Finite-Size Effects at Asymmetric First-Order Phase Transitions

Christian Borgs
Institute of Theoretical Physics, Freie Universität Berlin, Arminallee 14, 1 Berlin 33, Germany

Roman Kotecký
Department of Mathematical Physics, Charles University, V Holešovičkách 2, Praha 8, Czechoslovakia and Center for Theoretical Studies, Charles University, Ovocný trh 5, Praha 1, Czechoslovakia
(Received 9 April 1990; revised manuscript received 17 December 1990)

We present a rigorous description of finite-size effects for a large class of models with an asymmetric first-order transition, assuming that all phases contributing to the transition have a finite correlation length. If the model describes the coexistence of two phases, it is shown that, at sufficiently low temperatures, the shift of the transition point due to finite-size effects in a volume L^d with periodic boundary conditions is $O(L^{-2d})$, in contrast to certain claims in the literature. We also discuss different ways to determine the transition point from finite-size data, which involve only exponentially small systematic errors in L.

PACS numbers: 64.60.−i, 05.50.+q

First-order phase transitions are characterized by discontinuities in the first derivatives of the free energy in the idealized infinite-volume limit. However, experimental or simulation data are taken from finite samples, where the discontinuity is smoothed. As a result, it is *a priori* not clear where to locate the transition point, and sometimes even the order of the transition is difficult to distinguish. A clear understanding of the details of the finite-size rounding is therefore important when interpreting experimental or simulation data.

A typical situation in Monte Carlo simulations is to consider a cubic lattice system of volume $V=L^d$ with periodic boundary conditions [1]. For a symmetric first-order transition, say an Ising magnet, the finite-size scaling is well understood [2,3]: The dependence of the partition function on an ordering field h is well approximated by the sum

$$e^{\beta hmV}+e^{-\beta hmV}, \tag{1}$$

where m is the (infinite volume) spontaneous bulk magnetization per lattice site and β is the inverse temperature, $\beta=1/kT$. This yields the magnetization (under periodic boundary conditions)

$$m_{\text{per}}(T,h;L)\approx m\tanh(\beta hmV), \tag{2}$$

which is rounded on the scale L^{-d}.

The situation is less clear for models with asymmetric (field- or temperature-driven) transitions. An attempt to understand the finite-size scaling for such models was made by approximating the equilibrium probability distribution $P_L(s)$ of an order parameter s [for Ising-like systems $s=(1/V)\sum\sigma_i$ is the magnetization per lattice site] by a sum of two Gaussian distributions. Unfortunately, this approach leads to a controversy [3–5]. Two different results are obtained if, on one hand [4], the relative height of these Gaussians is chosen in such a way that the *area* under both peaks of P_L is equal at the transition point $h=h_t(\beta)$, or if, on the other hand [5], it is chosen in such a way that both peaks of P_L have equal *height* at the transition point [6,7].

Our aim in this Letter is not only to resolve this controversy, but in general, to put the theory of finite-size effects on a rigorous footing. Restricting ourselves, for the moment, to the case of field-driven transitions, we will consider the magnetization

$$m_{\text{per}}(h,L)=\frac{L}{\beta L^d}\frac{d}{dh}\ln Z_{\text{per}}(h,L), \tag{3}$$

and the susceptibility

$$\chi_{\text{per}}=\frac{dm_{\text{per}}(h,L)}{dh}, \tag{4}$$

where $Z_{\text{per}}(h,L)$ is the partition function in a volume $V=L^d$ with periodic boundary conditions. Our results will cover a large class of models [8], including perturbed Ising models at low temperatures, large-N lattice Higgs models, lattice $P(\phi)_d$ and continuum $P(\phi)_2$ models, and more generally, all those models which can be treated by the Pirogov-Sinai theory [9]. One important class of models that do not have a contour representation and to which neither the Pirogov-Sinai theory, nor our results apply, are Heisenberg-like systems with continuous symmetries.

The theory presented here starts from the fact, proven in Ref. [8], that the partition function of a model describing the coexistence of N phases at $h=h_t$ is, at low temperatures, very well approximated [10] by

$$Z_{\text{per}}(h,L)\approx\sum_{q=1}^{N}\exp\{-f_q(h)\beta L^d\}, \tag{5}$$

where $f_q(h)$ is some sort of "metastable free energy" of the phase q. The quantity $f_q(h)$ is equal to the free energy $f(h)$ of the model whenever q is stable, and $f_q(h)>f(h)$ if q is unstable. While it is not expected that $f_q(h)$ can be chosen as an analytic function [11], it still may be introduced in such a way that it is differentiable [in Ref. [8] the differentiability of the fourth order is needed and, indeed, it is shown that $f_q(h)$ may be chosen to be C^4].

There are two interrelated problems concerning the approximation (5). One is the meaning of the words "well approximated," and the second is a definition of metastable free energies $f_q(h)$, $q=1,\ldots,N$. Without entering into technicalities, we describe here only the main ideas. First, representing the configurations in a geometrical fashion as regions of ground states separated by collections of energetically unfavorable contours [12], one expresses $Z_{\rm per}(h,L)$ as a partition function of a "contour model." If we disregard all configurations with contours wrapped around the torus, thus committing an error of the order $e^{-c\beta h}$, we get $Z_{\rm per}(h,L)$ as a sum of N terms, $Z_q(h,L)$, each describing a "gas of excitations" immersed in the qth phase. While all excitations about a stable phase are exponentially damped, excitations about an unstable phase which are larger than a certain critical size $L_c(h)\sim 1/(|h-h_t|)$ are in fact favorable if they introduce a transition from an unstable into a stable phase [13]. Following an idea originally appearing in Ref. [14], we now introduce a modified partition function, $Z_q^{\rm trunc}(h,L)$, where these excitations are artificially suppressed (see Ref. [8] for details). Its logarithm may then be analyzed, at low temperatures, by a convergent expansion and one may show that the corresponding free energy $f_q(h)$ is equal to the free energy $f(h)$ of the full model if q is stable, while $f_q(h)-f(h)\geq {\rm const}|h-h_t|$ if q is unstable.

Let us now assume that $|h-h_t|L\ll 1$. Then $L<L_c(h)$, all excitations are exponentially damped, $Z_q(h,L)=Z_q^{\rm trunc}(h,L)$, and its logarithm may be analyzed by a convergent expansion whether q is stable or unstable. Since $Z_q(h,L)$ is defined in a periodic box, the expansion for its logarithm contains no surface terms and $Z_q(h,L)$ is equal to $\exp\{-f_q(h)\beta L^d\}$, except for an exponentially small error $\exp\{-f_q(h)\beta L^d\}O(e^{-c\beta L})$. This justifies the approximation (5) in the region [15] $|h-h_t|L\ll 1$, with an explicit error bound $\exp\{-f_q(h)\beta L^d\}O(e^{-c\beta L})$. Note that the restriction $|h-h_t|L\ll 1$ is actually no restriction in the context considered here, since the rounding of the infinite-volume jump takes place in the region where $|h-h_t|=O(L^{-d})\ll L^{-1}$. Let us mention that the use of cluster expansions requires low temperatures. For a model like the Ising model it means that our proofs are valid only for temperatures up to about half the critical temperature, even though we believe that the approximation (5) holds for all temperatures below the critical temperature (with the constant $c\beta$ in the above error bound replaced by a constant of the order $1/\xi$, where ξ is the correlation length).

Once the approximation (5) is justified, one can use it to evaluate the finite-size behavior of $m_{\rm per}(h,L)$. Restricting ourselves to the coexistence of two phases ($N=2$), one of them, say +, being stable for $h\geq h_t$ and the other one for $h\leq h_t$, we introduce $m_{\rm per}(h)$ and $\chi_{\rm per}(h)$ as the $L\to\infty$ limit of (3) and (4), respectively, and define $m_0=(m_++m_-)/2$, $m=(m_+-m_-)/2$, with $m_\pm=m_{\rm per}(h_t\pm 0)$; χ, χ_0, χ_+, and χ_- are defined in a similar way. Expanding $f_q(h)$ around $h=h_t$, and using the fact that $f_q(h)=f(h)$ if h is stable to calculate the coefficients, we obtain

$$f_q(h)=f(h_t)+m_q(h-h_t)+\tfrac{1}{2}\chi_q(h-h_t)^2+O((h-h_t)^3)\,. \qquad (6)$$

As a consequence (see Ref. [8]), one can prove that

$$m_{\rm per}(h,L)=m_0+\chi_0(h-h_t)+[m+\chi(h-h_t)]\tanh\{L^d\beta[m(h-h_t)+\tfrac{1}{2}\chi(h-h_t)^2]\}+R(h,L)\,, \qquad (7)$$

where

$$|R(h,L)|\leq e^{-b_0\beta L}+K_1(h-h_t)^2 \qquad (8)$$

for some constants $b_0>0$, $K_1<\infty$. Notice that (7) is equivalent to formula (25) of Ref. [4] and thus offers its rigorous justification, while it disagrees with the corresponding formula of Ref. [5].

Concerning the shift of the transition due to finite-size effects, we will consider several definitions of the finite-volume transition point: (a) The point $h_{\max}(L)$ where the susceptibility $\chi_{\rm per}(h,L)$ is maximal, (b) the point $h_0(L)$ where $m_{\rm per}(h,L)=m_0$, and (c) the point $h_t(L)$ that is defined in the following way. Consider the quantity

$$N(h)=\lim_{L\to\infty}Z_{\rm per}(h,L)e^{\beta f(h)L^d}. \qquad (9)$$

It turns out that the limit $N(h)$ exists and that it equals the number of stable phases [16] [that means $N(h)=2$ for $h=h_t$ and $N(h)=1$ for $h\neq h_t$]. Indeed, this can be immediately seen from (5) and the property that $f_q(h)\geq f(h)$, with the equality if and only if the phase q is stable. Since the number of stable phases discontinuously increases at the coexistence point, it seems natural to define $h_t(L)$ as that point where a suitable finite-volume approximation to $N(h)$, say

$$N(h,L)=\left[\frac{Z_{\rm per}(h,L)^{2^d}}{Z_{\rm per}(h,2L)}\right]^{1/(2^d-1)}, \qquad (10)$$

attains its maximum. By an explicit calculation, it is easy to see that $h_t(L)$ may be equivalently defined as the point where

$$m_{\rm per}(h,L)=m_{\rm per}(h,2L)\,. \qquad (11)$$

Using (7) one may easily calculate $h_{\max}(L)$, yielding the shift

$$h_{\max}(L)=h_t+(3\chi/2\beta^2m^3)L^{-2d}+O(L^{-3d})\,. \qquad (12)$$

On the other hand, $|m_{\rm per}(h_t,L)-m_0|\leq O(e^{-b_0\beta L})$ and

$|N(h_t)-N(h_t,L)| \le O(e^{-b_0\beta L})$. One therefore should expect that the values $h_0(L)$ and $h_t(L)$ differ from h_t only by an exponentially small error. Indeed, such a statement has been proven in Ref. [8]. While $h_0(L)$ is difficult to determine from simulation data because the value m_0 is not known *a priori*, $h_t(L)$ can be easily determined, either using the available methods [17] to calculate partition functions, or using the equivalent definition (11). We therefore propose $h_t(L)$ as a new way to determine h_t from finite-size data [18].

For the sake of concreteness, we finally state our results in the form of two theorems. As an explicit model we take a perturbed Ising model with Hamiltonian

$$H=\sum_{|i-j|=1}\sigma_i\sigma_j+\sum J_A\prod_{i\in A}\sigma_i+h\sum\sigma_i\,, \tag{13}$$

where $A\subset\mathbb{Z}^d$ are finite sets of lattice sites, $J_A=0$ if $\mathrm{diam}A>R$ for some fixed R, and $\sup_A|J_A|\le\epsilon$ for a sufficiently small ϵ. We then have the following two theorems [8].

Theorem 1.—For some constants K_0, K_1, and b_0 and temperatures low enough, we have the formula (7) with the bound (8), and further (i)

$$m_{\rm per}(h)=\lim_{N\to\infty}m_{\rm per}(h,L)=\begin{cases}m_-(h) & \text{for } h<h_t\,,\\ \frac{1}{2}[m_-(h)+m_+(h)] & \text{for } h=h_t\,,\\ m_+(h) & \text{for } h>h_t\,,\end{cases} \tag{14}$$

where $m_\pm(h)=-df_\pm(h)/dh$, and [19] (ii)

$$|m_{\rm per}(h,L)-m_{\rm per}(h)|\le e^{-b_0L}+K_0\exp\{-b_0|h-h_t|L^d\}\,. \tag{15}$$

Theorem 2.—For a fixed constant δ and sufficiently low temperatures, one has (i)

$$h_{\max}(L)=h_t+(3\chi/2\beta^2m^3)L^{-2d}+O(L^{-3d})\,; \tag{16}$$

(ii) in the interval $[h_t-\delta,h_t+\delta]$, there exists a unique $h_0(L)$ for which $m_{\rm per}(h,L)=m_0$, and for this $h_0(L)$ one has $h_0(L)=h_t+O(e^{-b_0\beta L})$; and (iii)

$$h_t(L)=h_t+O(e^{-b_0\beta L})\,. \tag{17}$$

We conclude with three remarks.

(1) The formula (5) is valid for $N>2$ as well. As a consequence, one may analyze the finite-size behavior of the magnetization for the coexistence of more than two phases as well (see Sec. 5 of Ref. [8]). Even though the rigorous statement is proven only for the class of models mentioned above and at low temperatures, we believe that it is valid whenever a finite number of phases having finite correlation length take part in the transition. Thus we are led to the following conjecture:

Consider a first-order transition with driving parameter t and with the order parameter $X(t)=-df/dt$ jumping from X_- to X_+ at $t=0$. Assume that N_1 phases coexist for $t<0$ and N_2 for $t>0$, all of them coexisting at $t=0$ and all of them having finite correlation length. Then the finite-volume order parameter, under periodic boundary conditions, should scale like

$$\frac{X_-+X_+}{2}+\frac{X_--X_+}{2}\tanh\left[\frac{X_--X_+}{2}t|V|+\ln\frac{N_1}{N_2}\right], \tag{18}$$

if $t|V|$ stays constant and t goes to 0. The volume V should be nearly cubic (e.g., a parallelepiped with sides L_i such that $|V|\exp\{-\tau\min L_i\}\le 1$).

(2) Note that our work does not cover the case of long cylinders (not satisfying the condition above) that show a different finite-size scaling. See the Privman and Fisher paper [2] for a discussion of the case where two phases are related by a symmetry, and Ref. [20] for a rigorous version discussing the general case of N phases not necessarily related by a symmetry.

(3) The methods presented here do not depend on the fact that the phase transition is field driven. In fact, they have been applied to the q-state Potts model [21] which undergoes a temperature-driven first-order transition if q is large enough (for $d=2$, q must be larger than 4). It has been shown [22] for sufficiently large [23] q (thus verifying in this case the above conjecture) that the mean energy can be approximated by

$$E_{\rm per}(\beta,L)\approx E_0+E\tanh\{E(\beta-\beta_t)L^d+\tfrac{1}{2}\ln q\}\,. \tag{19}$$

As a consequence, the inverse temperature $\beta_{\max}(L)$, where the slope of $E_{\rm per}(\beta,L)$ is maximal is shifted by

$$-[\ln(q)/2E]L^{-d}+O(L^{-2d})\,, \tag{20}$$

while the inverse temperature $\beta_t(L)$ for which $N(\beta,L)$ is maximal differs from β_t only by an exponentially small error $O(q^{-bL})$, where $b>0$ is a constant [24]. Considering also the inverse temperature $\beta_V(L)$ where the so-called Binder parameter attains its minimum, one can show that its shift is again of the order L^{-d}, but with the coefficient differing from that in (22). Calculating the coefficients in the particular case when $d=2$ and $q=10$, we get for the shifts of temperatures

$$\begin{aligned}kT_{\max}(L)&=0.7012+1.63L^{-2}+O(L^{-4})\,,\\ kT_V(L)&=0.7012+2.39L^{-2}+O(L^{-4})\,,\end{aligned} \tag{21}$$

in very good agreement with the numerical data from Refs. [5] and [6].

[1] In fact, the conclusions of the present Letter remain true

not only for the considered block geometry, but also for any parallelepiped with sides growing simultaneously to ∞, each as a (possibly fractional) power of L.

[2] M. E. Fisher and A. N. Berker, Phys. Rev. B **26**, 2507 (1982); V. Privman and M. E. Fisher, J. Stat. Phys. **33**, 385 (1983).

[3] For a recent review, see V. Privman, in *Finite Size Scaling and Numerical Simulation of Statistical Systems,* edited by V. Privman (World Scientific, Singapore, 1990).

[4] K. Binder and D. P. Landau, Phys. Rev. B **30**, 1477 (1984).

[5] Murty S. S. Challa, D. P. Landau, and K. Binder, Phys. Rev. B **34**, 1841 (1986).

[6] Very recently, this probability distribution was carefully discussed by Jooyoung Lee and J. M. Kosterlitz, Phys. Rev. Lett. **65**, 137 (1990).

[7] It was pointed out that these controversies also arise in transfer-matrix studies; see Sec. 4 in V. Privman and J. Rudnick, J. Stat. Phys. **60**, 551 (1990), and references therein.

[8] C. Borgs and R. Kotecký, J. Stat. Phys. **61**, 79 (1990).

[9] S. Pirogov and Ya. Sinai, Theor. Math. Phys. Engl. Transl. **25**, 1185 (1975); **26**, 39 (1976).

[10] The errors involved in (5) can be bounded by $e^{-b_0\beta L}$ multiplied by the largest term on the right-hand side, that is, by $e^{-b_0\beta L}e^{-\beta|V|f(h)}$, where $b_0>0$ is a constant.

[11] This would be an analytic continuation of $f(h)$ over the point h_t that is expected [see M. E. Fisher, Physics (Long Island City, N.Y.) **3**, 255 (1967); J. S. Langer, Ann. Phys. (N.Y.) **41**, 108 (1967)] and sometimes known not to exist [see S. N. Isakov, Commun. Math. Phys. **95**, 427 (1984)].

[12] The fact that one pays for contours by an energy proportional to their length is the main assumption, the Peierls condition, of the underlying Pirogov-Sinai theory.

[13] In the droplet model, $L_c(h)$ would be the diameter of a critical droplet.

[14] M. Zahradník, Commun. Math. Phys. **93**, 559 (1984).

[15] The details of the proof of (5) for $|h-h_t|L \not\ll 1$ can be found in Ref. [8].

[16] This fact was already observed by C. Borgs and J. Imbrie, Commun. Math. Phys. **123**, 305 (1989).

[17] The method of multiple histograms seems to be useful for that purpose. See A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).

[18] Recently different versions of this criterion were successfully tested in numerical simulations of two-dimensional Potts ferromagnets: C. Borgs and W. Janke (to be published).

[19] Notice that the bounds (8) and (15) are valid for all h. However, while (8) is useful only for $|h-h_t|\ll 1$, the bound (15) is nontrivial for $|h-h_t|\gg L^{-d}$.

[20] C. Borgs and J. Imbrie (to be published).

[21] For a general review of the Potts model, see, e.g., F. Y. Wu, Rev. Mod. Phys. **54**, 235 (1982).

[22] C. Borgs, R. Kotecký, and S. Miracle-Solé, J. Stat. Phys. (to be published). We present here only the formula with first-order approximations in $\beta-\beta_t$. Similar formulas for Potts models were very recently proposed by A. Billoire *et al.*, Phys. Rev. B **42**, 6747 (1990), and Nucl. Phys. (Proc. Suppl.) **B17**, 230 (1990).

[23] Both results (19) and (20) are proven for large q. We were not trying to find a concrete estimate of how large a q one actually has to take.

[24] Again we expect an error $O(e^{-L/\xi})$, with ξ of the order of the correlation length for all $q>q_c$ (recall that $q_c=4$ for $d=2$). For $d=2$ and $q=8,10$, this was numerically verified in Ref. [18].