Journal of Statistical Physics, Vol. 52, Nos. 1/2, 1988

Percolation Cluster Numbers

Joan Adler[1,2] **and Amnon Aharony**[1]

Received March 31, 1988; revision received March 31, 1988

It has recently been suggested that there may be an infinite number of independent exponents hidden in the tails of the probability distribution of average percolation cluster numbers. A simple approximation of non-Gaussian effects was used to deduce this result and we show that this approximation is questionable. Extensive simulations of the cluster distribution have been made and an interesting dependence of the cumulants on concentration and range of summation has been observed.

KEY WORDS: Percolation; clusters; critical exponents; phase transition; Potts model; simulation.

The Kasteleyn–Fortuin formula[1] for the partition function Z_q of the q-state Potts model gives

$$Z_q = \langle q^{\sum_s N_s} \rangle = \left\langle \exp\left(\ln q \sum_s N_s \right) \right\rangle \tag{1}$$

where N_s is the number of clusters of occupied sites of size s on a lattice whose sites are occupied with probability p. The free energy of the q-state Potts model F_q, where

$$\ln(q)\, F_q \sim \ln Z_q \tag{2}$$

has a singular part

$$F_{q,\text{sing}} \sim A(q)\, |p - p_c(q)|^{2-\alpha(q)} \tag{3}$$

Stauffer and Coniglio[2] (SC) recently proposed that knowledge of the details of the distribution of the N_s, $P(N_s)$, should suffice for the

[1] School of Physics and Astronomy, Tel Aviv University, Ramat Aviv, 69978, Israel.
[2] Department of Physics, Technion, Haifa, 32000, Israel.

0022-4715/88/0700-0509$06.00/0

calculation of all the critical exponents [including $\alpha(q)$] and of the critical points $p_c(q)$ for the q-state Potts model. In order to determine where the (infinite number of) Potts exponents are located in the $P(N_s)$, SC proceeded to make a cumulant expansion for $\ln Z_q$:

$$F_q = \frac{\ln Z_q}{\ln q} = \left\langle \sum_s N_s \right\rangle + \frac{1}{2}(\ln q)\left\langle \left(\sum_s N_s - \left\langle \sum_s N_s \right\rangle \right)^2 \right\rangle + \frac{1}{6}(\ln q)^2 \left\langle \left(\sum_s N_s - \left\langle \sum_s N \right\rangle \right)^3 \right\rangle + \cdots = \sum_{k=1}^{\infty} \frac{1}{k!} c_k (\ln q)^{k-1} \tag{4}$$

where c_k is the kth cumulant of the distribution of $\sum_s N_s$. As SC show, neglecting the fluctuations in N_s would yield

$$Z_q = \exp\left(\ln q \sum_s \langle N_s \rangle \right), \qquad \text{or} \qquad \ln(Z_q) = \ln q \sum_s \langle N_s \rangle$$

incorrectly implying that the free energy of all Potts models is proportional to that of percolation, $\sum_s \langle N_s \rangle$. Thus, a more sophisticated way of calculating $\ln Z_q$ is required.

In order to simplify the cumulant expansion of Eq. (4), SC proposed the use of a result of Coniglio *et al.*[3] (CSS). CSS found upper and lower bounds that coincided to give

$$(\langle N_s^2 \rangle - \langle N_s \rangle^2)/\langle N_s \rangle \to 1 \tag{5}$$

for $s \to \infty$, $p \leqslant p_c$. They also obtained Monte Carlo data that showed that this result holds for small clusters ($s > 1$) and $p = p_c$. We shall present further evidence confirming this result below.

To this point we are in complete agreement with CS. However, during the simplification of the cumulants in Eq. (4), SC factorized

$$c_2 = \left\langle \left(\sum_s N_s - \left\langle \sum_s N_s \right\rangle \right)^2 \right\rangle = \sum_s \langle (N_s - \langle N_s \rangle)^2 \rangle \tag{6}$$

and

$$c_3 = \left\langle \left(\sum_s N_s - \left\langle \sum_s N_s \right\rangle \right)^3 \right\rangle = \sum_s \langle (N_s - \langle N_s \rangle)^3 \rangle \tag{7}$$

These simplifications were made in order to be able to invoke Eq. (5) to replace $\langle (N_s - \langle N_s \rangle^2 \rangle = \langle N_s^2 \rangle - \langle N_s \rangle^2$ with $\langle N_s \rangle$. A similar

replacement of $\langle (N_s - \langle N_s \rangle)^3 \rangle$ with $f_s N_s^{3/2}$, where $f_s = \langle (N_s - \langle N_s \rangle)^3 \rangle / \langle (N_s - \langle N_s \rangle)^2 \rangle^{3/2}$, was also proposed. The result of these factorizations and replacements is the conclusion that the free energy per site is

$$\frac{\ln Z_q}{L^d} = \sum_s n_s + \frac{\ln q}{2} \sum_s n_s + \frac{(\ln q)^2}{6} \sum_s n_s^{3/2} f_s L^{d/2} + \cdots \tag{8}$$

where $n_s = \langle N_s \rangle / L^d$, and L is the linear dimension. From this SC deduced that only the third and higher terms can contribute to the crossover toward Potts critical behavior. Since $f_s \sim L^{-d/2}$, they also claimed that the Potts exponents must be hidden in the finite-size deviations of $P(N_s)$ from a Gaussian distribution.

We find the use of the factorizations (6) and (7) somewhat questionable. From Eq. (4) it is clear that c_k/k is the $(k-1)$th derivative of F_q with respect to $\ln q$ at $q=1$. Since F_1 is the free energy of the percolation problem, all c_k must exhibit singularities only at p_c (percolation). For example, Eq. (4) implies that

$$\begin{aligned} c_2 \sim \left.\frac{\partial F_q}{\partial \ln q}\right|_{q=1} \sim & -A(1)[p - p_c(1)]^{1-\alpha(1)} \left.\frac{\partial p_c}{\partial q}\right|_{q=1} [2 - \alpha(1)] \\ & - A(1)[p - p_c(1)]^{2-\alpha(1)} \ln[p - p_c(1)] \left.\frac{\partial \alpha(q)}{\partial q}\right|_{q=1} \\ & + \left.\frac{\partial A}{\partial q}\right|_{q=1} (p - p_c)^{2-\alpha(1)} + \cdots \end{aligned} \tag{9}$$

Only the last explicitly presented term has the critical behavior of $\langle N_s \rangle$. The first two terms have a *more* singular p dependence. Thus, the factorization is wrong even for $k=2$ and would be worse for larger k.

Extensive numerical investigations on lattices of sizes up to 500×500 with 500 different lattices used for each estimate have been carried out in order to elucidate this matter. We present detailed results below. Our data show that within error bounds

$$\begin{aligned} \sum_{s=k}^{m} & \langle (N_s - \langle N_s \rangle)^2 \rangle \\ & = \sum_{s=k}^{m} \langle (N_s - \langle N_s \rangle)^3 \rangle \\ & = \sum_{s=k}^{m} \langle (N_s - \langle N_s \rangle)^4 \rangle - 3 \sum_{s=k}^{m} \langle (N_s - \langle N_s \rangle)^2 \rangle^2 \\ & = \sum_{s=k}^{m} \langle N_s \rangle \end{aligned} \tag{10}$$

for large enough k. In Eq. (10), m is the size of the largest finite cluster in each 500-lattice grouping and k takes selected values in the range $2 \leqslant k \leqslant 100$. We write the quantities appearing in Eq. (10) and the extreme right of Eqs. (6) and (7) as

$$N_\Sigma(j, k, p, L) = \sum_{s=k}^{m} \langle (N_s - \langle N_s \rangle)^j \rangle = A_\Sigma(j, k, p, L) \sum_{s=k}^{m} \langle N_s \rangle \quad (11)$$

Quantities appearing in the center of Eq. (4) and in the center of Eqs. (6) and (7) are henceforth denoted by

$$N_P(j, k, p, L) = \left\langle \left[\sum_{s=k}^{m} (N_s - \langle N_s \rangle) \right]^j \right\rangle = A_P(j, k, p, L) \sum_{s=k}^{m} \langle N_s \rangle \quad (12)$$

Finally, we also define

$$N_\Sigma^2(k, p, L) = \sum_{s=k}^{m} \langle (N_s - \langle N_s \rangle)^2 \rangle^2 = A_\Sigma^2(k, p, L) \sum_{s=k}^{m} \langle N_s \rangle \quad (13)$$

and note that fourth cumulants are given by

$$N_\Sigma^C(4, k, p, L) = N_\Sigma(4, k, p, L) - 3N_\Sigma^2(k, p, L) \quad (14a)$$

or

$$N_P^C(4, k, p, L) = N_P(4, k, p, L) - 3[N_P(2, k, p, L)]^2 \quad (14b)$$

The c_i of Eq. (4) are equivalent to $N_P(i, 1, p, \infty)$ $(i < 4)$, or to $N_P^C(4, 1, p, \infty)$, in the new notation. While the $A_\Sigma(j, k, p, L)$ are close to unity, the $A_P(j, k, p, L)$ are not, in general. The detailed behavior of the coefficients $A_P(j, k, p, L)$ will be discussed below, but we note in passing that they have the expected L dependence, an interesting k dependence, and a totally surprising, strong p dependence.

We have evaluated the various sums as well as powers of the sums for occupation probabilities between 0.35 and 0.8 on the square lattice. We have considered lattice sizes between 10^2 and 500^2 sites, with periodic boundary conditions. We counted the clusters with programs developed by Adler *et al.*[4] for termite diffusion and based on the Hoshen–Kopelman[5] algorithm. The calculations were carried out on an IBM 3090. Even with the large memory of this machine it was found to be more efficient to generate each group of 500 samples twice. On the first run we obtained $\langle N_s \rangle$, and the second time we obtained the other averages. This was because storing the data and recalling it was slower than generating it when needed, because of paging problems. Preliminary calculations were

carried out on the individual averages in an attempt to determine whether we could evaluate $(N_s - \langle N_s \rangle)^2$ and $(N_s - \langle N_s \rangle)^3$ as functions of s and then sum with the sum both within and without the expectation. Fluctuations for larger s values were so large that it was decided to collect data on sums with different cutoffs directly. We used 500 lattices for each evaluation and made between one and 15 independent evaluations for each p and L value. All sums and cutoffs were generated for each p and L choice.

In Fig. 1 we present the different sums as functions of L for $p = 0.55$, 0.59, and 0.62 with the cutoff $s \geqslant 20$. These results are typical of those for the other cutoffs and nearby p values, but perhaps exhibit slightly better convergence than the lower cutoffs. We chose to present $s \geqslant 20$, because we hoped that relation (5) would be valid here and had doubts about its validity for smaller s. Convergence is actually even better for higher cutoffs,

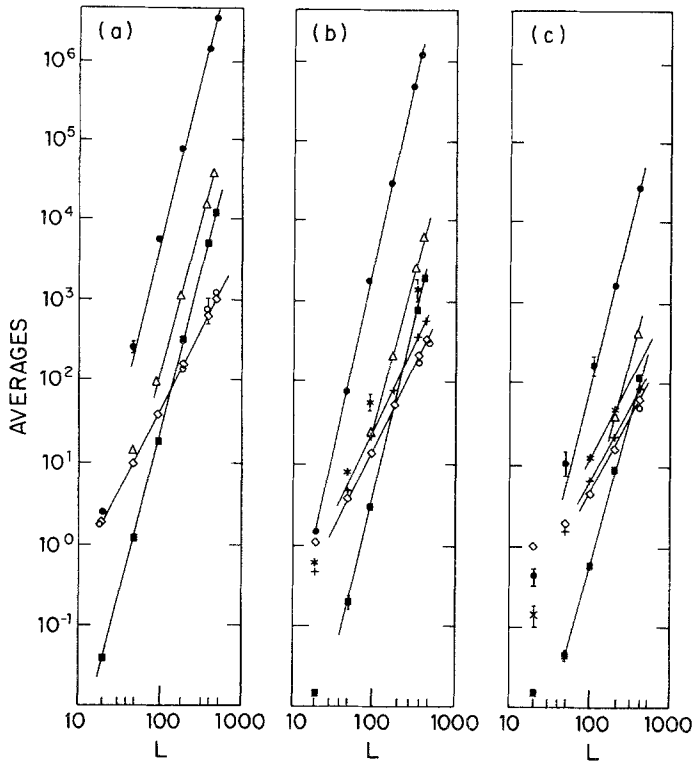

Fig. 1. Graphs of different averages as functions of system size for concentrations (a) $p = 0.55$, (b) $p = 0.59$, and (c) $p = 0.62$. The different averages are as follows: ($\diamond$) $\sum_{s=20}^{m} \langle N_s \rangle$, $N_\Sigma(3, 20, p, L)$, $N_\Sigma(3, 20, p, L)$; ($\diamond$ or $\circ$) $N_\Sigma^C(4, 20, p, L)$; ($+$) $N_P(2, 20, p, L)$; ($*$) $N_P(3, 20, p, L)$; ($\triangle$) $N_\Sigma(4, 20, p, L)$; ($\blacksquare$) $N_\Sigma^2(20, p, L)$; ($\bullet$) $N_P(2, 20, p, L)$. Solid lines are drawn to guide the eye only.

but these sums contain less information. In Fig. 1 the $N_\Sigma(2, 20, p, L)$, $N_\Sigma(3, 20, p, L)$, and $\sum_{s=20}^{m}\langle N_s\rangle$ are all indicated by the same symbol, since the differences are within the error bars, which are smaller than the symbols. Some $N_\Sigma^C(4, 20, p, L)$ estimates also fell within the error bars and some without, and only the latter are separately indicated. The $N_P(2, 20, 0.55, L)$ estimates of Fig. 1a also fell on the $\sum_{s=20}^{m}\langle N_s\rangle$ values, but this was not true at other p choices, where these are separately indicated. In the other cases where N_P estimates fail to be given, poor convergence was seen. The N_Σ and N_P estimates that do converge clearly show that both have the same L dependence as that of $\sum_{s=20}^{m}\langle N_s\rangle$ and we see that $N_\Sigma(4, 20, p, L)$, $N_P(2, 20, p, L)$, $N_\Sigma^2(20, p, L)$, and $[N_P(2, 20, p, L)]^2$ all have different L dependences.

From Fig. 1 and our other similar plots we became convinced that A_P values are certainly not unity. For example, at $p = 0.59$, (Fig. 1b), for $j = 2$ we observe that $N_\Sigma/N_P \approx (A_P)^{-1}$ is greater than unity for small values of L and approaches unity for $50 < L < 100$. For $L > 100$ this ratio approaches a constant value which is less than unity. This constant is a function of j, and by comparison with Fig. 1c we see that the constant appears to be p dependent as well. We decided to explore this matter further by carrying out an extensive set of calculations for different k and p choices. From Fig. 1 we may conclude that asymptotic behavior has set in by $L = 100$. Since we found that for some p choices convergence of A_P values required many different runs, we decided that $L = 100$ was a good size for extensive calculations, since the larger samples required several hours for each 500-sample calculation.

We made extensive calculations of 15 different sets of 500 samples for k choices of 2, 5, 10, 20, and 100 for various p choices. Some results of these analyses are plotted in Figs. 2a and 2b, where we give plots of $A_P(2, k, p, 100)$ and $A_P(3, k, p, 100)$, respectively. Unfortunately, the fluctuations in the $A_P(4, k, p, 100)$ data were too large to draw any conclusions and not even the $A_P(3, k, p, 100)$ points were convergent for lower p values.

The A_P estimates show a complex and interesting p and k dependence. In addition, it does appear that the second and third cumulants have different dependences on these variables, although there may be a simple scaling between these. This behavior would seem to suggest that the factorizations of Eqs. (6) and (7) are not correct. We have been unable to develop arguments analogous to those of SC with replacements for these equations because this dependence is so very complex.

As a first step in the further investigation of this singular behavior, we looked at some of the N_P values to search for the divergences in the c_i that are suggested by Eq. (9). In Fig. 3 we graph $\sum_{s=k}^{m}\langle N_s\rangle$ and

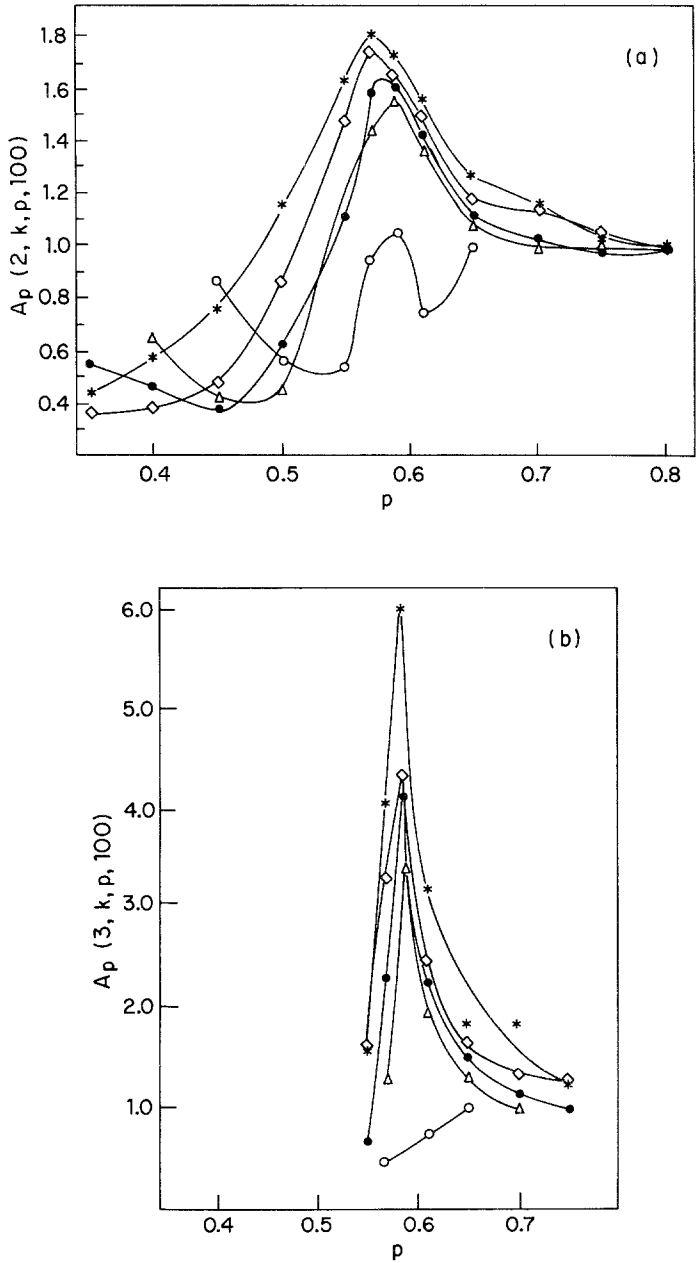

Fig. 2. Graphs of (a) $A_P(2, k, p, 100)$ and (b) $A_P(3, k, p, 100)$ as functions of p for different cutoff k choices: (*) $k=2$, (◇) $k=5$, (●) $k=10$, (△) $k=20$, (○) $k=100$. Solid lines are drawn to guide the eye only.

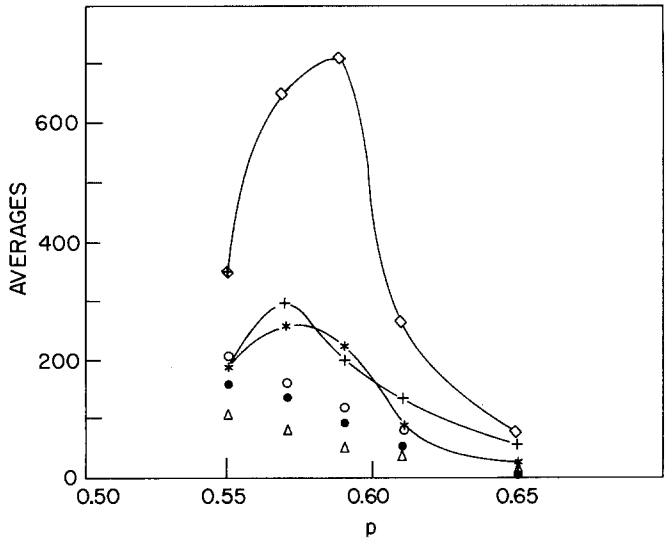

Fig. 3. Graph of N_P and N_Σ values indicating evidence for a potential divergence in c_3. The different k and j choices are indicated as follows: (○) $\sum_{s=2}^{m}\langle N_s\rangle$, (△) $\sum_{s=5}^{m}\langle N_s\rangle$, (+) $N_P(2, 2, p, 100)$, (●) $N_P(2, 5, p, 100)$, (◇) $N_P(3, 2, p, 100)$, (*) $N_P(3, 5, p, 100)$. Solid lines are drawn to guide the eye only.

$N_P(j, k, p, 100)$ for $j=2$ and 3 and observe that for $j=3$ there is indeed an indication of a possible divergence in the infinite system. The $N_P(3, k, p, 100)$ for the two smallest k values exhibit peaks centered in the region of p_c. In order to confirm that these peaks can be interpreted as a divergence of the infinite system, we would need to carry out a comprehensive finite-size scaling analysis. This would require the generation of data of this quality for larger sizes and this is not feasible at present. The $N_\Sigma(2, k, p, 100)$ and $N_\Sigma(3, k, p, 100)$ values were also calculated, but are not presented on the graph for reasons of clarity of presentation. They did deviate slightly from the $\sum_{s=k}^{m}\langle N_s\rangle$ values, but, as discussed below, these deviations may not be significant.

Table I. Different $A_\Sigma(2, k, p, 100)$ Values

p	$k=2$	$k=5$	$k=10$	$k=20$	$k=100$
0.57	1.039	1.014	1.003	1.002	0.998
0.59	1.049	1.015	1.009	1.000	1.000
0.61	1.042	1.010	0.997	0.998	1.045
0.65	1.025	1.004	1.006	1.007	0.985
0.70	1.028				

Since we generated many more samples and larger systems than those of CSS, we decided to also use our data to investigate the A_Σ. Some of these results are presented in Table I. Here we observe only a slight deviation from unity, which may not be significant, although there do appear to be some consistent trends for smaller k choices.

The main conclusion from our analysis is that while we do agree that for larger s values $(\langle N_s^2\rangle - \langle N_s\rangle^2)/(\langle N_s\rangle) \to 1$, the fact that in general

$$\left\langle \left(\sum_s N_s - \left\langle \sum_s N_s \right\rangle \right)^2 \right\rangle \neq \sum_s \langle (N_s - \langle N_s\rangle)^2\rangle$$

$$\left\langle \left(\sum_s N_s - \left\langle \sum_s N_s \right\rangle \right)^3 \right\rangle \neq \sum_s \langle (N_s - \langle N_s\rangle)^3\rangle$$

prevents this observation from providing a direct path to Eq. (8) and thence to the deductions of SC. In other words, cross correlations between different cluster sizes preclude use of their simple factorization. Instead, we see a complex, cumulant-dependent behavior already in the $\langle(\sum_s N_s - \langle\sum_s N_s\rangle)^2\rangle$ and $\langle(\sum_s N_s - \langle\sum_s N_s\rangle)^3\rangle$. It might be conjectured that somewhere within this dependence Potts critical exponents are hidden.

ACKNOWLEDGMENTS

We thank D. Stauffer and A. Coniglio for discussions and the consultants at the Tel Aviv University Computer Centre for assistance in organizing the computations on the IBM 3090. This work was supported in part by grants from the U. S.–Israel Binational Science Foundation and the Israel Academy of Sciences and Humanities.

REFERENCES

1. C. M. Fortuin and P. W. Kasteleyn, *J. Phys. Soc. Japan (Suppl.)* **26**:11 (1969).
2. D. Stauffer and A. Coniglio, *Physica* **143**:326 (1987).
3. A. Coniglio, H. E. Stanley, and D. Stauffer, *J. Phys. A* **12**:L323 (1979).
4. J. Adler, A. Aharony, and D. Stauffer, *J. Phys. A* **18**:L129 (1985).
5. U. Hoshen and R. Kopelman, *Phys. Rev. B* **14**:3438 (1976).

Communicated by J. L. Lebowitz