FIRST-ORDER PHASE TRANSITIONS IN FINITE SYSTEMS I: PERIODIC BOUNDARY CONDITIONS

Igor Medved'
Department of Physics, Constantine the Philosopher University, Tr. A. Hlinku 1, 949 74 Nitra
E-mail: imedved@ukf.sk

Summary We briefly review rigorous results on the finite-size effects near first-order phase transitions at which a two-phase coexistence takes place. We consider a large class of statistical mechanical models in (hyper)cubic volumes with periodic boundary conditions at low temperatures. The results show a universal behavior of the asymptotic smoothing of the phase transition discontinuities. The determination of the transition point from the finite-size data is presented.

Abstrakt. Stručne popisujeme rigorózne výsledky o efektoch konečného objemu blízko fázových prechodov prvého druhu, kde prebieha koexistencia dvoch fáz. Uvažujeme veľkú triedu modelov štatistickej mechaniky v (hyper)kubických objemoch s periodickými okrajovými podmienkami pri nízkych teplotách. Výsledky ukazujú univerzálne chovanie asymptotického vyhladenia nespojitostí fázových prechodov. Uvádzame aj určenie bodu prechodu z konečnoobjemových dát.

1. INTRODUCTION

First-order phase transitions are determined by discontinuities of the first derivatives of some thermodynamic potential, for example, of the specific free energy. The derivatives correspond to macroscopic observables, like the specific internal energy and the specific magnetization.

In order that such observables may possibly exhibit a discontinuity, it is necessary to go to the *idealized* infinite volume, that is, to take the thermodynamic limit. However, *real* macroscopic systems are always finite (although they contain a huge number of constituting particles — atoms, molecules, etc.), and no discontinuities can appear: thermodynamic potentials in finite volumes are, as a rule, analytic functions.

As a matter of fact, the infinite-volume jumps are in finite volume smoothed out into rounded transitions. The larger the system is, the more abrupt the roundings become. The positions of the rounded transitions are in general shifted with respect to those in the thermodynamic limit, see Fig. 1. Moreover, the second-order derivatives (the specific heat capacity, the susceptibility, etc.) and the higher-order derivatives that have singularities of the δ-function type in infinite volumes are changed into sharp but finite spikes once the system is finite. The points where the spikes are maximal are natural (though not the only) candidates to describe the shifts of the rounded transitions: at these points the transitions are steepest.

The phenomena connected with such roundings of relevant physical quantities are commonly referred to as the *finite-size effects* near (or at) first-order phase transitions. The aim of this paper is to give a brief review of rigorous results of Borgs and Kotecký[1] that describe in detail these effects for

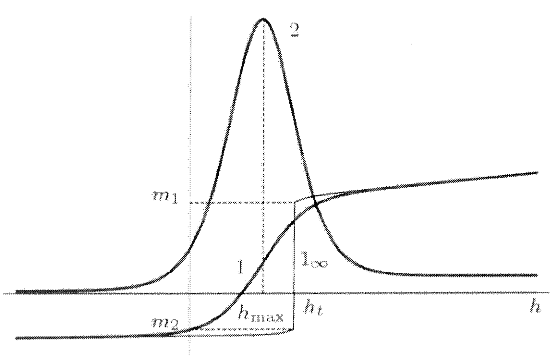

Fig. 1: An infinite-volume observable 1_∞ that is the first derivative of a thermodynamic potential exhibits a jump between m_1 and m_2. However, the jump is smoothed out in its finite-volume version 1. The second derivative of a thermodynamic potential 2 has the shape of a spike in a finite volume, while it has a singularity of the δ-function type (not shown). Its maximum position $h_{\max}$ is in general shifted with respect to the infinite-volume transition point h_t.

large but finite d-dimensional cubic systems of size L^d with periodic boundary conditions. Since there are no phase transitions in d=1, we take d≥2 in the sequel. For the sake of simplicity, we restrict ourselves to the situations, where there are just two coexisting phases. The general coexistence of several phases is discussed in Ref. [1].

The analysis of Borgs and Kotecký is based on the powerful methods and techniques of the Pirogov-Sinai theory [2,3] of first-order phase transitions. Therefore, their results can be applied to all the models that can be treated by this theory. Roughly speaking, these models are those whose configurations can be rewritten in a geometrical fashion as regions of ground states mutually separated by collections of energetically unfavorable barriers (usually called contours), see Eq. (5) below. They include, for instance, spin lattice models with a finite number of ground states and finite-range interactions (such as the Ising model) at sufficiently low temperatures or the q-state Potts models with q sufficiently large. A prominent group of models to which the results cannot be applied are Heisenberg-like systems with continuous symmetries.

It should be remarked that the power of the Pirogov-Sinai theory consists partly in the fact that it allows one to treat asymmetric transitions when, arguably, no serious alternatives to the theory are available. Moreover, both the cases of field- as well as temperature-driven phase transitions can be handled.[4]

Before stating and discussing the very results on the finite-size effects from Ref. [1] in Section III, we first introduce, in Section II, a simple toy model that catches and, to some extent, gives an insight into the essential features of these effects. A few concluding remarks are appended in Section IV.

2. A TOY MODEL

It is often very illuminating to examine first a toy model that extremely simplifies the problem under consideration, yet provides a good amount of insight. Let us therefore consider the following "two-level" model.

The model has a finite number of microscopic configurations $\sigma_1,\ldots,\sigma_{n_1+n_2}$, where $n_1, n_2 \geq 1$. We want to study the statistical mechanical properties of an extensive observable M (for example, the magnetization or the internal energy) that we assume to be a sum of translation invariant local functions of the configurations. The Hamiltonian H of the model contains interaction potentials that are also translation invariant and of finite-range.

In order to make the model extremely simple, we will assume that H is a constant (it has the same value E_0 for all microscopic configurations), whereas M attains two values: M_1 for the first n_1 configurations and $M_2 \neq M_1$ for the remaining n_2 configurations. This situation may be thought of as the zero-temperature approximation of a system that has two ground states with the degeneracy n_1 and n_2, respectively, and the corresponding values M_1 and M_2 of M. An approximation of this sort may be expected, under appropriate circumstances, to give a plausible picture even when the system's temperature is slightly raised above zero.

Confining the system to a cube of size L^d with L large and imposing periodic boundary conditions, we first observe that $M_i = m_i L^d$, i=1,2, where m_i is the specific counterpart of M_i (specific magnetization, for instance). Similarly, $E_0 = e_0 L^d$. Using now the standard procedure, we modify the original Hamiltonian H of the system by adding the term $-hM$ to it, where h is a parameter (physically, it is a conjugate quantity to M, like an external magnetic field). Thus obtained generalized model is suitable for studying the properties of M. To get back to the original model, it suffices to take h=0 in the end.

The partition function of the generalized model is

$$Z_{toy}(h) \equiv \sum_{k=1}^{n_1+n_2} e^{-\beta[H(\sigma_k)-hM(\sigma_k)]} = n_1 e^{-\beta e_1(h)L^d} + n_2 e^{-\beta e_2(h)L^d}, \tag{1}$$

where $\beta = \frac{1}{k_B T}$ is the inverse temperature and $e_i(h) \equiv e_0 - hm_i$. Hence, the specific mean value of M is

$$m_{toy}(h) \equiv \frac{1}{L^d}\langle M \rangle_{toy} = \frac{1}{Z_{toy}} \sum_{k=1}^{n_1+n_2} \frac{M(\sigma_k)}{L^d} e^{-\beta[H(\sigma_k)-hM(\sigma_k)]} = \frac{1}{\beta L^d}\frac{\partial}{\partial h}\ln Z_{toy}(h) = \frac{m_1+m_2}{2} + \frac{m_1-m_2}{2}\tanh[\beta\frac{m_1-m_2}{2}(h-h_{\max}^{toy})L^d], \tag{2}$$

where

$$h_{\max}^{toy} = \frac{\ln \frac{n_2}{n_1}}{(m_1 - m_2)\beta L^d}. \tag{3}$$

Similarly, the specific variance of M (corresponding to the specific susceptibility, say) is

$$\chi_{toy}(h) \equiv \frac{1}{L^d}\left(\langle M^2\rangle_{toy} - \langle M\rangle_{toy}^2\right) = \frac{1}{\beta}\frac{\partial m_{toy}(h)}{\partial h}$$
$$= \left(\frac{m_1 - m_2}{2}\right)^2 L^d \cosh^{-2}\left[\beta\frac{m_1 - m_2}{2}(h - h_{\max}^{toy})L^d\right]. \tag{4}$$

Higher-order derivatives of $\ln Z_{toy}(h)$ may be readily obtained in a similar manner. Notice that $h_{\max}^{toy}$ is the point at which $\chi_{toy}(h)$ attains its maximum.

From Eqs. (2) to (4) we may draw these conclusions (c.f. Fig. 2). The specific mean value of M (that is, the first derivative of the thermodynamic potential with respect to h) interpolates between m_1 and m_2 as the function tanh, and the change from one value to the other takes place within an extremely narrow region of the order L^{-d}. The specific variance (the second derivative) exhibits a very tall and sharp spike determined by the function $\cosh^{-2}$. The height of the spike is of the order L^d and its width is of the order L^{-d}. If the ground-state degeneracies are equal ($n_1 = n_2$), the terms containing $\ln\frac{n_2}{n_1}$ vanishes from the above equations, leading, in particular, to $h_{\max}^{toy} = 0$.

It is worth pointing out that the energy of the system can also be chosen as the observable M. Formally, this corresponds to first setting β=1 and $E_0 = 0$ and then substituting -β for h and two energy levels E_1 and E_2 for M_1 and M_2, respectively (thus, e_i for m_i, i=1,2).

3. THE SETTING AND RESULTS

In this section we will see that most of the behavior of the toy model remains true even for models that exhibit a first-order phase transition with a two-phase coexistence. In particular, it will not be essential whether the phase transition is temperate-driven or field-driven, both the cases can be analyzed analogously. To be specific, however, we will have in mind the latter case. A rigorous analysis of a temperature-driven transition can be found in Ref. [4].

As stated in the Introduction, the models we consider have the property that their configurations can be rewritten as regions of ground states that are separated by a collection of barriers called contours. To be somewhat more precise, we consider the models that are defined on a d-dimensional torus **T** (periodic boundary conditions) with sides of length L in each direction and whose partition function has the form

$$Z_L = \sum_{\{\gamma_\alpha\}} e^{-\beta e_1 R_1 - \beta e_2 R_2} \prod_\alpha \rho(\gamma_\alpha). \tag{5}$$

Here the sum goes over collection of non-overlapping contours (connected unions of closed unit cubes in $\boldsymbol{R}^d$), R_1 and R_2 are the regions of $\mathbf{T} \setminus \bigcup_\alpha \gamma_\alpha$ occupied by the first and the second ground state, respectively, e_1 and e_2 are the corresponding specific ground-state energies, and $\rho(\gamma_\alpha)$ is a translation invariant weight of the contour γ_α. For a full and precise description of Z_L we refer the reader to Ref. [1]. We will again use $n_1, n_2 \geq 1$ to denote the degeneracies of the ground states.

Thus, contours indeed play the role of boundaries between the regions of different ground states (put differently, they represent perturbations of the ground states), and Z_L is simply the partition function of a hard-core gas of contours. Remarkably, there is a large class of statistical mechanical models whose partition function can be put into the form from Eq. (5), see the Introduction.

The energies e_i as well as the contour weights $\rho(\gamma_\alpha)$ are assumed to be smooth functions of a parameter h. In addition, we assume that there is a unique point h_0 at which $e_1(h)$ and $e_2(h)$ coincide. Finally, the label i=1,2 is chosen in such a way that, say,

$$\frac{\partial}{\partial h}(e_1 - e_2) < 0. \tag{6}$$

The assumption (6) that the derivatives of e_1 and e_2 differ enables one to prove that the model exhibits a first-order phase transition.

The main point of the Pirogov-Sinai theory is that whenever the energy paid for a contour γ_α separating the ground-state regions is proportional to its size $|\gamma_\alpha|$ (that is, to the number of unit cubes in γ_α),[1] then the zero-temperature behavior of the model (to be read off from the ground-state energies $e_1(h)$ and $e_2(h)$) prevails also when the temperature is slightly increased above zero. Each ground state then gives rise to a single low-temperature phase that is just a small perturbation of this ground state. As a consequence, these results follow whenever β and *L* are sufficiently large.[1]

One can introduce "meta-stable" specific free energies $f_1(h)$ and $f_2(h)$ of the first and second phase, respectively, through which a very detailed control over the partition function Z_L can be established:

1. $f_1(h)$ and $f_2(h)$ are both smooth functions of *h*;
2. $f_1(h)$ and $f_2(h)$ are very well approximated by the ground-state energies, namely,

$$f_i(h) = e_i(h) + O(e^{-const\,\beta}), \tag{7}$$

   and similar relations are true for the derivatives of $f_i(h)$;
3. the true specific free energy of the model $f(h) \equiv -\frac{1}{\beta}\lim_{L\to\infty}\frac{1}{L^d}\ln Z_L(h)$ exists and $f(h) = \min\{f_1(h), f_2(h)\}$;
4. we have

$$Z_L(h) = (n_1 e^{-f_1(h)L^d} + n_2 e^{-f_2(h)L^d})[1 + O(e^{-const\beta L})]. \tag{8}$$

The symbol $O(x)$ stands here and below for an error term that can be bounded by *x* with the constant depending only on the dimension *d* (the bounds are uniform in β, *L*, and *h*). Using this detailed control over Z_L, the following consequences can be derived in a straightforward way.[1]

First, in view of (6) and (7), there is a unique point h_t at which $f_1(h)$ and $f_2(h)$ coincide, and $f(h) = f_1(h)$ for $h \geq h_t$, while $f(h) = f_2(h)$ for $h \leq h_t$. Moreover,

$$\frac{\partial}{\partial h}(f_1 - f_2)(h_t) < 0. \tag{9}$$

Therefore, h_t is the infinite-volume transition point, and, by (7), one has $h_t = h_0 + O(e^{-const\,\beta})$ (here the error term can be explicitly evaluated to an arbitrary precision, if desired).

Second, let us introduce the finite-volume quantities

$$m_L(h) \equiv \frac{1}{\beta L^d}\frac{\partial}{\partial h}\ln Z_L(h),\; \chi_L(h) \equiv \frac{1}{\beta}\frac{\partial m_L(h)}{\partial h} \tag{10}$$

and the numbers $m_1 \equiv -\frac{\partial f(h_t^+)}{\partial h}$, $m_2 \equiv -\frac{\partial f(h_t^-)}{\partial h}$, $\chi_1 \equiv -\frac{\partial^2 f(h_t^+)}{\partial h^2}$, and $\chi_2 \equiv -\frac{\partial^2 f(h_t^-)}{\partial h^2}$. These numbers are universal for a given model (independent of boundary conditions). There is a unique point $h_{max}(L)$ at which χ_L attains its maximum, and

$$h_{max}(L) = h_t + \frac{\ln\frac{n_2}{n_1}}{(m_1 - m_2)\beta L^d} + \frac{6(\chi_1 - \chi_2)}{(m_1 - m_2)^3\beta^2 L^{2d}} + O(L^{-3d}). \tag{11}$$

If *h* is in the interval $|h - h_t| \leq \frac{C}{L}$ (the value of the constant *C* is not essential), then

$$m_L(h) = \frac{m_1 + m_2}{2} + \frac{m_1 - m_2}{2}\tanh[\beta\frac{m_1 - m_2}{2}(h - h_{max}(L))L^d] + O(L^{-1}) \tag{12}$$

---

[1] This assumption is called the Peierls condition and it means an exponential decay of every contour weight $\rho(\gamma_\alpha)$ in $|\gamma_\alpha|$, see Refs. [1-3] for details.

$$\chi_L(h) = (\frac{m_1 - m_2}{2})^2 L^d$$
$$\times \cosh^{-2}[\beta \frac{m_1 - m_2}{2}(h - h_{\max}(L))L^d] + O(L^{d-1}). \quad (13)$$

On the other hand, if $| h - h_t | \geq \frac{C}{L}$, then m_L and χ_L are very well approximated by their infinite-volume limits,

$$m_L(h) = -\frac{\partial f(h)}{\partial h} + O(e^{-const\,\beta L}), \quad (14)$$

$$\chi_L(h) = -\frac{\partial^2 f(h)}{\partial h^2} + O(e^{-const\,\beta L}). \quad (15)$$

Eqs. (11) through (15) may be in short worded as follows (see Fig. 2). The quantity $m_L(h)$ *smoothly* interpolates between the values m_1 and m_2 according to the function tanh, whereas $\chi_L(h)$ has the shape of a spike approximated by the function $\cosh^{-2}$. The width of the region within which m_L abruptly (but smoothly) changes from m_1 to m_2 is of the order L^{-d}. Thus, the slope of m_L in this region, that is, the height of the spike exhibited by χ_L, is of the order L^d. This immense height (using probabilistic arguments, one would expect the height of χ_L, being the specific variance of M, see Eq. (4), to be of the order $L^0 = 1$) is an aftermath of the presence of the phase transition. Measuring the position of the maximum point $h_{\max}(L)$, from Eq. (11) we may trace the position of the transition point h_t. Notice that the terms containing $\ln \frac{n_2}{n_1}$ disappear from the equations once $n_1 = n_2$. In this case the shift $h_{\max}(L) - h_t$ is of the order L^{-2d} instead of L^{-d}.

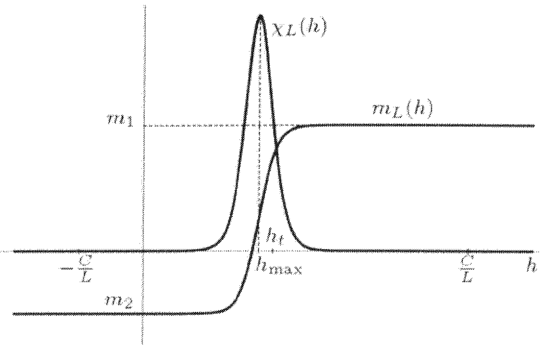

Fig. 2: In the interval $| h - h_t | \leq \frac{C}{L}$ *the finite-volume quantities* $m_L(h)$ *and* $\chi_L(h)$ *are determined by the function tanh and* $\cosh^{-2}$*, respectively. The shift* $h_{\max}(L) - h_t$ *is of the order* L^{-d} *(or* L^{-2d} *when* $n_1 = n_2$*).*

4. CONCLUDING REMARKS

In this paper we briefly reviewed the rigorous results obtained in Ref. [1] on the finite-size effect near first-order phase transitions for *d*-dimensional cubic systems with periodic boundary conditions. We considered the situation of a field-driven transition at which two phases coexisted. Actually, we presented the results in a slightly different, perhaps more convenient form than in Ref. [1]. In particular, we took into account possible degeneracies n_1 and n_2 of the two ground states, whereas in Ref. [1] $n_1 = n_2 = 1$, and all the terms containing $\ln \frac{n_1}{n_2}$ are missing (they are identically zero).

REFERENCES

[1] C. Borgs and R. Kotecký. A rigorous theory of finite-size scaling at first-order phase transitions. *J. Stat. Phys.*, **61**:79–119, 1990.

[2] C. Borgs and J. Z. Imbrie. A unified approach to phase diagrams in field theory and statistical mechanics. *Commun. Math. Phys.*, **123**:305–328, 1989.

[3] M. Zahradník. An alternate version of Pirogov-Sinai theory. *Commun. Math. Phys.*, **93**:559–581, 1984.

[4] C. Borgs, R. Kotecký, and S. Miracle-Solé. Finite-size scaling for the Potts models. *J. Stat. Phys.*, **62**:529–552, 1991.