that if we consider several (m, say) independent but communicating replicas, then equidistribution over them will cease to be metastable if $\rho > \rho_g$, whatever ν. It will cease to be absolutely stable for some $\rho < \rho_g$ if $m > \nu_P/\nu$; this then never occurs in the case $\nu = 0$.

The character of criticality is also different in this case. If $\nu > 0$, then $R_\nu(\rho)$ is strictly convex for $\rho > \rho_g$ and so for ρ in this range there is one supercritical replica complemented by $m - 1$ subcritical replicas with a common (but positive) density. One may say that the gel phase is concentrated in a single replica (i.e., there is only one gel molecule, or giant component) and there is matter in the sol phase, distributed uniformly over remaining replicas. In the case $\nu = 0$ matters are different.

Proposition 15. If $\rho > \rho_g$ and $\nu = 0$, then there are no subcritical replicas, and the distribution of units over replicas is indifferent in the sense that, if ρ_i is the density of units in the ith replica, then all configurations for which $\sum_i \rho_i = m\rho$ and $\rho_i \geqslant \rho_g$ for all i are equally probable. Otherwise expressed, the gel phase is distributed indifferently over replicas.

The 'indifferent' distribution is quite different from a 'uniform' distribution over replicas. In the latter case equidistribution over replicas constitutes an equilibrium which is at least metastable. In the former, *all* distributions of matter in the set indicated are equilibria, but neutrally stable. Below criticality there is no gel fraction and the sol fraction is dispersed uniformly over replicas, presumably by migration. Above criticality the gel fraction is dispersed over replicas, but not with the uniformity that the sol phase shows below criticality. Rather than one gel molecule (giant

component), there are infinitely many, and the surplus to a critical density stays where it is rather than migrates.

Proof. Since $R_0(\rho)$ is concave, a most probable distribution is that in which the distribution is uniform:

$$\sum_{i=1}^{m} R_0(\rho_i) \leqslant m R_0(\rho) \tag{9.1}$$

where $\rho = m^{-1} \sum_i \rho_i$. Suppose $\rho > \rho_g$. Then inequality (9.1) will be strict if any ρ_i is on the strictly concave part of R_0; i.e., in the subcritical range. The first assertion thus follows. On the other hand, equality will hold in (9.1) if all the ρ_i are in the range where R_0 is linear; i.e., in the supercritical range. ∎

The evaluation of R_0 as the greatest concave minorant of R is consistent with the analytic approach to the problem. Several points are raised in such an approach which we shall analyze elsewhere, but there are points which are worth making now heuristically.

We see from relations (2.4), (2.6) that

$$\tilde{Q}_v(\theta, V) = \exp\left[\sum_{s=0}^{\infty} v^s V^{1-s} G_s(\theta)\right]$$

where $G_s(\theta)$ is the sum of $\Gamma_r \theta^{n_r}$ over all r-mers with exactly s excess bonds, i.e., for which $b_r + 1 - n_r = s$. Thus $A_0(N, V)$ is the coefficient of θ^N in $\exp[V G_0(\theta)]$, which we can write as

$$A_0(N, V) = \frac{1}{2\pi i} \int e^{V G_0(\theta)} \theta^{-N-1} \, d\theta \tag{9.2}$$

where the path of integration is an appropriate circuit in the complex plane. A saddle-point evaluation of the integral leads to the evaluation

$$R_0(\rho) = \inf_{\theta}[G_0(\theta) - \rho \log \theta] \tag{9.3}$$

where the infimum is over real, nonnegative θ. This is a concave function of ρ, identifiable with $R(\rho)$ in the range $\rho < \rho_g$. Criticality is revealed by the fact that, as ρ approaches ρ_g, then θ approaches a branch point θ_g of $G_0(\theta)$. For $\rho > \rho_g$ the infimum in (9.3) must be regarded as attained at the fixed value $\theta = \theta_g$, and one has then in this range

$$R_0(\rho) = G_0(\theta_g) - \rho \log \theta_g = R(\rho_g) + R'(\rho_g)(\rho - \rho_g)$$

This is exactly the continuation of R into the supercritical region by its greatest concave minorant.

Finally, if the replicas we have been considering really are regarded as communicating compartments in space, then there is an inconsistency in the specification of the process if new effects (such as Potts criticality) can occur as one considers more compartments. For self-consistency of the compartmental version of the process as an embryonic spatial version one must require that R_v be independent of v. This is just a way of saying that the freeing of communication between compartments in space should not be able to achieve anything more than can be achieved by passage to the thermodynamic limit. This invariance can hold only if R is concave. For the first-shell model this can be the case only when 'excess bonding' (and so cycle formation) is forbidden.

However, such a demand is unnecessarily stringent. If we consider more general models than the first-shell model, then cycles can be allowed so long as there is some constraint of steric hindrance or the like which sets a sufficiently strong limit on the internal bonding pattern of a polymer molecule. If such constraints are sufficiently effective, then $R(\rho)$ will be concave and the specification spatially self-consistent.

REFERENCES

1. M. Aizenmann, J. T. Chayes, L. Chayes, and C. M. Newman, Discontinuity of the magnetisation in one-dimensional $1/|x - y|^2$ Ising and Potts models, *J. Stat. Phys.* **50**:1–40 (1988).
2. H. W. J. Blöte, M. P. Nightingale, and B. Derrida, Critical exponents of two-dimensional Potts and bond percolation models, *J. Phys. A* **14**:L45–L49 (1981).
3. P. J. Flory, *Principles of Polymer Chemistry* (Cornell University Press, Ithaca, New York, 1953).
4. I. J. Good, Cascade theory and the molecular weight averages of the sol fraction, *Proc. R. Soc. A* **272**:54–59 (1963).
5. P. W. Kasteleyn and C. M. Fortuin, Phase transitions in lattice systems with random local properties, *J. Phys. Soc. Jpn.* **26**(Suppl.): 11–14 (1969).
6. T. C. Lubensky and J. Isaacson, Field theory for the statistics of branched polymers, gelation and vulcanisation, *Phys. Rev. Lett.* **41**:829–832 (1978).
7. D. Stauffer, A. Coniglio, and M. Adam, Gelation and critical phenomena, in *Advances in Polymer Science* (Springer, 1982).
8. W. H. Stockmayer, Theory of molecular size distribution and gel formation in branched chain polymers, *J. Chem. Phys.* **11**:45–55 (1943).
9. W. H. Stockmayer, Theory of molecular size distribution and gel formation in branched polymers. II. General cross-linking, *J. Chem. Phys.* **12**:125–131 (1944).
10. P. Whittle, Statistical processes of aggregation and polymerisation, *Proc. Camb. Phil. Soc.* **61**:475–495 (1965).
11. P. Whittle, The equilibrium statistics of a clustering process in the uncondensed phase, *Proc. R. Soc. A* **285**:501–519 (1965).

12. P. Whittle, Polymerisation processes with intrapolymer bonding, Parts I, II and III, *Adv. Appl. Prob.* **12**:94–115, 116–134, 135–153 (1980).
13. P. Whittle, A direct derivation of the equilibrium distribution for a polymerisation process, *Teoriya Veroyatnostei* **26**:360–361 (1981).
14. P. Whittle, *Systems in Stochastic Equilibrium* (Wiley, New York, 1986).
15. F. Y. Wu, The Potts model, *Rev. Mod. Phys.* **54**:235–268 (1982).