Cluster number scaling in two-dimensional percolation

1986 J. Phys. A: Math. Gen. 19 291

(http://iopscience.iop.org/0305-4470/19/2/022)

J. Phys. A: Math. Gen. 19 (1986) 291-304. Printed in Great Britain

Cluster number scaling in two-dimensional percolation

D C Rapaport†

Department of Physics, Bar-Ilan University, Ramat-Gan, Israel

Received 4 June 1985

Abstract. The cluster distribution for site percolation on the triangular lattice is studied at and close to p_c using Monte Carlo techniques. Lattice sizes range as high as 2.56×10^{10} sites, and periodic boundary conditions are imposed instead of the customary free boundaries in order to reduce residual finite-size effects. The results support scaling to leading order both at and near p_c. Corrections to scaling are also examined, but though a value for the associated exponent is obtained, the results suggest that a description based on a single exponent may not be adequate.

1. Introduction

In order to examine the critical behaviour of two-dimensional site percolation clusters (Stauffer 1979, Essam 1980) recent Monte Carlo simulations have been compelled to consider lattices of ever-increasing size (Hoshen *et al* 1979, Margolina *et al* 1984). In a recent letter (Rapaport 1985) it was demonstrated that the use of extremely large lattices—containing over 10^{10} sites—was not in itself particularly helpful unless accompanied by periodic boundary conditions rather than the free boundaries customarily used. The results for a large simple quadratic (SQ) lattice at the critical probability p_c obtained using both periodic and free boundaries were compared and, whereas in the periodic case the evidence in support of scaling (Stauffer 1979) was convincing, the same could not be said for the free boundaries where the support was far from conclusive.

The present paper describes a Monte Carlo study of site percolation on the triangular (TRI) lattice, again using periodic boundaries. The advantage of the TRI lattice over the SQ is that its p_c is known exactly ($p_c = \frac{1}{2}$). The emphasis is on a range of p lying very close to p_c, and in fact one of the conclusions is that, provided periodic boundaries are used, the convergence of the cluster distribution as lattice size is increased is sufficient to suggest that very large lattices are not a necessity away from p_c—a conclusion that does not follow from the free boundary results. Away from p_c there is therefore little point in investing undue effort in considering such large systems and the effort is better expended on generating multiple realisations of smaller systems in order to improve the statistics.

Scaling theory (Stauffer 1979) predicts that at p_c the s dependence of the number of s-site clusters is $n_s(p_c) \propto s^{-\tau}$, while for $p \neq p_c$ but sufficiently close, $n_s(p)/n_s(p_c) = f_0(z)$, with the scaling variable z equal to $(p - p_c)s^{\sigma}$. There are also corrections which are required for small s; a more detailed discussion appears later. Monte Carlo techniques will be used to derive numerical estimates for $n_s(p)$ which will then be

† Present address: IBM Department 48B/428, Neighborhood Road, Kingston, NY 12401, USA.

0305-4470/86/020291+14$02.50

used to test the accuracy of the scaling predictions. The computational techniques—in part well known and in part new—are discussed in § 2. The results themselves together with the scaling tests appear in § 3.

It is interesting to observe that while periodic boundaries are the accepted norm in essentially all condensed matter simulations with the purpose of reducing the undesirable effects of finite system size, for some reason this has not been the case for percolation. In fact, doubts have even been raised as to whether periodic boundaries would lead to any modification of the cluster distribution at intermediate cluster sizes where the scope for testing scaling is greatest (Margolina *et al* 1984). The SQ lattice simulations (Rapaport 1985) establish beyond doubt that the choice of boundary conditions has a significant effect on the cluster distribution, a result which holds true in general. The computations involved in treating the periodic case are slightly more complicated, but the improved results more than justify the effort involved.

2. Computational methods

The techniques involved in generating the cluster distributions for lattice percolation have been described at length in the literature (Hoshen and Kopelman 1976, Margolina *et al* 1984). In this section we will therefore only briefly review those aspects of our method which can be found elsewhere, and in greater detail the extensions that have been made for the present calculations. Only the two-dimensional lattice is addressed, but similar techniques could be used in three dimensions.

The basic scheme for constructing the clusters of a single lattice realisation is to scan the lattice sites systematically, row by row, with a series of random numbers obtained from a reliable source being used to declare the sites occupied with a given probability p and vacant otherwise. The previously scanned neighbour sites of an occupied site are then examined; if the occupied neighbours all belong to the same cluster then the current site is assigned to this cluster, while if several distinct clusters are involved a merging of clusters must take place.

Since the feature of interest here is the distribution of cluster sizes (i.e. the numbers of sites per cluster), and not the metrical properties, the only information that must be retained for each cluster that is only partially constructed, apart from its current size, is a list of the sites belonging to it in the most recently scanned row of the lattice. It follows from this that in order to process a row of sites the only information required about the lattice (apart from the cluster sizes) is contained in the immediately preceding row; indeed the total number of lattice sites of interest at any instant is equal to the row length, and comprises the sites already scanned in the current row and those sites of the previous row adjacent to sites of the current row yet to be scanned. Periodic boundaries additionally require that information concerning sites on all the other lattice edges be retained. This represents an enormous saving in the amount of storage required for the problem and makes possible the analysis of large lattices.

In order to handle the potentially time-consuming task of merging existing clusters that occur in the course of the generation, Hoshen and Kopeman (1976) proposed a scheme known as 'multiple labelling' which can be summarised as follows: each occupied site in the current/previous row is assigned a label indicating the serial number of the cluster to which it belongs. Different label values can correspond to the same cluster, the smallest such value being the 'proper' label. For each assigned label m—a new label value is used whenever a site is encountered that is a potential

start of a new cluster—a quantity E_m is defined so that if m is a proper label E_m represents the current cluster size, whereas for the other label values E_m is given a negative value chosen to ensure that the set $\{-E_m\}$ for a given cluster forms a sequence of pointers eventually leading to the proper label. This approach means that two partial clusters can be merged by simply altering a single pointer and no relabelling of the sites is necessary. The pointer sequences turn out to be surprisingly short; measurements made at p_c (the results are dependent on p) indicate that an average of only 0.1 pointer references are required per site, while the maximum number of references amounted to only four.

Due to the frequent merging of partial clusters many more cluster labels will be assigned than the actual number of clusters (i.e. proper labels), and it will be necessary to compress the set of labels at regular intervals in order to conserve storage. This compression is actually carried out together with a test for complete clusters, i.e. clusters no longer present in the most recently scanned row; the final sizes of these clusters are included in a summary histogram and their labels freed for re-use. It is at this stage that different processing is required for free and periodic boundaries; while clusters that extend to a free boundary are deemed to terminate there, periodic boundaries imply a toroidally mapped lattice and cluster wraparound must be allowed for.

The generation technique used here differs from the earlier approaches in that the lattice is produced as a series of slabs which are subsequently combined to form lattices of various sizes with either free or periodic boundaries. Those clusters formed during slab construction which meet any of the slab edges cannot be completed until the slabs are combined to create the required lattice. Slab combination entails scanning pairs of slab edges that are brought into adjacency and merging those clusters that extend over more than one slab. The multiple labelling technique is used to handle this stage of the calculation as well. Four slabs of edge B combine to produce a region of edge $2B$, four such regions produce one with edge $4B$, and so on. When the required lattice size is reached, either the opposite edges are brought into adjacency to produce periodic boundaries, or the clusters are truncated at the extreme edges corresponding to free boundaries. While generating a single lattice with edge L, 4 lattices of edge $L/2$ are obtained at no additional cost, 16 with edge $L/4$, etc. The use of slabs as an intermediate step in the generation means that the most time-consuming part of the calculation, slab generation, requires relatively modest amounts of computer storage; large amounts of storage are only required for combining the slabs which is a relatively short process. If the entire lattice is generated at once (Margolina *et al* 1984) the computations are at the same time both time and storage consuming, a not especially desirable feature.

There is another advantage inherent in the slab approach. The 32 bit integer word length of most (although not all) computers limits the maximum cluster size to $2^{31}-1$ sites; a floating-point value (of suitable precision) could of course be used but this would add to the computation time. Provided the slab size is chosen to prevent such large clusters from appearing there will be no problem with integer overflow; large clusters can then only appear at the slab combination stage and these are accommodated by representing the sizes as double-precision floating-point quantities (permitting 56-bit accuracy) during these relatively brief calculations. An alternative means of overcoming this problem is of course to find a computer with an enlarged range of integers, a solution adopted by Margolina *et al* (1984).

The simulations were carried out on an IBM 3081 computer. Most of the software was written in FORTRAN, with the exception of the code responsible for scanning a

single row which was written in assembly language for a 40% gain in speed. Generation of a single slab with edge length $B = 10^4$ requires 5 min of processor time at $p \approx 0.5$ (the times are dependent on p); this is slightly faster than the SQ lattice calculation primarily due to the fact that the p values of interest here are lower, and this appears to outweigh the increased coordination number of the TRI lattice. An $L = 16 \times 10^4$ lattice contains 256 slabs and therefore requires about 21 hours of computation; the computations required to join the slabs amount to a couple of minutes only. The calculation of a single large lattice realisation yields a series of smaller lattices at the same time.

The biggest lattice generated contains 2.56×10^{10} sites and is half as large again as the largest TRI lattice previously reported (Margolina *et al* 1984), although it will become apparent subsequently that it is not so much the extreme size that is of importance but the ability to incorporate periodic boundary conditions. The computation rate is approximately double that of the earlier work; a program typical of that used previously is to be found in Binder and Stauffer (1984), although the present program is considerably longer.

The reliability of the random number generation is a vital element in a calculation of this kind. A high-speed generator capable of producing the large batches of uniformly distributed random variates needed here is one based on the shift register principle. The generator described by Kirkpatrick and Stoll (1981) was used after modification to produce integers in the range $(0, 2^{24}-1)$ rather than floating-point numbers, again with a view to maximising speed. The site occupancy test was then based on a comparison of a random variate with that integer value closest to $2^{24}p$.

Two separate realisations of the largest $L = 16 \times 10^4$ lattice were constructed at p_c. At other values of p, as described in the following section, the largest L was 8×10^4 and only a single realisation was produced for each. While it is possible to process even larger lattices, it will become apparent from the results already available that a mere doubling of L will not lead to any significant improvement in the quality of the predictions.

3. Results

3.1. Scaling theory

When applied to the cluster number distribution, scaling theory (Stauffer 1979) states that the average number of clusters of s sites, normalised per lattice site, at site occupation probability p is

$$n_s(p) = s^{-\tau}[f_0(z) + s^{-\Omega} f_1(z) + \ldots], \tag{1}$$

where the scaling functions f_0 and f_1 are analytic for sufficiently small z and where z, the scaling variable, is defined as

$$z = (p - p_c)s^{\sigma}. \tag{2}$$

The second term in (1) is the first of a series of higher-order corrections that enter at small s. When $p = p_c$

$$n_s(p_c) = s^{-\tau}[f_0(0) + s^{-\Omega} f_1(0) + \ldots]. \tag{3}$$

Two of the exponents introduced in (1) and (2), σ and τ, are believed to be known exactly in two dimensions. The thermal and magnetic eigenvalues for the $d = 2$, $Q = 1$

case of the Q-state Potts model are $y_t = \frac{3}{4}$, $y_h = \frac{91}{48}$ (den Nijs 1979, Pearson 1980). These eigenvalues are related to the usual percolation critical exponents by $d/y_t = 2-\alpha$, $(d-y_h)/y_t = \beta$, and since $2-\alpha$ and β are the exponents of the singular parts of the zeroth and first moments of n_s (Stauffer 1979) it follows that $\sigma = \frac{36}{91}$, $\tau = \frac{187}{91}$. (Note that $\tau > 2$ in order that the first moment $\Sigma s n_s$ be convergent; its value is p.) The $Q = 1$ Potts model corresponds to bond percolation but the exponents are expected to apply to the site problem as well. The third exponent in (1), Ω, has no well established value associated with it, although a number of suggestions have been put forward (Margolina *et al* 1984). In the analysis that follows, the 'exact' values of σ and τ will be used; prior to their becoming available the best estimates were 0.39 and 2.05 respectively (Hoshen *et al* 1979) derived from the series values of the exponents β and γ (Sykes *et al* 1976), estimates which are very close to the rational fractions.

In order to improve the count statistics of the larger clusters, the individual n_s are grouped along the lines proposed by Hoshen *et al* (1979); the quantities actually studied are thus

$$G_s(p) = \sum_{s'=2^{k-1}}^{2^k-1} n_{s'}, \qquad s = 2^k. \tag{4}$$

The scaling prediction for this partial sum at p_c can be derived by integrating (3) over the appropriate range of s; the result to leading order is

$$G_s(p) = (2^{(\tau-1)/2} - 2^{-(\tau-1)/2}) s_{AV}^{1-\tau} f_0(0)/(\tau-1), \tag{5}$$

where $s_{AV} = [2^{k-1}(2^k-1)]^{1/2} \approx s/\sqrt{2}$ is the geometric mean of the size range covered.

3.2. Results at p_c

A double logarithmic plot of $G_s(p)$ as a function of s_{AV} is shown in figure 1 for $p = p_c$ and a nearby value on either side. At p_c the data points lie on a straight line for a range of s_{AV} covering four orders of magnitude. The measured gradient, which from (5) is the value of $1-\tau$, corresponds to $\tau = 2.06$, a value close to the exact $\frac{187}{91} = 2.055$, while the intercept with the vertical axis at $\log(s_{AV}) = 0$ leads to $f_0(0) = 0.030$. The deviations from linearity at small s ($<10^2$) signal the appearance of scaling corrections, while the large s ($>10^6$) breakdown corresponds to a distortion in the distribution of the larger clusters due to finite lattice size (in this instance $L = 8\times10^4$). Figure 1 also makes it clear that there is no linearity when p differs from p_c by even 1%; this is to be expected because $f_0(z)$ is implicitly dependent on s, but it means that an alternative approach to estimating τ will be required for $p \neq p_c$ (see further).

The Monte Carlo estimates for $G_s(p_c)$ can be compared with the exact enumeration values (Sykes and Glen 1976). The average number of isolated sites ($k = 1$) for the two $L = 16\times10^4$ realisations differs from the exact value by 0.01%. A similar difference was found by Margolina *et al* (1984) for free boundaries; while the choice of boundary affects the entire cluster distribution, the significant changes (for the lattice sizes considered) are first noticed for clusters of a few hundred sites while for $k = 1$ the difference is typically 0.004% (Rapaport 1985). The deviations for $k = 2$ and 3 from the exact values are also very small. Agreement does not hold for the total cluster count however, which, in the case of exact enumeration, must be obtained by numerical extrapolation. The Monte Carlo values (normalised per site) for the two largest realisations are 0.017 625 and 0.017 627; these are just below the Margolina *et al* (1984) estimate of $0.017\,630 \pm 0.000\,02$, presumably due to the absence of fragmented boundary

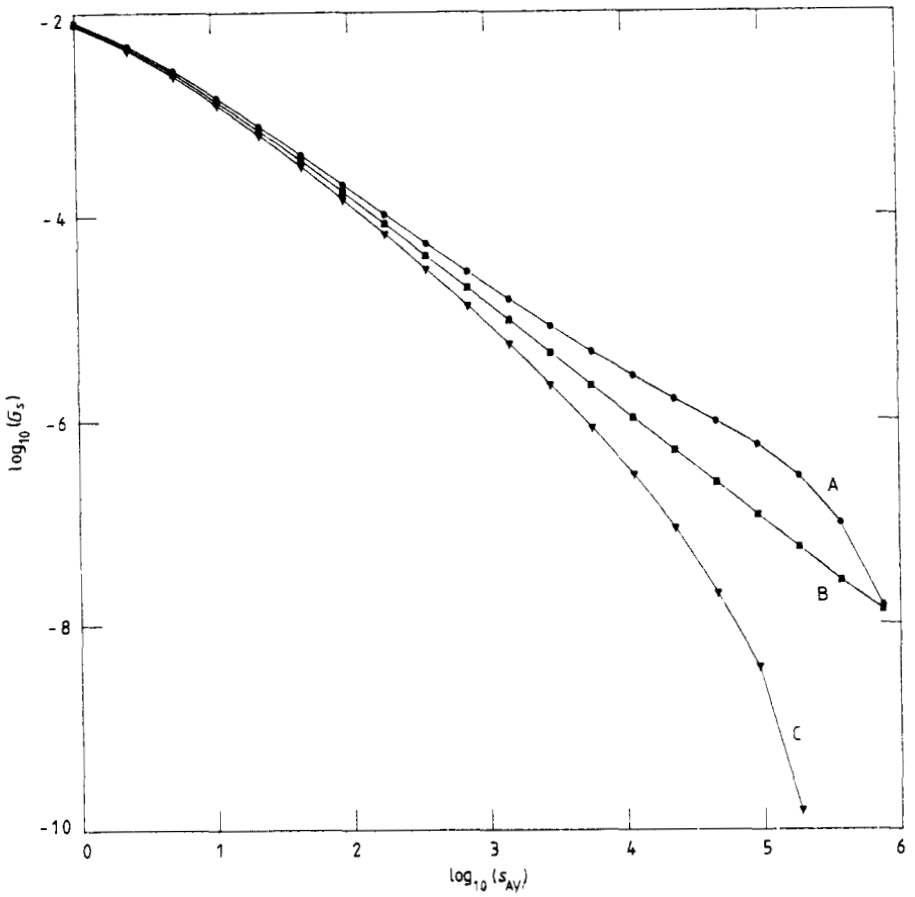

Figure 1. Log-log plot of $G_s(p)$ as a function of s_{AV} for $L = 8 \times 10^4$ and several p values: $p =$ 0.496 (A), 0.5 (B), 0.504 (C).

clusters, but significantly different from the series estimate 0.0168 ± 0.0002 (Domb and Pearce 1976).

Previous Monte Carlo simulations have employed an alternative form of data presentation that emphasises any (dis)agreement with scaling and clearly shows the range of cluster sizes over which scaling applies (Hoshen *et al* 1979, Margolina *et al* 1984, Rapaport 1985). Consider the quantity

$$N_s(p) = s^{\tau-1} \sum_{s' \geqslant s} n_{s'}(p) \tag{6}$$

which includes all clusters from s to the largest present. The expected asymptotic behaviour again follows from integration of (3),

$$N_s(p_c) = s^{\tau-1} \int_s^{\infty} n_{s'}(p_c)\, ds' \tag{7}$$

$$= \frac{f_0(0)}{\tau - 1} + \frac{f_1(0) s^{-\Omega}}{\Omega + \tau - 1}. \tag{8}$$

In the actual numerical evaluation corresponding to (7) the factor $s^{\tau-1}$ was replaced by $s_{AV}^{\tau-1}$ of the lowest range (i.e. k value) present in the sum; the estimate of $f_0(0)$

obtained in this way will be a factor of $(\sqrt{2})^{\tau-1}$ greater than that of Margolina *et al* (1984). In order to maintain consistency with the earlier work, s in (8) will be taken to mean s_{AV}.

Equation (8) implies the existence of a plateau in $N_s(p_c)$ provided s is sufficiently large. Figure 2 shows that the plateau indeed occurs and extends over an almost 1000 fold range of cluster sizes for the largest lattice ($L = 16 \times 10^4$), although the plateau is well established even for $L = 2 \times 10^4$ (note that the cluster sizes are shown on a logarithmic scale). The analysis uses the exact value of τ. The deviations from the plateau at both extremes reflect the same breakdown of scaling observed in figure 1.

The present results should be contrasted with those of Margolina *et al* (1984) where free boundaries were used; there only a hint of a plateau is observed even for the largest lattice size ($L = 130\,000$) and the size dependence is much more pronounced. In our previous treatment of the SQ lattice (Rapaport 1985) a similar dependence on boundary conditions was found.

Figure 3 examines the sensitivity of the plateau to the choice of p ($L = 8 \times 10^4$ and fewer realisations than in figure 2). A shift away from p_c by only 1 part in 5000 is seen to destroy the plateau; the plateau at p_c is an indication of the onset of the infinite cluster. On the SQ lattice the plateau provides a useful criterion for estimating p_c.

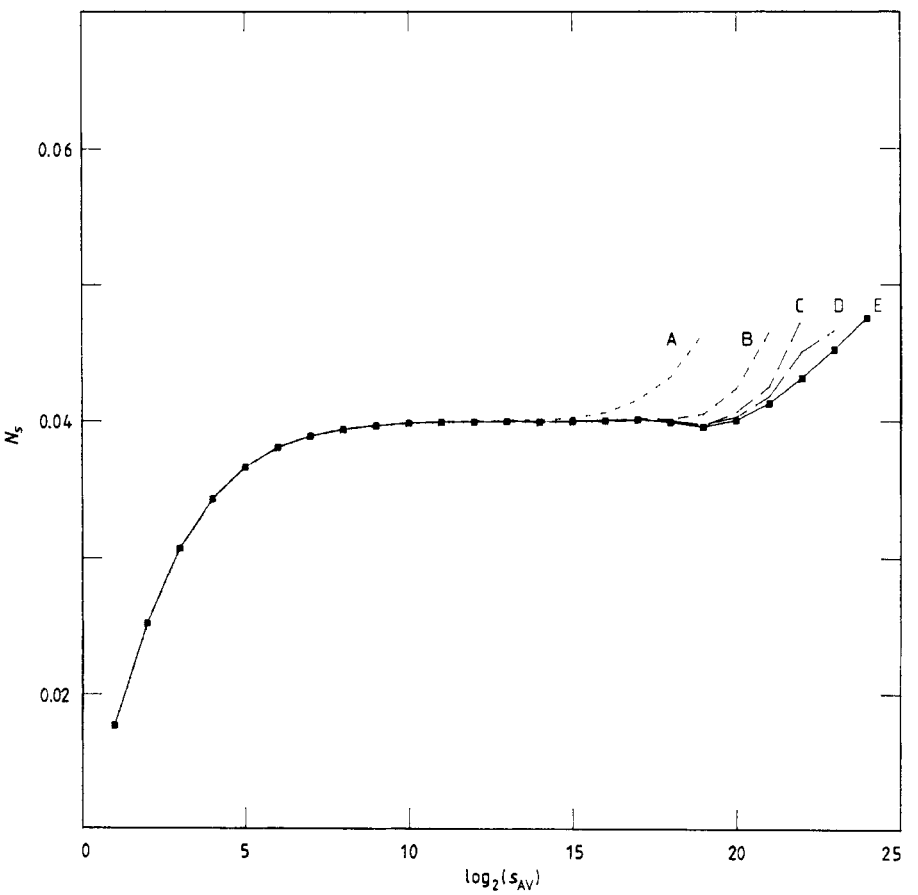

Figure 2. Plot of $N_s(p_c)$ against s_{AV} (note the logarithmic scale) for several lattice sizes: A, $L = 10^4$, B, 2×10^4, C, 4×10^4, D, 8×10^4, E, 16×10^4.

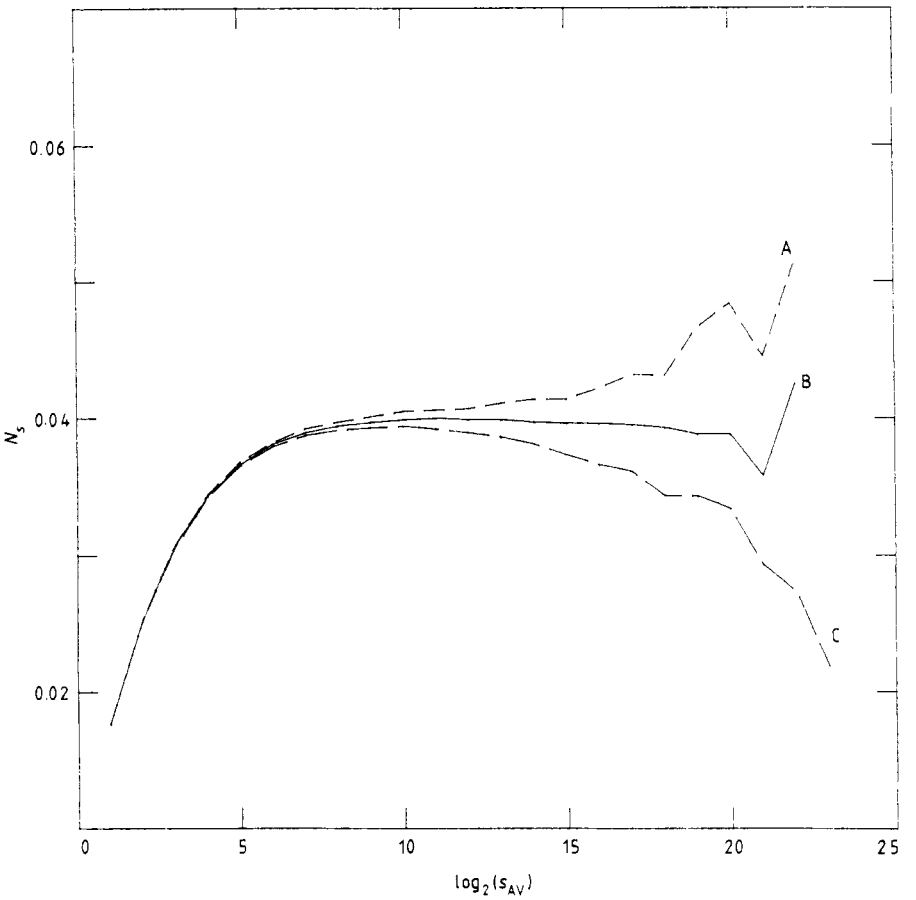

Figure 3. Plot of $N_s(p)$ for $L = 8 \times 10^4$ and values of p very close to p_c: A, $p = 0.4999$; B, 0.5000; C, 0.5001.

The measured plateau height is 0.04; this leads to $f_0(0) = 0.029$ which agrees with the value from figure 1 and the estimate of Margolina *et al* (1984) for free boundaries. However, unlike the present calculation which yields $f_0(0)$ directly, the lack of a well established plateau for the free boundary case means that $f_0(0)$ can only be obtained by fitting (8)—including the unknown scaling correction exponent Ω as well as both $f_0(0)$ and $f_1(0)$—to the data; the fit succeeds because boundary effects are unimportant over the range of s involved.

The scaling correction itself can be studied by introducing the quantity

$$\Delta N_s = f_0(0)/(\tau - 1) - N_s \tag{9}$$

which measures the deviation from the plateau. It follows from (8) that

$$\ln(\Delta N_s) = \ln\left(\frac{-f_1(0)}{\Omega + \tau - 1}\right) - \Omega \ln s \tag{10}$$

and so both Ω and $f_1(0)$ can be determined graphically.

Figure 4 shows these log-log plots for several possible values of plateau height close to that observed: the results are essentially independent of lattice size (for the largest lattices). The most linear of the plots shown corresponds to height 0.040 15,

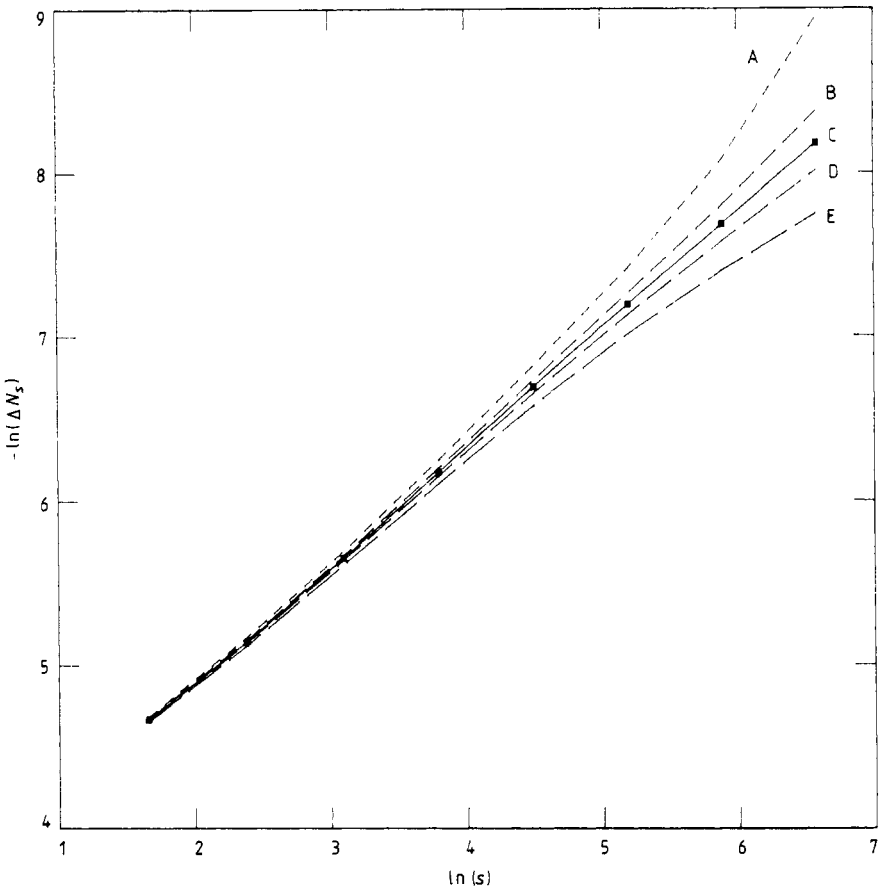

Figure 4. Log-log plot of the deviations from the plateau (ΔN_s) at small s; the plateau heights shown are A, 0.0400; B, 0.0401; C, 0.040 15; D, 0.0402; E, 0.0403.

slightly above the measured value. Even assuming this to represent the true plateau height (implying that the measured height is still 1 part in 300 away from convergence) there remains a certain amount of curvature which, as with the SQ lattice (Rapaport 1985), cannot be eliminated. The points plotted cover the range $k = 3$ (cluster sizes 4-7) to 10 (sizes 512-1023); the range is limited at small s by the fact that even if (8) is correct it omits expected higher-order terms, while at larger s the deviations $\Delta N_s \to 0$ and so errors in n_s will be magnified here. A linear fit to the points $k = 5$-8 yields $\Omega = 0.74$ whereas for $k = 7$-10, $\Omega = 0.71$. These values are above the SQ lattice value (0.64) but, as was pointed out there, it is not clear that the results are describable in terms of only a single correction exponent and that a more complex behaviour may be involved. The values of $f_1(0)$ corresponding to the two Ω estimates are -0.030 and -0.027 respectively, leading to ratios $|f_1(0)|/f_0(0)$ equal to 1.03 and 0.93. The uncertainty here is considerable, just as in the previous analysis (Margolina *et al* 1984) where the results obtained by series analysis are also surveyed, but there is no basic disagreement.

3.3. Results near p_c

In the neighbourhood of p_c the entire scaling expression (1) is open to examination. Ideally, it ought to be possible to demonstrate that the exponent τ (and perhaps even

Ω) do not vary with p and that the functions f_0 and f_1 depend on z alone. Previous work (Hoshen *et al* 1979, Margolina *et al* 1984) has tended to support these claims (although Ω is apparently p dependent), but it is difficult, if not impossible, to establish the scaling result (1) with any degree of certainty given that it may only represent a reasonable numerical fit to the data sufficiently close to p_c.

From the preceding results at p_c it is clear that once it is recognised that improved convergence is obtained with periodic boundary conditions, there is little point in considering extremely large lattices unless very close to p_c. Indeed a systematic study of scaling for various 2D lattices could probably be carried out using lattices no larger than, say, $L = 4 \times 10^4$; such a study has yet to be undertaken. In the present paper we confine our attention to a very narrow region around p_c and ask what can be learned about scaling from the very large lattices.

That the simple $s^{-\tau}$ dependence of $n_s(p)$ no longer applies for even the slightest shift of p away from p_c is clear from figures 1 and 3, and it will not therefore be possible to study the exponent τ without accounting for the scaling function $f_0(z)$. The quantity to be considered then is the ratio

$$v_s(p) = n_s(p)/n_s(p_c) \tag{11}$$

which, assuming τ to be p independent, and s to be sufficiently large to eliminate corrections, should be a function of the single variable z.

The variation of $\ln v_s(p)$ with z, assuming the result $\sigma = \frac{36}{91}$, is shown in figures 5 and 6 for p above and below p_c. To reduce known large- and small-cluster deviations only values of s in the range 2^6-2^{19} are included. The prediction is that $v_s(p) = f_0(z)$, a result that seems to hold quite well over the range $-0.15 < z < 0.1$; the fact that $v_s(p)$ depends only on z implies that τ is independent of p as hoped. The results here apply for p within 1% of p_c, but similar results for a much smaller lattice ($L = 4000$) exist for a range of p extending 10% to either side of p_c (Hoshen *et al* 1979). Near $z = 0$ the variation of $\ln f_0(z)$ is linear, suggesting that to leading order $f_0(z) = e^{-a_1 z}$, with a measured gradient corresponding to $a_1 = 7.0$ above p_c and 6.7 below—there is a ± 0.2 uncertainty in both values. Clearly, if $f_0(z)$ is analytic at the origin then the same initial gradient should apply for p on either side of p_c and the results certainly allow for this possibility. The curvature that appears for larger $|z|$ can be attributed to $\ln f_0(z)$ being a polynomial in z, and in fact Hoshen *et al* (1979) have carried out a fit using a cubic polynomial in which the coefficient of the linear term is 7.1 in agreement with the a_1 obtained here.

The z dependence of $v_s(p)$ close to $z = 0$ is shown on an expanded scale in figure 7. In order to demonstrate the breakdown of pure z dependence at small s, the graph includes data from even the smallest clusters. The data points for small s no longer lie on a common curve and the curves for different p no longer extrapolate to the origin; note that the corresponding points are omitted from a similar graph in Hoshen *et al* (1979). The deviations are to be expected given the scaling assumption (1). The gradient estimated from this graph corresponds to $a_1 = 7.2$, consistent with the earlier values.

Given the difficulty in obtaining a meaningful value of the correction exponent Ω at p_c for both TRI and SQ lattices, the chances of successfully studying Ω at other values of p seem remote because of the additional variable z. After attempting an analysis of this kind, Margolina *et al* (1984) arrived at a similar conclusion; while a value of about 0.7 was consistent with the data it proved impossible to improve the accuracy or determine the p dependence.

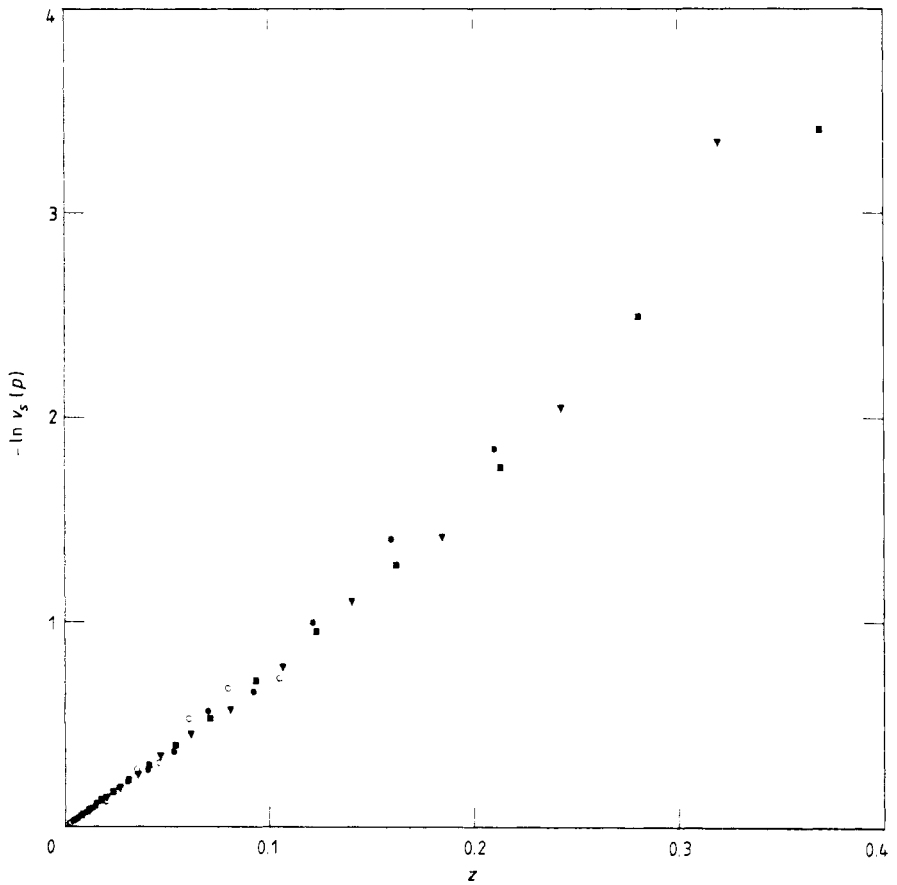

Figure 5. Plot of $v_s(p)$ (on a logarithmic scale) as a function of the scaling variable z for values of p above p_c: $p = 0.5005$ (○), 0.501 (●), 0.502 (▼), 0.504 (■). Only clusters with sizes between 2^6 and 2^{19} are included.

3.4. Other aspects

In addition to showing that $\ln f_0(z)$ could be fitted satisfactorily with a cubic polynomial in z, Hoshen *et al* (1979) also found that above p_c, $\ln v_s \propto s^{\zeta}$, with $\zeta = \frac{1}{2}$. This result is inconsistent with an analytic scaling function since it impies that $\ln f_0(z) \propto z^{1/2\sigma}$. However, if one examines the data on which this conclusion is based it is obvious that the range of s involved (approximately 1-500) lies in the region where scaling corrections are important and hence no simple z dependence need be expected. If corrections are allowed there is no longer any objection to retaining the analytic scaling functions. The fact that it can be proved that $\zeta = 1 - 1/d = \frac{1}{2}$ (for $d = 2$) at sufficiently large p (Kunz and Souillard 1978) has no *a priori* bearing on the critical behaviour.

Below p_c each $n_s(p)$ has a maximum at some value $s = s_{max}(p)$; this reflects the initial appearance of clusters of any specified size as p increases and their eventual disappearance as they merge into even larger clusters (including the infinite cluster). In the present work p has been restricted to within 1% of p_c and the lowest value of s_{max} observed is approximately 10^5; a factor of ten closer to p_c results in a shift of s_{max} to $\approx 3 \times 10^6$. The statistics for these large clusters do not allow an analysis along the

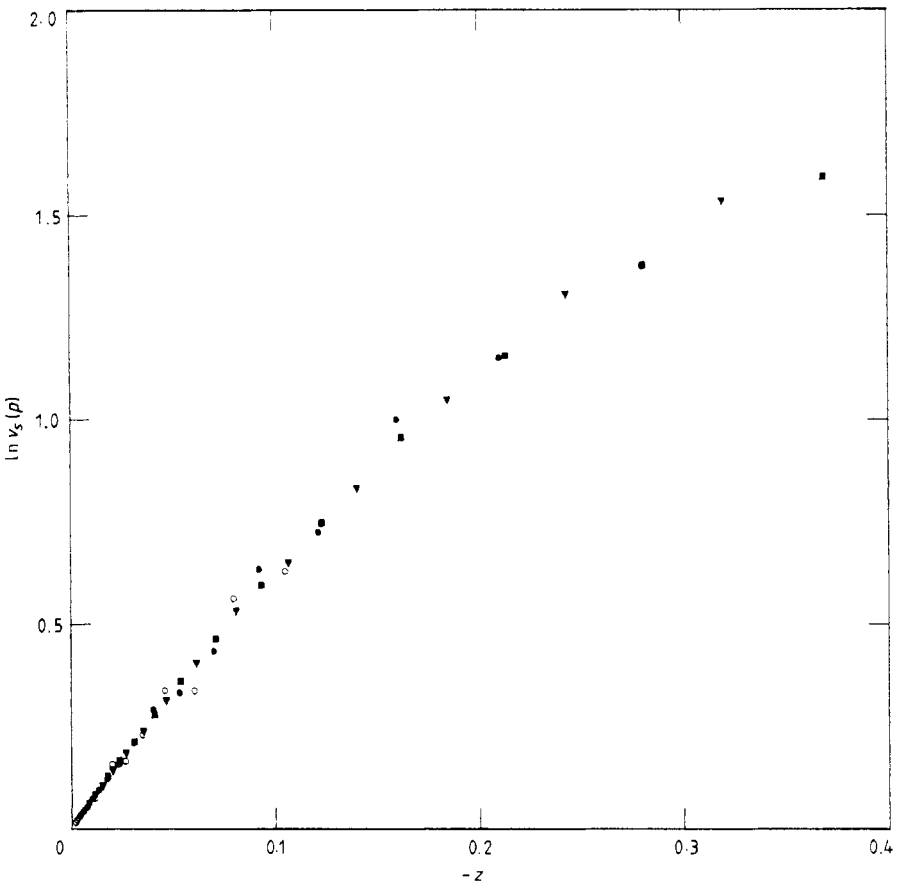

Figure 6. Plot of $v_s(p)$ as in figure 5 but for p below p_c: $p = 0.4995$ (○), 0.499 (●), 0.498 (▼), 0.496 (■).

lines of previous work (Margolina *et al* 1984), which was conducted for p more than 10% below p_c where much smaller values of s_{max} were involved. Here again a detailed study of this kind does not require extremely large lattices provided periodic boundaries are used.

One last comment is appropriate here. In order to overcome the problem of the absence of a definite plateau when free boundaries are used, Margolina *et al* (1984) carry out a finite-size scaling calculation in an attempt to recover the plateau (by introducing an additional L-dependent scaling function that involves the cluster fractal dimensionality as an exponent), while at the same time expressing doubts whether the use of periodic boundaries would improve the situation. The present results clearly put these doubts to rest; free boundaries strongly distort the cluster distribution, whereas for periodic boundaries there is so little lattice-size dependence that a finite-size scaling treatment is neither necessary nor possible.

4. Summary

The use of periodic boundaries has been shown to lead to a cluster size distribution that is in close agreement with scaling at p_c. The leading order correction at p_c is not

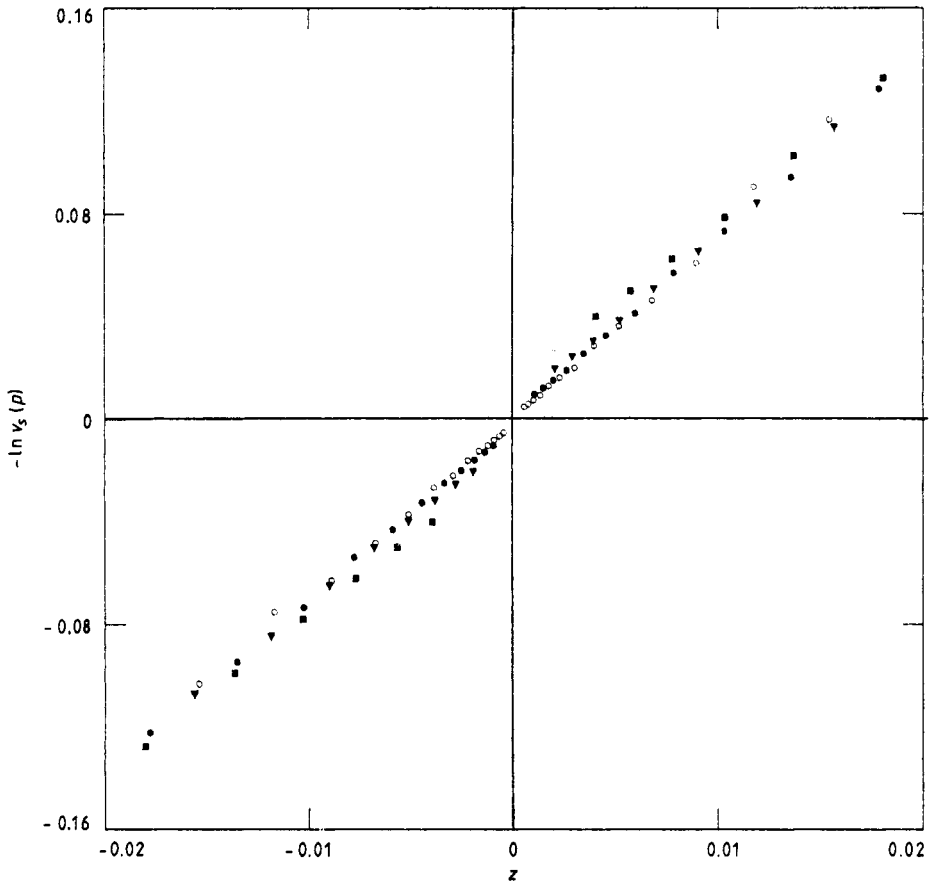

Figure 7. Plot of $v_s(p)$ as a function of z over a limited range of z with p both below and above p_c; the different sets of points are for $|p - p_c| = 0.0005$ (○), 0.001 (●), 0.002 (▼), 0.004 (■).

so readily analysed, and from the results it is not clear that a description that uses a single additive correction term provides a satisfactory explanation of the data. Close to p_c the results again support scaling to leading order. The exponents τ and σ that arise in the scaling theory are set to their conjectured exact values, and it is for these values that the agreement is obtained. A new computational technique based on dividing the problem into a set of smaller subproblems was developed and the resulting implementation shown to be considerably faster than previously described methods.

Acknowledgment

We thank D Stauffer for discussion and correspondence.

References

Binder K and Stauffer D 1984 in *Applications of the Monte Carlo Method in Statistical Physics* ed K Binder (Berlin: Springer) p 24

den Nijs M P M 1979 *J. Phys. A: Math. Gen.* **12** 1857

Domb C and Pearce P A 1976 *J. Phys. A: Math. Gen.* **9** L137
Essam J W 1980 *Rep. Prog. Phys.* **43** 833
Hoshen J and Kopelman R 1976 *Phys. Rev.* B **14** 3438
Hoshen J, Stauffer D, Bishop G H, Harrison R J and Quinn G D 1979 *J. Phys. A: Math. Gen.* **12** 1285
Kirkpatrick S and Stoll E P 1981 *J. Comput. Phys.* **40** 517
Kunz H and Souillard B 1978 *J. Stat. Phys.* **19** 77
Margolina A, Nakanishi H, Stauffer D and Stanley H E 1984 *J. Phys. A: Math. Gen.* **17** 1683
Pearson R B 1980 *Phys. Rev.* B **22** 2579
Rapaport D C 1985 *J. Phys. A: Math. Gen.* **18** L175
Stauffer D 1979 *Phys. Rep.* **54** 1
Sykes M F, Gaunt D S and Glen M 1976 *J. Phys. A: Math. Gen.* **9** 97, 725
Sykes M F and Glen M 1976 *J. Phys. A: Math. Gen.* **9** 87