PHYSICAL REVIEW B VOLUME 14, NUMBER 8 15 OCTOBER 1976

Percolation and cluster distribution. I. Cluster multiple labeling technique and critical concentration algorithm*

J. Hoshen and R. Kopelman
Department of Chemistry, The University of Michigan, Ann Arbor, Michigan 48109
(Received 1 June 1976)

A new approach for the determination of the critical percolation concentration, percolation probabilities, and cluster size distributions is presented for the site percolation problem. The novel "cluster multiple labeling technique" is described for both two- and three-dimensional crystal structures. Its distinctive feature is the assignment of alternate labels to sites belonging to the same cluster. These sites are members of a simulated finite random lattice. An algorithm useful for the determination of the critical percolation concentration of a finite lattice is also presented. This algorithm is especially useful when applied in conjunction with the cluster multiple labeling technique. The basic features of this technique are illustrated by applying it to a small planar *square* lattice. Numerical results are given for a *triangular* subcrystal containing up to 9000000 sites. These results compare favorably with the exact value of the infinite lattice critical percolation concentration.

I. INTRODUCTION

The dynamic and static aspects of percolation phenomena are well-known and to a large extent understood. The concept of percolation has been closely associated with the permeation of fluid through porous media. The formation of an infinite cluster of identical molecules in a multicomponent crystal can also be described within the framework of percolation theory. Various examples of the phenomenon and its applications are enumerated in the literature. These include among others: Spontaneous magnetization of dilute ferromagnets,[1] spreading of diseases in orchards,[2] formation of polymer gels,[3] electrical conductivity of amorphous semiconductors and metal-ammonia solutions,[4] distribution of grain size in photographic emulsions,[5] and exciton percolation in isotopically disoredered crystals and photosynthetic units.[6] It should be noted that this list includes only a small "random sample" of a variety of percolation processes occurring in nature.

Various mathematical and computational approaches to the problem, including their scope and limitations, have been outlined in numerous papers and review articles.[7] The computational simulation methods, broadly termed as Monte Carlo methods,[2,8] are applied to finite sublattices, whereas percolation theory, basically, is concerned with the probability for locating an infinite cluster in an infinite lattice. The question is how to treat an infinite cluster within the framework of a finite lattice. Furthermore, surface effects which are encountered in a finite lattice do not pertain to an infinite lattice.

In this paper we shall introduce a novel method suitable for computer simulations and applicable for a fast and accurate determination of cluster distributions, critical percolation concentration, and percolation probabilities. The success of the method, and its versatility in solving complex percolation problems, is based on the application of alternate labels to sites belonging to the same cluster. The basic features of the "cluster multiple labeling technique" and its application to the site percolation problem[9] are described in Sec. II. In Sec. II we shall also elaborate on some variations of the basic method. In Sec. III we discuss a new practical algorithm appropriate for the determination of the critical percolation concentration for a finite randomly mixed binary crystal. A simple demonstration of the application of the method is presented in Sec. IV. Also in this section, results are given for the critical site percolation concentration, percolation probabilities, and cluster size distributions of a triangular lattice. The results for the critical percolation concentration are compared with the known exact value of this parameter for an infinite triangular crystal.[10]

II. DETERMINATION OF CLUSTER DISTRIBUTIONS

A crystal containing two types (A and B) of randomly distributed molecules is considered, where the concentration of type A is c. The probability that a randomly selected site is occupied by an A molecule is c, whereas the probability for a B molecule is $1-c$. We shall focus our attention on a finite section of the crystal which we shall denote as the subcrystal. It will be our goal to label subcrystal sites, and to classify and count clusters of A molecules in the subcrystal.

Each crystal site i, which is occupied by an A molecule, will be assigned a cluster label m_i^{α}, where α is a symbolic name for the cluster in question. A cluster α may be assigned several

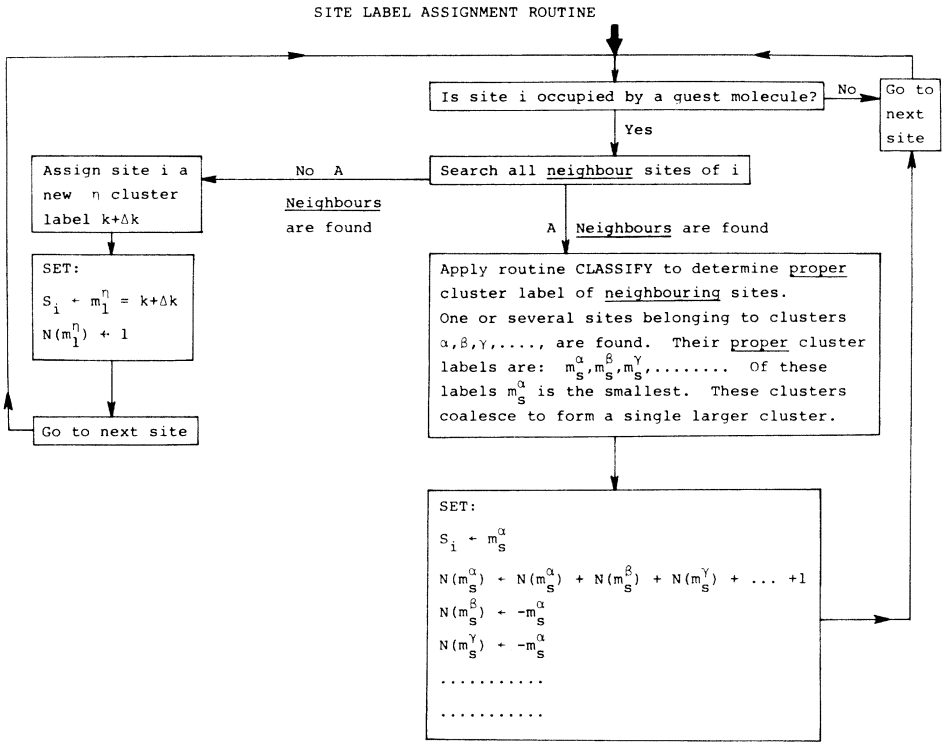

FIG. 1. Site label assignment routine. This routine labels sites occupied by A molecules and readjusts the $N(m_t^\alpha)$ numbers. The notation $D \leftarrow E$ denotes the transfer of the content of E into D, without affecting E.

cluster labels. These are given as a set of natural numbers:

$$\{m_1^\alpha, m_2^\alpha, \ldots, m_s^\alpha, \ldots, m_t^\alpha, \ldots\}. \quad (1)$$

In this set only one number is regarded as a *proper* cluster label which we shall designate as m_s^α. This is the smallest number of the set Eq. (1). The following set of integers provides the connections between the m_t^α labels:

$$\{N(m_1^\alpha), N(m_2^\alpha), \ldots, N(m_s^\alpha), \ldots, N(m_t^\alpha), \ldots\}. \quad (2)$$

In Eq. (2), $N(m_s^\alpha)$ is the only positive integer member of the set, and denotes the number of A molecules in the cluster. The remaining members of the set Eq. (2) are negative integers, providing links between the other m_t^α labels and the *proper* label m_s^α. The m_t^α labels are related to the m_s^α label by the following type of equation set:

$$m_r^\alpha = -N(m_t^\alpha), \quad m_q^\alpha = -N(m_r^\alpha), \quad \ldots, \quad \ldots,$$
$$m_s^\alpha = -N(m_p^\alpha). \quad (3)$$

These equations are solved from left to right (see example in Sec. IV). We have learned from our experience that the hierachy of Eq. (3), in most cases, consists of one or two levels (one or two equations only).

The disordered subcrystal is simulated by generating pseudorandom numbers in the range $0 < X < 1$. A site i is occupied by an A molecule if $X < c$; otherwise it is occupied by a B molecule. Sites are filled with molecules in a consecutive order, site after site, column after column (and layer after layer in a three-dimensional lattice), until all the subcrystal sites are occupied. Site labeling and cluster classification can be performed simultaneously with subcrystal creation. Sites occupied by B molecules are labeled with zeros, whereas sites occupied by A molecules are labeled with natural numbers in accordance with Eq. (1). The assignment of cluster labels to subcrystal sites is illustrated in Fig. 1. The first time an A molecule is encountered during the inspection of the subcrystal the label counter k is set to an initial value k_0. The increment Δk is a small positive integer, whose value is set in accordance with the computer language used. Previously labeled sites, in the *neighborhood* of a newly created A molecule at a site i, are searched by routine CLASSIFY which is described in Fig. 2. The *neighbors* of site i are defined as those sites connected by bonds[11] to site i. The number of *neighbors* is equal to the bond order of site i. The *neighborhood* includes all the neighbors of site i. If there are no A *neighbors* in the vicinity of site i, the site is assigned a new label after k is incremented. If there is only one A neighbor (at site n) site i gets the same label as site n (see Fig. 1). The unique features of the cluster multiple labeling technique become apparent when the site i, which is occupied by an A molecule, links two or

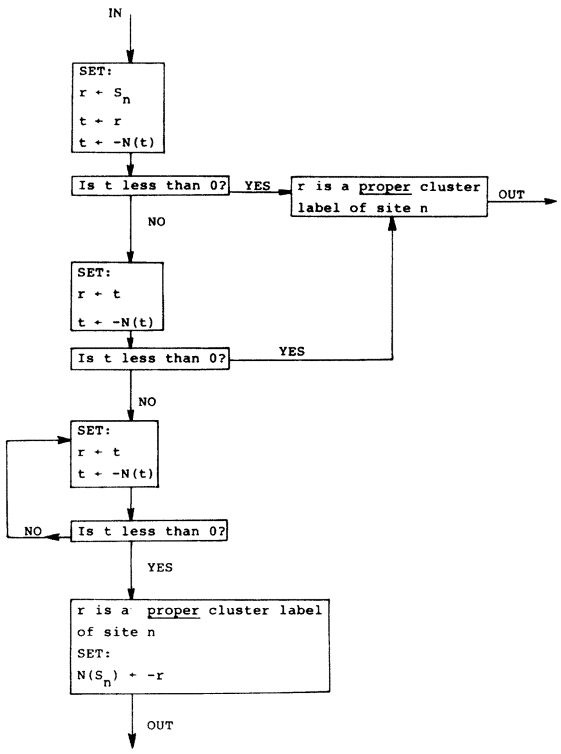

FIG. 2. Routine CLASSIFY. This routine determines the *proper* cluster labels for sites containing A molecules. S_n is *neighbor* site n.

more previously labeled cluster fragments into a single cluster. No site belonging to any of these cluster fragments is relabeled and once a subcrystal site is labeled, it retains its original label throughout the simulation process. The actual readjustments occur within the $N(m_i^\alpha)$ sets [see Eq. (2)], as clusters are newly created, grown, or coalesced in the simulation process. The number of readjusted $N(m_i^\alpha)$ numbers for site i is equal to the number of coalescing cluster fragments at the site i (see Fig. 1).

In summation, site labeling and cluster size classification can be accomplished by a single scan of the subcrystal. The sizes of the clusters are given by the positive $N(m_s^\alpha)$ numbers. The additional advantage of the single scan process is that only a fraction of the columns (or layers) have to reside in the computer memory concurrently. The number of stored columns (or layers) in the computer memory is a function of the site bond order.[11] The entire subcrystal need not be stored on an external input-output device because after a segment of columns (or layers) have been processed they are not referred to again by the program.

There are many possible variations or versions of the basic method described above. The choice of a particular version depends on the specific application and the computer facilities. The following list describes some of the available options:

(a) It is possible to select the largest member of the set Eq. (2) as the *proper* cluster label, rather than the smallest member of the set. This could reduce the range of labels applied to a particular subcrystal column (or layer) at program execution time.

(b) For ASSEMBLER IBM 360 usage, it is preferable to set $N(m_s^\alpha)$ to be negative, the remaining members of set Eq. (2) to be positive, and $\Delta k = 4$. Cluster labels can be equated with actual core addresses.

(c) Subcrystal columns (or layers) can be divided into sections where in each section the same set of labels can be used. This reduces the number of labels required for a very large subcrystal and also reduces the computer storage allocation for the $N(m_i^\alpha)$ sets. Clusters extending from one section to the next section should be assigned new *proper* cluster labels upon entering the new section.

(d) Cyclic boundary conditions for the subcrystal can be incorporated into the method in a simple manner. (Additional internal or external storage would be required for the boundary columns or layers).

III. CRITICAL PERCOLATION CONCENTRATION ALGORITHM

The difficulties in defining the critical percolation concentration of a finite lattice were briefly discussed in Sec. I. A comprehensive discussion of this topic was given by Dean.[8] We have attempted to give a practical solution to the problem, by utilizing the *reduced averaged cluster size*, which we define as

$$I'_{\rm av} = \left(\sum_{n=1}^{n_{\max}} i_n n^2 \right) / G - n_{\max}^2 / G, \qquad (4)$$

where G denotes the total number of A molecules in the subcrystal, and i_n is the number of clusters of size n in the subcrystal. We have subtracted from the summation the contribution of the largest cluster, whose size is $n_{\max}$. By plotting $I'_{\rm av}$ versus the concentration of A we find that $I'_{\rm av}$ exhibits a very sharp maximum in the region of the critical percolation concentration. Fisher and Essam[12] introduced for an infinite crystal the mean cluster size density $S(c)$ function, which resembles in shape our $I'_{\rm av}$ function. In their approach the contribution of the infinite cluster is excluded from $S(c)$ for all concentrations *above* the percolation concentration. We, on the other hand, exclude the contribution of the largest cluster from $I'_{\rm av}$ for *all*

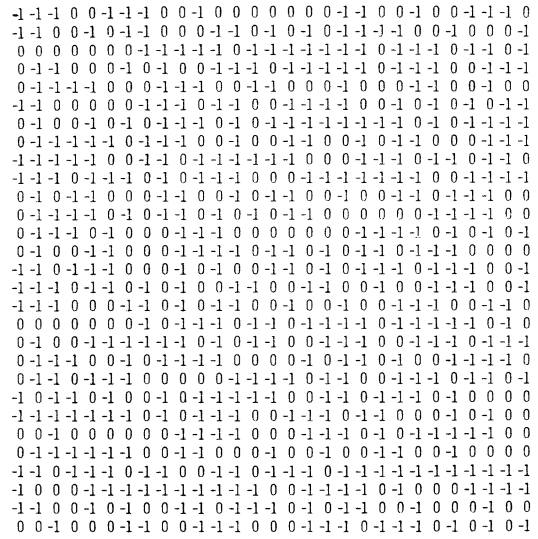

FIG. 3. A binary, substitutionally random square lattice containing 29×29 sites, where the nominal concentration of the A molecules is 0.57. Here -1 denotes an A molecule, and 0 denotes a B molecule.

concentrations, since we cannot specify an "infinite" cluster (in our subcrystal) that can be mathematically distinguished from the finite clusters. However, it should be noted that the difference resulting from the exclusion of the contribution of a single *finite* cluster, including the largest, from an average pertaining to an *infinite* crystal, is negligible. It should be small even in the case of a finite subcrystal *below* the critical percolation concentration. Thus below this critical point I'_{av} would increase with concentration. Above the critical concentration the largest cluster grows rapidly, leading to a sharp decline in the value of I'_{av}. This decline resembles the decline of[12] $S(c)$ above the critical percolation concentration, but is steeper due to our different "normalization," i.e., division by G throughout the full concentration region (while Fisher and Essam exclude from G the contribution of the infinite cluster).

IV. NUMERICAL EXAMPLES

As the cluster multiple labeling technique involves some complex logics, a simple example may illustrate the important sequence of steps invoked by the technique. We shall treat a square lattice, shown in Fig. 3, containing 29×29 molecular lattice sites, where the concentration of the A species is 0.57. In Fig. 3, a symbol 0 denotes a B molecule in the disordered lattice, whereas an A molecule is symbolized by a -1. Due to fluctuations in the pseudorandom numbers the actual concentration of A is not exactly 0.57. The labeling of A sites, which is illustrated in Fig. 4, begins with the leftmost column. The first molecule in this column is of type A and is labeled 1. The second molecule is also labeled 1. The next three molecules are of the B type and their label is 0. These B molecules are followed by a single A molecule which is not within the bond range of cluster 1 and hence it is labeled by 2. The possibility of cluster coalescence does not exist for the first column, so its labeling is very simple. The first and the second sites of the second column contain molecules of type A, and the search of their *neighbor* sites in the first column reveals the label 1, so these two sites are also labeled by 1, and $N(1)$ is now set to 4. The next two sites are labeled by 7, as all previously labeled sites in their *neighborhood* are labeled with 0. Cluster fragment linking begins at the next site, since one *neighbor* site is labeled by 7 and the other by 2. The *proper* cluster label is the smaller of the two and is set to 2. The linking site is labeled by 2. The readjustments of the N set [see Eq. (2)], involving labels 2 and 7 are [see Eq. (3)] $N(7)=-2$ and $N(2)=4$. The cluster fragments labeled 3 and 4 also coalesce with 2 during the inspection of the second column. In the succeeding columns the cluster fragments labeled 11 and 14 link with 2. Finally, the total number of A molecules belonging to the cluster, which is denoted by the *proper* label 2, can be determined from $N(2)$ by observing that $N(2)=45$. The largest cluster in the subcrystal retains the *proper* label 5. Here $N(5)=209$. In order to obtain a better perception of the cluster structure of the subcrystal the reader is referred to Fig. 5.

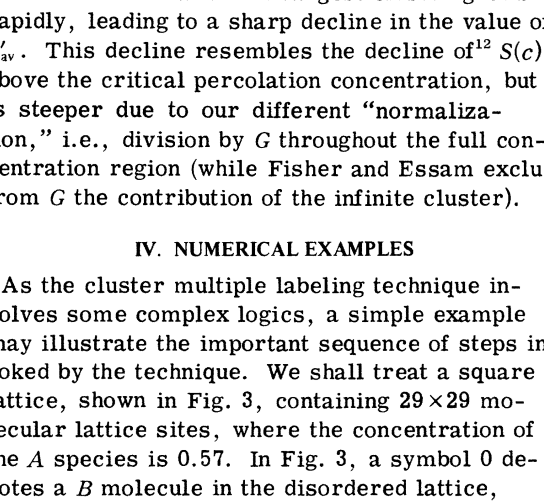

FIG. 4. The A sites of Fig. 3 are labeled with natural numbers by utilizing the cluster multiple labeling technique. The bond order for the sites is 4. The increment $\Delta k=1$ and $k_0=0$.

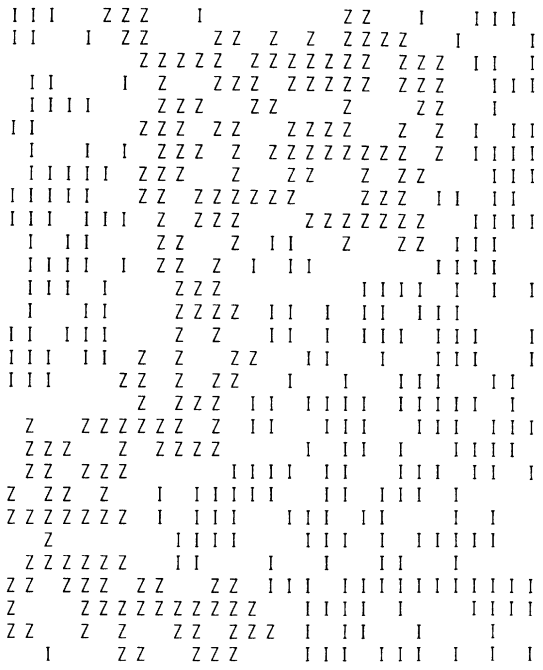

FIG. 5. A display of the cluster structure for the lattice given in Fig. 3. A Z denotes a site occupied by a molecule belonging to the largest A cluster. Other A clusters are represented by I's. The B molecules are denoted by blanks.

Figures 6, 7, and 8 of this paper display the results of an actual test of the technique. These results are also valuable in providing an insight to the actual operation of the technique. Figures 6,

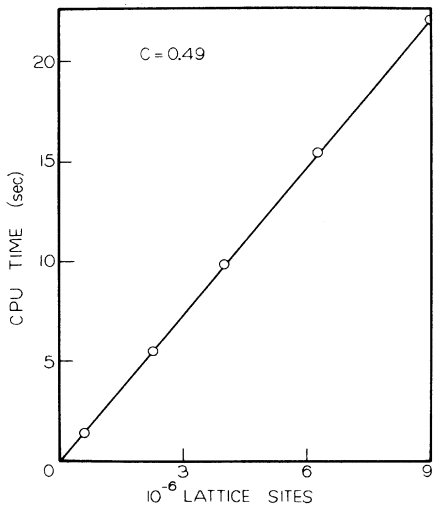

FIG. 6. A plot of the computer central-processor-unit (CPU) time *versus* the number of sites processed in the run for a triangular subcrystal. The circles denote data points for 750 × 750, 1500 × 1500, 2000 × 2000, 2500 × 2500, and 3000 × 3000 subcrystals. The Michigan computer is an Ahmdal 470.

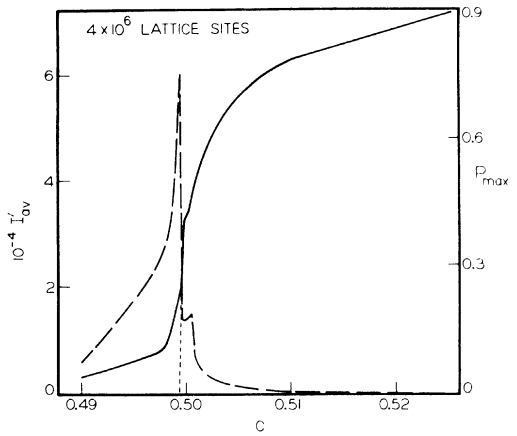

FIG. 7. I'_{av} [see Eq. (4)] and $P_{max} = n_{max}/G$, denoted by the long-dashed line and the solid line, respectively, versus c, the concentration, for a triangular crystal containing 2000 × 2000 sites. The short-dashed line denotes the interpolation range of P_{perc}.

7, and 8 apply to a planar triangular lattice where each site is surrounded by six nearest neighbors. The order of the cluster bonds is chosen to be six, corresponding to the six nearest neighbors. Because we are concerned with short-ranged bonds only, only two lattice columns have to be stored concurrently in the computer core. The basic method together with options (a), (b), and (c) as described in Sec. II are utilized. The application of ASSEMBLER language has been preferred because the program written in this language ran about fifteen times faster than an equivalent FORTRAN program (on the Michigan Ahmdal 470 computer).

The efficiency of the method is clearly demonstrated in Fig. 6, where the central processor time required for the subcrystal simulation *and* cluster classification is plotted versus the number of sites in the subcrystal. While the mere 22 sec required to process a subcrystal containing 9 000 000 sites illustrate the extreme efficiency of the method, it is the almost perfect linearity of the time curve that represents the most impressive feature of the technique. It should be noted that the computer run corresponding to the 9 000 000 sites utilizes 6002 4-byte words for a temporary storage of sites and 25 000 4-byte words for the storage of the $N(m_t^\alpha)$ vector. The 25 000 labels m_t^α, which comprise an arithmetic series with $\Delta k = 4$, have been used 27 times for this particular run in accordance with option (c) of Sec. II.

The $P_{max}(c)$ and the $I'_{av}(c)$ curves, which provide the pertinent information on the percolation probability and the critical percolation concentration c^*, respectively, are displayed in Fig. 7. The I'_{av} curve exhibits a sharp maximum at $c = 0.4995$,[13]

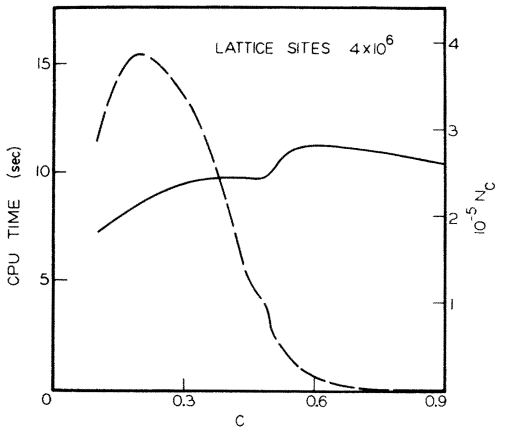

FIG. 8. CPU time (solid line) and the total number of clusters (dashed line) in the subcrystal, N_c, *versus* the concentration c. Triangular lattice (bond order six).

very close to the exact infinite lattice value of 0.5 for c^* given by Sykes and Essam.[10] The decline of the I'_{av} curve is very steep after it achieves its maximum value. There is a small kink at $c = 0.5005$, reflecting strong fluctuations in the sizes of clusters of intermediate size in the region of c^*. If we assume $c = 0.4995$ to be the critical percolation concentration c^* for our particular subcrystal, then P_{max} can represent the percolation probability P_{perc} for $c > c^*$. As $P_{perc}(c^*) = 0$ by definition we can interpolate P_{perc} in the region $c^* < c < c^* + \delta c$, where δc denotes a small increment. The interpolation is represented in Fig. 7 by the short-dashed line.

In Fig. 8, data are given on the total number of clusters N_c in the subcrystal and on the central-processor-unit (CPU) time required to process the subcrystal as function of c. As can be observed from the figure there is very little correlation between the two curves. The maximum number of clusters of all sizes and shapes in the subcrystal occurs at a concentration close to $c = 0.20$. The effort required to process the subcrystal is reflected in the CPU time curve (solid line). The CPU time increases slowly in the region 0.1–0.35. This region is followed by a plateau. A sharp rise in the CPU time can be observed at a concentration somewhat above c^*, indicating that the largest cluster in this region is becoming more complex.

The cluster distribution for a subcrystal containing 4 000 000 sites is given in Table I. The significant feature of this distribution is that there are many clusters of intermediate and large size in the region of c^*. For $c = 0.525$, clusters above 1000 are absent with the exception of the largest cluster which contains 1 874 653 A molecules.

V. DISCUSSION

Several years ago, it was stated by Hammersley and Handscomb[14] that a solution to the percolation problem utilizing the direct simulation approach " is out of the question," as they estimated that it would be necessary to keep a high-speed computer

TABLE I. Cluster size distribution for a 2000×2000 triangular lattice (bond order six).

	Number of clusters within a specified cluster size range [a]									
c	1 to 1000	1001 to 2000	2001 to 6000	6001 to 15 000	15 001 to 50 000	50 001 to 100 000	100 001 to 200 000	200 001 to 500 000	500 001 to 1 000 000	1 000 001 to 4 000 000
0.40	209 072	0	0	0	0	0	0	0	0	0
	1 599 456									
0.45	134 450	34	0	0	0	0	0	0	0	0
	1 758 222	41 479								
0.48	93 526	190	117	15	1	0	0	0	0	0
	1 151 159	271 436	373 307	108 058	15 643					
0.495	76 315	67	62	23	12	4	1	0	0	0
	681 904	90 443	222 831	229 447	315 708	293 580	145 683			
0.50	71 118	51	18	14	7	0	1	0	1	0
	567 824	71 674	64 531	136 815	198 652		138 955		821 087	
0.505	66 287	32	5	3	2	0	0	0	0	1
	470 284	43 452	13 671	26 668	57 792					1 407 500
0.525	50 091	0	0	0	0	0	0	0	0	1
	224 364									1 874 653

[a] The first line for each concentration c denotes the number of clusters within the specified cluster size range, whereas the underlined numbers denote the total number of molecules of type A belonging to the clusters of the specified size range.

busy for 50 yr to accomplish the task. Hammersley and Handscomb came to the above conclusion by estimating the computer time that would be required to determine the portion of "wet sites" (sites belonging to the largest cluster, in our terminology) contained in a simulated cube of 8 000 000 sites. Their estimates were made for the bond percolation problem, considering approximately 300 separate computer experiments.[15]

The results of Sec. IV, as well as our experience with a variety of percolation simulations for two- and three-dimensional lattices[16] disprove Hammersley and Handscomb's[14] strong statement on the feasibility of the direct simulation approach to the percolation problem. Given the improvement in computer technology in recent years, which would shorten the original 50-yr estimate[14,15] to about a year, the reduced time span would still be several orders of magnitude longer than the *minutes* it would take to perform a complete concentration scan on a simulated lattice containing several million sites, utilizing the cluster multiple labeling technique.

As was pointed out in Sec. I., there have been available two basically different simulation approaches to the percolation problem[2,8] besides the one given in this paper. The method presented here shares some common features with Dean's Monte Carlo method.[8] Furthermore, the *entire cluster distribution* is determined by both methods while Broadbent and Hammersley's[2] Monte Carlo method is concerned only with the largest cluster. Dean[8] proposed to use the maximum slope of the "modified second moment of cluster distribution" as a criterion for determining the critical percolation concentration c^* of a finite crystal. His reasoning for choosing this criterion is very similar to our rationale for selecting the peak of I'_{av} to define c^*. The difficulties associated with the application of Dean's method are related to the assignment of a single label to each cluster. The use of a single cluster label involves repeated scanning and relabeling of subcrystal sites,[3] which may be impractical for large subcrystals as such a process requires both considerable storage facilities and an extended computer execution time.

Broadbent and Hammersley's[2] Monte Carlo method offers some advantages for the calculation of percolation probabilities above c^*. However, in the vicinity of c^* this method suffers from some difficulties.[17] The existence of relatively large clusters just below c^*, as can be observed from Table I, may explain some of the difficulties reported by Frisch *et al.*[17] for two-dimensional lattices.

In a future paper[16] we shall present results on a percolation problem with large bond order (which was utilized for triplet exciton migration in molecular crystals[18]). In addition, data will be given for three-dimensional subcrystals with varied thicknesses and on several cluster distribution functions.

ACKNOWLEDGMENT

One of the authors (Hoshen) would like to thank David Frank for his linguistic assistance.

*Supported by NSF grant DMR75-07832 A01 and NIH grant NS08116-08.

[1]R. J. Elliott, B. R. Heap, D. J. Morgan, and G. S. Rushbrooke, Phys. Rev. Lett. 5, 366 (1960).

[2]S. R. Broadbent and J. M. Hammersley, Proc. Camb. Philos. Soc. 53, 629 (1957).

[3]P. J. Flory, *Principles of Polymer Chemistry* (Cornell U.P., Ithaca, N.Y., 1953), Chap. 9.

[4]T. P. Eggarter and M. H. Cohen, Phys. Rev. Lett. 27, 129 (1971); M. H. Cohen, I. Webman and J. Jortner, J. Chem. Phys. 64, 2013 (1976).

[5]F. Kottler, J. Franklin Inst. 250, 339 and 419 (1950).

[6]R. Kopelman, E. M. Monberg, F. W. Ochs, and P. N. Prasad, Phys. Rev. Lett. 34, 1506 (1975); R. Kopelman, J. Lumin. 12, 775 (1976); J. Phys. Chem. (to be published).

[7](a) H. L. Frisch and J. M. Hammersley, J. Soc. Ind. Appl. Math. B 11, 894 (1963); (b) M. E. Fisher, Proceedings of the IBM Scientific Computing Symposium on Combinatorial Problems (New York, 1964); (c) J. W. Essam, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and M. S. Green (Academic, New York, 1973).

[8]P. Dean, Proc. Camb. Philos. Soc. 59, 397 (1963).

[9]The bond percolation problem can always be transformed to the site problem on another lattice, as shown by Dean (see Ref. 8).

[10]M. F. Sykes and J. F. Essam, Phys. Rev. Lett. 10, 3 (1963).

[11]A cluster defines a set of identical entities, occupying crystal sites, connected by a set of bonds (line segments drawn between the sites). Out of each equivalent lattice site originates the same number of bonds. This number is denoted as the site bond order. When sites on both ends of a bond are occupied by the same entity, these two entities are said to belong to the same cluster.

[12]M. E. Fisher and J. W. Essam, J. Math. Phys. 2, 609 (1961).

[13]The same set of "random numbers" are used for each concentration scan, so that as c increases to $c+\Delta c$, some B molecules transform into A molecules, whereas the original A molecules (at concentration c) retain their lattice positions. Utilizing two other sets of random numbers for two 2000×2000 sites triangular subcrystals, we obtained the following values for c^*:

0.50045, 0.5011.
[14]J. M. Hammersley and D. C. Handscomb, *Monte Carlo Methods*, (Methuen, London, 1964), p. 135.
[15]J. M. Hammersley, Methods Comput. Phys. 1, 281 (1963).
[16]J. Hoshen, R. Kopelman, and E. M. Monberg (unpublished).
[17]H. L. Frisch, J. M. Hammersley, and D. J. A. Welsh, Phys. Rev. 126, 949 (1962).
[18]R. Kopelman, E. M. Monberg, F. W. Ochs, and P. N. Prasad, J. Chem. Phys. 62, 292 (1975); J. Hoshen and R. Kopelman, J. Chem. Phys. (to be published); R. Kopelman, in *Topics in Applied Physics, Vol. 15: Radiationless Processes in Molecules and Crystals*, edited by F. K. Fong (Springer, New York, 1976); R. Kopelman, E. M. Monberg, and F. W. Ochs (unpublished).