

10

## CHAPTER 5

### *Towards a Rigorous Molecular Theory of Metastability\**

O. PENROSE

*The Open University, Milton Keynes  
Bucks, England*

JOEL L. LEBOWITZ

*Departments of Mathematics and Physics  
Rutgers University, New Brunswick, N.J., USA*

\*Supported in part by USAFOSR Grant no. 78-3522.

## *Contents*

1. Outline of chapter	295
2. Introduction	296
3. Three dynamical models	301
4. The restricted ensemble	303
5. How to define $R$	305
6. Estimation of the escape rate	307
7. Statistical mechanics of clusters	312
8. Clusters and metastability	319
9. Nucleation rates	322
10. Slowly-varying ensembles	328
11. Concluding discussion	333
Appendix 1: The Widom – Rowlinson model	335
Appendix 2: An upper bound on $Q_N$ for the plane square lattice	336
Appendix 3: A lower bound on $Q_N$ for the plane square lattice	337
Note added in proof	338
References	339

## 1. Outline of chapter

2. *Introduction*: An outline of the physical ideas used in theories of metastability and some difficulties in basing rigorous treatments on them: can the van der Waals–Maxwell and Bragg–Williams treatments be reconciled with exact equilibrium theory in which the free energy function must be convex? Can the droplet (nucleation) picture be freed from the lack of precision about basic definitions which makes it so hard to test accurately against real experimental data and computer simulations?
3. *Three dynamical models*: The importance of dynamics in the study of metastability: the Glauber and Kawasaki dynamical versions of the Ising model for ferromagnets and for gas–vapor or binary alloy systems, and the differences to be expected in their metastable behavior.
4. *The restricted ensemble*: Three criteria for characterizing metastable states; the idea of calculating the static (reversible) properties of these states from an ensemble in which the system is artificially restricted to a particular set  $R$  of dynamical configurations chosen in accordance with these criteria.
5. *How to define  $R$* : For systems of particles,  $R$  may be defined either by dividing space into imaginary cells and restricting the number of particles in each cell to a specified range of values, or by grouping the particles into clusters, defining sizes for the clusters, and restricting each cluster to a specified maximum size. For binary alloys and ferromagnets there are analogous definitions.
6. *Estimation of the escape rate*: An outline of the work of Penrose and Lebowitz on the van der Waals model of a fluid, and of Capocaccia, Cassandro and Olivieri on the Glauber model of a ferromagnet, showing that under suitable conditions a system started in a configuration in a suitably defined set  $R$  is unlikely to escape rapidly from  $R$ .
7. *Statistical mechanics of clusters*: A precise definition is given for the partition function of a cluster of given size in the lattice-gas version of the Ising model. Using this we can write down rigorous upper bounds on the partition function of the system and on the equilibrium density of clusters of given size. These upper bounds are also good low-density approximations.

The approximations provide a connection with standard nucleation theory for this model at sufficiently low vapor densities.

8. *Clusters and metastability*: The work of the preceding section is extended to the restricted ensemble and used to show that with Kawasaki dynamics the escape rate from  $R$  is very small under suitable conditions on temperature and density.
9. *Nucleation rates*: The Becker – Döring kinetic equations, describing the way cluster sizes change by absorption and evaporation of particles, are formulated and the approximations used in their derivation are discussed; formulae for calculating the coefficients in these equations from first principles are given, and the Becker – Döring derivation of a formula for nucleation rate is described.
10. *Steady-state ensembles*: Proposes a generalization of the Becker – Döring definition for nucleation rate which, unlike that definition, is not tied to any particular dynamical model or way of defining the set  $R$ .
11. *Concluding discussion*

## 2. Introduction

Statistical mechanics has well-defined “canonical” formalisms for obtaining (stable) equilibrium properties of macroscopic matter, and also, to a lesser extent, non-equilibrium properties of systems sufficiently close to stable equilibrium, from a knowledge of the interactions between the microscopic constituents of the material. It has however no such general formalism for metastability and nucleation; it has only a collection of ad hoc methods, most of them approximate, for particular problems. Recently there have been some attempts (Penrose and Lebowitz 1971, Capocaccia et al. 1974, Cassandro et al. 1976, 1977) to put the study of metastability on a more rigorous basis. It is the purpose of this article to explain and extend this work and to relate it to other rigorous as well as non-rigorous approaches to the problems of metastability and nucleation. Since a rigorous treatment of metastability is worth attempting only where the underlying phase transition can be treated rigorously, our discussion is confined to metastability associated with particular phase transitions, such as the liquid – vapor transition, ferromagnetism, and transitions in binary mixtures and alloys. We do not discuss metastability in glasses, superfluids or superconductors. The reader is referred to chapter 4 by Metiu et al. in this volume for a more phenomenological approach to many of the problems discussed here.

The main body of the article is in two parts. In the first we discuss the

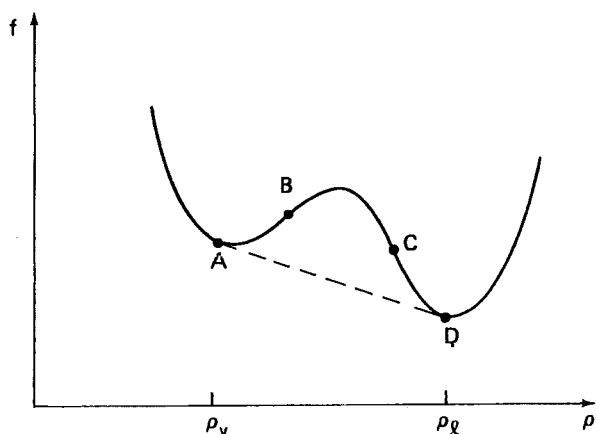


Fig. 1. Helmholtz free energy density in the van der Waals-Maxwell theory.

problem of showing rigorously that the system under study has a metastable state under suitable conditions, and of how to calculate the equilibrium (quasi-thermodynamic) properties of such a state; in the second we consider what can be said rigorously, as well as non-rigorously, about its non-equilibrium properties, such as the rate of nucleation. But before going into detail about rigor we outline some of the physical ideas and models that have been used in the theory of metastability.

The simplest and best-known approach appears in the van der Waals-Maxwell theory (Maxwell 1965) of the liquid-vapor transition. In this theory, an approximation which treats the system as a uniform fluid phase leads (at some suitable fixed temperature  $T$ ) to a relationship between  $f$ , the Helmholtz free energy per unit volume, and  $\rho$ , the density, whose graph has the general character of the solid curve shown in fig. 1.

This figure also shows a dashed straight line (the Gibbs double-tangent construction) touching the curve at two points A and D. The points on this line correspond to two-phase states, one phase (the vapor) having density  $\rho_v$ , the other (the liquid) having density  $\rho_l$ . Maxwell saw that for densities between  $\rho_v$  and  $\rho_l$  the stable equilibrium states would be the two-phase states, but thought that, by careful experimentation, it should also be possible to realize experimentally the parts of the curve labeled AB and CD; these describe the one-phase metastable states. The part of the curve between the inflection points B and C is impossible to realize because it has  $d^2f/d\rho^2 < 0$ , so that the states it describes are mechanically unstable.

The same free energy curve also appears in the mean-field theory of

binary alloys (Bragg and Williams 1934, Fowler and Guggenheim 1939, Yamauchi and de Fontaine 1974). In this case the horizontal axis is the fraction of sites occupied by one of the two types of atoms composing the alloy. The states on AB and CD are again metastable, but this time the states on BC are not mechanically unstable; the theory, due to Hillert (1961) and Cahn (1961, 1962), does, however, predict that they are unstable against the process of "spinodal decomposition" in which the two types of atoms diffuse in such a way that they ultimately separate themselves into small aggregations of the two different phases. Further developments of this type of theory are described by Langer et al. (1975) and in the article by Metiu et al. in chapter 4 of this volume. Thus, both for fluids and for alloys, if we accept the van der Waals expression for the free energy and Maxwell's kinetic interpretation of it, then we know precisely for what values of density (or concentration) and temperature the system can exist in a one-phase metastable state.

The loci of  $\rho_v$  and  $\rho_l$  in a density - temperature diagram are called coexistence curves, and the loci of the densities corresponding to the points B and C in fig. 1 are called the spinodal curves (see fig. 2). In metallurgy the name "spinodal" is given to the corresponding curves in a concentration - temperature diagram, and the region enclosed between them may be called the spinodal region, since it represents the set of temperatures and concentrations for which the theory predicts that spinodal decomposition will take place. (In some materials the spinodal curve is thought not to touch the coexistence curve.)

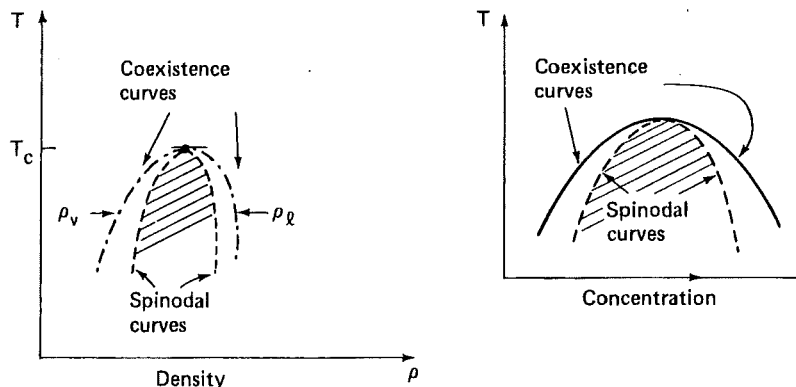


Fig. 2. Spinodal curves (dashed) and spinodal regions (shaded), according to van der Waals - Maxwell or Bragg - Williams theory. The metastable regions lie between the spinodal curves and the coexistence curves.

Although these approximate mean-field theories give metastable states easily, it is more difficult to see how such states can arise in an exact theory; for it follows from the general principles of statistical mechanics that the exact free energy density of any large enough system, calculated from the partition function, is a convex function of  $\rho$  and therefore cannot include arcs such as AB and CD, which do not lie on the same convex curve. To obtain the metastable states, some extension of the usual formalism of equilibrium statistical mechanics is necessary going beyond the normal methods based on the partition function.

One possibility, suggested by the preceding example, is that the thermodynamic functions for the metastable states can be obtained by extrapolation from the nearby stable one-phase states, so that, for example, the arc AB in fig. 1 would be obtained by extrapolation from the curve to the left of A. It is likely, however, that for real systems, for which van der Waals' theory is only an approximation, a singularity of the thermodynamic functions blocks the extrapolation (Fisher 1967, Langer 1967, Lanford and Ruelle 1969); and even when the extrapolation is possible, one still has to justify the assumption that the extrapolated thermodynamic functions really describe metastable states. (Recently Newman and Schulman (1977) considered alternative methods of analytic continuation, using the eigenvalues of the transfer matrix, from the equilibrium phase to a metastable one. Their analysis is still in a preliminary stage and we shall not comment on it further here; see however McCraw and Schulman (1978).)

The difficulty with these approaches is that they try to treat metastability by purely static methods, whereas it is really a dynamic phenomenon. Some of the basic ideas underlying this dynamics are already contained in Maxwell's own discussion. Maxwell (1965) recognized the importance of nucleation; he saw that to set up the metastable state we must be sure that none of the new phase is present. The idea of nucleation was developed further by Volmer and Weber (1926) and Becker and Döring (Becker and Döring 1935, Döring 1937) who used quasi-thermodynamic arguments, involving the surface tension, to find the smallest "liquid droplet" in a supersaturated vapor that will grow to become part of the new liquid phase. For droplets smaller than this minimum size, the "free energy" increases with size and hence they will tend to shrink. Using the Einstein relation between probability and free energy and some kinetic arguments about the rate of collisions between molecules and droplets which might lead to an increase in the size of the latter, this theory gives an estimate for the probability per unit volume of forming such a critical droplet in the homogeneous supersaturated phase. (More details are given later in this

article). This probability is found to be extremely small for some values of the temperature and pressure, which explains the metastability of the supercooled vapor. The basic ideas of this theory have been used with great success by many authors for a wide variety of phenomena: see, for example, Frenkel (1955), Langer and Fisher (1967), Langer and Ambegaokar (1967), Katz (1970), Skripov (1974).

Theories of this type do not give, like the van der Waals–Maxwell theory, a mathematically sharp separation (the spinodal line) between the metastable and unstable regions of the phase diagram since the probability of nucleating the new phase varies continuously as a function of the thermodynamic parameters of the system. Nevertheless, the transition from a vanishingly small nucleation rate per unit volume to an enormously large one occurs over such a small range of the parameters (e.g. temperature and pressure) that these theories lead to a fairly well-defined metastable region. The abruptness of the change in the probability of nucleating the new phase has the effect that in practice any reasonably-sized macroscopic system will either remain entirely in one phase, such as a supercooled vapor, or will quickly form many droplets of the other phase. The further development of these droplets will then be affected by various mechanisms, such as gravity, hydrodynamic effects, etc., which are not allowed for in the usual droplet theory.

Despite its successes, however, the classical droplet theory does not answer all our questions. It does not contain a precise definition of a droplet or of the free energy of a droplet, a central quantity in the theory; and the kinetic equations governing the growth and decay of the droplets are derived as if the droplets were spherical liquid drops of macroscopic size. That this lack of precision is not a purely aesthetic defect can be seen from the controversy, initiated by the work of Lothe and Pound (1962), about a large factor in the theory of the spontaneous nucleation rate of supersaturated vapor. While more recent work on nucleation theory (Binder and Müller-Krumbhaar 1974) avoids some of these oversimplifications, it is still very far from being a precisely defined theory.

The situation becomes even more complicated when one tries to use the theory to interpret computer simulations of the time evolution of a binary alloy quenched into the coexistence region (Bortz et al. 1974, Marro et al. 1975, Rao et al. 1976, Sur et al. 1977, Binder et al. 1979, Kalos et al. 1978). Here one observes events on a “molecular” time-scale, and therefore it is not always possible to observe them for long enough to decide whether the system should be said to be in a metastable or an unstable state. Consequently there is a need for a microscopic theory which deals directly with the quantities observed in these computer “experiments”.



### 3. Three dynamical models

One of the points which becomes clear as soon as we look into the microscopic theory of nucleation is the important part played by the dynamics of the system in determining its metastability properties. Most work on the statistical mechanics of metastability has been done on three dynamical models. One is the normal interacting particle model used in discussing the liquid-vapor transition. (A variant of this is the binary fluid mixture model of Widom and Rowlinson (1971), for which a proof of metastability was given recently by Cassandro and Olivieri (1977). This model is described in Appendix 1.) The other two are dynamical versions of the Ising model which we now describe briefly.

The Ising model (see, for example, Thompson 1972) is defined by considering some regular lattice at each site of which there is a tiny magnet (usually called a "spin") which can point either up or down. The spins interact with the vertical component of the external magnetic field. In the cases we shall consider there is a ferromagnetic interaction between neighboring spins, i.e. the parallel alignment of a pair of spins on nearest-neighbor lattice sites gives a lower energy than the anti-parallel alignment. Below the critical temperature  $T_c$  this interaction produces, in zero magnetic field, a spontaneous magnetization (alignment of the spins) which may point either up or down. If the magnetization is constrained to take a value less than the spontaneous value, the spins segregate into two phases, in one of which most of them point up, and in the other most point down.

Glauber's (1963) dynamical version of the Ising model consists of permitting the spins to reverse their directions in a random way. The probabilities of these spin reversals are chosen in accordance with the detailed balancing condition. To express this condition mathematically, suppose that a reversal of the  $j$ th spin converts the configuration  $\xi_1$ , with energy  $U_1$ , to a configuration  $\xi_2$  with energy  $U_2$ , and that  $W(\xi_2, \xi_1)$  denotes the conditional probability per unit time of a transition to the new configuration  $\xi_2$ , given that the old configuration is  $\xi_1$ ; then the detailed balancing condition is

$$W(\xi_2, \xi_1)e^{-U_1/kT} = W(\xi_1, \xi_2)e^{-U_2/kT}.$$

For simplicity it is usually assumed that, at a given temperature,  $W(\xi_2, \xi_1)$  depends only on the energy change  $\Delta U = U_2 - U_1$ ; this energy change is determined by the configuration of the spins on the lattice sites adjacent to  $j$ . Under this assumption the detailed balancing condition becomes

$$W(\Delta U) = e^{-\Delta U/2kT} w(\Delta U)$$

where  $w(\Delta U)$  is an even function. In the absence of specific information about microscopic transition rates we might conveniently choose  $w(\Delta U)$  to be a constant (although this "constant" would of course depend on temperature, perhaps as  $\exp(-E_0/kT)$  where  $E_0$  really is a constant). Another possible choice, used by Glauber, is  $w(\Delta U) = \text{const}/\cosh(\Delta U/2kT)$ .

The Ising model can also be used as a model of a binary alloy consisting of atoms of two types, A and B, when the "up" spins are identified with A atoms and "down" spins are identified with B atoms (see, for example, Hill 1956). For this interpretation the Glauber dynamics is inappropriate since it would not conserve particle number. The simplest appropriate dynamical assumption, first studied seriously by Kawasaki (1966, 1972), is that the spin configuration changes by random interchanges of the spins (atoms) on neighboring sites. As in the Glauber model the probability of an interchange that increases the energy by  $\Delta U$  is taken to be proportional to  $\exp[-\Delta U/2kT]$  times an even function of  $\Delta U$ . This is the third of the dynamical models we shall consider in this article.

As a variant of this alloy model, we may consider the "lattice gas", for which instead of B atoms we have unoccupied lattice sites. The rigorous study of the equilibrium properties of this model, based on its equivalence to the Ising model, was started by Lee and Yang (1952), although the model was used much earlier by Becker and Döring (1935) in a discussion of the process of crystallization. The dynamics appropriate to the lattice gas are the analogue of Kawasaki dynamics: the configuration changes by random steps in which an atom moves from some site to a neighboring empty site, with probability equal to  $\exp[-\Delta U/2kT]$  times an even function of  $\Delta U$ , as before.

The importance of dynamics in the study of metastability can be seen by comparing the Glauber and Kawasaki versions of the Ising model. Since every configuration has the same energy in both models, their (stable) equilibrium thermodynamic properties are the same, up to a reinterpretation of the relevant variables, even though the natural equilibrium ensembles for the two models are equivalent only in the thermodynamic limit (the appropriate ensemble for the Glauber model corresponds to a specified magnetic field, whereas that for the Kawasaki model corresponds to a specified concentration—the analogue of magnetization). But their non-equilibrium properties, including those related to metastability and spinodal decomposition, are profoundly different, because the local magnetization in the Glauber model can change rapidly, whereas the local concentration in the Kawasaki model can only change relatively slowly (as the result of diffusion of A or B atoms from another locality). For example,

when a ferromagnet is suddenly brought (by quenching) from a one-phase part of its magnetization-temperature diagram into the two-phase part, the nuclei of the new phase, once formed, can grow rapidly as spins reverse just outside the boundary of the nucleus; but if the analogous quenching is carried out for a binary alloy or a liquid mixture, the nuclei will grow much more slowly, since the atoms they require for growth must diffuse from other parts of the system. One may say that the alloy or mixture is exhibiting spinodal decomposition, but that the ferromagnet is not. Striking experimental demonstrations of such phase segregation in liquid mixtures have been given recently (Huang et al. 1974, Schwartz et al. 1975, Huang and Goldburg 1976, Wong and Knobler 1977).

The behavior of a (chemically pure) liquid-vapor system when quenched is likely to be different again, because although matter is conserved, as in an alloy, it can be transferred from one place to another rapidly by kinetic and hydrodynamic flow, instead of slowly by diffusion.

#### 4. The restricted ensemble

To begin our discussion of rigorous results related to metastability in these models, we list three properties (from Penrose and Lebowitz 1971) which, we believe, characterize metastable thermodynamic states:

(i) Only one thermodynamic phase is present, although the intensive thermodynamic parameters have values such that the equilibrium state would consist of more than one phase, or possibly a single but different phase. Under small slow changes of the thermodynamic parameters, the metastable state responds with small reversible changes obeying the usual laws of thermodynamics; but under large or rapid changes of the parameters the system may respond with a large irreversible change which takes it right out of the metastable state.

(ii) If the system is isolated the metastable state decays very slowly as a result of fluctuations which, if large enough, can nucleate the growth of a new phase; but these fluctuations are so improbable that the rate of decay is very slow, i.e. there is a high probability that a system in a metastable state at time  $t = 0$  will still be in that state at  $t = \tau$ , where  $\tau$  can be very large (say years).

(iii) Escape from the metastable state is an irreversible process: once the nuclei of the new phase are sufficiently large or numerous, the probability that they will disappear again is negligibly small.

A complete theory of metastability should describe both the static (reversible) properties relevant to criterion (i) and the dynamic or irreversible properties relevant to criteria (ii) and (iii). That is, it should

combine the equilibrium-like view of the metastable state coming from theories like that of van der Waals and Maxwell with the kinetic view coming from nucleation theory. In this and the next two sections of this chapter we concentrate on the static or quasi-equilibrium view of the metastable state; then in the later sections we turn to the dynamic or non-equilibrium view.

The main idea of the static formalism is to consider a modified system in which nucleation is prevented by a restriction or constraint, not present in the real system, on the motion of its representative point in phase space. The properties of this modified system when it is in (stable) equilibrium are then used as an approximation to those of the real system. Our main guide in deciding on a suitable form for the restriction is the first of the three criteria for a metastable state: that only one thermodynamic phase should be present. That is, the restriction should prevent any large nuclei of the new phase forming anywhere in the metastable phase. In the case of supercooled vapor, for example, the restriction would be contrived so as to keep the density from becoming too large in any locality, so that no nucleating droplets of liquid phase could form. In the case of an alloy, the restriction should keep the local composition from deviating too much from the overall composition; a thought-experiment using semi-permeable membranes which can do just this is discussed by Reiss (1975).

The effect of such a restriction would be to confine the system (or, strictly speaking, its representative point) to a particular region of configuration space, which we shall call  $R$ . (We shall discuss later how to specify  $R$ .) The restriction corresponds to adding to the Hamiltonian  $H$  a potential energy term which takes the value 0 for configurations in  $R$  and the value  $+\infty$  for configurations outside  $R$ . Then the equilibrium properties of the modified system can be calculated from a *restricted equilibrium ensemble* with probabilities given by

$$p(\xi) = \begin{cases} \text{const } e^{-H(\xi)/kT} & \text{if the configuration } \xi \text{ is in } R \\ 0 & \text{if not} \end{cases} \quad (4.1)$$

for all points  $\xi$  in phase space.

We have already mentioned that for this ensemble to give a good representation of the static properties of the metastable state it is necessary for  $R$  to be chosen so that no large nuclei of the "wrong" phase are present if the configuration is inside  $R$ . This choice meets our first criterion for a metastable state. To meet our second criterion, that the metastable state decays very slowly, we want to ensure that a system started somewhere in  $R$  is likely to take a long time to escape. The probability that it will have

escaped after a time  $t$  can in principle be calculated by studying the time evolution of the restricted ensemble under the action of the unmodified Hamiltonian  $H$ , which permits nucleation and will therefore move some of the systems of the ensemble from inside  $R$  to outside.

Let us define  $P(t)$  as the probability that a system started in the restricted ensemble (4.1) at time 0 will have escaped from  $R$  by time  $t$  (with  $t > 0$ ). Then we would expect that, after any initial transients had died out, the rate of increase  $\lambda(t) = dP(t)/dt$  will be equal to the rate at which systems escape from the metastable state  $R$  by nucleation. A simple argument due to Penrose and Lebowitz (1971, referred to below as PL) shows that  $\lambda(t) \leq \lambda(0)$ ; and hence if we can show that the initial escape rate  $\lambda(0)$  is small, the actual rate of decay of the metastable state described by the restricted ensemble (4.1) is also small. Some methods of showing that  $\lambda(0)$  is small for particular cases will be developed later.

Finally, to satisfy the third criterion, that the escape from  $R$  should be an irreversible process, it is (at least) necessary that the probability, in the true equilibrium ensemble, for finding the configuration in  $R$  be negligibly small. For the cases that have been considered in detail this necessary condition is always satisfied because the free energy of the restricted ensemble is less than that of the unrestricted by an amount proportional to the size of the system, so that the (true equilibrium) probability of a configuration in  $R$  goes to zero exponentially as the system is made larger and larger.

## 5. How to define $R$

In this section we describe in more detail some of the methods that have been used to define regions  $R$  in configuration space to represent metastable states in the model systems described earlier.

One such method is the one used by PL in the discussion of the liquid-vapor transition for a system with very long range attractive interactions—the van der Waals system. There, the volume occupied by the system is partitioned into imaginary cubical cells, and  $R$  is defined to consist of all configurations for which the number of particles in every cell is within a specified range of values. This range of values must, of course, include the value corresponding to a uniform distribution of particles, but excludes the value corresponding to the cell's being completely filled with the wrong phase. Moreover, the cell should not be so large that the particles in it can separate into two macroscopically distinguishable phases.

For an Ising system—ferromagnet, alloy, or lattice gas—the corresponding method of defining  $R$  would be to divide the lattice into cells and require that the number of up spins, or A atoms, or occupied sites within each cell

should lie within a specified range of values. That is to say,  $R$  would consist of all phase-space points or configurations  $\xi$  such that  $n_i(\xi)$ , the number of up spins or A atoms or occupied sites in the  $i$ th cell, lies, for every  $i$ , between two specified integers,  $n_-$  and  $n_+$ . In symbols, this definition can be written

$$R = \{\xi: n_- \leq n_i(\xi) \leq n_+ \quad (i = 1, 2, \dots, M)\} \quad (5.1)$$

where  $M$  is the number of cells.

This method of defining  $R$  is related to the discussion of the spinodal decomposition of an alloy given by Reiss (1975). He introduces into the system two sets of imaginary membranes, one permeable only to A atoms and the other only to B atoms. As long as the membranes do not move, no atom (of either kind) can cross any membrane, and so the membranes divide the volume occupied by the system into cells within each of which the number of each type of atom is restricted to a particular value. This is a particular case of the type of constraint defined in eq. (5.1), with both  $n_-$  and  $n_+$  equal to the number of particles required to be in each cell. Reiss also considers the effect of moving the membranes, but this does not seem to be necessary as long as we are discussing metastability rather than spinodal decomposition.

For alloys or real fluids, where the range of the interaction potential responsible for the phase transition is comparable with the interparticle separation in the system, the PL cell description is less satisfactory than for very long-range interactions. This is because the cell size has to satisfy two incompatible requirements. On the one hand, it should be large enough to permit the use of macroscopic concepts, such as the free energies of the bulk liquid and gas phases, to estimate its properties; on the other hand, it should be small enough to counteract the tendency for the particles inside each cell to form two phases. This incompatibility leads to great technical problems in proving that our second criterion of metastability is satisfied. These difficulties in the way of extending the results of PL to realistic systems have not yet been overcome in general. Such an extension has only been made so far for lattice models and for the Widom - Rowlinson model; in both cases the results are restricted to temperatures well away from the critical temperature and to densities (of up spins, or A atoms, or occupied sites) close to those at coexistence.

The first treatment of such a model was given by Capocaccia, Cassandro and Olivieri (1974) (referred to below as CCO), who studied metastability in the Ising model of a ferromagnet with Glauber dynamics, using a definition of  $R$  somewhat different from the one in (5.1). Their definition is particularly well adapted to lattice systems but, as indicated, only yields useful results for a very limited range of the thermodynamic parameters. Taking the

applied magnetic field to be in the down direction, they defined  $R$  to consist of all configurations for which all the down spins could be grouped into connected sets, which we call clusters, separated by up spins, in such a way that all the clusters were less than some specified size. By "connected" we mean that no member of one cluster is a nearest neighbor of a member of another, but that if a cluster is divided into two subsets then at least one member of one subset is a nearest member of the other. By the "size" of a cluster we mean the number of sites belonging to the cluster or completely enclosed by it. As with the cell method of defining  $R$ , we must not make the specified maximum size too large, for too large a cluster would be essentially the same as a macroscopic part of the system magnetized in the "down" direction and so would violate our single-phase criterion.

An analogous definition of  $R$  is possible for an alloy where the concentration of one type of atom is small enough, though larger than at coexistence: we just interpret the "down" spins as atoms of the less numerous kind and the "up" spins as atoms of the more numerous kind. The true equilibrium state at this concentration would have two phases, one rich in each type of atom. One of these phases would correspond, in the cluster description, to a single very large cluster, and the other would consist of a large number of small clusters.

For lattice systems, the definition of  $R$  in terms of clusters has the advantage of using only one arbitrary quantity, the size of the largest allowed cluster, instead of the three ( $n_-$ ,  $n_+$ , and the cell size) used in the PL method. A disadvantage is that (in three or more dimensions) the presence of very large clusters is not necessarily a sign that a second phase is present. It is found, for example, that if a concentration of more than 31% of "down" spins is arranged completely at random (as if at infinite temperature) on a simple cubic lattice then an *infinite* cluster will almost certainly be present, even though at infinite temperature there is no question of phase separation (Sykes and Essam 1964, Essam 1972). However, at low enough densities this difficulty does not arise. A possible way to remedy this situation for higher densities, by changing the method of defining a cluster, has been discussed recently by Binder and Stauffer (1976). We shall not comment on their analysis here except to note that whatever the definition of clusters the study of metastability by means of the restricted ensemble (5.1) can in principle still be carried out by a suitable definition of  $R$ .

## 6. Estimation of the escape rate

Having seen how to define a region  $R$  in configuration space consistently

with our first criterion of metastability, that only one phase should be present, we now want to check whether it is consistent with the second and third criteria. These are that the rate of escape from  $R$  should be slow, and that escape should be an irreversible process. In this section we outline the work of PL and CCO which, for the particular models they considered, shows that the escape rates are small and makes it plausible that escape is irreversible.

The model studied by PL is the van der Waals model of a fluid, with an interaction potential that is the sum of a short-range mainly repulsive part  $q(r)$  and a weak long-range mainly attractive part  $\gamma^v \phi(\gamma r)$ , where  $v$  is the number of dimensions (so that  $v = 3$  for real systems) and  $\gamma$  is a small parameter equal to the inverse range of the attractive potential. It was shown by Lebowitz and Penrose (1966) that in the limit  $\gamma \rightarrow 0$  the equilibrium behavior of such a system includes a phase transition of the type discovered by van der Waals and Maxwell. The proof involves dividing the system into cells whose size is much smaller than the long range of the attractive force  $\phi$ , and yet much larger than the short range of the repulsive force  $q$  or of the interparticle separation; these two conditions are compatible only in the "van der Waals limit" in which  $1/\gamma$ , which is the range of the attractive force  $\phi$ , becomes very large (i.e.  $1/\gamma \rightarrow \infty$ ).

To study the metastable states of this system we again divide the volume  $\Omega$  occupied by the system into  $M$  cells  $\omega_1, \dots, \omega_M$  and define  $n_i$  to be the number of particles in the cell  $\omega_i$ . We may take  $\Omega$  and the  $\omega_i$  to be cubes of volume  $|\Omega|$  and  $|\omega|$  respectively, where  $|\Omega| = M|\omega|$ . Then, as we have seen, we can take the region of restriction  $R$  to be the part of configuration space for which

$$n_- \leq n_i \leq n_+$$

with  $n_-$  and  $n_+$  suitably chosen integers. It is shown in PL that the initial escape rate  $\lambda(0)$  is the product of two factors: a "probability factor", giving the probability that the number of particles in some cell is at the edge of the allowed range, i.e. that

$$n_i = n_- \quad \text{or} \quad n_i = n_+ \quad \text{for some } i \quad (6.1)$$

and a "kinetic factor" giving the (conditional) probability per unit time that  $n_i$  will then go beyond the allowed range if the constraint is lifted. This product, when multiplied also by  $dt$ , gives the probability that a system of the ensemble is near the edge of the region  $R$  in configuration space at the moment the constraint is lifted and that it will cross the boundary of  $R$  in the first  $dt$  of time thereafter. The first of the two factors, the probability factor, has an upper bound of the form

$$\text{prob} \{n_i = n_- \quad \text{or} \quad n_i = n_+\} \leq \exp \{-C|\omega| + o(|\omega|)\} \quad (6.2)$$



where  $C$  is a constant,  $|\omega|$  means the volume of one cell, and  $o(|\omega|)$  means a quantity that increases less rapidly than  $|\omega|$  as  $|\omega|$  becomes large. The important thing is that (provided certain conditions are satisfied)  $C$  can be chosen positive (Penrose and Lebowitz 1971, Millard and Lund 1973) if the mean density  $\rho$  and the two local density bounds  $n_-/|\omega|$  and  $n_+/|\omega|$  are on the same *locally* convex part of the solid curve in fig. 1. This curve is the graph of the function  $f(\rho) + \frac{1}{2}\alpha\rho^2$  where  $f_0$  is the free energy density when the long-range part of the potential is set equal to zero, and  $\alpha$  is the space integral of this long-range part,

$$\alpha = \gamma^v \int \phi(\gamma r) d^v r = \int \phi(s) d^v s. \quad (6.3)$$

The sign of  $\alpha$  is negative, since this potential is attractive. The condition for the solid curve in fig. 1 to be convex is  $f_0''(\alpha) + \alpha > 0$ , which is Maxwell's stability condition and is equivalent to saying that the compressibility calculated from the free energy density  $f_0(\rho) + \frac{1}{2}\alpha\rho^2$  is positive.

The probability estimate (6.2) tells us that if the density  $\rho$  satisfies Maxwell's stability condition

$$f_0''(\rho) + \alpha > 0$$

then we can make the probability factor as small as we please by making the box size  $|\omega|$  large enough. This condition on  $|\omega|$ , however, affects the choice of the size of the system and the range of the long-range potential, because the estimate (6.2) is only valid when  $|\Omega|^{1/v}$ , the linear dimension of the system, is much greater than  $\gamma^{-1}$ , the range of the long-range part of the potential, and  $\gamma^{-1}$  in turn is much greater than  $|\omega|^{1/v}$ , the linear dimension of one of the cells. These conditions of validity can be combined in the formula

$$|\Omega|^{1/v} \gg \gamma^{-1} \gg |\omega|^{1/v}. \quad (6.4)$$

The second factor in the escape rate, the kinetic factor, is shown in PL to have a bound of the form  $\text{const} |\Omega||\omega|^{-1/v}$  and hence the escape rate as a whole has a bound of the form

$$\lambda \leq \text{const} |\Omega||\omega|^{-1/v} \exp \{-C|\omega| + o(|\omega|)\}. \quad (6.5)$$

From this we see that the escape rate per unit volume,  $\lambda/|\Omega|$ , can be made as small as we please by choosing  $|\omega|$  large enough provided always that we are allowed at the same time to choose the range  $\gamma^{-1}$  of the long-range potential even larger and the size  $|\Omega|^{1/v}$  of the container larger still. Making  $\gamma$  very small in this way is in perfect harmony with the spirit of van der Waals' equation, which becomes exact only in the limit where  $\gamma$  is vanishingly small.

The result that  $\lambda/|\Omega|$  is small tells us that the rate of escape from the metastable state per unit volume is small, but not that the escape rate itself is small. Since the escape rate  $\lambda$  is proportional to  $\Omega$ , the volume of the system, [see eq. (6.5)] one would not normally expect to find  $\lambda$  itself small when the system is very large; but in this particular model it is possible to make  $\lambda$  as small as we please by taking the thermodynamic limit (the limit in which the size of the system becomes very large) in a peculiar way, with the range of the interaction becoming large at the same time as the size of the system (instead of afterwards as in the normal derivation of the generalized van der Waals equation). Specifically, we take this limit in such a way that

$$|\Omega| \gg \gamma^{-\nu} \gg |\omega| \gg \ln |\Omega| \quad (6.6)$$

(which means that  $|\Omega|/\gamma^{-\nu} \rightarrow \infty$ ,  $\gamma^{-\nu}/|\omega| \rightarrow \infty$ , etc.). This condition ensures that (6.4) is satisfied, and at the same time ensures, by (6.5), that the escape rate  $\lambda$  itself tends to 0 in the limit, so that by suitable (finite) choices of  $|\Omega|$ ,  $\gamma$  and  $|\omega|$  we can make not only  $\lambda/|\Omega|$  but  $\lambda$  as small as we please.

Whichever way the limit is taken, we have finally to check our third criterion, that of low return probability. As explained in sect. 4, it is necessary (but perhaps not sufficient) to choose  $R$  so that the restricted free energy density, which is  $f_0(\rho) + \frac{1}{2}\alpha\rho^2$ , exceeds the true free energy density obtained from the Gibbs double tangent construction illustrated in fig. 1, which is the convex envelope (CE) of the function  $f_0(\rho) + \frac{1}{2}\alpha\rho^2$ . This condition is satisfied for all  $\rho$  corresponding to two-phase equilibrium states in the van der Waals–Maxwell theory. We conclude that Maxwell's conditions for metastability in a vapor–liquid system

$$\begin{aligned} f_0(\rho) + \frac{1}{2}\alpha\rho^2 &> \text{CE} \{f_0(\rho) + \frac{1}{2}\alpha\rho^2\} \\ f_0''(\rho) + \alpha &> 0 \end{aligned}$$

are precisely those required by the present theory when applied to a phase transition caused by a very long-range very weak attractive potential.

Many of the same ideas appear in the treatments of metastability given by CCO for the Ising model with Glauber dynamics (1974) and with Kawasaki dynamics (1976). We describe here their method for Glauber dynamics; the extension of these results to Kawasaki dynamics will be outlined in sect. 8. As we have mentioned already in the preceding section, they consider a ferromagnet in a downward magnetic field, so that at equilibrium most of the spins point downwards, but define  $R$  to be a set of configurations in which downward spins are relatively rare, so that when they are grouped into connected clusters all the clusters are less than a specified size. Confining their discussion to a two-dimensional system, they take as their measure of size the area enclosed by the contour forming the

outer boundary of the cluster. This area is equal to the number of sites occupied by the cluster plus the number completely enclosed by it. They denote this maximum area by  $c^2$ . Just as in the model treated by PL, the initial escape rate can be expressed as the product of a probability factor and a kinetic factor. The probability factor is the probability, calculated in the restricted ensemble, that the system contains a cluster, or several clusters close together, which can be changed into a single cluster bigger than  $c^2$  by the reversal of just one spin; and the kinetic factor is

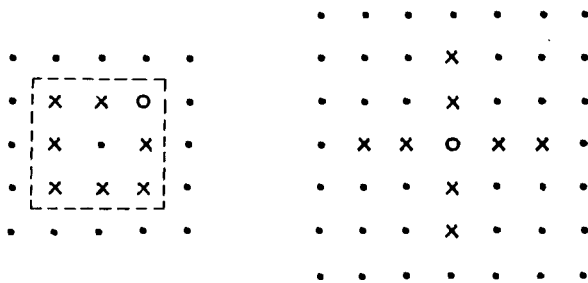


Fig. 3. If  $c^2 = 8$ , and down spins are represented by crosses, then the reversal of either of the encircled spins produces a cluster of "size" exceeding  $c^2$ . In the left-hand case, the actual number of down spins in the cluster does not exceed 8 even after the reversal, but the "size" of the new cluster is 9 because the area enclosed by the dashed contour is 9.

the probability per unit time that this particular spin will be reversed. CCO show that if

$$c \leq 4(J - kT \ln 3)/|h|$$

where  $h$  is the applied magnetic field, then the probability factor has an upper bound which is essentially of the form

$$\text{const } |A|^{3^{4c}} \exp[-4Jc/kT + |h|c^2/kT]$$

where  $A$  denotes the lattice and  $|A|$  the number of sites in it. The exponent is a bound on the energy of the "critical" cluster or set of clusters and  $|A|^{3^{4c}}$  is a bound on the number of such clusters. The kinetic factor is at most  $W_M$ , defined as an upper bound on the probability per unit time that a given spin will reverse its direction.

Provided the temperature is low enough [ $T < J/(k \ln 3)$ ], and  $h$  is small enough, we can use these bounds to show that the initial escape rate is very

small. For example, if we choose

$$c = 2(J - kT \ln 3)/|h|$$

the upper bound on the initial escape rate per lattice site is

$$\lambda/|A| \leq W_M \exp[-4(J - kT \ln 3)^2/|h|kT] \quad (6.7)$$

which goes to zero rapidly as the applied field  $|h|$  goes to zero. CCO also show that this model satisfies our other criteria of metastability and so we may conclude that, according to the static picture of metastability, this two-dimensional model of a ferromagnet has metastable states at sufficiently small magnetic fields for all temperatures below about  $0.2T_c$ .

The fact that we are able to prove rigorous results about metastability here only by considering a limiting case ( $\gamma \rightarrow 0$ , i.e. infinitely long range forces, in the van der Waals fluid;  $h \rightarrow 0$  in the ferromagnet) is probably not accidental. A necessary condition for a metastable state is that there should be two very different relaxation times: a short one for the establishment of metastable equilibrium, and a long one for the decay out of it. In the models we have considered we can make the ratio of these relaxation times as large as we like by a suitable choice of parameters such as  $\gamma$  or  $h$ . In a real system, however, we cannot always choose the physical parameters to suit our mathematical convenience; so although the ratio of the two relaxation times may be very large under suitable conditions it is not possible to make this ratio arbitrarily large, and there is no way of completely separating the process of establishment of metastable equilibrium from the process of decay out of it.

## 7. Statistical mechanics of clusters

The rigorous treatments described in the preceding sections are based on somewhat different concepts from those in the usual "droplet" theory of nucleation due to Becker and Döring (1935) and others (see Abraham 1974). One of the most important concepts in nucleation theory is the idea of the free energy of a droplet or nucleus of the new phase, which is used to estimate the expected number of such nuclei as a function of their size when the system is in equilibrium or close to it. To establish a connection between this concept and the concepts in the rigorous theory, we develop in this section some formulae and inequalities for the expected number of clusters of various types in a lattice system at equilibrium. We shall find that a fairly close connection can in fact be set up.

Since the droplet theory, as usually formulated, applies to a saturated

vapor, we give our discussion in a language appropriate to a lattice gas, rather than to a ferromagnet; this makes the connection with a real vapor-liquid system a little clearer. By a lattice gas we mean a system of particles which occupy some of the sites of a regular lattice, at most one particle being permitted at each site. Each configuration of an Ising system can be interpreted as a lattice gas configuration, by regarding the up spins as unoccupied sites and the down spins as occupied sites. The connected clusters of down spins considered by CCO can then be regarded as connected clusters of occupied sites (i.e. of particles of the lattice gas).

In the correspondence between the Ising ferromagnet and the lattice gas, the canonical ensemble for the ferromagnet at a given temperature and magnetic field corresponds to the grand canonical ensemble for the lattice gas at the same temperature and a chemical potential (or fugacity) which is related to the temperature and magnetic field in a simple way (Lee and Yang 1952). We shall use this grand canonical ensemble, where only the average rather than the exact number of particles in the system is specified, *only* for the computation of static properties; the dynamics of the lattice gas will always be taken to be particle-conserving. The legitimacy of using the grand canonical ensemble in this way is discussed by PL and CCO.

The grand partition function of the lattice gas, at fugacity  $z$  and reciprocal temperature  $\beta$ , is

$$\Xi(z) = \sum_{\xi} z^{N(\xi)} e^{-\beta U(\xi)} \quad (7.1)$$

where the sum goes over all possible configurations  $\xi$  of occupied and unoccupied sites on the given lattice,  $N(\xi)$  denotes the number of occupied sites in the configuration  $\xi$ , and  $U(\xi)$  denotes the potential energy of that configuration. The potential energy is given by

$$U(\xi) = -U_0 p(\xi) \quad (7.2)$$

where  $U_0$  is a positive constant and  $p(\xi)$  is the number of pairs of occupied sites that are nearest neighbors.

For each configuration we can partition the set of occupied sites into connected subsets, which we shall call clusters. By "connected" we mean (as in sect. 5) that no member of one cluster is a nearest neighbor of a member of another, but that if a cluster is subdivided into two subsets then at least one member of one subset is a nearest neighbor of a member of the other subset.

At sufficiently low values of the fugacity, i.e. at sufficiently low densities, the clusters will be fairly widely separated and we can, following Minlos and Sinai (1968), think of the system as a "gas" of clusters with an "interaction"

which prevents two clusters overlapping or touching (i.e. using neighboring sites). To express this idea mathematically we write  $N(\xi)$  and  $U(\xi)$  as sums of contributions from each cluster; that is

$$N(\xi) = \sum_i N(c_i), \quad U(\xi) = \sum_i U(c_i) \quad (7.3)$$

where  $c_1, c_2, \dots$  are the clusters constituting  $\xi$ ,  $N(c_i)$  is the number of particles in the  $i$ th cluster and  $U(c_i)$  its energy. Then we can rewrite the grand partition function as

$$\mathcal{E} = 1 + \sum_{\{c\}} \prod_i z^{N(c_i)} e^{-\beta U(c_i)} \quad (7.4)$$

where  $\{c\}$  denotes a possible arrangement of clusters  $c_1, c_2, \dots$  (i.e. one in which no two clusters overlap or use neighboring sites). The term 1 corresponds to the "arrangement" in which no clusters at all are present.

An approximation to the partition function (7.4) which is good at low densities, and gives an upper bound at all densities, can be obtained by including in the sum additional terms corresponding to all the impossible (overlapping or touching) arrangements of clusters. This gives an upper bound which can be written

$$\mathcal{E} \leq \prod_c (1 + z^{N(c)} e^{-\beta U(c)}) \quad (7.5)$$

where the product is over all possible single clusters. The corresponding upper bound on the grand canonical thermodynamic potential is

$$\begin{aligned} \log \mathcal{E} &\leq \sum_c \log(1 + z^{N(c)} e^{-\beta U(c)}) \\ &\leq \sum_c z^{N(c)} e^{-\beta U(c)}. \end{aligned} \quad (7.6)$$

In the formula (7.6) we can group together the terms corresponding to clusters of the same size, obtaining

$$\log \mathcal{E} \leq \sum_{N=1}^{|A|} z^N Z_N(A) \quad (7.7)$$

where  $|A|$  is the total number of lattice sites (analogous to the volume in a real system) and

$$Z_N(A) = \sum_{c: N(c)=N} e^{-\beta U(c)} \quad (7.8)$$

the sum going over all clusters containing  $N$  sites.

If  $|A|$  is large enough in comparison with  $N$ , then each cluster in the sum (7.8) is congruent under translations to about  $|A|$  other clusters. (This is

exactly so, for all  $N$ , if we use periodic boundary conditions on  $A$ .) Since all these congruent clusters have the same energy  $U(c)$ , the sum in (7.8) is approximately proportional to  $|A|$ , and in fact it is not hard to show rigorously that the following limit exists

$$\begin{aligned} Q_N &= \lim_{(\text{large } A)} (1/|A|) \sum_{c: N(c)=N} e^{-\beta U(c)} \\ &= \sum'_{c: N(c)=N} e^{-\beta U(c)} \end{aligned} \quad (7.9)$$

where the sum  $\sum'$  includes just one representative from each set of translationally equivalent clusters (for example, we might count only clusters whose "center of mass" falls in the unit cell containing the origin). We can think of  $Q_N$  as the internal partition function for clusters of size  $N$  (i.e. the partition function for non-translational degrees of freedom).

If the lattice  $A$  is rectangular in shape, and  $2A$  denotes a new lattice, also rectangular in shape, made by putting two copies of  $A$  next to each other, then the definition (7.8) implies  $Z_N(2A) \geq 2Z_N(A)$ . It follows from this that the limit in (7.9) is approached from below if a sequence  $A, 2A, 4A, \dots$ , is used in taking the limit, and hence that

$$Z_N(A) \leq |A| Q_N \quad (7.10)$$

for any rectangular-shaped lattice. (For periodic boundary conditions (7.10) is an equality.)

If  $z$  is small enough for the series

$$\sum_{N=1}^{\infty} z^N Q_N \quad (7.11)$$

to converge, then this series gives an upper bound on  $|A|^{-1}$  times the series in (7.7), and hence also on the thermodynamic limit of the grand canonical potential per lattice site, so that we obtain

$$\lim_{(\text{large } A)} (1/|A|) \log \Xi \leq \sum_{N=1}^{\infty} z^N Q_N \quad (7.12)$$

and this upper bound can be shown to give a good approximation to  $\log \Xi$  if the activity  $z$  is small enough. The right side of (7.12) has been used by authors such as Fisher (1967) for discussing the thermodynamics of the critical point, but the justification for using it as an approximation at such high values of the activity must rest on different grounds from those discussed here.

The convergence of the series  $\sum z^N Q_N$  can indeed be established for small enough  $z$ . For a plane square lattice,  $Q_N$  has (see Appendix 2) the upper

bound

$$Q_N \leq \frac{9}{8} \exp \{ (2 \log 3 + 2\beta U_0)N - 2\beta U_0 \sqrt{N} \}$$

so that the series (7.11) certainly converges if

$$z \leq \frac{1}{9} e^{-2\beta U_0}.$$

For a simple cubic lattice, the corresponding upper bound on  $Q_N$  is

$$Q_N \leq \frac{31}{20} \exp \{ (4 \log 3 + 3\beta U_0)N - 3\beta U_0 N^{2/3} \}$$

and so the series (7.11) certainly converges if

$$z \leq \frac{1}{81} e^{-3\beta U_0}.$$

The methods that gave us the upper bound (7.5) on the grand partition function also give us an upper bound on the probability  $P(c)$  that the cluster  $c$  is present in the system (Lebowitz and Penrose 1977). This upper bound is

$$P(c) \leq z^{N(c)} e^{-\beta U(c)}. \quad (7.13)$$

The formula (7.13) can be derived by noting that for every configuration with cluster  $c$  present, the configuration obtained by removing this cluster, while leaving the rest untouched, has a probability that is greater by the factor  $(z^{N(c)} e^{-\beta U(c)})^{-1}$ . In addition there are further configurations with cluster  $c$  absent, in which other clusters overlap or touch the sites used in  $c$ . Consequently the probability of not finding cluster  $c$  is more than  $(z^{N(c)} e^{-\beta U(c)})^{-1}$  times the probability of finding it; since these two probabilities add to 1, the probability of finding cluster  $c$  is less than  $1/[1 + (z^{N(c)} e^{-\beta U(c)})^{-1}]$ , which in turn is less than  $z^{N(c)} e^{-\beta U(c)}$ . This completes the derivation of (7.13).

From (7.13), (7.8) and (7.10) it follows that  $m_N$ , the expected number of  $N$ -particle clusters per unit volume in the system, has the upper bound

$$\begin{aligned} m_N &= (1/|\Lambda|) \sum_{c: N(c)=N} P(c) \leq (1/|\Lambda|) z^N Z_N(\Lambda) \\ &\leq z^N Q_N. \end{aligned} \quad (7.14)$$

It is also possible (Lebowitz and Penrose 1977) to derive lower bounds on  $m_N$  having the general form

$$m_N \geq z^N Q_N (1 + O(z)). \quad (7.15)$$

From this it follows that at very low densities the approximation

$$m_N \sim z^N Q_N \quad (7.16)$$

is a good one.



Sur et al. (1977) have carried out computer simulations from which concentrations of clusters can be estimated. The results of these simulations have been analyzed by Kalos et al. (1978). They found that the distribution of sizes of small clusters at equilibrium and also at low supersaturations, when the system appeared to be in a metastable state, could be represented by an empirical formula

$$\begin{aligned} m_1 &\simeq (1 - \rho)^3 w \\ m_N &\simeq (1 - \rho)^4 w^N Q_N, \quad (l \geq 2). \end{aligned}$$

Here  $\rho$  is the density (in lattice gas language) or fractional concentration of minority phase (in alloy language) and  $w$  is an adjustable parameter. The empirical formula is consistent with (7.16) since the latter refers only to very small values of  $\rho$ .

As we have mentioned, one can interpret  $Q_N$  as the internal partition function of a single droplet containing  $N$  particles. If we define an internal free energy  $F_N$ , related to this internal partition function by the usual formula  $Q_N = \exp[-\beta F_N]$ , we can write (7.16) in the form

$$m_N \simeq z^N \exp(-\beta F_N) = \exp[-\beta(F_N - NkT \log z)] \quad (\text{small } z). \quad (7.17)$$

We can look on  $F_N$  as the increase in the non-translational part of the free energy (i.e. the contribution from the internal partition function) that would be brought about by taking  $N$  single-particle droplets out of the system and putting in one  $N$ -particle droplet.

A reasonable first guess at the dependence of  $F_N$  on  $N$  for large  $N$ , in  $\nu$  dimensions, is given by the formula

$$F_N \simeq Nf + N^{(\nu-1)/\nu} s \quad (7.18)$$

where  $f$  is the equilibrium bulk free energy per particle for the dense phase and  $s$  is proportional to the surface tension. Our formulae are consistent with this approximation, at least in the case  $\nu = 2$  where for the plane square lattice we have (see Appendix 3) the bounds

$$F_N > -2(kT \log 3 + U_0)N + 2U_0 N^{1/2} - kT \log 9/8$$

and

$$F_N < -2U_0 N + 2U_0 N^{1/2} + U_0.$$

The dominant terms in these bounds are both of the form (7.15), suggesting that the approximation (7.18) may be valid, with a value for  $f$  somewhere between  $-2U_0$  and  $-2U_0 - 2kT \log 3$ , and with  $s = 2U_0$ . In particular at  $T = 0$  we are compelled to choose  $f = -2U_0$  and  $s = 2U_0$ , and these are

indeed the correct zero-temperature values for the free energy per particle in the dense phase and for the surface tension.

However at higher temperatures expressions of the form

$$F_N \simeq Nf + N^\sigma s$$

with  $\sigma$  decreasing to zero as the temperature  $T$  increases to the critical temperature  $T_c$ , give (Kalos et al. 1978) a better fit to the few values of  $F_N$  that can be calculated exactly (i.e. those for  $N \leq 10$ , in the case of the single cubic lattice). Expressions of the form first suggested by Fisher (1967)

$$F_N = Nf + N^\sigma s - \tau \ln N$$

with  $\sigma$  now independent of temperature ( $\sigma = 0.64$  for the simple cubic lattice) and  $s$  decreasing to zero as  $T$  increases to  $T_c$ , give an even better fit. Recently Fisher and Cagenalp (1977) have rigorously justified a formula of the type (7.18) for the thermodynamic free energy of an  $N$ -particle system on a lattice with a number of sites that increases in proportion to  $N$ , but there is no rigorous proof that any of the large- $N$  approximations we have discussed are valid for the cluster free energies  $F_N$ .

In nucleation theory the formula normally used is (Reiss 1975)

$$m_N \simeq \Gamma m_1 \exp(-W_N/kT). \quad (7.19)$$

In this formula the quantity  $W_N$  is defined by

$$W_N = N(\mu_d - \mu_v) + N^{2/3}s \quad (7.20)$$

with  $\mu_d$  and  $\mu_v$  the chemical potentials of the droplets and the vapor, and is an approximation to the reversible work required to form an  $N$ -particle cluster; and  $\Gamma$  is the controversial "replacement factor" introduced by Lothe and Pound (1962), to allow for differences of definition between the partition function of a cluster or droplet on the one hand and a portion of bulk liquid on the other. For a lattice system, the natural identifications are

$$\mu_d = f, \quad \mu_v = kT \log z.$$

If we also take the replacement factor to be

$$\Gamma = m^{-1} (\simeq z^{-1})$$

then for a lattice system eq. (7.19) is precisely the same as the simplest approximation suggested by the rigorous theory, namely the formula

$$m_N \simeq z^N \exp -\beta(Nf + N^{2/3}s)$$

obtained by combining the rigorously justified approximation corresponding to (7.17) with the reasonable first guess (7.18). For a continuous system,

of course, the rigorous theory is not so well developed, and so there is much more scope for controversy in the discussion of the replacement factor. A very neat exact formula for this factor is, however, given by Reiss (1974).

### 8. Clusters and metastability

In this section we show how to use the information about the equilibrium concentrations of various types of clusters, obtained in the preceding section, to find an upper bound on the rate of escape from a suitably defined set of configurations  $R$ . While following the main ideas of CCO's method we deviate from it in detail, considering Kawasaki rather than Glauber dynamics since it is to particle-conserving dynamics that the droplet model is normally applied, and defining  $R$  in a different way from CCO so that we can more conveniently use results from the preceding section. CCO defined their restricted ensemble by requiring all clusters to encompass at most some specified number  $c^2$  of sites; but we shall require the number of sites constituting the cluster itself to be bounded, by some number  $u$ . The difference is that if a cluster encloses some empty sites, or even some smaller clusters (see fig. 3), we do not reckon these in the "size" of the cluster, whereas CCO do. Under the conditions we are considering the change is numerically unimportant, since very few clusters enclose empty sites, but it simplifies the correspondence with standard nucleation theory, in which the "size" of a cluster or droplet is measured by the number of particles in it rather than by its volume.

As we have seen in earlier sections the main problem in showing that the given restricted ensemble describes a metastable state is to estimate the rate of escape. For the restricted ensemble we are discussing, escape is defined to take place if any cluster consisting of more than  $u$  particles forms. One way for this to happen is for a cluster containing exactly  $u$  particles to acquire one or more further particles. With particle-conserving dynamics, this acquisition can only happen if there is another cluster near by, and at low densities this other cluster very probably consists of just one particle. Alternatively two clusters, each comprising more than one but less than  $u$  particles, may join up to make one cluster of more than  $u$  particles; and there are also processes where three or more clusters join together. The rate of escape  $\lambda$  will be the sum of the contributions due to all these different processes.

To illustrate how these contributions may be estimated, we consider the first of the processes just mentioned, in which a  $u$ -particle cluster combines with a one-particle cluster. Its contribution to the total escape rate will be

denoted here by  $\lambda(u, 1)$ . As in the treatments by PL and CCO, described in sect. 6, we can write  $\lambda(u, 1)$  as the product of a kinetic factor and a probability factor. The probability factor is the probability (in the restricted ensemble) that the configuration includes a  $u$ -particle cluster and also a one-particle cluster, close enough to form a  $(u + 1)$ -particle cluster in one move. The kinetic factor is the probability per unit time that, given this configuration, the  $(u + 1)$ -particle cluster will actually form. The probability factor, which must of course be calculated in the restricted grand canonical ensemble, can be bounded above by the method we used in the preceding section to estimate the probability of finding a single cluster of given size in the unrestricted grand canonical ensemble. The upper bound has the form

$$g_u z^{u+1} Q_u |A|$$

where  $g_u$  is a geometrical factor giving the largest possible number of different ways of placing a single-particle cluster near enough to a given  $u$ -particle cluster to form a  $(u + 1)$ -particle cluster in one move; the other factor  $z^{u+1} Q_u |A|$  can be thought of as the product of  $z$ , which is an upper bound on the mean number of single-particle clusters per lattice site, of  $z^u Q_u$ , which is an upper bound on the mean number of  $u$ -particle clusters per lattice site, and of  $|A|$ , the total number of lattice sites. The kinetic factor has an upper bound of the form

$$g' W_M$$

where  $W_M$  is an upper bound on the probability per unit time that a given particle will move to a given vacant neighboring site, and  $g'$  is an upper bound on the number of different moves that will join the one-particle cluster to the  $u$ -particle cluster. Multiplying together our upper bounds on the probability factor and the kinetic factor we obtain

$$\lambda(u, 1) \leq g_u g' W_M z^{u+1} Q_u |A|.$$

To complete this calculation we need an upper bound on  $g_u g'$ . Suitable, though crude, bounds on  $g_u$  and  $g'$  are

$$g_u \leq qu, \quad g' \leq q^2$$

where  $q$  is the coordination number of the lattice (i.e. the number of nearest neighbors to each site). The contribution of  $\lambda(u, 1)$  to the total escape rate is therefore bounded by

$$[\lambda(u, 1)/|A|] \leq \text{const } uz^{u+1} Q_u.$$

In a similar way we can find bounds on the contributions from other types of process, for example the process where a  $(u - 2)$ -particle cluster combines

with a 3-particle cluster: its contribution would have an upper bound of the form

$$\begin{aligned} [\lambda(u-2, 3)/|A|] &\leq \text{const}(u-2)z^{u-2}Q_{u-2}z^3Q_3 \\ &= \text{const}(u-2)z^{u+1}Q_{u-2}Q_3. \end{aligned}$$

Numerical estimates indicate that this is roughly  $\frac{1}{5}$  of the corresponding upper bound for  $\lambda(u, 1)$ , so that these other processes are far from negligible.

The complication of having to add up contributions from many different types of processes can be partly avoided by an alternative way of estimating the escape rate, in which we consider together all the processes which lead to the formation of a  $(u+1)$ -particle cluster. Provided the transition probabilities satisfy the principle of detailed balancing, the equilibrium probability (per unit time) for each such process is the same as the equilibrium probability of the inverse process, calculated in the unrestricted ensemble. The total of all these probabilities of inverse processes equals the probability of a  $(u+1)$ -particle cluster breaking up. Now the equilibrium probability of finding a  $(u+1)$ -particle cluster is at most  $z^{u+1}Q_{u+1}|A|$ ; the number of transitions that can break it up is at most  $q(u+1)$  since none of the  $u+1$  particles in it has more than  $q$  places it can move to; and the probability per unit time of each such transition is at most  $W_M$ ; therefore the equilibrium probability per unit time of a  $(u+1)$ -particle cluster breaking up is at most

$$z^{u+1}Q_{u+1}|A|q(u+1)W_M.$$

This formula also provides an upper bound on the equilibrium probability per unit time of forming a  $(u+1)$ -particle cluster from smaller clusters. By an extension of the above argument, one can show that this last upper bound holds also for the restricted equilibrium ensemble. It follows that  $\lambda(u+1)$ , the probability of escape from the restricted ensemble by formation of a  $(u+1)$ -particle cluster, has the bound

$$z^{u+1}Q_{u+1}|A|q(u+1)W_M.$$

In the same way we can obtain bounds on  $\lambda(u+2)$ , the contribution from processes that produce a  $(u+2)$ -particle cluster, and so on up to the contribution from processes that produce clusters of size  $(q-1)u+1$ , which is the largest cluster than can be produced in a single step starting from a cluster compatible with our restricted ensemble. In this way we can obtain the upper bound

$$[\lambda/|A|] \leq W_M q \sum_{n=u+1}^{(q-1)u+1} nz^n Q_n.$$

This upper bound could probably be improved considerably, since the later terms in the sum are fairly crude estimates; but it serves to show how the rate of escape can be very small if  $u$  is not too large and if  $z$  exceeds the radius of convergence of the series  $\sum_1^\infty \tau^n Q_n$  by a small enough margin to make  $z^n Q_n$  very small over a range of values of  $n$ . Under these conditions the region  $R$  also satisfies the other two of our three conditions for metastability, and so the rigorous theory confirms the prediction of nucleation theory that these conditions make a metastable state possible.

## 9. Nucleation rates

We now turn from the problem of proving that a metastable state exists to the second main problem in the statistical mechanics of metastability, which is to obtain quantitative information about the rate at which the metastable phase goes over into the new phase. In nucleation theory the measure of this rate which is normally used is the *nucleation rate*, defined as the number of nuclei of the new phase formed per unit volume per unit time. The escape rate concept, which we have been using so far to help prove that a metastable state exists, does provide the upper bound  $\lambda/|\Omega|$  on the time rate of nucleation, but this upper bound is rather crude. We may expect that the nucleation rate is considerably less than the escape rate. This is partly because the escape rate calculations assume one-way traffic from small to large clusters, whereas in fact this is partially compensated by processes going in the opposite direction, such as the evaporation of a particle from a  $(u + 1)$ -particle cluster. Also the escape rate depends on the number of clusters of size  $u$  and less, which is probably overestimated in the restricted equilibrium ensemble since the processes that tend to convert such clusters into super-critical ones are not accounted for in the equilibrium theory of the restricted ensemble.

The first question to settle is what mathematical expression is to be used to define the rate of nucleation. In the present section, we base our discussion on the standard treatment due to Becker and Döring (1935), for which the necessary mathematical expression is known. This treatment, however, depends on simplifying assumptions which are unlikely to be accurate even at very low densities, and therefore in the section following this one we shall describe a more general way of defining nucleation rates.

We begin by outlining the treatment due to Becker and Döring, with a view to showing how it relates to the rigorous methods used in the preceding section. The main assumptions in this treatment are that clusters

are statistically independent and that processes in which a cluster gains or loses more than one particle may be ignored.

As we have noted in the preceding section, other processes, for example the absorption of a three-particle cluster into an  $n$ -particle cluster, can be comparable in importance with those involving one-particle clusters, so that the Becker–Döring theory will be only qualitatively correct, even at very low densities. The more complicated kinetic equation that is needed to take account of these other processes has been formulated by Miold and Binder (1978). If we accept the assumption that these processes can be ignored, however, then with particle-conserving dynamics the only way an  $n$ -particle cluster can grow is by absorbing a 1-particle cluster. The number of times this process happens per unit volume per unit time is proportional to  $m_n$ , the density of  $n$ -particle clusters, and also to  $m_1^{(n)}$ , the mean density of 1-particle clusters near an  $n$ -particle cluster; it is therefore taken to be  $a_n m_n m_1^{(n)}$  where  $a_n$  is independent of time and  $m_n$  and  $m_1^{(n)}$  are the numbers of  $n$ -particle and 1-particle clusters per unit volume. The only way an  $(n+1)$ -particle cluster can shrink is by emitting a 1-particle cluster; the number of times this happens per unit volume per unit time is taken to be  $b_{n+1} m_{n+1}$  where  $b_{n+1}$  is independent of time. The net rate at which  $n$ -particle clusters are being converted to  $(n+1)$ -particle clusters is therefore

$$J_n = a_n m_n m_1^{(n)} - b_{n+1} m_{n+1}, \quad (n \geq 1) \quad (9.1)$$

and the net rate of change of  $m_n$  is given by

$$dm_n/dt = J_{n-1} - J_n, \quad (n \geq 2). \quad (9.2)$$

This equation does not apply for  $n=1$  since one-particle clusters are not restricted to processes involving other one-particle clusters. In the original paper of Becker and Döring it was assumed, in effect, that  $m_1$  does not significantly change with time, so that

$$dm_1/dt = 0. \quad (9.3)$$

This assumption has the mathematical advantage of making the kinetic equations (9.1) and (9.2) a linear constant-coefficient system for the variables  $m_2, m_3, \dots$ ; moreover it is exact if we imagine the system coupled to a reservoir which can emit and absorb only one-particle clusters. For a more realistic treatment, however, we should replace (9.3) by the condition that  $\sum_1^\infty n m_n$ , the expectation of the total number of particles in the system, does not change with time; this condition, combined with (9.2) gives

$$\begin{aligned} dm_1/dt &= - \sum_{n=2}^{\infty} n(J_{n-1} - J_n) \\ &= -2J_1 - \sum_{n=2}^{\infty} J_n. \end{aligned} \quad (9.4)$$

To complete the system of eqs. (9.1, 9.2, 9.4), we need a relation connecting  $m_1^{(n)}$  with  $m_1$ . At equilibrium we have, in the low density limit, simply

$$m_1^{(n)} = m_1$$

Becker and Döring assumed that this relation was also true away from equilibrium, but it was later realized (Lifshitz and Slyozov 1961) that away from equilibrium the quantities  $m_1^{(n)}$  and  $m_1$  will differ because of the concentration gradient necessary to make one-particle clusters diffuse from infinity to the surface of an  $n$ -particle cluster. It is reasonable to assume that this concentration difference is proportional to  $J_n/m_n$ , the average rate at which  $n$ -particle clusters are increasing in size. This assumption gives

$$m_1^{(n)} = m_1 - k_n J_n / m_n$$

where, for each value of  $n$ ,  $k_n$  is inversely proportional to the diffusion coefficient. If this last equation is substituted into (9.1) we obtain

$$J_n = \frac{a_n}{1 + k_n a_n} m_n m_1 - \frac{b_{n+1}}{1 + k_n a_n} m_{n+1} \quad (9.1a)$$

an equation depending on cluster concentrations in the same way as the one analyzed by Becker and Döring when they assumed (ignoring the diffusion effect) that  $m_1^{(n)} = m_1$  in (9.1)

The equilibrium solutions of these equations are those for which all the "flows"  $J_n$  are zero. There is a family of such solutions, with one member for each value of  $m_1$ , given by

$$m_n = c_n (m_1)^n \quad (9.5)$$

where

$$c_1 = 1, c_2 = \frac{a_1}{b_2}, c_3 = \frac{a_1 a_2}{b_2 b_3}, \dots, c_n = \frac{a_1 \dots a_{n-1}}{b_2 \dots b_n} \quad (9.6)$$

In the lattice model we considered in sect. 6, the coefficients  $a_n, b_n, c_n$  can be expressed in terms of microscopic properties of the model. This is simplest for  $c_n$ : the low-density approximation corresponding to the rigorous inequality (7.13) is  $m_n \cong z^n Q_n$  and in particular (since  $Q_1 = 1$ )  $m_1 \cong z$ ; and since these approximations are correct to lowest order in  $z$  their substitution into (9.6) gives

$$c_n = Q_n \quad (9.7)$$

The other coefficients,  $a_n$  and  $b_n$ , can be determined from the type of



argument we used in sect. 8, if we make the further assumption that the densities of different types of clusters containing the same number of particles are always in the same ratios as they are at equilibrium (this further assumption is dubious, but is inherent in the form of kinetic equation we are working with. An accurate low-density kinetic equation would use a separate density  $m_c$  for each type of cluster  $c$ , instead of lumping together all clusters with the same number of particles in a single density). Having adopted this assumption, we can calculate the coefficients  $a_n$  and  $b_{n+1}$  by calculating the equilibrium values of  $a_n m_n m_1$  and  $b_{n+1} m_{n+1}$ . The equilibrium value of  $a_n m_n m_1$  is the rate per unit volume at which  $n$ -particle and 1-particle clusters are combining to give  $n+1$ -particle clusters; it was denoted by  $\lambda(n+1, 1)/|A|$  in sect. 8, and the type of argument used there shows that

$$\begin{aligned} a_n m_n m_1 &= b_{n+1} m_{n+1} \\ &= z^{n+1} \sum_{c, c'} W_{c, c'} e^{-\beta U(c')}/|A| \end{aligned} \quad (9.8)$$

where  $c$  is any  $(n+1)$ -particle cluster,  $c'$  is a configuration consisting of a  $n$ -particle cluster and a 1-particle cluster, such that  $c'$  can change to  $c$  in just one step, and  $W_{c, c'}$  is the probability per unit time that this step takes place given that the configuration of the system is such that the step is possible. The formula we have written in (9.8) is for the process  $n \rightarrow n+1$ , but the formula for the reverse process, obtainable by interchanging  $c$  and  $c'$ , is the same because of detailed balancing. Inserting the equilibrium formulae  $m_n = z^n Q_n$ , etc., in (9.8) we obtain

$$\begin{aligned} a_n &= (1/Q_n) \sum_{c, c'} W_{c, c'} e^{-\beta U(c')}/|A| \\ b_{n+1} &= (1/Q_{n+1}) \sum_{c, c'} W_{c, c'} e^{-\beta U(c')}/|A|. \end{aligned} \quad (9.9)$$

Thus although the Becker-Döring equations are far from being rigorously justified or even accurate, we can at least write down well-defined formulae for the kinetic coefficients appearing in them.

Let us now see how Becker and Döring calculated the nucleation rate from their kinetic equations. Their method is to assume that every time the number of particles in a cluster passes a specified fairly large value  $u$ , the cluster is removed from the system. This has the effect of replacing eq. (9.2) for  $n > u$  by the condition

$$m_n = 0, \quad \text{if } n > u. \quad (9.10)$$

There is then a steady-state solution of the eq. (9.2) for  $dm_n/dt$  ( $n \geq 2$ ) in

which  $J_n$  is independent of  $n$ . If we are using the Becker–Döring assumption (9.3) that  $m_1$  is constant, rather than the more realistic equation (9.4) for the time variation of  $m_1$ , then making  $J_n$  independent of  $n$  is also consistent with the equation for  $dm_1/dt$ . The common value of  $J_n$  can then be calculated from the eq. (9.1a), which it is convenient to rewrite in terms of the new variables  $\Phi_1, \Phi_2, \dots$  and  $R_1, R_2, \dots$  defined by

$$\Phi_n = m_n/(m_1)^{n-1} c_n, \quad (n \geq 2) \quad (9.11)$$

$$\Phi_1 = m_1$$

and

$$R_n = (1 + k_n a_n)/(m_1)^n c_n a_n, \quad (n \geq 1) \quad (9.12)$$

so that the equation (9.1a) takes the form

$$J_n R_n = \Phi_n - \Phi_{n+1}, \quad n \geq 1. \quad (9.13)$$

(As first pointed out by Becker and Döring these equations have the interpretation that  $R_n$  is the resistance of an electrical path from a node  $n$  to a node  $n+1$ ,  $J_n$  the current in it, and  $\Phi_n$  is the potential at the node  $n$ . See also Binder and Stauffer (1976)). Summing from  $n=1$  to  $u$ , using the conditions  $\Phi_1 = m_1$  [from eq. (9.11)] and  $\Phi_{u+1} = 0$  [from (9.10) and (9.11)], we obtain

$$\sum_{n=1}^u J_n R_n = m_1. \quad (9.14)$$

The nucleation rate  $J$  is defined as the common value of  $J_1, J_2, \dots$  and is therefore given by

$$J = m_1 / \sum_{n=1}^u R_n. \quad (9.15)$$

Equation (9.15), with (9.12), expresses the nucleation rate in terms of microscopic quantities (it is practically independent of  $u$  if  $u$  is chosen correctly—i.e. sufficiently large). However, its derivation depends on various assumptions which may not always be valid and therefore it is not a complete solution to the problem of defining nucleation rates microscopically.

A more realistic picture of nucleation would be obtained if we based our analysis on the more accurate kinetic eq. (9.4) for  $dm_1/dt$  and used (9.2) instead of the approximation (9.10) to determine  $m_n$  for  $n > u$ . A rigorous analysis is difficult because the resulting equations are non-linear, but we may expect that the results would be fairly similar to the Becker–Döring

analysis over times which, while long compared to the time for the values of  $J_1, J_2, \dots, J_u$  to become approximately equal are short compared to the time for  $m_1$  to change by an appreciable fraction of its original value.

To conclude this section let us compare the Becker–Döring estimate (9.15) of the nucleation rate with the escape rate estimated for the corresponding restricted ensemble by the methods used earlier in this article. To get the equilibrium properties of the restricted ensemble, we alter the dynamics so that no  $u$ -particle cluster can increase in size. This is not the same thing as removing all  $(u+1)$ -particle clusters, which we did in estimating the nucleation rate; instead it corresponds to replacing eq. (9.1) for  $n \geq u$  by

$$J_n = 0, \quad (n \geq u). \quad (9.16)$$

The equilibrium solutions (with  $J_n = 0$  for  $n = 1 \dots u-1$ ) for the modified kinetic equations, using either (9.3) or (9.4) for  $dm_1/dt$ , are

$$m_n = \begin{cases} c_n(m_1)^n, & \text{if } n \leq u \\ 0, & \text{if } n > u. \end{cases} \quad (9.17)$$

The escape rate is calculated by the usual method: we assume the above distribution at time 0, with unrestricted dynamics [i.e. with (9.16) replaced by the original kinetic equation (9.1)] and calculate the probability per unit time per unit volume of forming a cluster of size greater than  $u$ . This probability rate is

$$\begin{aligned} (dm_{u+1}/dt)_{t=0+} &= J_u - J_{u+1} = J_u \quad (\text{since } J_{u+1} = 0 \text{ at } t = 0+) \\ &= m_1 R_u \text{ (by (9.14) since } J_1 = \dots = J_{u-1} = 0 \text{ at } t = 0+). \end{aligned}$$

The optimum value of  $u$  (the one minimizing this escape rate) is the one that maximizes  $R_u$ , but comparison with (9.15) shows that even this escape rate is a crude overestimate of the nucleation rate  $J$ , corresponding to using the largest term in the series  $\sum_{n=1}^u R_n$  as an estimate of the entire series. This estimate could be wrong by a factor comparable with  $u$ , which could be several orders of magnitude in unfavorable cases. On the other hand, these calculations confirm that the escape rate does provide an upper bound on the estimated nucleation rate  $J$ , so that a proof that the state is metastable in the sense we have defined earlier (which implies that the escape rate is small) is also a proof that the estimate (9.15) of the nucleation rate is small.

Recently Mirolid and Binder (1977) described the onset of phase segregation in a binary alloy, following quenching into the miscibility gap,

by means of a non-linear set of equations for the concentration of clusters of the minority phase. Their equations are a generalization of the Becker – Döring equations in that they take into account also the coagulation of clusters of sizes  $n$  and  $m$  to form one cluster of  $n + m$  particles as well as the inverse process. They solved their equations numerically, using clusters up to size 150, and found results qualitatively similar to those obtained in the computer simulations of Sur et al. (1977) at high values of supersaturation.

### 10. Slowly-varying ensembles

In this section we suggest a possible way of improving on the restricted ensemble as a way of treating metastable lattice systems, which may make it possible to calculate nucleation rates from first principles instead of using kinetic equations, such as the Becker – Döring equations, whose reliability is open to some doubt. The method is a development of the idea of Becker and Döring of looking for a distribution of droplet sizes which is steady if every droplet reaching some specified super-critical size is removed from the system and replaced by small droplets. The main differences from their method are these: first, we consider an ensemble of systems in place of a collection of droplets within a single system; secondly we assume that every system of the ensemble in which a super-critical droplet forms is removed completely from the ensemble instead of being replaced by a system with the super-critical droplet broken up into one-particle droplets; and thirdly we look for a slowly varying ensemble in place of a steady one.

Our method applies to any system with a discrete set of microscopic states, whose dynamics has the form of a Markov process. It therefore applies to the Ising model with either Glauber or Kawasaki dynamics, but not to a system of particles, such as the Widom – Rowlinson model, obeying Newton's laws of motion.

We use the variables  $i, j$  to denote microscopic states of our system, and write  $W_{ij}$  for the probability per unit time that the system will make a transition from state  $j$  to state  $i$ . Then the rate of change of  $p_i(t)$ , the probability that the system is in state  $i$  at time  $t$ , is given for each  $i$  by the "master equation"

$$\frac{d}{dt} p_i(t) = \sum_j [W_{ij} P_j(t) - W_{ji} P_i(t)]. \quad (10.1)$$

Let us denote by  $e_i$  the equilibrium probability of the state  $i$ , not necessarily normalized. For a lattice model we would take  $e_i = \exp(-E_i/kT)$  where  $E_i$  is the energy of state  $i$ , but this formula for  $e_i$  does not

apply when the state  $i$  of the Markov process contains more than one microstate. The principle of detailed balancing, which we shall assume to hold, asserts that

$$W_{ij}e_j = W_{ji}e_i, \quad \text{for all } i, j \quad (10.2)$$

so that the only steady solution of (10.1) is  $p_i = \text{const } e_i$ , unless the states happen to split up into two or more groups between which no transitions are possible.

If the system has a (macroscopic) metastable state, then there should be a set  $R$  of microscopic states with the properties that escape from  $R$  is slow and that the probability of return by a system which has escaped from  $R$  is completely negligible. The second of these conditions implies that, for the probability distributions we are interested in, the terms in (10.1) of the form  $W_{ij}P_j(t)$  with  $i$  inside  $R$  and  $j$  outside can be neglected when computing  $P_i(t)$  for  $i \in R$ , if the system starts inside  $R$ ; hence (10.1) can be replaced by the approximation

$$\frac{d}{dt}P_i(t) = \sum_j [W_{ij}P_j(t) - W_{ji}P_i(t)], \quad (i \in R) \quad (10.3)$$

$$P_i(t) = 0, \quad (i \notin R)$$

which is our basic system of equations for this section.

To study the system of equations (10.3) we write them in the form

$$e_i \frac{d}{dt}v_i = \sum_j G_{ij}v_j, \quad (i \in R) \quad (10.4)$$

where  $\sum_j$  means a sum over the microscopic states in  $R$  only,  $v_i$  and  $G_{ij}$  are defined by

$$\begin{aligned} v_i &= v_i(t) = p_i(t)/e_i \\ G_{ij} &= W_{ij}e_j - \delta_{ij} \sum_k W_{ki}e_i, \quad (i, j \in R) \end{aligned} \quad (10.5)$$

and the sum  $\sum_k$  goes over all microscopic states. Because of detailed balancing [eq. (10.2)] the matrix  $G$  is symmetric. Moreover, assuming at least one transition out of  $R$  to be possible, the matrix  $G$  is negative definite, because for any non-zero vector  $\{v_i\}$  we have

$$\begin{aligned} \sum_i \sum_j v_i G_{ij} v_j &= \sum_i \sum_j v_i W_{ij} e_j v_j - \sum_i v_i^2 \sum_k W_{ki} e_i \\ &= -\frac{1}{2} \sum_i \sum_j W_{ij} e_j (v_i - v_j)^2 - \sum_i v_i^2 \sum_k W_{ki} e_i \\ &< 0 \end{aligned} \quad (10.6)$$

where  $\sum''$  means a sum over all the microscopic states outside  $R$ . It follows that all the eigenvalues  $\lambda^{(r)}$  of the equation

$$\lambda^{(r)} e_i v_i^{(r)} = - \sum_j G_{ij} v_j^{(r)} \quad (10.7)$$

are real and positive and that we can solve the initial-value problem for our basic equation (10.3) by expressing the initial vector as a linear combination of eigenvectors

$$v_i(t) = \sum_r A_r \exp[-\lambda^{(r)} t] v_i^{(r)}$$

where  $A_0, A_1, \dots$ , are chosen to make

$$v_i(0) = \sum_r A_r v_i^{(r)}, \quad \text{for all } i \text{ in } R.$$

After a long time, therefore, the probabilities in an ensemble that was started with any distribution over  $R$  will have the asymptotic form

$$P_i(t) = e_i v_i(t) \sim \text{const} \exp[-\lambda^{(0)} t] e_i v_i^{(0)} \quad (i \in R) \quad (10.8)$$

where  $\lambda^{(0)}$  is the smallest eigenvalue. From this fact we can see that  $\lambda^{(0)}$  is non-degenerate, unless there are two non-vanishing sets of states within  $R$  such that transitions from a state in one set to a state in another have zero probability – in which case  $R$  splits into two disjoint metastable sets of configurations.

Equation (10.8) shows that after the initial transients have died out the probability of finding the system in  $R$  decays exponentially at a rate  $\lambda^{(0)}$ . If the region  $R$  corresponds to a metastable equilibrium, it is therefore natural to interpret  $\lambda^{(0)}$  as the rate of decay of the metastable state, i.e. the *rate of nucleation*. (This corresponds exactly to the interpretation of a similarly defined eigenvalue in the problem of dissociation of diatomic molecules, as studied, for example, by Bak and Lebowitz (1962, 1963).) Our second criterion of metastability, that this rate of decay should be very slow, then tells us that  $\lambda^{(0)}$  must be very small in the case of metastability, but unless there is more than one metastable region inside  $R$  we would not expect any of the other eigenvalues to be very small.

One virtue of identifying the rate of nucleation with  $\lambda^{(0)}$  is that we can apply the Rayleigh–Ritz principle (Courant and Hilbert 1953). This principle can be written, using (10.6), in the form

$$\lambda^{(0)} \leq \frac{\frac{1}{2} \sum_i \sum_j W_{ij} e_j (v_i - v_j)^2 + \sum_i v_i^2 \sum_k W_{ki} e_i}{\sum_i e_i v_i^2} \quad (10.9)$$

for any vector  $\{v_i\}$ , with equality if and only if  $v_i$  is an eigenvector of  $G_{ij}$  belonging to the eigenvalue  $\lambda^{(0)}$ . Thus, with enough ingenuity in choosing trial vectors  $\{v_i\}$  we may be able to calculate very good estimates of the nucleation rate.

A possible disadvantage of defining the nucleation rate as the smallest eigenvalue of  $G_{ij}$ , or by the variational principle (10.9), is that it depends on the choice of the metastable region  $R$ , which is somewhat arbitrary. However, this disadvantage is inherent in the use of such a region. We hope that if  $R$  is large enough then  $\lambda^{(0)}$  will not depend seriously on the choice of  $R$ .

To illustrate the use of (10.9) we consider the choice  $v_i = 1$  for all  $i$  in  $R$ . This gives

$$\lambda^{(0)} \leq \frac{\sum_i' e_i \left( \sum_k'' W_{ki} \right)}{\sum_i' e_i} \quad (10.10)$$

The right-hand side is the average, calculated in the restricted equilibrium ensemble, of the probability per unit time that the system will escape from  $R$ ; in other words, it is the initial escape rate  $\lambda(0)$ , and we recover our previous conclusion that the nucleation rate is bounded above by the escape rate which in turn was shown to be small in the models considered by PL and CCO.

As a more substantial application of the variational principle (10.9), we show how the Becker – Döring nucleation rate formula (9.15) can be derived from it. For this, we take the states  $j$  to be the configurations of a lattice gas and give  $v_j$  the form

$$v_j = \prod_k f(c_k) \quad (10.11)$$

where  $\{c_1, c_2, \dots\}$  is the set of clusters constituting the configuration  $j$ , and  $f$  is a function which assigns a positive number to each cluster that can exist on the lattice. To make the metastable distribution  $p_i = e_i v_i$  agree as well as possible with the corresponding Gibbs equilibrium distribution when all the clusters are small, we would like  $v_i$  to be close to 1 when all clusters are small; that is, we would like  $f(c)$  to be close to 1 when  $N(c)$ , the number of particles in the cluster  $c$ , is small. We shall allow for this requirement in our calculation by setting

$$f(c) = 1, \quad \text{if } N(c) = 1. \quad (10.12)$$

At the same time, we want  $v_j$  to be zero outside the set  $R$  of configurations

defining the restricted ensemble; we do this by setting

$$f(c) = 0, \quad \text{if } N(c) > u \quad (10.13)$$

where  $u$  is the maximum cluster size.

In accordance with the basic approximation made by Becker and Döring, we consider only processes in which a  $n$ -particle cluster and a 1-particle cluster combine to give an  $(n+1)$ -particle cluster, together with their inverses. The contribution to the right-hand side of (10.9) from processes in which a specified cluster  $c$  combines with some one-particle cluster to give a specified cluster  $c'$  is bounded above by

$$\frac{W_{c',c} z^{n+1} e^{-\beta U(c)} (f(c') - f(c))^2 \left\{ 1 + \sum_{(c)}'' \prod_k z^{N(c_k)} e^{-\beta U(c_k)} f(c_k)^2 \right\}}{1 + \sum_{(c)}'' \prod_k z^{N(c_k)} e^{-\beta U(c_k)} f(c_k)^2 + \text{other terms}}$$

where  $\sum_{(c)}''$  denotes a sum over all configurations of contours which are compatible with (i.e. do not touch or overlap) the contour  $c$ . The given expression is in turn bounded above by

$$W_{c',c} z^{n+1} e^{-\beta U(c)} (f(c') - f(c))^2. \quad (10.14)$$

If we use a trial function  $f$  which depends only on the number of particles in the cluster we can sum this last formula over all  $n$ -particle clusters  $c$ , and all clusters  $c'$  obtainable from them by the absorption of a one-particle cluster, to obtain

$$\begin{aligned} & \sum_{c',c} W_{c',c} z^{n+1} e^{-\beta U(c)} (f(n+1) - f(n))^2 \\ &= |A| z^{n+1} Q_n a_n (f(n))^2 \end{aligned} \quad (10.15)$$

where we have used (9.9) in deriving the last line. Summing over all  $n \leq u-1$ , dropping the factor  $\frac{1}{2}$  in (10.9) to allow for the contribution of the inverse transitions, which by detailed balancing is precisely equal to that of the transitions we have considered, and adding in an  $n = u$  term to take care of the second sum on the top of (10.9), we obtain

$$|A| \sum_{n=1}^u z^{n+1} Q_n a_n (f(n+1) - f(n))^2 \quad (10.16)$$

for our approximation to the upper bound on  $\lambda^{(0)}$ .

The condition for this approximate upper bound to be a minimum with respect to changes in  $f(n)$  is

$$z^{n+1} Q_n a_n (f(n+1) - f(n)) = z^n Q_{n-1} a_{n-1} (f(n) - f(n-1)). \quad (10.17)$$

From this and the conditions (10.12) and (10.13) which require  $f(1) = 1$ ,



$f(u+1)=0$ , it follows that (in the notation of (9.7) and (9.12), with  $m_1$  set equal to  $z$ , an approximation which becomes exact in the low-density limit)

$$f(n) = \frac{\sum_{k=n}^u R_k (1 + k_n a_n)^{-1}}{\sum_{k=1}^u R_k (1 + k_n a_n)^{-1}} \quad (10.18)$$

Hence our approximation (10.16) to the upper bound on  $\lambda^{(0)}$  becomes

$$|A|z / \sum_{k=1}^u R_k (1 + k_n a_n)^{-1}. \quad (10.19)$$

To get a rigorous upper bound, however, we should add in the effects of the types of transitions neglected by Becker and Döring, in which more than one multi-particle cluster takes part.

In the low-density approximation, where  $m_1 \approx z$ , eq. (10.19) agrees with the Becker-Döring estimate (9.15) for the nucleation rate. The fact we noted at the end of sect. 9, that their formula (9.15) or (10.19) gives a nucleation rate which is less than the escape rate, now has a simple interpretation: both can be obtained from the variational principle (10.9), but in deriving (10.19) we choose the function  $f$  to make our upper bound on  $\lambda^{(0)}$  as small as possible, whereas our derivation of the escape rate is equivalent to the choice

$$f(c) = \begin{cases} 1 & \text{if } N(c) \leq u \\ 0 & \text{if } N(c) \geq u + 1 \end{cases}$$

which is far from the best and therefore gives a larger upper bound.

## 11. Concluding discussion

Metastability is a kinetic phenomenon. It is due to the fact that under suitable experimental circumstances the system finds itself in a configuration-space region  $R$  from which it has only a small probability of escaping, though an equilibrium system outside the region has an even smaller probability of getting into it. When the system is in such a region it has only one thermodynamic phase, e.g. super-cooled vapor, even though the relevant equilibrium Gibbs ensemble predicts that the system will, with overwhelming probability, consist of two phases. The exact specification of the boundary of the region  $R$  is not critical to the theory, since the system is extremely unlikely to be found near the boundary of  $R$  in any case: in

different parts of this paper we have used several different ways of specifying  $R$ , according to the mathematical needs of the particular situation under study. The important thing is to specify  $R$  in such a way that the configuration is in  $R$  if and only if there are no large droplets or nuclei of the new phase in the system.

The above statements require, of course, that we specify the probability distribution of the system inside  $R$ . For this purpose, as well as to calculate the static properties of the metastable state, we replace the equilibrium Gibbs ensemble by a modified Gibbs ensemble whose members are restricted to the region  $R$ . To justify the use of such an ensemble for the metastable state we would have to show that the dynamics of the system contains both a "fast" mechanism bringing the system to the (quasi-) equilibrium described by the restricted ensemble, and a "slow" mechanism for taking the system out of the configuration-space region where it is trapped. So far this problem has not been solved, but the work of PL and CCO described in the first part of this chapter does show that for certain model system (the van der Waals model, the two dynamical versions of the Ising model, and the Widom-Rowlinson model), if we assume that the fast mechanism exists, then we can demonstrate (by proving that the escape rate is small) that the other mechanism is "slow" for a suitable choice of the region  $R$  and of the thermodynamic parameters. The methods of proof rely heavily on equilibrium concepts, since we use the Gibbs equilibrium ensemble to give the relative probabilities of the different microstates in  $R$  - in other words we assume that the correct ensemble to describe a system started "somewhere in  $R$ " is the appropriate equilibrium Gibbs ensemble restricted to  $R$ . For greater realism, it would be desirable to consider more carefully the way in which the system reached  $R$  in the first place - perhaps as the result of the sudden quenching of a binary alloy from a higher temperature or the sudden cooling of a previously unsaturated vapor by adiabatic expansion. Such an investigation would provide some of the information about the "fast" mechanism which is taken for granted in the work of PL and CCO.

The work described in the second part of the chapter carries the investigation of one of our special models (the lattice gas with Kawasaki dynamics) a little beyond the stage of demonstrating that a metastable state exists, and provides a method of obtaining further information about both the "fast" and "slow" mechanisms. This work shows that if the density is very low and if processes involving more than one many-particle cluster can be neglected, then the Becker-Döring kinetic equations may be expected to apply to this model, and gives explicit formulae for the coefficients in these equations. We hope that these formulae will make it possible to make

accurate predictions of the metastable behavior of such models, which can be tested against computer simulations. However, it is unlikely that all the assumptions going into the Becker–Döring equations are valid, even at very low densities; in particular, better estimates of the effects of processes involving more than one many-particle cluster, and the effects of correlations between clusters, are needed. Some recent work on these problems has been reported by Binder and Stauffer (1976), and by Miold and Binder (1978), but it is clear that much more work remains to be done on the formation and derivation of kinetic equations for droplets.

The last part of the paper discusses ways of obtaining a numerical measure of the “degree of metastability” of a system which would be well-defined, simply related to experimental data and, we hope, essentially independent of the arbitrary features in the definition of  $R$ . When the Becker–Döring equations apply, such a measure is provided by the nucleation rate, but we would like to have a more general formulation valid also for higher densities and temperatures where these equations do not apply. A generalization which may help to answer this requirement is described in sect. 10, and is shown to give the same answer as the Becker–Döring nucleation rate formula when the latter applies.

Since so little work has been done up to now on the rigorous treatment of metastability, it is hardly surprising that so many questions remain unanswered. What is encouraging is that a start has been made and that further progress appears to be possible.

#### Acknowledgements

We would like to thank K. Binder, J. Langer, H. Reiss, D. Capocaccia, M. Cassandro and E. Olivieri for helpful discussions and correspondence.

Part of this work was done while both authors were visitors at the Institute des Hautes Etudes Scientifique, Bures-sur-Yvette, France. We thank Prof. N. Kuiper for his kind hospitality there.

#### Appendix 1: The Widom–Rowlinson model

This model (Widom and Rowlinson 1971) consists of a mixture of two kinds of particles. There is no interaction between like particles, but unlike particles repel by means of a hard-sphere potential. In other words, the Hamiltonian is the sum of a kinetic energy term for each particle and a potential energy, which is equal to 0 if all the unlike particles are separated by a distance exceeding some fixed value  $a$ , and to  $+\infty$  if not.

It was shown rigorously by Ruelle (1971) that in two dimensions this system has a phase transition: below a certain temperature, if the density is high enough, the system separates into two phases, each of them rich in one of the two types of particle. The proof by Cassandro and Olivieri (1977) that this model has metastable states is based on the techniques developed by Ruelle for his proof. For two dimensions, the fundamental idea of Ruelle's method is to define, for each configuration of the particles, a system of polygonal contours which enclose particles of one kind only. The particles inside each polygon may be regarded as a cluster of that type of particle, and the proofs of the existence of a phase transition and of metastable states for this model are very similar in spirit to the corresponding proofs for lattice systems given respectively by Peierls (1936) and Griffiths (1964), and by CCO (1974).

#### Appendix 2: An upper bound on $Q_N$ for the plane square lattice

Equations (7.2) and (7.9) give  $Q_N = \sum_{c: N(c)=N} e^{\beta U_0 p(c)}$ . Now  $p(c)$  is related to  $b(c)$ , the number of neighboring pairs of sites, one of which is occupied and the other not, by  $qN = 2p + b$  where  $q$  is the coordination number; so the formula can be written, for the plane square lattice which has  $q = 4$

$$Q_N = e^{2\beta U_0 N} \sum_c e^{-\beta U_0 b(c)/2}.$$

Since  $b \geq 4\sqrt{N}$  (every cluster having a perimeter  $b$  at least as large as that of a square containing the same number of sites) this formula gives

$$Q_N \leq A_N e^{2\beta U_0 N - 2\beta U_0 \sqrt{N}}$$

where  $A_N$  is the number of terms in the sum  $\sum_c$ , that is, the number of translationally inequivalent clusters with  $N$  occupied sites. To estimate  $A_N$ , we note that every cluster can be converted to a simply connected cluster by making, if necessary, some cuts in the bonds joining its sites. The resulting simply connected cluster has a boundary consisting of a polygon made out of line segments of unit length, each bisecting one of the bonds on its boundary perpendicularly (see fig. 4). The resulting polygon has at most  $2N + 2$  line segments in its sides, and the number of such polygons with  $2k + 2$  sides is at most  $3^{2k}$ , because in drawing it segment by segment we make  $2k$  choices, each from 3 alternatives, of which direction to draw the next segment in. Hence the number of such polygons with at most  $2N + 2$  sides is less than  $3^2 \dots + 3^{2N} = \frac{1}{8}(3^{2N+2} - 3^2) < \frac{9}{8} 3^{2N}$ . It follows that  $A_N \leq \frac{9}{8} 3^{2N}$  and hence that  $Q_N \leq \frac{9}{8} \exp[2(\log 3 + \beta U_0)N - 2\beta U_0 \sqrt{N}]$ .



Fig. 4. At left: cutting one bond so as to make the cluster simply connected (crosses denote particles); at right, the "boundary" of the resulting simply connected cluster.

### Appendix 3: A lower bound on $Q_N$ for the plane square lattice

One of the terms in the formula (7.9) for  $Q_N$  is an integer multiple of  $\exp \beta U_0 p_{\max}(N)$  where  $p_{\max}(N)$  is the largest possible number of nearest-neighbor pairs in a cluster of  $N$  particles. If  $N$  is a perfect square, say  $r^2$ , then a square array gives  $r(r-1)$  vertical bonds and an equal number of horizontal bonds, so that

$$p_{\max}(r^2) = 2r(r-1).$$

Removing any number up to  $r-1$  of particles from one side of the square we find, since  $k$  removals destroy at least  $2k$  bonds

$$p_{\max}(r^2 - k) = \left. \begin{aligned} &= 2r(r-1) - 2k \\ &= 2(r^2 - k) - 2r \end{aligned} \right\} \quad \text{if } 0 \leq k \leq r-1.$$

Alternatively we may add up to  $r$  particles to one side of the square. The first adds one bond, but after that each can add two bonds, and so we find

$$p_{\max}(r^2 + k) = \left. \begin{aligned} &= 2r(r-1) + 2k - 1 \\ &= 2(r^2 + k) - 2r - 1 \end{aligned} \right\} \quad \text{if } 1 \leq k \leq r.$$

Every integer  $N$  can be written in one of the forms  $r^2$ ,  $r^2 - k$ , or  $r^2 + k$  with  $k$  restricted as indicated, and all three of the above formulae are consistent with the inequality

$$p_{\max}(N) \geq 2N - 2\sqrt{N} - 1, \quad (N = 1, 2, \dots).$$

We conclude that

$$Q_N \geq \exp [2\beta U_0 N - 2\beta U_0 \sqrt{N} - \beta U_0].$$

**Note added in proof**

We add here a few words about some recent developments in the areas mentioned in this chapter.

*1. Equilibrium properties of clusters*

Binder and Kalos (1979) have studied clusters of supercritical size on a lattice by computer simulations in which these clusters were stabilized by confining the vapor surrounding them to a sufficiently small region. This work provides information about quantities such as the cluster partition functions and free energies defined in sect. 7. Further references are also given in that section. McGraw and Reiss (1979) have discussed corrections to the low-density theory discussed in sect. 7, arising from the fact that clusters cannot overlap, concentrating on the effect of these corrections on metastability of a gas-liquid system near its critical temperature.

*2. The Becker-Döring equations*

Some confirmation of the validity of these equations, for densities below about 7.5% (five times the saturation vapor density) in a lattice gas with Kawasaki dynamics, is provided by the work of Penrose et al. (1978) and of Penrose and Buhagiar (1979). In the first of these papers an approximate solution of the equations is obtained using ideas due to Lifshitz and Slyozov (1961) and used to explain the time-dependent cluster distributions obtained after the breakdown of metastability. In the second, the coefficients in the Becker-Döring equations are calculated for small clusters using an extension of the methods described in sect. 9 and then extrapolated to large clusters; the resulting system of differential equations is integrated numerically, and the predicted time dependent cluster distributions compared with the cluster distributions found in computer simulations.

*3. Fluids and fluid mixtures*

Further experiments on spinodal decomposition in binary fluid mixtures have been carried out by Goldburg et al. (1978). A theoretical analysis of the various competing processes determining the nucleation kinetics in a fluid near its critical point is given by Siggia (1979). Molecular dynamics computations on the formation and decomposition of clusters in a two-dimensional gas with square-well hard-core interactions are described by Zurek and Schieve (1978); and on spinodal decomposition in a three-

dimensional gas with Lennard-Jones interactions, by Muzrik et al. (1978). A review of the theoretical, experimental and computer simulation work on this topic is given by Abraham (1979).

### References

- Abraham, F., 1974, *Homogeneous Nucleation Theory* (Academic Press, New York).
- Bak, T. and J.L. Lebowitz, 1962, *Disc. Farad. Soc.* **33**, 189.
- Bak, T. and J.L. Lebowitz, 1963, *Phys. Rev.* **B1**, 1138.
- Becker, R. and W. Döring, 1935, *Ann. der Phys.* **24**, 719-752.
- Binder, K. and Müller-Krumbhaar, 1974, *Phys. Rev.* **B9**, 2328.
- Binder, K. and D. Stauffer, 1976, *Ann. Phys. (N.Y.)* **98**, 390-417.
- Binder, K., M.H. Kalos, J.L. Lebowitz and J. Marro, 1979, *Advances in colloid and interface science* **10**, 173.
- Bortz, A.B., M.H. Kalos, J.L. Lebowitz and M.S. Zendejas, 1974, *Phys. Rev.* **B10**, 535.
- Bragg, W.L. and E.J. Williams, 1934, *Proc. Roy. Soc.* **A145**, 699.
- Cahn, J.W., 1961, *Acta Metall.* **9**, 795.
- Cahn, J.W., 1962, *Acta Metall.* **10**, 179.
- Cassandro, M. and E. Olivieri, 1976, private communication.
- Cassandro, M. and E. Olivieri, 1977, *J. Stat. Phys.* **17**, 229.
- Capocaccia, D., M. Cassandro and E. Olivieri, 1974, *Commun. Math. Phys.* **39**, 185.
- Courant, R. and D. Hilbert, 1953, *Methods of mathematical physics*, vol. I.
- Döring, W., 1937, *Z. Phys. Chem.* **36**, 371.
- Essam, J.W., 1972, in *Phase transition and critical phenomena*, vol. 2 (C. Domb and M.S. Green eds.) (Academic Press, New York) pages 197-269.
- Fisher, M.E., 1967, *Physics* **3**, 255.
- Fisher, M.E. and G. Cagenalp, 1977, *Wall and boundary free energies I*; Cornell preprint.
- Fowler, R.H. and E.A. Guggenheim, 1939, *Statistical thermodynamics* (Cambridge University Press, London) p. 572.
- Frenkel, J., 1955, *Kinetic theory of liquids* (Dover, New York), ch. 7.
- Glauber, R.J., 1963, *J. Math. Phys.* **4**, 294-307.
- Griffiths, R.B., 1964, *Phys. Rev.* **136A**, 437.
- Hill, T.L., 1956, *Statistical mechanics* (McGraw-Hill, New York) p. 293.
- Hillert, M., 1961, *Acta Metall.* **9**, 525.
- Huang, J.S. and W.I. Goldburg, 1976, *Bull. Am. Phys. Soc.* **21**, no. 1, 55.
- Huang, J.S., W.I. Goldburg and A.W. Bjerkaas, 1974, *Phys. Rev. Lett.* **32**, 921.
- Kalos, M.H., J.L. Lebowitz, O. Penrose and A. Sur, 1978, *J. Stat. Phys.*, **18**, 39.
- Katz, J.L., 1970, *J. Stat. Phys.* **2**, 137.
- Kawasaki, K., 1966, *Phys. Rev.* **145**, 224.
- Kawasaki, K., 1972, in *Phase transitions and critical phenomena*, vol. 2 (C. Domb and M.S. Green, eds.) (Academic Press, New York).
- Langer, J.S. and V. Ambegaokar, 1967, *Phys. Rev.* **164**, 498.
- Lanford, O. and D. Ruelle, 1969, *Commun. Math. Phys.* **13**, 194.
- Langer, J.S., 1967, *Ann. Phys. N.Y.* **41**, 108.
- Langer, J.S. and M.E. Fisher, 1967, *Phys. Rev. Lett.* **19**, 560.
- Langer, J.S., M. Bar-on and H.D. Miller 1975, *Phys. Rev.* **A11**, 1417.
- Lebowitz, J.L. and O. Penrose, 1966, *J. Math. Phys.* **7**, 98.

- Lebowitz, J.L. and O. Penrose, 1977, *J. Stat. Phys.* **16**, 321.
- Lee, T.D. and C.N. Yang, 1952, *Phys. Rev.* **87**, 410.
- Lifshitz, I.M. and V.V. Slyozov, 1961, *J. Phys. Chem. Solids* **19**, 35.
- Lothe, J. and G.M. Pound, 1962, *J. Chem. Phys.* **36**, 2080.
- Marro, J., A.B. Bortz, M.H. Kalos and J.L. Lebowitz, 1975, *Phys. Rev.* **B12**, 2000.
- Maxwell, J.C., 1965, *Scientific Papers* (ed. W.D. Niven) (Dover, New York), p. 425.
- McCraw, R.J. and L.S. Schulman, 1978, *J. Stat. Phys.* **18**, 293.
- Millard, K. and L.H. Lund, 1973, *J. Stat. Phys.* **8**, 225.
- Minlos, R.A. and Ya. Sinai, 1968, *Trans. Moscow. Math. Soc.* **19**, 121.
- Mirolid, P. and K. Binder, 1978 *Acta. Met.* **25**, 4535.
- Newman, C.M. and L. Schulman, 1977, *J. Math. Phys.* **18**, 23.
- Peierls, R., 1936, *Proc. Camb. Phil. Soc.* **32**, 477.
- Penrose, O. and J.L. Lebowitz, 1971, *J. Stat. Phys.* **3**, 211.
- Rao, M., Kalos, M.H., J.L. Lebowitz and J. Marro, 1976, *Phys. Rev.* **B13**, 4325.
- Reiss, H., 1974, The replacement free energy in nucleation theory, in *Nucleation II*, (A.C. Zettlemoyer ed.) (special issue of advances in colloid and interface science).
- Reiss, H., 1975, The existence of the spinodal, *Proceedings of the Bunsengesellschaft für Physikalische Chemie*, Vienna, Austria.
- Ruelle, D., 1971, *Phys. Rev. Lett.* **27**, 1040.
- Schwartz, A.J., J.S. Huang and W.I. Goldberg, 1975, *J. Chem. Phys.* **62**, 184.
- Skripov, V.P., 1974, *Metastable liquids* (Wiley, New York).
- Stoll, E. and T. Schneider, 1972, *Phys. Rev.* **A6**, 429.
- Sur, A., J.L. Lebowitz, J. Marro and M.H. Kalos, 1977, *Phys. Rev.* **B15**, 535.
- Sykes, M.F. and J.W. Essam, 1964, *Phys. Rev.* **I33A**, 310-315, and *J. Math. Phys.* **5**, 1117-1127.
- Thompson, C.J., 1972, *Mathematical statistical mechanics* (MacMillan, New York), ch. 5 and 6.
- Volmer, M. and A. Weber, 1926, *Z. Phys. Chem.* **119**, 277-301.
- Widom, B. and J.S. Rowlinson, 1971, *J. Chem. Phys.* **52**, 1670.
- Wong, N.C. and C.M. Knobler, 1977, *J. Chem. Phys.* **66**, 4707.
- Yamauchi, H. and D. de Fontaine, 1974, in *Order-disorder transformations in alloys* (H. Warlimont, ed.) (Springer, Berlin).

### References added in proof

- Abraham, F.F., 1979, On the thermodynamics, structure and phase stability of the nonuniform fluid state, to appear in *Phys. Rep.*
- Binder, K. and M.H. Kalos, 1979, Critical clusters in a super-saturated vapor: theory and Monte Carlo simulation, preprint.
- Goldberg, W.I., C.H. Shaw, J.S. Huang and M.S. Pilant, 1978, *J. Chem. Phys.* **68**, 484.
- McGraw, R. and H. Reiss, 1979, *J. Stat. Phys.* **20**, 385.
- Muzrik, M.R., F. Abraham and G.M. Pound, 1978, *J. Chem. Phys.* **69**, 3462.
- Penrose, O., J.L. Lebowitz, J. Marro, M.H. Kalos and A. Sur, 1978, *J. Stat. Phys.* **19**, 243.
- Penrose, O. and A. Buhagiar 1979, preprint.
- Siggia, E.D., 1979, Late stages of spinodal decomposition in binary mixtures, preprint.
- Zurek, W.H. and W.C. Schieve, 1978, Clustering and nucleation: a molecular dynamics study, preprint.