On the cluster size distribution for critical percolation

1976 J. Phys. A: Math. Gen. 9 L9

(http://iopscience.iop.org/0305-4470/9/1/003)

J. Phys. A: Math. Gen., Vol. 9, No. 1, 1976. Printed in Great Britain. © 1976

LETTER TO THE EDITOR

On the cluster size distribution for critical percolation

George D Quinn, George H Bishop and Ralph J Harrison
Army Materials and Mechanics Research Center, Watertown, Massachusetts 02172, USA

Received 3 December 1975

Abstract. A Monte Carlo study of clustering in the vicinity of the critical percolation probability p_c has been carried out on the body-centered cubic lattice in which both first- and second-neighbour sites were counted. The total number of clusters shows an inflection point when plotted against concentration of occupied sites in the vicinity of p_c suggesting that a characteristic of p_c may be a maximum rate of loss of clusters due to aggregation. An empirical form for the cluster size distribution in the critical region has been found which matches the present data as well as previous results on other lattice geometries obtained by Dean and Bird. At large cluster sizes the distribution at p_c becomes asymptotically an inverse power law with exponent slightly larger than two. An argument is presented which relates this exponent to the linear extent of the cluster.

This letter reports certain features of the cluster size distribution in the vicinity of the critical percolation probability p_c obtained in a Monte Carlo study of clustering on the body-centred cubic lattice in which first- and second-neighbour sites were counted (the BCC 1–2 site problem). As will be indicated however, much of the discussion is applicable to more general geometries. The computational method uses a cubic cell of 128 000 lattice sites, each site randomly occupied or unoccupied with probability p or $(1-p)$ respectively. A search then obtains the size† and topology of each cluster, utilizing periodic boundary conditions to continue the search across the cell faces. The method is similar to that of Dean and Bird (1966) in overall approach but differs in the way the clusters are generated and examined; Dean and Bird do not examine cluster topology. A minimum of three statistically independent repetitions of the computation were made at each value of p. Fifteen repetitions were made at $p = 0{\cdot}175$ which is the value given by Domb and Dalton (1966) for the most probable value of p_c. Further details of the computational method as well as the topological results will be presented separately (Quinn *et al* 1976).

Averages for quantities such as $c(n, p, N)$, the number of clusters of size n, given p and the sample size N, and for the total number of clusters $T(p, N)$ and the mean cluster size $S(p) = \Sigma n^2 c(n, p, N)/\Sigma nc(n, p, N)$ were obtained from the Monte Carlo sampling procedure. These quantities can also be evaluated using theoretical results for $\chi(n, p) = \lim_{N\to\infty} c(n, p, N)/N$. Polynomial expressions in p and $(1-p)$ have been obtained by enumerating the various cluster configurations for each n for a number of lattices for relatively small values of n (Sykes and Glen 1976). For the BCC 1–2 site problem these polynomials are available up to $\chi(8, p)$ (Domb and Sykes, private communication). We have obtained good agreement between our Monte Carlo data and the theory.

† As is conventional in percolation studies size is taken as the number of sites in the cluster. There is currently a growing interest in the shape, lineal extent and topology of clusters (Domb *et al* 1975, Quinn *et al* 1976). Recent reviews of percolation theory include those of Shante and Kirkpatrick (1971) and Essam (1972).

Computational results for the average of the total number of clusters $T(p, N)$ are given in figure 1. As p increases unoccupied sites are progressively occupied. A newly occupied site may fall in one of three categories: (i) it may be isolated, increasing T by

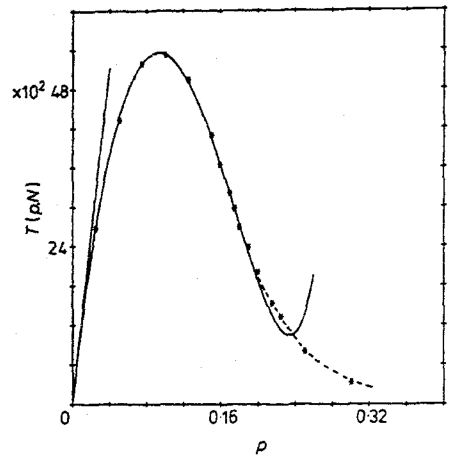

Figure 1. Computational results for the average of the total number of clusters $T(p, N)$. The full curve gives the series expansion and the broken curve with crosses gives the Monte Carlo data. $N = 128\,000$.

one; (ii) it may be a neighbouring site to a single existing cluster so that T is unchanged; or (iii) it can link two to six existing clusters, decreasing T by one or more. At small p the first possibility predominates and $T \cong pN$ (the straight line). As p increases, the probability of forming larger clusters increases and T passes through a maximum when the rate of new cluster formation is just balanced by the rate of cluster linkage. The total number of clusters then decreases, has an inflection point, and approaches zero asymptotically. The rapid decrease of T with p beyond the maximum is due to the increased probability that newly occupied sites will link existing clusters. An expansion of T as a power series in p to order m may be obtained from a knowledge of the polynomials $\chi(n, p)$ through order $n = m$. This expansion in T may also be obtained more directly (Essam 1972). To order 10 it yields

$$T/N = p-7p^2+12p^3-3p^4-6p^5+77p^6-206p^7+1144p^8-3180p^9+17367p^{10}. \qquad (1)$$

This polynomial which is plotted as the full curve in figure 1 yields good agreement with the Monte Carlo data for p less than 0·20.

An interesting feature of the T–p curve is the inflection which occurs very close to, and possibly precisely at, p_c. This suggests the following properties: the rate of loss by linking of clusters is a maximum at p_c and therefore for a region around p_c this rate is nearly constant. In turn this suggests that the cluster size distribution characteristic of p_c is also nearly constant in some range of p above and below p_c, in agreement with features of the size distribution data to be presented shortly. The evidence that the inflection occurs exactly at p_c can be only suggestive on the basis of numerical computations; we may point out that the inflection points obtained from equation (1) by truncations at the fourth through tenth powers of p are $p = 0·218$, 0·254, 0·185, 0·227, 0·180, 0·199, 0·177 respectively. The even-order terms appear to approach the value 0·175 more quickly and closely than the odd terms.

Characterization of the cluster size distribution near p_c by numerical techniques is difficult since most interest centres on the large clusters for which the statistical fluctuations are most troublesome. The large clusters are also difficult to investigate by evaluation of the cluster polynomials since the number of possible configurations increases exponentially with n. In the present work with 128 000 sites and with the order of 15 trials the individual values of $c(n, p, N)$ are unreliable for n much greater than 15. The statistics can be greatly improved however by examining the partial sums $C_r(i, p, N) = \Sigma(n = r^{i-1}, r^i - 1)c(n, p, N)$. We found that the choice $r = 2$ provides a good compromise between resolution and scatter. Figure 2 shows the data for $p = 0{\cdot}175$ and for a few values of p above and below this.

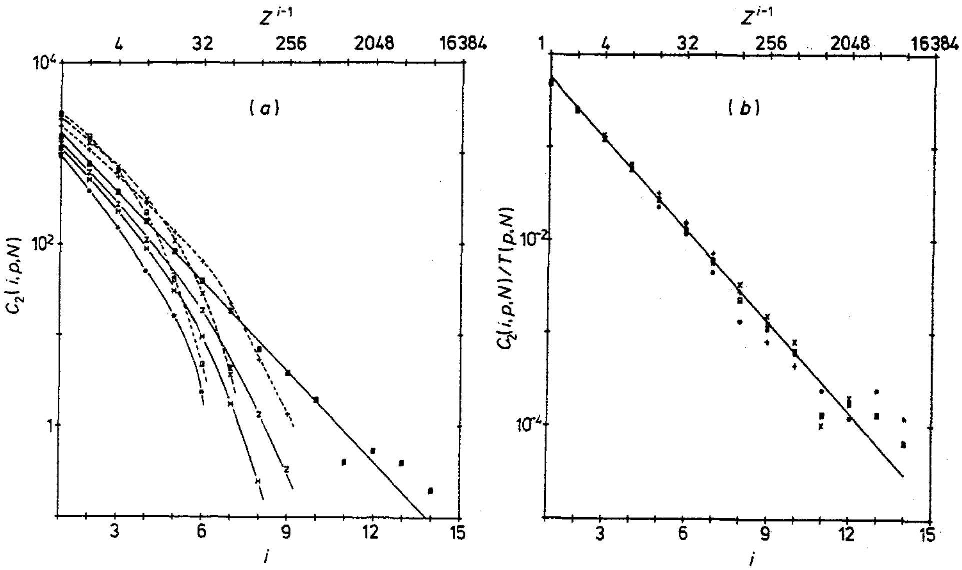

Figure 2. (*a*) Plot of $C_2(i, p, N)$ against i for $N = 128\,000$ and various values of p: □, 0·10; ×, 0·125; +, 0·15; ■, 0·175; Z, 0·19; H, 0·20; ∗, 0·215. (*b*) Plot of $C_2(i, p, N)/T(p, N)$ for various values of p: +, 0·16; ×, 0·17; □, 0·175; ∗, 0·18.

If we refer to the distribution at precisely p_c as canonical, we see in figure 2(*a*) that for small values of p the curves fall off† rather sharply from the canonical curve at relatively small sizes. As p increases a flatter region develops at lower cluster sizes before falling off at larger sizes. As p_c is approached one approaches the canonical distribution which appears quite linear in this plot. For p in the immediate vicinity of p_c the distribution curves are very nearly parallel to the canonical distribution up to quite large cluster sizes, see figure 2(*b*) in which $C_2(i, p, N)$ has been normalized by division by $T(p, N)$. For values of p greater than p_c the distribution falls off from canonical quite similarly to the way it does below p_c. The fact that the distributions both above and below the critical probability fall off much more rapidly for large n than does the

† This fall off is consistent with the exponential increase of the average cluster size $S(p)$ near critical. The BCC 1–2 lattice data give $S(p) \cong A(p_c - p)^{-\gamma}$, with $A \sim 160$ and $-\gamma = 1{\cdot}7$, and the exponent agrees well with estimates based on series expansions (Essam 1972).

canonical distribution makes the latter the distribution of 'maximum richness' for large clusters. Although in the numerical computations the finite value of N will cause a truncation of the larger clusters, one might expect the above limiting behaviour of the canonical distribution in the limit of infinite N. For the data near p_c two effects of the finite sample size are that no clusters are observed in groups with $i > 14$, and the $C_2(i, p, N)$ for groups $i = 13, 14$ appear to be too high, presumably because truncation of large clusters pushes them into lower size groups.

The linearity of the canonical distribution in figure 2 reveals what appears to be a general property of the partial sums $C_r(i, p, N)$, namely that they are approximated quite closely for large i and for N much larger than r^i by a simple law of the form

$$C_r(i, p_c, N) = \text{constant}/\theta^i \tag{2}$$

in which θ is a constant dependent on r and geometry which can be expressed as $\theta = r^{(1+\delta)}$, with δ a small positive constant varying with geometry. The best fit to the BCC 1–2 site data yields $\delta = 0{\cdot}0938$. For $i = 1$ and 2, with $r = 2$, the Monte Carlo data deviate from the power law relationship of equation (2). This appears to be an intrinsic property of the canonical cluster size distribution, since the exact values calculated from $\chi(n, p)$ and the Monte Carlo data agree almost precisely here. For these values of i only simple clusters are possible and they differ from the typical highly ramified (Domb *et al* 1975) larger clusters which show highly multiple connected topology (Quinn *et al* 1976).

An approximate inversion of equation (2) to extract the form of the cluster size distribution itself may be obtained in the form

$$c(n, p_c, N) = \frac{aN}{n(n+1)[\frac{2}{3}(2n+1)]^\delta}. \tag{3}$$

With $a = 0{\cdot}0283$ the partial sums of this inversion, plotted as the straight line in figure 2 deviate from equation (2) by less than 0·1% and the individual values for $c(n)$ are in excellent agreement with the Monte Carlo data for n greater than 3 and up to the point where statistical fluctuations become large. In fact for very small n equation (3) provides a slightly better fit to the canonical distribution than would an exact inversion of equation (2).

Thus far the discussion of the data has been restricted to the site problem for the BCC 1–2 lattice. Equations (2) and (3) appear however to be more generally valid. Dean and Bird (1966) have published results of Monte Carlo calculations on a number of other lattices. Because their data for larger cluster sizes were grouped in powers of ten we examined the partial sums $C_{10}(i, p, N)$ and again found both equations to give a good fit to data at the critical percolation probability. Moreover, it was found that the constants a and δ could be correlated with the coordination number Z of the various lattices. For two-dimensional lattices the value of δ is constant at $0{\cdot}0476 \pm 0{\cdot}0166$ to within the scatter in the data, while the value of a varied approximately as $Z^{-1/2}$. Thus the size distribution for two-dimensional lattices can be written as

$$c(n, p_c, N) = \frac{0{\cdot}0609Z^{-1/2}N}{n(n+1)(2n+1)^\delta}. \tag{4}$$

For three-dimensional lattices it is found that a varies approximately as Z^{-1} while δ also varies with Z, being given approximately by $2{\cdot}24Z^{-1{\cdot}2}$. The three-dimensional size distribution is given then by

$$c(n, p_c, N) = \frac{0{\cdot}398Z^{-1}N}{n(n+1)[\frac{2}{3}(2n+1)]^{\delta(Z)}}. \tag{5}$$

Equations (4) and (5) agree with the data of Dean and Bird and with our own results within experimental error for n larger than 7. It may be noted that while the empirical expressions were obtained using data on the site problems they also fit the 'SQ 1' bond data of Dean and Bird.

The primary qualitative result of the present empirical study is that the critical size distribution approximates a simple power law of the form $c(n, p_c, N) \sim n^{-(2+\delta)}$ for large n. The exponent of this law can be related to the degree of linear spreading of large clusters by the argument we shall now outline.

One may define the critical percolation probability as that fraction of randomly occupied sites for which the probability of finding a cluster spanning the system becomes finite, in the limit of large systems. For three-dimensional lattices, one can be more precise and define the shape of the bounding surface. One usually considers a cube, bounded by a pair of square surfaces which are to be connected by the spanning cluster. In principle the critical probability may thus be a function of the orientation of the bounding surfaces, but for the moment we shall neglect this effect. First imagine an infinite system at the critical percolation concentration, with cubes of increasing sizes inscribed in this system. For large enough cubes there is a finite probability that they will contain a spanning cluster. These clusters will tend, on the average, to be of increasing size as the cubes themselves increase in size. It is to be expected therefore that the existence of a distribution function of cluster size which does not fall off too rapidly with increase in cluster size is crucial to the existence of a critical percolation path. How rapidly is the question.

Let us assume that the linear extent L of a cluster in the direction of a normal to the bounding surface varies as some power ϕ of the number n in the cluster. We shall discuss reasonable limits for the latter. Since the distance between the cube faces is of the order $N^{1/3}a_0$, with a_0 the lattice parameter, the characteristic size of a cluster large enough to span the system is proportional to $N^{1/(3\phi)}$. If we integrate the canonical cluster size distribution over all sizes larger than this we should obtain a result proportional to the probability of finding a spanning cluster. The condition for critical percolation, which must be independent of the system size N, will be expressed in the condition that the result of the integration is N independent and does not go to zero with any power of N.

With our assumption that the cluster size distribution at large n and N is given by a constant$\times Nn^{-(2+\delta)}$ our integration result will be proportional to $N(N^{1/(3\phi)})^{-(1+\delta)}$ The condition that this does not decrease to zero with large N is then $1-(1+\delta)/(3\phi) \geqslant 0$ or $\delta \leqslant 3\phi - 1$. On the other hand for convergence of the cluster size distribution function, δ must be non-negative, so that ϕ must be greater than $\frac{1}{3}$. This limiting value is just what one might expect if the clusters were completely uniform, i.e., with no statistical correlation between the fact of their existence at some point and their existence at neighbouring points. Correlation exists of course from the very fact of connectivity. For non-percolation clusters where correlation effects have been analysed, such as the random walk or the self-avoiding random walk (SAW), ϕ has the values $\frac{1}{2}$ and $\frac{3}{5}$ respectively (see e.g. Domb 1969). Percolation clusters differ from these in that there is no preferred origin and they may be expected to be somewhat more uniformly distributed over their volume, leading to values of ϕ closer to $\frac{1}{3}$.

The correlations causing ϕ to be greater than $\frac{1}{3}$ come from the fact that the requirements for connectivity tend to give somewhat larger than average densities 'inside' the cluster as opposed to their extremities which may be expected to be somewhat straggly. Where the number of neighbours Z is large so that the continuity of the cluster can be satisfied at any one of a large number of neighbouring sites the correlation effects will be smaller. The Monte Carlo evidence presented earlier indicates that the value of δ for three-dimensional lattices tends toward the value 0 with increasing Z, consistent with the value of ϕ tending toward $\frac{1}{3}$.

A similar argument to the above for two-dimensional lattices yields the condition $0 < \delta \leqslant (2\phi - 1)$, or $\phi > \frac{1}{2}$. Both the uniform uncorrelated cluster and the simple random walk give $\phi = \frac{1}{2}$, while the self-avoiding walk gives $\phi = \frac{3}{4}$. In analogy to the three-dimensional case one might also argue that δ should approach zero with increasing Z, but the Monte Carlo results indicate δ to be approximately constant although smaller than in three dimensions. The value of δ independent of Z in two dimensions fits in with the idea of universality, and is in contrast with the apparent variation of δ with Z in three dimensions. Further investigation is desirable to see if true asymptotic behaviour has been reached in both cases.

A more detailed account of this work together with numerical data is currently in preparation.

The authors are grateful to Professor C Domb for helpful correspondence and discussion.

References

Dean P and Bird N F 1966 *National Physical Laboratory Report* Ma 61

Domb C 1969 *Adv. Chem. Phys.* **15** 229–59

Domb C and Dalton N W 1966 *Proc. Phys. Soc.* **89** 859–71

Domb C, Schneider T and Stoll E 1975 *J. Phys. A: Math. Gen.* **8** L90–4

Essam J W 1972 *Phase Transitions and Critical Phenomena* vol 2, eds C Domb and M S Green (New York: Academic Press) chap 6

Quinn G D, Bishop G H and Harrison R J 1976 *Proc. NBS Conf. on Computer Simulation for Materials Applications* (Washington: National Bureau of Standards) to be published

Shante V K and Kirkpatrick S 1971 *Adv. Phys.* **20** 325–57

Sykes M F and Glen M 1976 *J. Phys. A: Math. Gen.* **9** 87–95