



# Becoming Large, Becoming Infinite: The Anatomy of Thermal Physics and Phase Transitions in Finite Systems

David A. Lavis<sup>1,2</sup> · Reimer Kühn<sup>1</sup> · Roman Frigg<sup>2</sup>

Received: 23 July 2020 / Accepted: 14 June 2021  
© The Author(s) 2021

## Abstract

This paper presents an in-depth analysis of the anatomy of both thermodynamics and statistical mechanics, together with the relationships between their constituent parts. Based on this analysis, using the renormalization group and finite-size scaling, we give a definition of a large but finite system and argue that phase transitions are represented correctly, as incipient singularities in such systems. We describe the role of the thermodynamic limit. And we explore the implications of this picture of critical phenomena for the questions of reduction and emergence.

**Keywords** Scaling · Renormalization · Large systems · Incipient singularities · Reduction · Emergence

## 1 Introduction

Thermodynamics and statistical mechanics coexist in a collaborative relationship within the envelope of thermal physics. In many presentations of the subject, particularly in undergraduate texts, it is heuristically advantageous to intermingle the macroscopic concepts of thermodynamics with the micro-picture provided by statistical mechanics. And it is, of course, self-evident that statistical mechanics<sup>1</sup> needs the basic structure of thermodynamics with inter-theory connecting relationships defining the thermodynamic quantities like internal energy, temperature and entropy. On the other hand, there are some advantages, both aesthetic and mathematical, in producing an account of thermodynamics which makes no reference to the underlying microstructure of the system, as would seem to be one of the aims of (among

<sup>1</sup> At least in its application to physics, rather than to its more modern application to sociological and geographical problems.

✉ David A. Lavis  
david.lavis@kcl.ac.uk

<sup>1</sup> Department of Mathematics, King's College London, London WC2R 2LS, UK

<sup>2</sup> Centre for Philosophy of Natural and Social Science, London School of Economics, London WC2A 2AE, UK

others) the books of Giles [39] and Buchdahl [17] and the papers of Lieb and Yngvason.<sup>2</sup> For Buchdahl we have the first law implying the existence of the internal energy function  $U$  and Carathéodory's [23] version of the second law yielding the entropy  $S$  and temperature  $T$ ; for Lieb and Yngvason three sets of axioms accomplish the same task. This, together with an account of the nature of adiabatic processes (as described, for example, in [17, Chaps. 5, 6, 75, Sect. 2.1, 68, Sect. 2.1.1]) provides the basic framework into which the models of statistical mechanics are embedded.

This raises the question of how statistical mechanics and thermodynamics relate to each other. Attempts to answer this question run up against a problem. The neat labels 'statistical mechanics' and 'thermodynamics' mask the fact that neither theory is a monolithic bloc. Indeed, each has a complicated internal structure with several layers of different theoretical postulates and assumptions. So the question of how statistical mechanics and thermodynamics relate ought to be interpreted as the more complex question of (a) what the internal structure of each theory is and of (b) how the various parts of each theory relate to the various other parts of the other theory. The complexity of the internal structures of both theories, as well as the intricacy of their interrelations, seems to have been somewhat under-appreciated in the philosophical literature on the subject, and so the first aim of this paper is to present an in-depth analysis of the anatomy of both theories and the connections between their parts.<sup>3</sup>

Figure 1 provides a schematic advance summary of the analysis that we develop in this paper. It sees statistical mechanics and thermodynamics as parallel developments, each decomposed into separate levels representing the stages of theory-based development in which features are added to the system. The cross-interactions between the levels in the two columns contain interventions integral to this development. On the left are the levels for thermodynamics, as described in detail in Sect. 2. These levels are related to each other by adopting special assumptions, beginning at the bottom with *basic thermodynamic theory* (labelled TD1). Adding the extensivity assumption to this theory takes us to the next level, the *density representation of thermodynamics* (labelled TD2). Augmenting TD2 with the notion of phase transitions and critical phenomena (PTCP) gives *thermodynamics with PTCP* (labelled TD3). Finally, supplementing TD3 with a version of the Kadanoff scaling hypothesis leads us to *thermodynamics with scaling theory* (labelled TD4).

The parallel development for statistical mechanics is represented on the right of Fig. 1, as described in detail in Sect. 3. The picture here is a little more complicated, involving, as we explain in our discussion, three different paths. At the bottom is the fundamental theory, which we here take to be *Gibbsian statistical mechanics*

<sup>2</sup> Lieb and Yngvason [75] is the most comprehensive account of their work, with briefer versions in Lieb and Yngvason [74] and Lieb and Yngvason [76]. The extension to non-equilibrium is given in Lieb and Yngvason [77].

<sup>3</sup> For surveys of the philosophical discussions about statistical mechanics and thermodynamics see, for instance, Sklar [117], Uffink [121] and Frigg [35].

(labelled SM1).<sup>4</sup> Assuming that the systems to which the theory is applied are large leads us to the next layer, *large statistical mechanical systems* (labelled SM2). This marks a branching point in the structure of the theory: three different additions can be made to SM2, resulting in three different branches. Adding the thermodynamic limit to SM2 leads to the *statistical mechanics of infinitely large systems* (labelled SM3). Adding renormalization group techniques to SM2 leads to the *renormalization group approach to statistical mechanics* (labelled SM4). Finally, adding the analysis of phase transitions for finite systems<sup>5</sup> to SM2 leads to the *statistical mechanics of finite-system phase transitions* (labelled SM5).

It is our aim in this work to keep the developments of thermodynamics and statistical mechanics as separate as possible, in order to make visible the internal structure of each separate theory. However, as indicated above, on close examination it becomes evident that there are in fact some ‘messages’, both implicit and explicit, sent *from statistical mechanics* (FSM), that is to say from the microstructure, to thermodynamics, which provides the macrostructure. These are spelled out in FSM–1, FSM–2, FSM–3, FSM–4. In the other direction the connecting relationships *from thermodynamics* (FTD), labelled FTD–1, FTD–2, FTD–3, identify quantities in statistical mechanics with thermodynamic variables. As we shall see FSM–1 also plays a role in the connecting process and can be seen as in dialogue with FTD–3. The remaining interventions FSM–2, FSM–3, FSM–4, can be viewed as an aid to the clarification of a number of important issues. We discuss these links between elements of both theories in the appropriate places in Sects. 2 and 3.

Much of the recent interest in the relationship between thermodynamics and statistical mechanics has concentrated on PTCP. It is the second aim of this paper to revisit the issue of PTCP in the light of our analysis of the internal structure of the two theories and their interrelations. Doing so will lead us to some unexpected, and we think important, conclusions.

In the modern theory of critical phenomena, dating from the middle of the 1960s,<sup>6</sup> critical exponents, which classify the type of singular behaviour in the approach to a critical region, play an important role. In our development of thermodynamics in Sect. 2 scaling theory is the final destination with scaling laws relating these critical exponents. However, as already indicated and as described below, thermodynamics is a structured shell into which particular models are embedded, either by the assumption of a phenomenological form for the entropy function or from statistical mechanics. In the absence of such an embedding it is not possible to calculate values for critical exponents, nor to discuss universality. This is the idea [54] that all critical situations<sup>7</sup> can be divided into universality classes, characterized by the values of their critical exponents and differentiated by a small number of properties

<sup>4</sup> We set aside Boltzmannian statistical mechanics. For discussion of the relation between Gibbsian and Boltzmannian statistical mechanics see Lavis [66] and Frigg and Werndl [36].

<sup>5</sup> Where, as described in Sect. 4, phase transitions are defined in a way which avoids the involvement of singularities.

<sup>6</sup> For an historic account see Domb [28].

<sup>7</sup> Of which there may be more than one in any model.

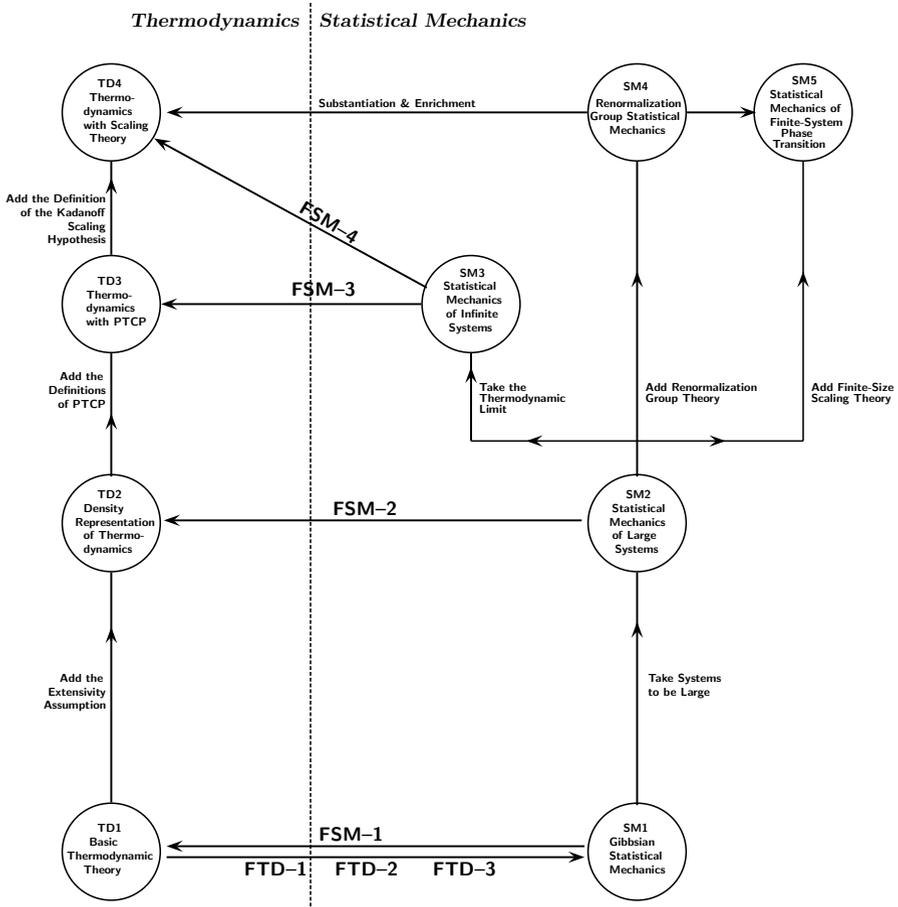


Fig. 1 Schematic representation of the relationship between thermodynamics and statistical mechanics

of which the most important are the (physical) dimension  $d$  of the system and the symmetry group of the order parameter. The first, but not the second, of these plays an important role in our discussions,<sup>8</sup> in particular in the case of the Ising model, which we shall use as an illustrative example throughout this work. This, the most well-known and thoroughly investigated model in the statistical mechanics of lattice systems, is briefly described in Appendix 2. With the list of critical exponents given there for  $d = 2$ ,  $d = 3$  and  $d \geq 4$ , it provides an example of the dependence of these exponents and hence the universality class on the dimension of the system. The dimension  $d$  is also of importance, in our discussion of scaling theory in Sect. 2.4, of finite-size scaling in Sect. 3.4.2 and of phenomenological renormalization in Sect. 3.4.3(c).

<sup>8</sup> For an account of the role of the order parameter in critical phenomena see, for example, Binney et al. [15, Sect. 1.2].

These observations concerning universality classes together with the inter-theory connecting relationships FSM–2, FSM–3, FSM–4, provide the impetus to investigate, and clarify a number of important issues relating to PTCP. These are (not necessarily in the order in which they arise in the discussion):

- (i) Are infinite systems really necessary in thermodynamics or statistical mechanics and:
  - (a) If so, what for?
  - (b) If they are, is this solely because extensivity is not exactly true in most cases in statistical mechanics?
  - (c) Is the thermodynamic limit irrelevant to thermodynamics or has it already been implicitly applied?<sup>9</sup>
  - (d) Is the thermodynamic limit in statistical mechanics necessary for the implementation of the procedures of the renormalization group?
  - (e) Is there a meaningful way to represent PTCP in finite systems?
- (ii) Given that, in thermodynamics, critical behaviour involves discontinuities in densities and singularities in response functions, is this necessarily still the case in statistical mechanics?
- (iii) Are the ideas of enrichment and substantiation helpful in describing the relationship between thermodynamics and statistical mechanics?
- (iv) Where do reduction and emergence feature in the accounts of the relationship between thermodynamics and statistical mechanics?

As indicated, in the title of this work and by the progression between levels in the statistical mechanical column in Fig. 1, we will discuss these issues with a special focus on large systems and infinite systems. In particular we shall address the question as to where realism is to be found, in the study of large systems, because real systems are finite but large (in the sense that they typically have  $\sim 10^{23}$  constituents), or in the thermodynamic limit of an infinite system, because singular behaviour (in susceptibilities and compressibilities) is believed to be experimentally observed, and in theories this arises only in the thermodynamic limit. This broad categorization of large systems is refined in Sect. 4. The process of taking the thermodynamic limit is the determination of the asymptotic properties of a system as it becomes infinitely large. In general this will involve taking  $d$  limits in each of the linear dimensions of the system and such a  $d$ -dimensionally infinite system, which where appropriate we call a *fully-infinite system*, is implicitly the object of investigation by scaling theory in Sect. 2.4.<sup>10</sup> However, relevant to our discussions is the case of a *partially-infinite*

<sup>9</sup> It is an interesting observation that discussions of PTCP in the context purely of thermodynamics (e.g. Pippard [105, Chap. 9], Buckingham [18]) rarely if ever feel the need to invoke or even refer to the thermodynamic limit.

<sup>10</sup> Underlying this description is, of course, the question of the *existence* of the thermodynamic limit and whether it depends on the boundary conditions of the erstwhile finite system. For a discussion of these questions see, for example, Griffiths [41], Griffiths [43, pp. 10–41] and Ruelle [110, Chaps. 2, 3].

system, where the limit is taken in only  $\mathfrak{d} < d$  dimensions. Here it is  $\mathfrak{d}$  rather than  $d$  which should count for the critical behaviour as the dimension of the system. The idea underlying our approach to PTCP is that reality lies with *fully-finite systems* ( $\mathfrak{d} = 0$ ) and that the judgment as to whether the large system will show behaviour which in practical terms is indistinguishable from singular behaviour is based on comparing the behaviour of systems of ever increasing size to see whether their properties indicate convergence towards those of the infinite system. In principle, as described in Sect. 4 this limiting process is in all  $d$  dimensions. In practice, as we see in our discussion of  $d = 2$  transfer matrix calculations in Sect. 3.3, it also has relevance to the case where one limit has already been taken and increasing size is in the remaining dimension.

Thus, as we have indicated, Sects. 2 and 3 trace the steps in our developments of thermodynamics and statistical mechanics with the inter-theory connections between them; with Sect. 3.5 addressing different proposed resolutions to the contradiction between the finiteness of real systems and the perceived necessity of phase transitions being portrayed as singularities in infinite systems. Section 3.6 discusses the proposal of Mainwood [80] for representing the occurrence of phase transitions in finite systems. Using the account of finite-size scaling in Sect. 3.4.2 we propose in Sect. 4 our alternative quantitative account for phase transitions in finite *large* systems. Section 5 contains some after-thoughts on enrichment, substantiation, reduction and emergence and our conclusions are in Sect. 6.

## 2 From Classical Thermodynamics to Scaling Theory

Accounts of thermodynamics range from those designed for the practical needs of engineers to those which aim for a degree of mathematical rigour. However, all share some common features and assumptions some of which are at variance with the insights gained in statistical mechanics. As indicated above, we flag these differences in the form of messages from statistical mechanics (FSM–1 to FSM–4).

### 2.1 The Structure of Thermodynamics

All accounts of thermodynamics contain (in some form or another) the first law, which establishes the existence of the internal energy function  $U$  and the second law which establishes the existence of the entropy  $S$  and temperature  $T$ . Details are not necessary for the present discussion. The only thing we need to carry forward is the fundamental thermodynamic differential form. Given a thermodynamic system with:

- (i) One mechanic extensive/intensive<sup>11</sup> conjugate variable pair  $(X, \xi)$ , where  $X$  could stand for the volume  $V$  or magnetic moment  $\mathcal{M}$  with conjugate intensive

<sup>11</sup> The extensive variables  $U, X$  and  $N$  scale with the size of the system, intensive variables  $T, \xi$  and  $\mu$  are invariant with respect to such scaling.

variables, which in the case of  $V$  is the (negative) pressure  $-P$  and in the case of  $\mathcal{M}$  is the magnetic field  $\mathcal{H}$ ;

- (ii) A (dimensionless) extensive variable  $N$  which counts the number of units of mass in the system with a conjugate (intensive) energy  $\mu$ , called the chemical potential, carried by each unit of mass;<sup>12</sup>

for a differential change in the space  $\Xi_0$  of the variables  $(U, X, N)$  the differential change in the entropy  $S$  satisfies<sup>13</sup>

$$dS = \zeta_1 dU - \zeta_2 dX - \zeta_3 dN, \tag{1}$$

where

$$\zeta_1 := \epsilon/T, \quad \zeta_2 := \xi/T, \quad \zeta_3 := \mu/T, \tag{2}$$

are couplings. It is clear that the couplings are intensive *and* dimensionless. That the variables  $(U, X, N) \in \Xi_0$  appear as differentials on the right of (1) should be understood as signifying that they are independent variables. This means that the system is thermally, mechanically and chemically isolated with  $U, X$  and  $N$  fixed by an experimenter. Legendre transformations can be used to replace  $U$  and  $X$  successively as independent variables by  $\zeta_1$  and  $\zeta_2$ . Firstly, with Helmholtz free energy

$$\Phi_1 := \zeta_1 U - S, \tag{3}$$

we have

$$d\Phi_1 = U d\zeta_1 + \zeta_2 dX + \zeta_3 dN, \tag{4}$$

so that the independent variables are  $(\zeta_1, X, N) \in \Xi_1$ . The system is in contact with a source of thermal energy at temperature  $T = \epsilon/\zeta_1$ . Secondly, with Gibbs free energy

$$\Phi_2 := \zeta_1 U - \zeta_2 X - S, \tag{5}$$

we have

$$d\Phi_2 = U d\zeta_1 - X d\zeta_2 + \zeta_3 dN, \tag{6}$$

<sup>12</sup> In most presentations of thermodynamics  $N$  is simply taken to be the number of particles in the system. Our usage is designed to avoid reference to the microstructure of the system and to allow  $N$  to have non-integer values.

<sup>13</sup> At this point it is convenient: (i) To clarify the dimensionality of the thermodynamic variables. It is straightforward to show that, by scaling with respect to suitable constants,  $T$  and  $\xi$  can be made of the dimensions of energy ( $J := m^2 kg s^{-2}$ ) and  $U, S$  and  $X$  made dimensionless. In the case of  $U$  this is achieved by factoring out an energy constant  $\epsilon > 0$ . This is the field-extensive variable representation of Lavis [67, Sect. 1.1], where scaling for  $S$  and  $T$  is effected using Boltzmann’s constant  $k_B$ . The further change to the coupling-extensive variable representation is achieved by taking ratios of  $\epsilon, \xi$  and  $\mu$  with respect to  $T$  as shown in (2). (ii) To observe that the generalization to more than one mechanical variable pair is straightforward. (iii) To emphasise that this differential form should not be understood as some sort of equilibrium process in the space  $\Xi_0$  [68].

so that the independent variables are  $(\zeta_1, \zeta_2, N) \in \Xi_2$ . The system is now, through  $\zeta_2$ , also in mechanical contact with its environment, be it a fluid system subject to a pressure  $P$  or a magnetic system subject to a field  $\mathcal{H}$ . The couplings  $\zeta_1$  and  $\zeta_2$  are referred to as the *thermal and field (or mechanical) couplings* respectively.

It is tempting to suppose that this process could be taken one step further, interchanging the roles of  $N$  and  $\zeta_3$ . However, it is not difficult to see that the Legendre transformation implementing this would involve a free energy  $\Phi_3$  which is constant and can thus without loss of generality be taken to be identically zero. A viable form of thermodynamics must retain (at least) one extensive variable (here we choose that to be  $N$ , although we could have used  $X$ ) which registers the size of the system.

Observing that in thermodynamics the uncontrolled variables remain constant when the corresponding controlled variables are held constant, this is now the point for the first message from statistical mechanics:

**FSM–1** Unlike in thermodynamics, extensive variables in statistical mechanics that are uncontrolled quantities *fluctuate* even when the corresponding controlled variables are kept constant. (In  $\Xi_1$  the energy corresponding to the internal energy  $U$  fluctuates, and in  $\Xi_2$  the variable corresponding to  $X$ , be it the volume or the magnetic moment, fluctuates. This is born out by experiment [79].) The variances of the fluctuations are given in terms of response functions and are  $\mathcal{O}(N)$ . This means that standard deviations of fluctuations are  $\mathcal{O}(\sqrt{N})$  and become negligibly small compared to  $\mathcal{O}(N)$  variables only in the thermodynamic limit  $N \rightarrow \infty$ .

For fixed  $N$  let  $(U, X, N) \xrightarrow{A} (U', X', N)$  denote an adiabatic process. It can be shown [69], from Carathéodory's first version of the second law [23],<sup>14</sup> that thermodynamic systems are of four types according to whether the adiabatic process gives  $U \leq U'$  or  $U \geq U'$  and  $S \leq S'$  or  $S \geq S'$  corresponding, respectively, to the possibilities of the temperature and heat capacity being positive or negative.<sup>15</sup> Standard accounts of thermodynamics concentrate solely on the case where both internal energy and entropy increase, which is the situation where both temperature and heat capacity are positive. We shall restrict our attention to that case.<sup>16</sup>

## 2.2 Extensivity and the Thermodynamic Limit

Departing from the formulation TD1 of the structure of thermodynamics we ascend the left-hand column in Fig. 1, where it is now useful to consider the embedding of particular models. In this context they are of two types, ones which posit a phenomenological

<sup>14</sup> Or from some extensions to the approach of Lieb and Yngvason [75].

<sup>15</sup> The Clausius version of the second law needs modification to include negative temperatures [64, 108] and both the Kelvin–Planck and Clausius versions need modification to accommodate negative heat capacities (Lavis, 69).

<sup>16</sup> It is a matter of dispute (see, for example, Lavis [69] and references therein) whether statistical mechanical models support the existence of negative temperatures, and the experimental evidence is also questioned. The same is the case for negative heat capacities.

equation of state and ones derived from some microstructure according to the procedures of statistical mechanics. Most examples in the first category, the perfect gas equation, the Weiss-field equation for ferromagnetism and the van der Waals equation<sup>17</sup> introduce the models in terms of an equation relating the mechanical variable pair  $(X, \xi)$  and  $N$  to the temperature. However, it is more consonant with our approach to begin with a *defining* relationship for the entropy surface  $S(U, X, N)$ , from which  $T, \xi$  and  $\mu$ , or equivalently the couplings  $\zeta_1, \zeta_2$  and  $\zeta_3$  can be calculated using (1). Thus:

- For the *perfect gas*

$$S(U, V, N) := Nc + \frac{3}{2}N \ln\left(\frac{U}{N}\right) + N \ln\left(\frac{V}{N}\right), \tag{7}$$

for some constant  $c$ ,<sup>18</sup> giving<sup>19</sup>

$$T = \frac{2U\varepsilon}{3N}, \quad P = \frac{NT}{V}. \tag{8}$$

- For the *van der Waals fluid*

$$S(U, V, N) := Nc + \frac{3}{2}N \ln\left(\frac{U}{N} + \frac{N}{V}\right) + N \ln\left(\frac{V}{N} - 1\right), \tag{9}$$

giving

$$T = \frac{2}{3}\varepsilon\left(\frac{U}{N} + \frac{N}{V}\right), \quad P = \frac{NT}{V-N} - \frac{\varepsilon N^2}{V^2}. \tag{10}$$

The entropy (7) is a concave function of  $(U, V)$ , but for (9) it is necessary to take the concave envelope. This is, of course, equivalent in the case of the van der Waals [122] fluid and other phenomenological equations of state to the application of Maxwell’s equal areas rule [81], which avoid the inclusion of unstable states and leads to a first-order gas-liquid phase transition (see Sect. 2.3).

It will be noted that, for both the perfect gas and van der Waals fluid with densities  $u := U/N$  and  $v := V/N$ , there exists an entropy density  $s$  satisfying

$$s := \frac{S(uN, vN, N)}{N} = s(u, v), \tag{11}$$

for all  $N > 0$ , which avoids any reference to the size  $N$  of the system. But, of course, these are rather special models and the question arises as to whether entropy, in general, when  $X$  replaces  $V$  and  $x := X/N$  replaces  $v$ , satisfies

$$s := \frac{S(uN, xN, N)}{N} = s(u, x), \quad \forall N > 0. \tag{12}$$

For this question the following result is important:

<sup>17</sup> And a number of lesser known relationships like the Redlich–Kwong and Dieterici equations of state.

<sup>18</sup> Which can be evaluated using statistical mechanics but whose value is unimportant here.

<sup>19</sup> Remember that  $\zeta_2 := -P/T$ .

**Theorem 1** Equation (12) is true iff

$$S(\lambda U, \lambda X, \lambda N) = \lambda S(U, X, N), \quad \forall \lambda > 0, \tag{13}$$

is true.

**Proof** That (12) follows from (13) is easily seen by taking  $\lambda = 1/N$  and defining  $s(u, x) := S(u, x, 1)$ .

In the reverse direction, this last relationship  $s(u, x) = S(u, x, 1)$  in fact follows from (12) by setting  $N = 1$ . Then from (12)  $S(U, X, N) = NS(U/N, X/N, 1)$  and again setting  $\lambda = 1/N$  recovers (13).  $\square$

Equation (13) is the condition that  $S$  is an extensive function and it is easily shown from (3) and (5) that the free energies  $\Phi_1$  and  $\Phi_2$  are extensive functions if and only if the entropy is an extensive function. But, as pointed out by Menon and Callender [83, Sect. 2] and show in Sect. 3.3,

**FSM-2** The extensivity of entropy and of free energies assumed in thermodynamics is not exactly true for all systems in statistical mechanics, but is approximately true for large systems.

For entropy the thermodynamic limit in statistical mechanics, assuming it exists,<sup>20</sup> is given by

$$\lim_{N \rightarrow \infty} \frac{S(uN, xN, N)}{N} = s(u, x). \tag{14}$$

But for thermodynamics the corresponding formula is (12), without the need for the limiting process. Exact extensivity in thermodynamics can be regarded as unnecessary or trivially true.

Differentiating (13) with respect to  $\lambda$ , and substituting from (12) gives

$$S = \zeta_1 U - \zeta_2 X - \zeta_3 N, \tag{15}$$

when  $\lambda$  is put equal to 1. From (1) to (12),

$$u d\zeta_1 - x d\zeta_2 - d\zeta_3 = 0, \tag{16}$$

which is a version of the *Gibbs–Duhem relationship*. In terms of densities (15) becomes

$$s = \zeta_1 u - \zeta_2 x - \zeta_3, \tag{17}$$

and substituting into (1)–(6)

$$\begin{aligned} ds &= \zeta_1 du - \zeta_2 dx - (s - \zeta_1 u + \zeta_2 x + \zeta_3) dN/N \\ &= \zeta_1 du - \zeta_2 dx. \end{aligned} \tag{18}$$

<sup>20</sup> See footnote 10.

Then, for free-energy densities  $\phi_1 := \Phi_1/N$  and  $\phi_2 := \Phi_2/N$ ,

$$\phi_1 = \zeta_1 u - s = \zeta_2 x + \zeta_3, \quad d\phi_1 = u d\zeta_1 + \zeta_2 dx, \tag{19}$$

$$\phi_2 = \zeta_1 u - \zeta_2 x - s = \zeta_3, \quad d\phi_2 = u d\zeta_1 - x d\zeta_2. \tag{20}$$

These are the fundamental size-free thermodynamic relationships in terms of density variables and density functions. They are exact in thermodynamics but approximately true only for large systems in statistical mechanics. The question of large systems and the thermodynamic limit in statistical mechanics is treated in Sects. 3.3, 3.5 and 4.

### 2.3 Thermodynamics with PTCP

Having arrived at a formulation of thermodynamics in terms of densities and couplings the modern theory of PTCP is largely concerned with an investigation and classification of the singular properties of systems (see e.g., [18]). Specifically the singularities which could occur on the hypersurface of the entropy density, or the appropriate free-energy density, which defines the state of the system. However we should be forewarned that the account of statistical mechanics in Sect. 3 concludes that:

**FSM-3** The association of PTCP with singularities in the entropy and free-energy densities which is made in thermodynamics can be made in statistical mechanics only for infinite systems.

The association of PTCP with singularities in both TD3 and SM3 leads to a tendency for them to be mistakenly conflated. (We shall discuss this in more detail in relation to limit reduction in Sect. 5.1).

We now consider three thermodynamic spaces,  $\tilde{\Xi}_0, \tilde{\Xi}_1$  and  $\tilde{\Xi}_2$ , which correspond respectively to the spaces  $\Xi_0, \Xi_1$  and  $\Xi_2$  defined in Sect. 2.1 except that now densities replace extensive variables. In reverse order, since this is more heuristically transparent:

- (i) *In the space  $\tilde{\Xi}_2$  of the vector  $\zeta := (\zeta_1, \zeta_2)$  the free-energy density  $\phi_2(\zeta_1, \zeta_2)$  is a surface with normal in the direction  $(1, -u, x)$  and phases are separated by lines of transitions. The simplest example is a line  $\mathcal{L}^*$  across which there is a discontinuity of the gradient  $\nabla\phi_2 = (u, -x)$ ; an isothermal section ( $\zeta_1$  constant) of this surface is shown in Fig. 2. The point  $\zeta^* := (\zeta_1^*, \zeta_2^*) \in \mathcal{L}^*$ , with  $\zeta_3^* = \phi_2(\zeta^*)$ .  $\mathcal{L}^*$  can be regarded as representing the *coexistence of two phases* with different densities. As  $\zeta$  is varied across  $\mathcal{L}^*$  through  $\zeta^*$  there is a *first-order phase transition* where the densities change discontinuously. In the case of both fluid and magnetic systems a first-order transition will involve a discontinuity of the internal energy density  $u$ . In a fluid system there will be a discontinuity of the (physical) density as the system changes between a liquid*

and a gas. In a magnetic system there will be a discontinuity in the magnetization (or equivalently the magnetization density) as shown for the Ising model in Fig. 9.

- (ii) In the space  $\tilde{\Xi}_1$  of the vector  $(\zeta_1, x)$  the free-energy density  $\phi_1(\zeta_1, x)$  is a surface convex with respect to  $x$  with normal in the direction  $(1, -u, -\zeta_2)$ , as shown by an isothermal  $(\zeta_1 = \zeta_1^*)$  section in Fig. 3. A first-order transition corresponds to the part of the isotherm, labelled  $C^*$ , which is linear with respect to  $x$ . At the ends of  $(\zeta_1^*, x^{*(+)})$  and  $(\zeta_1^*, x^{*(-)})$  of  $C^*$  all three couplings  $\zeta_1, \zeta_2$  and  $\zeta_3$  have the same values as is otherwise shown in Fig. 2. Typically, as  $\zeta_1^*$  varies along  $\mathcal{L}^*$  the ends of  $C^*$  converge to a *critical point* where the system exhibits a *second-order transition*. There the densities are continuous but one or more of the response functions (that is to say the curvature components of the free-energy surface) is singular.<sup>21</sup> A projection of the linear coexistence region in Fig. 3 is shown in Fig. 4, and the situation where the corresponding transition line  $\mathcal{L}^*$  terminates is shown in Fig. 5.
- (iii) The space  $\tilde{\Xi}_0$  of the vector  $(u, x)$ , in which the entropy density  $s(u, x)$  is a concave surface is similar to that for  $\phi_1(\zeta_1, x)$ ,<sup>22</sup> except that now the linear generator  $C^*$  of the coexistence region has endpoints  $(u^{*(+)}, x^{*(+)})$  and  $(u^{*(-)}, x^{*(-)})$ . As  $\zeta^*$  varies along  $\mathcal{L}^*$ ,  $C^*$  traces out the boundary of a ruled<sup>23</sup> region on the entropy surface with  $C^*$  converging in one direction to the critical point described in (ii).

*Critical exponents* at the critical point are associated with the curvature of the coexistence curve in  $\tilde{\Xi}_1$  and the coexistence line in  $\tilde{\Xi}_2$ , and the asymptotic singular behaviour of the (per particle) heat capacities  $c_x$  and  $c_\xi$  at constant density and field respectively and a response function  $\varphi_T$ , which in a fluid corresponds to the compressibility and in a magnet to the susceptibility. It will also be useful to include the coefficient of thermal expansion  $\alpha_\xi$ . These are defined together with their critical exponents in Appendix 1. The heat capacities  $c_x$  and  $c_\xi$  are normally positive and from (105) it follows that, if  $\varphi_T > 0$ , then  $c_\xi$  dominates both  $c_x$  and  $\alpha_\xi^2/\varphi_T$  as  $T \rightarrow T_c$ . For the critical exponents  $\sigma$  and  $\sigma'$  characterizing the singularity of  $c_x$  on approach to the critical point from above and below  $T_c$ , and the analogously defined critical exponents  $\alpha$  and  $\alpha'$  characterizing the singularity of  $c_\xi$ , and  $\gamma$  and  $\gamma'$  characterizing the singularity of  $\varphi_T$ , as well  $\beta$  characterizing the curvature of the coexistence curve, this means that

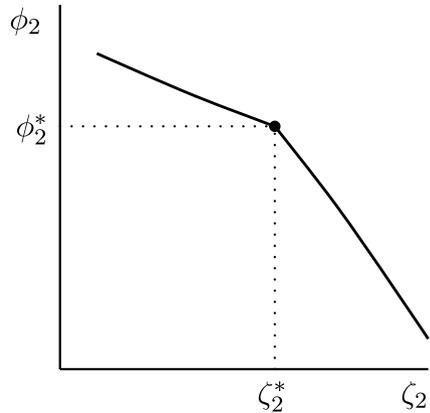
$$\sigma \geq \alpha, \quad \sigma' \geq \alpha', \quad \sigma' + 2\beta + \gamma' \geq 2. \tag{21}$$

<sup>21</sup> Such critical points can also occur as lines. A line of first-order transitions can terminate on a line of second-order transitions at a point called a *critical end-point*, or be continued as a line of second-order transitions at a point called a *tricritical point*.

<sup>22</sup> That convexity is replaced by concavity is clear from the negative sign of  $s$  in (19).

<sup>23</sup> A ruled surface (like, for example, the surface of a cylinder) is one densely covered by a set of straight lines.

**Fig. 2** A first-order transition showing as a discontinuity of slope in an isothermal section ( $\zeta_1 = \zeta_1^*$ ) of  $\phi_2 = \zeta_3$  plotted against  $\zeta_2$



The condition  $\varphi_T > 0$  is true for a magnetic system and in this case the third inequality in (21) was first established by Rushbrooke [113]. The stronger condition

$$\alpha' + 2\beta + \gamma' \geq 2, \tag{22}$$

was obtained by Griffiths [42] for both magnetic and fluid systems using the convexity properties of the free energy. In fact it is a consequence of scaling theory (Sect. 4) that, for systems with a special symmetry which is present in magnetic systems where, as for the Ising model in Appendix 2, the coexistence curve coincides with the zero field axis,  $\sigma' = \alpha'$  and inequalities (21) and (22) become identical. Otherwise  $\sigma' = \gamma'$ . Griffiths [42] also derived a number of other inequalities. In particular

$$\gamma' \geq \beta(\delta - 1), \tag{23}$$

where  $\delta$ , given by (109), is the exponent characterizing the (critical) equation of state.

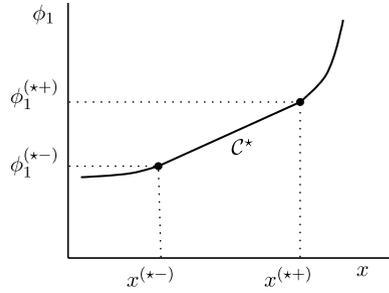
### 2.4 Thermodynamics with Scaling Theory

In view of our aim to keep as distinct as possible the developments of thermodynamics and statistical mechanics, we choose here to present scaling theory as a mathematical axiomatization of the properties of PTCP in thermodynamics. Although, as we see below, it has deep roots in, and is substantiated by, statistical mechanics, in particular renormalization group theory,<sup>24</sup> where, in almost all cases,<sup>25</sup> the realization of this picture of scaling involves approximations and yields scaling forms of only local validity.

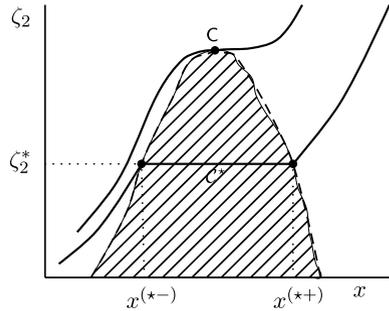
<sup>24</sup> The assumption (24) that, near to a critical region, the free-energy density can be divided into smooth and singular parts is justified in terms of the form of the Hamiltonian [44, p. 3519].

<sup>25</sup> An exception being the one-dimensional Ising model (see e.g. [67, Sect. 15.5.1]).

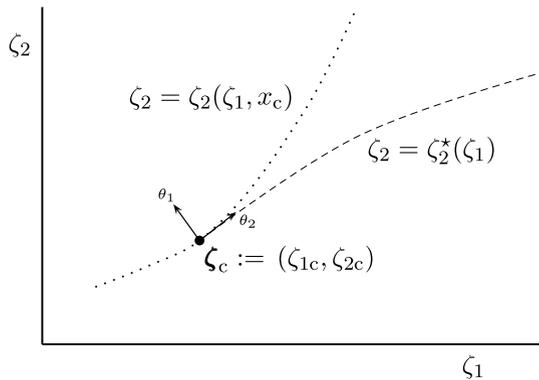
**Fig. 3** A first-order transition showing as the linear section  $C^*$  in an isothermal section of the  $\phi_1$  surface



**Fig. 4** A first-order transition showing as a horizontal part  $C^*$  of an isotherm of  $\zeta_2$  plotted against  $x$  together with the isotherm through the critical point C. As  $\zeta_1$  varies the ends of  $C^*$  trace the boundary of the coexistence region (shaded)



**Fig. 5** A critical point  $(\zeta_{1c}, \zeta_{2c})$  in  $\tilde{\Xi}_2$ . The first-order transition (coexistence curve)  $\zeta_2 = \zeta_2^*(\zeta_1)$  is represented by a broken line and the critical isochore, along which the density  $x$  takes its critical value  $x = x_c$  by a dotted line. The directions of the axes of the two relevant scaling fields at the critical point, as described in Sect. 2.4, are shown



Originating in the work of (among others) Widom [126, 127] and Kadanoff [53] our approach is essentially that of Hankey and Stanley [44]. Given here in brief outline<sup>26</sup> it is sufficient for an analysis of power-law singularities in the critical region.<sup>27</sup>

Suppose we have the free-energy density of a system in terms of its maximum number of independent couplings. In the discussion above that maximum number

<sup>26</sup> For a more detailed account see, for example, Lavis [67, Chap.4].

<sup>27</sup> For statistical mechanical systems like the Ising model with  $d = 2$  which exhibit logarithmic singularities, it has been shown by Nightingale and 'T Hooft [95] that a slight generalization needs to be used.

was two, but for the moment we generalize to  $n$  couplings so the free-energy density is  $\phi_n(\boldsymbol{\zeta})$ , where  $\boldsymbol{\zeta} := (\zeta_1, \zeta_2, \dots, \zeta_n)$ , which is represented as a hypersurface of dimension  $n$  in the  $(n + 1)$ -dimensional space  $(\phi_n, \boldsymbol{\zeta})$ . Now suppose that there is a critical region  $\mathcal{C}$  of dimension  $n - s$ . Although  $\phi_n(\boldsymbol{\zeta})$  itself is continuous and finite across and within  $\mathcal{C}$  it may have discontinuous first-order derivatives, meaning that  $\mathcal{C}$  is a region of phase coexistence with a first-order transition when, as is shown in Fig. 2, the phase point crosses through  $\mathcal{C}$ , or it may have singular second-order derivatives in  $\mathcal{C}$ , as is the case in the situation described above where a line of first-order transition terminates at a critical point.<sup>28</sup>

With respect to some origin  $\boldsymbol{\zeta}^\circ \in \mathcal{C}$  a system of orthogonal curvilinear coordinates  $\theta_1, \theta_2, \dots, \theta_n$  called *scaling fields* is constructed. These are smooth functions of the couplings which parameterize  $\mathcal{C}$  so that  $\theta_1 = \dots = \theta_s = 0$  within  $\mathcal{C}$ . The scaling fields in this subset are called *relevant* with those in the remaining subset  $\theta_{s+1}, \theta_{s+2}, \dots, \theta_n$ , called *irrelevant*, acting as a local set of coordinates within  $\mathcal{C}$ .<sup>29</sup> The free-energy density  $\phi_n(\boldsymbol{\zeta})$  is separated into two parts

$$\phi_n(\boldsymbol{\zeta}) = \phi_{\text{smth}}(\boldsymbol{\zeta}) + \phi_{\text{sing}}(\Delta\boldsymbol{\zeta}), \tag{24}$$

where  $\phi_{\text{smth}}(\boldsymbol{\zeta})$  is a regular function and, with  $\Delta\boldsymbol{\zeta} := \boldsymbol{\zeta} - \boldsymbol{\zeta}^\circ$ ,  $\phi_{\text{sing}}(\Delta\boldsymbol{\zeta})$ , for which  $\phi_{\text{sing}}(\mathbf{0}) = 0$ , contains all the non-smooth parts of  $\phi_n(\boldsymbol{\zeta})$  in  $\mathcal{C}$ . It is now assumed that  $\phi_{\text{sing}}(\Delta\boldsymbol{\zeta})$  can be re-coordinated in terms of the scaling fields so that it is a generalized homogeneous function satisfying the *Kadanoff scaling hypothesis*<sup>30</sup>

$$\phi_{\text{sing}}(\lambda^{y_1}\theta_1, \dots, \lambda^{y_n}\theta_n) = \lambda^d \phi_{\text{sing}}(\theta_1, \dots, \theta_n), \tag{25}$$

for all real  $\lambda > 0$ , where  $d$  is the physical dimension of the system, and  $y_j$ ,  $j = 1, 2, \dots, n$  are *scaling exponents* satisfying

$$y_j > 0, \quad j = 1, \dots, s, \quad y_j < 0, \quad j = s + 1, \dots, n. \tag{26}$$

The exponents in the first subset are, like the corresponding scaling fields, called *relevant* and those in the latter subset are called *irrelevant*.<sup>31</sup> Of the assumptions made here, that scaling fields can be derived is not particularly demanding; at the very least it is usually straightforward to obtain their linear parts near to the origin. And the division of the free-energy density (24) into smooth and singular parts has

<sup>28</sup> Or there may be discontinuities or singularities in higher-order derivatives; but we shall for simplicity concentrate solely on cases involving first- and second-order derivatives.

<sup>29</sup> ‘Relevance’ here refers to their role in an understanding of the nature of the criticality in  $\mathcal{C}$ .

<sup>30</sup> The physical dimension  $d$  of the system is not something which plays a significant role in most of thermodynamics. It is included here to bring compatibility with the discussion of statistical mechanics. It could be removed by redefining  $\lambda$ .

<sup>31</sup> We have for the sake of simplicity excluded the possibility of a zero exponent; such an exponent is called *marginal*. Marginal exponents are associated in renormalization group theory with an ‘underlying’ parameter of the system, often resulting in lines of fixed points as we see in our treatment of the one-dimensional Ising model in Sect. 3.4.3. It will also be assumed that no exponent is complex. In practice this is not always the case (see e.g., [67, Sect. 15.5.2]), but situations arising from complex exponents are not difficult to interpret in particular examples.

very little content until we explore in more detail the consequences of the scaling hypothesis (25) which we now do for the case of a critical point terminating a coexistence curve.

There are many general accounts of scaling theory, treating a variety of critical phenomena. Here we restrict attention to the case of a critical point terminating a line of first-order transitions, as shown in Fig. 5. So we have two critical regions. The first is the critical point with two relevant scaling fields and scaling exponents with axes chosen perpendicular to and along the coexistence curve. For this we shall show that the critical exponents defined in Appendix 1, can be expressed in terms of the two scaling exponents. The second is the coexistence curve which has one relevant and one irrelevant scaling field constructed with respect to some chosen origin (not shown in Fig. 5) on the coexistence curve.

For the sake of further simplifying our presentation we restrict attention to a simple ferromagnetic system with  $\xi := \mathcal{H}$ , the magnetic field,  $X := \mathcal{M}$ , the magnetization and  $x := m = \mathcal{M}/N$ , the magnetization density. The coupling  $\zeta_1$  is the thermal coupling so we relabel it as  $\zeta_T = \epsilon/T$  and  $\zeta_2$  is the field coupling which we relabel as  $\zeta_{\mathcal{H}} = \mathcal{H}/T$ . This model, of which an example in statistical mechanics is the Ising model described in Appendix 2, has the advantage of having the special symmetry that the coexistence curve lies along the zero-field axis in an interval  $T \in [0, T_c]$  with  $\mathcal{H}_c = m_c = 0$ . This axis with  $T > T_c$  is the critical isochore. Thus (referring to Fig. 5) the coexistence curve lies along the  $\zeta_{\mathcal{H}} = 0$  axis in an interval  $[\zeta_{Tc}, \infty)$ . This same phase diagram for the Ising model, now plotted with respect to the temperature  $T$  and the magnetic field  $\mathcal{H}$ , is shown in Fig. 8.

We consider separately the critical point and the coexistence curve, beginning with the critical point where we can take the scaling fields to be

$$\theta_T := \zeta_T - \zeta_{Tc} = \epsilon \left( \frac{1}{T} - \frac{1}{T_c} \right) \geq 0, \quad \theta_{\mathcal{H}} := \zeta_{\mathcal{H}} = \frac{\mathcal{H}}{T}. \tag{27}$$

The scaling hypothesis (25) becomes

$$\phi_{\text{sing}}(\lambda^{y_T} \theta_T, \lambda^{y_{\mathcal{H}}} \theta_{\mathcal{H}}) = \lambda^d \phi_{\text{sing}}(\theta_T, \theta_{\mathcal{H}}), \tag{28}$$

and, from (24) and (103),

$$m = - \frac{\partial \phi_{\text{smth}}}{\partial \zeta_{\mathcal{H}}} - \frac{\partial \phi_{\text{sing}}}{\partial \theta_{\mathcal{H}}}, \tag{29}$$

$$\frac{\partial \phi_{\text{sing}}}{\partial \theta_{\mathcal{H}}}(\lambda^{y_T} \theta_T, \lambda^{y_{\mathcal{H}}} \theta_{\mathcal{H}}) = \lambda^{d-y_{\mathcal{H}}} \frac{\partial \phi_{\text{sing}}}{\partial \theta_{\mathcal{H}}}(\theta_T, \theta_{\mathcal{H}}). \tag{30}$$

Since  $m_c = 0$ ,  $\partial \phi_{\text{smth}}/\partial \zeta_{\mathcal{H}} = 0$  at the critical point. For an approach to the critical point along the coexistence curve  $\theta_{\mathcal{H}} = 0$  and setting  $\lambda = \theta_T^{-1/y_T}$  in (30) and substituting into (29) gives

$$m \simeq -\theta_T^{(d-y_{\mathcal{H}})/y_T} \frac{\partial \phi_{\text{sing}}}{\partial \theta_{\mathcal{H}}}(1, 0) \sim (T_c - T)^{(d-y_{\mathcal{H}})/y_T}, \tag{31}$$

which, when comparing with (107) establish the identification

$$\beta = (d - y_{\mathcal{H}})/y_T. \tag{32}$$

At this point we could carry out a similar procedure for the response functions in (104) and (105) to determine the critical exponents defined in (106)–(110). However, the analysis can be shortened by a closer examination of the way that the expression (32) for  $\beta$  was obtained. From this we see that the scaling exponent  $y_{\mathcal{H}}$  in the numerator indicates that differentiation was once with respect to  $\zeta_{\mathcal{H}}$ . And that the approach was in the direction of varying  $\zeta_T$  is indicated by the scaling exponent  $y_T$  in the denominator. So with the same reasoning it follows from (109) that

$$\delta = y_{\mathcal{H}}/(d - y_{\mathcal{H}}), \tag{33}$$

and bearing in mind that the analysis yields singularities for response functions so  $\phi_{\text{smth}}$  can play no role, from (108),

$$\gamma = \gamma' = (2y_{\mathcal{H}} - d)/y_T. \tag{34}$$

When we come to consider  $c_{\xi} := c_{\mathcal{H}}$ , given by (104), the situation becomes a little more complicated, since there are three terms and we need to know which dominates as the critical point is approached. This will depend on the relative magnitudes of  $y_T$  and  $y_{\mathcal{H}}$  and it can be shown (Lavis, [67], Sect. 4.5.1) that, in general for a critical point terminating a line of first-transitions, the exponent associated with approaches tangential to the coexistence curve is smaller (less relevant) than that associated with an approach at a non-zero angle to this curve. These are called respective *weak and strong approaches* and in the present context we have  $y_{\mathcal{H}} > y_T$ , these being respectively the weak and strong exponents. Returning to the formula for  $c_{\mathcal{H}}$  in (104) we see that the third term on the right-hand side would be the one that dominates meaning that, from (110),  $\sigma = \sigma' = \gamma$ . However, because of the symmetry of the magnetic model  $\zeta_{2c} = \zeta_{\mathcal{H}c} = 0$  and the only remaining term is the first, meaning that

$$\sigma = \sigma' = (2y_T - d)/y_T. \tag{35}$$

Finally we need to determine the asymptotic form for  $c_x := c_m$  using (105). Here the situation need a more detailed analysis, when it can be shown (Lavis, [67], Sect. 4.5.4) that, *whether or not the magnetic symmetry applies* cancellation of coefficients leads to an asymptotic form equivalent to that of a second-order derivative with respect to  $\zeta_T$ ; that is,

$$\alpha = \alpha' = (2y_T - d)/y_T. \tag{36}$$

This means that it is the asymptotic form of the heat capacity with constant intensive variable (pressure or magnetic field) which is dependent on symmetry. In the magnetic system the exponent is the same as that of the heat capacity with constant

extensive variable (the magnetization) and in a fluid, where there is no symmetry it is equal to that of  $\varphi_T$ , which is the compressibility. Equations (32)–(36) are formulae for the exponents  $\alpha, \beta, \gamma$  and  $\delta$  in terms of  $y_T$  and  $y_{\mathcal{H}}$ . They are, therefore, not independent and two relationships exist between them. These can be expressed in the form  $\alpha + 2\beta + \gamma = 2$ , called the *Essam–Fisher scaling law* [30], which is a strengthening of the inequality (22) and  $\gamma' = \beta(\delta - 1)$ , called the *Widom scaling law* [126], which is a strengthening of the inequality (23).

For the coexistence curve, scaling fields, chosen with respect to some arbitrary origin  $\zeta_T = \zeta_T^\circ, \zeta_{\mathcal{H}} = 0$  are

$$\theta'_T := \zeta_T - \zeta_T^\circ, \quad \theta'_{\mathcal{H}} := \zeta_{\mathcal{H}} = \mathcal{H}/T, \tag{37}$$

with  $y'_T$  and  $y'_{\mathcal{H}}$  irrelevant and relevant exponents respectively. In general it can be shown that relevant exponents are less than or equal to  $d$  meaning in this case that  $0 < y'_{\mathcal{H}} \leq d$ . With primes attached to the exponents and fields (29) and (30) continue to applied to the magnetization density. If  $y'_{\mathcal{H}} < d$

$$\frac{\partial \phi_{\text{sing}}}{\partial \theta_{\mathcal{H}}}(0, 0) = 0, \tag{38}$$

and  $m$  is continuous at the origin; there is no first-order phase transition. If  $y'_{\mathcal{H}} = d$  then (38) does not necessarily hold. There may be a contribution to (29) from the derivative of  $\phi_{\text{sing}}$ . This will be the only way in which the magnetization can be discontinuous across the coexistence curve. So a scaling exponent equal to  $d$  is a necessary, but not sufficient condition for a first-order transition. An example of such a first-order transition with an exponent of  $d$  is at zero temperature in the one-dimension Ising model (Sect. 3.4.3(a)). Discontinuities in higher-order derivatives can be treated in a similar way.

### 2.5 Dimensionality and Phase Transitions

Although, as we have seen, thermodynamics, and particularly its treatment of PTCP, assumes that the system is infinite, the dimension  $d$  of the system entered into the discussion in Sect. 2.4. And once dimensionality has entered then finiteness has also appeared. Thus, for example, a two-dimensional system can be viewed as a three-dimensional system of ‘thickness’ one in the third dimension and it is only a small step from there to increase the thickness to two. In Sect. 1 we referred to the classification of singularities in terms of universality classes. This, as we asserted, can be discussed only in the context of statistical mechanics, with  $d$  one of the factors determining the universality class of an occurrence of singular behaviour. If the number of directions in which the system is infinite is increased, then its critical behaviour will change from one universality class to another. This is an example of what in scaling and renormalization group theory is called ‘cross-over’.<sup>32</sup> The dimension of

<sup>32</sup> Of course, such a change of universality class is counter-factual [50], in the sense that one cannot change the dimension or extensivity properties of a real system.

the system affects not just the universality class of singular behaviour but whether it occurs at all. However, that dimension is not  $d$  but  $\mathfrak{d} \leq d$ , the number of directions in which the system is infinite.<sup>33</sup> And the final message sent from statistical mechanics to thermodynamics is that:

**FSM-4** There exists a *lower-critical dimension*  $d_{LC}$  such that, if  $\mathfrak{d} \leq d_{LC} < d$  singular behaviour can occur in the fully-infinite system but not in the partially-infinite system. If  $d > \mathfrak{d} > d_{LC}$  then singular behaviour can occur in both, but in different universality classes.

### 3 From Gibbsian Statistical Mechanics to the Renormalization Group

The move from thermodynamics to statistical mechanics is, we shall argue, an *enrichment* and *substantiation* of the picture we have of any system under investigation. This operates at two levels. The first is structural, where renormalization group theory embedded in statistical mechanics provides a fuller picture in terms of renormalization group transformations and fixed points than scaling theory embedded in thermodynamics. The second is in the provision of specific models which arise from assumptions about the microstructure of the system. We now consider the development represented by the right-hand column in Fig. 1, beginning with the basic structure of statistical mechanics.

#### 3.1 Inter-Theory Connecting Relationships

Let the microstate of the system be given by a value of the vector variable  $\sigma$  in the phase space  $\Gamma$ . In the case of a fluid system  $\sigma$  will be a set of values for the positions and momenta of all the particles; for a spin system on a lattice, like the Ising model in Appendix 2,  $\sigma$  will be the set of values of all the spin variables. The microscopic and macroscopic structure of the system is then determined by the Hamiltonian. This is an explicit function of the independent couplings with the independent extensive variables imposing constraints on  $\sigma$ . Thus we have three cases:

- (i) When  $(U, X, N) \in \Xi_0$  are the independent variables the Hamiltonian is  $\widehat{H}_0(\sigma; X, N)$ , with values constrained by

<sup>33</sup> The connection between the thermodynamic limit and extensivity is retained in a partially-infinite system with  $N_k$  sites in the  $k$ -direction and  $N_1 N_2 \cdots N_d = N$ , when, in the case, for example, of entropy, (13) is replaced by

$$S(\lambda' U, \lambda' X, \lambda_1 N_1, \lambda_2 N_2, \dots, \lambda_{\mathfrak{d}} N_{\mathfrak{d}}, N^{(\mathfrak{b})}) = \lambda' S(U, X, N_1, N_2, \dots, N_{\mathfrak{d}}, N^{(\mathfrak{b})}),$$

where  $N^{(\mathfrak{b})} := N_{\mathfrak{d}+1} N_{\mathfrak{d}+2} \cdots N_d$  and  $\lambda' := \lambda_1 \lambda_2 \dots \lambda_{\mathfrak{d}}$ .

$$\hat{H}_0(\boldsymbol{\sigma}; X, N) = U, \tag{39}$$

and  $\boldsymbol{\sigma}$  constrained, according to the nature of the particular model by  $X$  and  $N$ .<sup>34</sup>

- (ii) When  $(\zeta_1, X, N) \in \Xi_1$  are the independent variables the Hamiltonian  $\hat{H}_1(\boldsymbol{\sigma}; \zeta_1, X, N)$  is a linear function of  $\zeta_1$ . The constraint (39) is removed but  $\boldsymbol{\sigma}$  remains constrained by  $X$  and  $N$ .
- (iii) When  $(\zeta_1, \zeta_2, N) \in \Xi_2$  are the independent variables the Hamiltonian  $\hat{H}_2(\boldsymbol{\sigma}; \zeta_1, \zeta_2, N)$  is a linear function of  $\zeta_1$  and  $\zeta_2$ . The only remaining constraint is from  $N$ .

Connecting relationships are now invoked in three stages:

**FTD–1** The independent variables in  $\Xi_0$ ,  $\Xi_1$  and  $\Xi_2$  are endowed with their thermodynamic meanings.

To proceed to the next stage of the inter-theory connecting process we need to give a form in cases (i), (ii) and (iii), respectively, for the entropy, and the free energies  $\Phi_1$  and  $\Phi_2$ . Case (i) gives the *microcanonical distribution*<sup>35</sup> and cases (ii) and (iii) give, respectively, the *canonical distribution* and the *constant pressure or magnetic field distribution*. For the sake of simplicity we concentrate exclusively on case (iii), where the Gibbs free energy is defined by

$$\Phi_2(\zeta_1, \zeta_2, N) := -\ln\{Z_2(\zeta_1, \zeta_2, N)\}, \tag{40}$$

where

$$Z_2(\zeta_1, \zeta_2, N) := \sum_{\{\boldsymbol{\sigma}\}} \exp\{-\hat{H}_2(\boldsymbol{\sigma}; \zeta_1, \zeta_2, N)\}, \tag{41}$$

is the Gibbs partition function.<sup>36</sup> Then

**FTD–2**  $\Phi_2$  is endowed with its thermodynamic properties and, using (3)–(6),

$$\Phi_1 = \Phi_2 + \zeta_2 X, \quad S = \zeta_1 U - \Phi_1, \tag{43}$$

$$U = \frac{\partial \Phi_2}{\partial \zeta_1}, \quad X = -\frac{\partial \Phi_2}{\partial \zeta_2}, \quad \zeta_3 = -\frac{\partial \Phi_2}{\partial N}, \tag{42}$$

establishes the connection between  $U$ ,  $X$ ,  $\zeta_3$ ,  $\Phi_1$  and  $S$  and their thermodynamic equivalents.

<sup>34</sup>  $N$  fixes the number of particles in  $\boldsymbol{\sigma}$ . If the system is a fluid, with  $X := V$  the volume of its container, then this constrains the range of the configuration component of  $\boldsymbol{\sigma}$ . Rather less physically achievable, if the system is a magnet, with  $X := M$  the magnetization, this will constrain the spin configuration of the microsystems.

<sup>35</sup> Where there is still some dispute about the appropriate form for the entropy (see, for example, [69], and references therein).

<sup>36</sup> Of course, according to the nature of the system this could involve an integral rather than a sum.

This completes a *sufficient* set of the connecting relationships. However, we can make some further links. Suppose that

$$\hat{H}_2(\boldsymbol{\sigma}; \zeta_1, \zeta_2; N) := \hat{U}(\boldsymbol{\sigma})\zeta_1 - \hat{X}(\boldsymbol{\sigma})\zeta_2. \tag{44}$$

Then, from (40)–(43),

$$U = \langle \hat{U}(\boldsymbol{\sigma}) \rangle, \quad X = \langle \hat{X}(\boldsymbol{\sigma}) \rangle. \tag{45}$$

**FTD–3**  $U$  and  $X$  are identified respectively as the expectation values of  $\hat{U}(\boldsymbol{\sigma})$  and  $\hat{X}(\boldsymbol{\sigma})$  with respect to the probability distribution with density

$$\rho(\boldsymbol{\sigma}; \zeta_1, \zeta_2) := \frac{\exp[-\hat{H}_2(\boldsymbol{\sigma}; \zeta_1, \zeta_2; N)]}{Z_2(\zeta_1, \zeta_2, N)}. \tag{46}$$

And it further follows from (40)–(46) that

$$\text{Var}[\hat{X}(\boldsymbol{\sigma})] = \frac{\partial^2 \Phi_2}{\partial \zeta_2^2} = N\varphi_T, \tag{47}$$

where  $\varphi_T$  is the response function given by (104). This is an example of a *fluctuation–response function* relationship. Similar relationships apply to  $\hat{U}(\boldsymbol{\sigma})$  and all uncontrolled extensive variables.

### 3.2 Correlation Function and Correlation Length

As is already evident, thermodynamics is a ‘black-box’ theory with a set of macro-variables some of which are independent and controllable and others whose values change in response to the changes in the independent variables. The only concession made to internal structure was, in Sect. 2.1, to allow a counting of the number  $N$  of mass units of the system. Now with the ‘enrichment’ provided by statistical mechanics we are able to record the microstate  $\boldsymbol{\sigma}$  of the system, which is simply the aggregate of the states of the individual microsystems.

Suppose that we take the  $d$ -dimensional hypercubic lattice<sup>37</sup>  $\mathcal{N}_d$  with sites  $\mathbf{r} := (n_1, n_2, \dots, n_d)\mathbf{a}$ , for  $n_k = 1, 2, \dots, N_k$  with  $N = N_1N_2 \dots N_d$ , where  $\mathbf{a}$  is the lattice spacing.<sup>38</sup> Then, given that the states of the microsystems on sites  $\mathbf{r}$  and  $\mathbf{r}'$  of the lattice are  $\sigma(\mathbf{r})$  and  $\sigma(\mathbf{r}')$ , respectively, how does the state of one effect the state of the other; that is to say, how are their states *correlated*? More specifically, how is the correlation between  $\sigma(\mathbf{r})$  and  $\sigma(\mathbf{r}')$  affected by:

<sup>37</sup> The restriction of this presentation to a hypercubic lattice is in the interests of simplicity. It can easily be generalized to other lattices.

<sup>38</sup> It is convenient for our discussions to suppose that the microsystems are confined to the sites of a lattice. It is, of course, the case that a whole area of statistical mechanics concerned with fluid systems treats the case of microsystems/molecules moving in a continuum.

- (i) the distance  $|\mathbf{r} - \mathbf{r}'|$  between the sites?
- (ii) the closeness of the thermodynamic state of the system to a critical region?

To begin to answer these questions suppose that, as for the Hamiltonian (44) in the Ising model in Appendix 2,  $\hat{X}(\boldsymbol{\sigma})$  is a linear sum of the states on the sites of  $\mathcal{N}$ . And (temporarily) suppose that the coupling  $\zeta_2$  takes different values  $\zeta_2(\mathbf{r})$  at the sites. Then, denoting the set of couplings  $\zeta_2(\mathbf{r})$  by the vector  $\boldsymbol{\zeta}_2$ ,

$$\hat{H}_2(\boldsymbol{\sigma}; \zeta_1, \boldsymbol{\zeta}_2; \mathcal{N}) := \hat{U}(\boldsymbol{\sigma})\zeta_1 - \sum_{\{\mathbf{r}\}} \sigma(\mathbf{r})\zeta_2(\mathbf{r}) \tag{48}$$

and from (46), the expectation values of  $\sigma(\mathbf{r})$  is

$$\langle \sigma(\mathbf{r}) \rangle = \sum_{\{\boldsymbol{\sigma}\}} \sigma(\mathbf{r}) \rho(\boldsymbol{\sigma}; \zeta_1, \boldsymbol{\zeta}_2) = -\frac{\partial \Phi_2}{\partial \zeta_2(\mathbf{r})}. \tag{49}$$

If the states  $\sigma(\mathbf{r})$  and  $\sigma(\mathbf{r}')$  are uncorrelated  $\langle \sigma(\mathbf{r})\sigma(\mathbf{r}') \rangle$  will factor into  $\langle \sigma(\mathbf{r}) \rangle \langle \sigma(\mathbf{r}') \rangle$ . So

$$\Gamma(\mathbf{r}, \mathbf{r}'; \zeta_1, \{\zeta_2(\mathbf{r})\}) := \langle \sigma(\mathbf{r})\sigma(\mathbf{r}') \rangle - \langle \sigma(\mathbf{r}) \rangle \langle \sigma(\mathbf{r}') \rangle = -\frac{\partial^2 \Phi_2}{\partial \zeta_2(\mathbf{r}) \partial \zeta_2(\mathbf{r}')}, \tag{50}$$

called the *pair correlation function* is a measure of the degree of correlation between  $\sigma(\mathbf{r})$  and  $\sigma(\mathbf{r}')$ . If all the couplings  $\zeta_2(\mathbf{r})$  are set equal to  $\zeta_2$ , it follows from (104) that

$$\sum_{\{\mathbf{r}, \mathbf{r}'\}} \Gamma(\mathbf{r}, \mathbf{r}'; \zeta_1, \zeta_2) = N\varphi_T, \tag{51}$$

which is a *fluctuation-response function relationship*. If translational invariance is assumed, then  $\Gamma(\mathbf{r}, \mathbf{r}'; \zeta_1, \zeta_2) = \Gamma(\bar{\mathbf{r}}; \zeta_1, \zeta_2)$ , where  $\bar{\mathbf{r}} := \mathbf{r} - \mathbf{r}'$  and

$$\sum_{\{\bar{\mathbf{r}}\}} \Gamma(\bar{\mathbf{r}}; \zeta_1, \zeta_2) = \varphi_T \quad \text{with} \quad \Gamma^*(\mathbf{0}; \zeta_1, \zeta_2) = \varphi_T, \tag{52}$$

where  $\Gamma^*(\mathbf{k}; \zeta_1, \zeta_2)$  is the Fourier transform of  $\Gamma(\bar{\mathbf{r}}; \zeta_1, \zeta_2)$ .<sup>39</sup> The correlation length  $\xi(\zeta_1, \zeta_2)$ , given by<sup>40</sup>

$$\xi^2(\zeta_1, \zeta_2) := c(\mathbf{d}) \frac{\sum_{\{\bar{\mathbf{r}}\}} |\bar{\mathbf{r}}|^2 \Gamma(\bar{\mathbf{r}}; \zeta_1, \zeta_2)}{\sum_{\{\bar{\mathbf{r}}\}} \Gamma(\bar{\mathbf{r}}; \zeta_1, \zeta_2)} = -c(\mathbf{d}) \frac{\nabla_{\mathbf{k}}^2 \Gamma^*(\mathbf{0}; \zeta_1, \zeta_2)}{\Gamma^*(\mathbf{0}; \zeta_1, \zeta_2)}, \tag{53}$$

is a measure of distance over which microscopic degrees of freedom are statistically correlated.

<sup>39</sup> Given by

$$\Gamma^*(\mathbf{k}; \zeta_1, \zeta_2) := \sum_{\{\bar{\mathbf{r}}\}} \Gamma(\bar{\mathbf{r}}; \zeta_1, \zeta_2) \exp(-i\mathbf{k} \cdot \bar{\mathbf{r}}), \quad \Gamma(\bar{\mathbf{r}}; \zeta_1, \zeta_2) = \frac{1}{N} \sum_{\{\mathbf{k}\}} \Gamma^*(\mathbf{k}; \zeta_1, \zeta_2) \exp(i\mathbf{k} \cdot \bar{\mathbf{r}}).$$

<sup>40</sup> The prefactor  $c(\mathbf{d})$  is dependent on the number of dimensions  $\mathbf{d}$ , in which the system is infinite. It can be show from Ginzburg-Landau theory that  $c(\mathbf{d}) = 1/(2\mathbf{d})$  [67, Sect. 5.6].

We are now able to augment the scaling theory, described in Sect. 2.4, by applying it to the correlation function and correlation length. Again adopting the magnetic model used in of Sect. 2.4, suppose that near a critical point these functions can be re-expressed in terms of the scaling fields  $\theta_T$  and  $\theta_{\mathcal{H}}$ ;  $\bar{r}$  and  $\mathbf{k}$  can also be treated as scaling fields which, on dimensional grounds will have exponents  $-1$  and  $+1$  respectively. Then the relationships (52) between the correlation function and the response function  $\varphi_T$ , together with the formula (112) derived from Ginzburg–Landau theory suggests a scaling form<sup>41</sup>

$$\Gamma(\lambda^{-1}\bar{r}; \lambda^{y_T}\theta_T, \lambda^{y_{\mathcal{H}}}\theta_{\mathcal{H}}) = \lambda^{\eta+d-2}\Gamma(\bar{r}; \theta_T, \theta_{\mathcal{H}}), \tag{54}$$

for the correlation function, and, hence

$$\Gamma^*(\lambda\mathbf{k}; \lambda^{y_T}\theta_T, \lambda^{y_{\mathcal{H}}}\theta_{\mathcal{H}}) = \lambda^{\eta-2}\Gamma^*(\mathbf{k}; \theta_T, \theta_{\mathcal{H}}), \tag{55}$$

for its Fourier transform. Then, from (53), the scaling form for the correlation length is

$$\xi(\lambda^{y_T}\theta_T, \lambda^{y_{\mathcal{H}}}\theta_{\mathcal{H}}) = \lambda^{-1}\xi(\theta_T, \theta_{\mathcal{H}}). \tag{56}$$

From (52), (55) and (104),  $d - 2y_{\mathcal{H}} = \eta - 2$  and, setting  $\lambda = |\theta_2|^{-1/y_T}$  in (56) gives, from (111)

$$\nu = \nu' = 1/y_T. \tag{57}$$

Then, from (34) and (36),  $\nu(2 - \eta) = \gamma$ , which is the *Fisher scaling law* [32] and  $d\nu = 2 - \alpha$ , which is the *Josephson hyper-scaling law* [52].<sup>42</sup>

### 3.3 Transfer-Matrix Methods

As we have already shown  $S$ ,  $\Phi_1$  and  $\Phi_2$  are all extensive functions of their extensive variables or none of them is. The message FSM–2 sent from statistical mechanics to thermodynamics is that the latter is the case, and in particular that

$$\phi_2 := \frac{\Phi_2(\zeta_1, \zeta_2, N)}{N} = \phi_2(\zeta_1, \zeta_2) \tag{58}$$

is true only as an approximation for large systems.<sup>43</sup> We shall now substantiate this claim by considering a particular way to develop statistical mechanical models,

<sup>41</sup> The exponent of minus one for  $\bar{r}$  is chosen on dimensional grounds. It is also equivalent to the rescaling of length in the renormalization group (item (iii) in Sect. 3.4.1).

<sup>42</sup> This is the only scaling law which involves the dimension  $d$  of the system. For reasons which become evident if Ginzburg–Landau theory is used in the Gaussian approximation [67, Sect. 5.6] it becomes invalid when  $d > d_{UC}$ , the upper-critical dimension. This is the dimension such that, when  $d \geq d_{UC}$ , critical exponents become dimensionally independent with the classical values given by, for example, the van der Waals fluid. For the Ising and similar non-quantum systems (see Appendix 2)  $d_{UC} = 4$ .

<sup>43</sup> In fact the Sackur–Tetrode formula for the entropy of a perfect gas given by (7) and treated there as an assumption is, when derived from statistical mechanics, also not completely extensive. This condition is achieved only when  $N$  is large and the Stirling formula for  $N!$  is applied.

namely the method of transfer matrices. Although, of course, statistical mechanics can model systems of microsystems (molecules) moving, as in a fluid, through a continuum of points, transfer matrix methods are restricted to microsystems confined to the points of a lattice. In principle lattices of any dimension can be considered, but we shall, for easy of presentation, consider only the two-dimensional case. A virtue of this development is that it can be clearly seen how it unfolds as the two lattice directions in which the system gets larger and then infinite are applied separately.

Consider a square lattice, of lattice spacing  $\mathbf{a}$ , with  $N_H$  sites in the horizontal direction,  $N_V$  in the vertical direction, so that  $N = N_H N_V$ . This situation is like the one considered for finite-size scaling in Sect. 3.4.2, when extensivity can be considered separately in the two directions. Periodic boundary conditions are applied so that the lattice forms a torus with horizontal rings of  $N_H$  sites and rings in a vertical plane of  $N_V$  sites.<sup>44</sup> We suppose that the sites of the lattice are occupied by identical microsystems having  $\nu$  possible states.<sup>45</sup> The state of the whole system is  $\boldsymbol{\sigma} := (\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \dots, \boldsymbol{\sigma}_{N_H})$ , where  $\boldsymbol{\sigma}_i$ , the state of the  $i$ -th vertical ring of sites, has one of  $N_R := \nu^{N_V}$  values. Given that contributions to the Hamiltonian arise (at least in the horizontal direction) only between first-neighbour sites the Hamiltonian can be decomposed into interactions between neighbouring rings of sites and within rings. The latter can be distributed between interacting pairs of rings so that the Hamiltonian takes the form of the sum of contributions of interactions between rings and it is straightforward to show that the partition function is expressible in the form

$$Z_2(\zeta_1, \zeta_2, N) = \text{Trace}\{\mathbf{V}^{N_H}\}, \tag{59}$$

where  $\mathbf{V}$  is the  $N_R$ -dimensional transfer matrix with elements consisting of the exponentials of the negatives of the inter-ring interactions. Assuming that  $\mathbf{V}$  is diagonalizable,<sup>46</sup> it is an elementary algebraic result that its trace is equal to the sum of its eigenvalues, which in decreasing order of magnitude we denote as  $\Lambda^{(\ell)}(\zeta_1, \zeta_2, N_V)$ ,  $\ell = 1, 2, \dots, N_R$ . Then, from (40) and (59),

$$\Phi_2(\zeta_1, \zeta_2, N) = -\ln\{[\Lambda^{(1)}(\zeta_1, \zeta_2, N_V)]^{N_H} + \dots + [\Lambda^{(N_R)}(\zeta_1, \zeta_2, N_V)]^{N_H}\}. \tag{60}$$

As we can see the factors  $N_H$  and  $N_V$  of  $N$  are ‘buried’ at different places in this expression and it is clear that the extensivity condition (58) is not satisfied and the negative aspect of the message FSM–2 from statistical mechanics to thermodynamics is justified. However, we can make some progress because, if all the elements of  $\mathbf{V}$  are strictly positive, as will usually be the case, an important theorem of Perron [104] (see also, [37, p. 64], [67, p. 673]) states that the largest eigenvalue of  $\mathbf{V}$  is real, positive and non-degenerate. This means that, in the approximation when  $N_H$  becomes large,

<sup>44</sup> The point we are establishing with respect to extensivity is even more evident in systems with open boundaries.

<sup>45</sup> The Ising model of Appendix 2 is an example of such a model with  $\nu = 2$ .

<sup>46</sup> The condition for this to be the case is that  $\mathbf{V}$  is *simple* [63, p. 146].

$$\Phi_2(\zeta_1, \zeta_2, N) \simeq -N_H \ln\{A^{(1)}(\zeta_1, \zeta_2, N_V)\} \tag{61}$$

with extensivity achieved in the horizontal direction. Two strategies emerge at this point:

The first is to calculate an expression of the form

$$A_1(\zeta_1, \zeta_2, N_V) := [\psi(\zeta_1, \zeta_2)]^{N_V}, \tag{62}$$

valid in the limit  $N_V \rightarrow \infty$  and giving

$$\phi_2(\zeta_1, \zeta_2) = -\ln\{\psi(\zeta_1, \zeta_2)\} \tag{63}$$

in the limit  $N \rightarrow \infty$ . If this calculation can be carried out it is an effective proof of the existence of the thermodynamic limit,<sup>47</sup> which achieves complete extensivity, with free-energy density given by (63). It is, however, a strategy that has been successfully applied in only a few cases, of which Onsager’s [98] solution of the two-dimensional zero-field Ising model and Baxter’s [11] solution of the eight-vertex model are the most well-known instances.

In the absence of a complete solution as represented by (63), the strategy most often adopted is to treat  $N_V$  as a parameter indexing a sequence of models. That is

$$\Psi^{(N_V)}(\zeta_1, \zeta_2) := A^{(1)}(\zeta_1, \zeta_2, N_V) \tag{64}$$

and

$$\phi_2^{(n)}(\zeta_1, \zeta_2) \simeq -\frac{\ln\{\Psi^{(n)}(\zeta_1, \zeta_2)\}}{n}. \tag{65}$$

In the case of the Ising and similar semi-classical models it can be shown by a method due to Peierls [103] that  $\phi_2^{(n)}(\zeta_1, \zeta_2)$  is a smooth function for all  $n > 0$  which exhibits maxima in response functions. A quantitative analysis using finite-size scaling theory (see Sect. 3.4.2) shows that such maxima become increasingly steep for increasing values of  $n$ , with convergence to the singularity associated with the transition in the two-dimensionally infinite system as  $n \rightarrow \infty$ . In particular to the corresponding singularities in Onsager’s solution of the two-dimensional zero-field Ising model. However, in view of the discussion later in this work it should be noted that the limiting process is singular. Although the maxima in the finite- $N_V$  models converge to the singularities in the  $N_V = \infty$  model they remain of a different (non-singular) character however large  $N_V$  becomes.

The pair correlation function and correlation length were defined in Sect. 3.2. In terms of this transfer matrix formulation it can be shown [67, Sect. 11.1.3] that in the limit  $N_H \rightarrow \infty$

$$\xi(\zeta_1, \zeta_2, N_V) \simeq -\mathbf{a} \left\{ \ln |\Omega_2(\zeta_1, \zeta_2, N_V)| \right\}^{-1}, \tag{66}$$

<sup>47</sup> Although, of course, the *current* absence of such a calculation is not a proof of the contrary assertion.

where  $\mathbf{a}$ , the lattice spacing, is now the distance between neighbouring rings of sites,

$$\Omega_2(\zeta_1, \zeta_2, N_V) := A^{(2)}(\zeta_1, \zeta_2, N_V) / A^{(1)}(\zeta_1, \zeta_2, N_V) \tag{67}$$

and

$$\Gamma_2(\mathbf{r}, \mathbf{r}'; \zeta_1, \zeta_2, N_V) \sim \exp\{-|\mathbf{r} - \mathbf{r}'| / \xi(\zeta_1, \zeta_2, N_V)\}, \tag{68}$$

in the limit  $|\mathbf{r} - \mathbf{r}'| \rightarrow \infty$ , where  $\mathbf{r}$  and  $\mathbf{r}'$  lie on the same vertical ring of sites which establishes an asymptotic form for  $f_d(|\bar{\mathbf{r}}|/\xi)$  in (112).

The situation where  $N_H \rightarrow \infty$  and  $N_V$  is finite corresponds to that to be discussed in Sect. 3.4.2, below, for finite-size scaling, where here  $\mathfrak{d} := 1$  and the thickness of the lattice  $\aleph := N_V$ , with a maximum in  $\varphi_T$  and in other response functions signalling an incipient singularity.<sup>48</sup> The eigenvalue ratio  $\Omega_2(\zeta_1, \zeta_2, N_V)$  can also be used as a means of detecting an incipient singularity, but in a slightly different way. Since, in Onsager’s solution for the Ising model, the largest eigenvalue is degenerate along the first-order transition line below the critical temperature [27, p. 194], we expect that  $\Omega_2(\zeta_1, \zeta_2, N_V)$  will begin, as  $N_V$  is increased, to form a ‘plateau’ with small (negative) slope for small temperatures. The end of this plateau, where the negative curvature is a maximum can then be construed as the location of an incipient singularity.<sup>49</sup> The finite-size scaling argument of Sect. 3.4.2 can be applied to all these quantities showing that the maxima converge towards the infinite-system critical value as  $N_V$  increases. However, of course, for finite  $N_V$  we cannot expect these locations to exactly coincide. These perceptions are given further weight by the phenomenological renormalization group procedure described in Sect. 3.4.3(c).

As we have already indicated, the use of transfer matrix methods to determine exact solutions for infinite systems leads into our discussion in Sect. 3.5.1 of the thermodynamic limit. In a similar way our account of incipient singularities resulting from an analysis of systems with  $N_V$  finite leads into our discussion of phase transitions in finite systems is Sect. 3.6.

### 3.4 The Renormalization Group Method

Once it became evident, around the turn of the twentieth century that the exponents associated with a critical point, both in experimental systems and theoretical models were not those derived from classical models, like van der Waals equation, an interest developed in determining their exact values, in experimental systems and also in theoretic models, where of course it was also necessary in many cases to derive the critical temperature. Before the advent of renormalization group methods the most successful way to do this was by using high and low temperature series. These were very successful in obtaining critical temperatures and exponents at second-order critical points. However, although they can be adapted to deal with first-order

<sup>48</sup> In Sect. 3.4.2 we are primarily concerned with fully-finite systems, although as we indicated there, the analysis also applies to cases where, like here,  $0 < \mathfrak{d} \leq d_{LC} < d$ .

<sup>49</sup> Similar arguments apply to the three-state Potts model [67, Sect. 11.32].

transitions, this is not their main strength and they are also not designed to map out the whole picture of phase transition curves in thermodynamic space. This contrasts with the renormalization group methods developed in the late sixties–early seventies. They are able (when they work) not only to deal with critical points but also curves of first-order and second-order transitions. However, any account of these methods should be preceded by some words of warning, like those of John Cardy. As he says [25, pp. 28–29]:

Not only are the words ‘renormalization’<sup>50</sup> and ‘group’<sup>51</sup> examples of unfortunate terminology, the use of the definite article ‘the’ which usually precedes them is even more confusing. It creates the misleading impression that the renormalization group is a kind of universal machine through which any problem may be processed, producing neat tables of critical exponents at the other end. This is quite false. It cannot be stressed too strongly that the renormalization group is merely a framework, a set of ideas, which has to be adapted to the nature of the problem at hand. In particular, whether or not a renormalization group approach is quantitatively successful depends to a large extent on the nature of the problem, but lack of success does not necessarily invalidate the qualitative picture it provides.

Here we shall concentrate solely on the approach to the renormalization group which is usually referred to as happening in ‘real space’; in contrast to the approach initiated by Wilson [128] where renormalization is performed in wave-vector space resulting in expansions in the parameter  $\epsilon := d - 4$ .<sup>52</sup>

The core of real-space renormalization group (RSRG) methods is the construction of a semi-group of transformations on the independent couplings, or functions thereof. There is a variety of procedures for doing this. Many are based on the block-spin method of Kadanoff [53], and another popular technique is decimation, where the states of a proportion of the microsystems is summed out of the partition function. In fact decimation applied to the one-dimensional Ising model, or related models like the Potts model, (see, e.g. [67] Sect. 15.5.1, and Sect. 3.4.3(a) below) is one of the few examples of an exact RSRG transformation. Most transformations involve approximations, which thus means that the critical exponents are approximations with, in many cases no obvious way to make improvements, unlike series methods where, in principle and often with a great deal of labour, improvements are made by extending the series.

In essence the RSRG transformation involves some fractional reduction in the number of degrees of freedom. It would, therefore, seem to follow that there must

<sup>50</sup> A carry-over on the part of Wilson [128] from his work on the high-energy behaviour of renormalized quantum electrodynamics.

<sup>51</sup> It is in fact a semi-group since it has no inverse.

<sup>52</sup> A comprehensive collection of articles on both types of renormalization group methods is contained in the articles in Domb and Green [29] and on real-space methods in the volume edited by Burkhardt and van Leeuwen [19]. For a description of renormalization methods in wave-vector space the reader is referred to Ma [78] and Amit [1]. Accounts of both approaches are given by Goldenfeld [40], Binney et al. [15] and Cardy [25].

have been a prior application of the thermodynamic limit. Whether this is required for the renormalization group and, more generally, whether it is needed at all in the statistical mechanics of critical phenomena is a question that we return to in Sect. 3.5, following a brief account of the ideas involved in the RSRG.

### 3.4.1 General Theory

Underlying the semigroup of transformations on couplings, which is the real-space renormalization group, is a mapping from a lattice  $\mathcal{N}$  to a lattice  $\tilde{\mathcal{N}}$ . For the sake of simplicity we suppose that both are hypercubic lattices with periodic boundary conditions. Then:

- (i) The number of sites  $N$  and  $\tilde{N}$  of  $\mathcal{N}$  and  $\tilde{\mathcal{N}}$  are related by  $\tilde{N} = N/\lambda^d$ , where  $\lambda > 1$ .
- (ii) The lattice spacings  $\mathbf{a}$  and  $\tilde{\mathbf{a}}$  of  $\mathcal{N}$  and  $\tilde{\mathcal{N}}$  are related by  $\tilde{\mathbf{a}} = \lambda\mathbf{a}$ .
- (iii) The size of  $\tilde{\mathcal{N}}$  is reduced by a length scaling  $|\tilde{\mathbf{r}}| = |\mathbf{r}|/\lambda$ .

The renormalization group is constructed by imposing onto the lattice transformation a statistical mechanical transformation. To do this we modify the Hamiltonian (44) to

$$\hat{H}'_2(\boldsymbol{\sigma}; \zeta_0, \boldsymbol{\zeta}; N) := N\zeta_0 + \hat{H}_2(\boldsymbol{\sigma}; \boldsymbol{\zeta}; N), \tag{69}$$

where, for reasons that will become evident below, we have added a term including a trivial coupling  $\zeta_0$  and, as in the presentation of scaling theory at the beginning of Sect. 2.4, generalized the number of non-trivial couplings from two to  $n$ , with  $\boldsymbol{\zeta} := (\zeta_1, \zeta_2, \dots, \zeta_n)$ .<sup>53</sup> The terminology ‘trivial’ signals the fact that, if in (46)  $\hat{H}_2$  is replaced by  $\hat{H}'_2$  and  $Z_2$  by

$$Z'_2(\zeta_0, \boldsymbol{\zeta}, N) := \sum_{\{\boldsymbol{\sigma}\}} \exp\{-\hat{H}'_2(\boldsymbol{\sigma}; \zeta_0, \boldsymbol{\zeta}, N)\}, \tag{70}$$

then the probability density function is left unchanged and

$$\Phi_2(\boldsymbol{\zeta}, N) := -\ln\{\exp(N\zeta_0)Z'_2(\zeta_0, \boldsymbol{\zeta}, N)\}. \tag{71}$$

Bearing in mind the remarks of Cardy, given above, a successful application of this method depends on being able to construct relationships between the couplings  $\zeta_0, \boldsymbol{\zeta}$  in the system on  $\mathcal{N}$  and the couplings  $\tilde{\zeta}_0, \tilde{\boldsymbol{\zeta}}$  in the system on  $\tilde{\mathcal{N}}$ , done in such a way that the values for the couplings for  $\mathcal{N}$  place it in a critical region if and only the same is the case for the values of the couplings for  $\tilde{\mathcal{N}}$ . Since the critical properties of a system are contained within the partition function the invariance

<sup>53</sup> As we see in the example of the two-dimensional Ising model in Sect. 3.4.3(b) below, the renormalization group transformation will in many cases generate further couplings beyond the set dictated by the physics of the model under consideration.

$$Z'_2(\tilde{\zeta}_0, \tilde{\zeta}, \tilde{N}) = Z'_2(\zeta_0, \zeta, N) \tag{72}$$

of that function is a sufficient guarantee; and this is achieved by the relationship

$$\exp\{-\hat{H}'_2(\tilde{\sigma}; \tilde{\zeta}_0, \tilde{\zeta}, \tilde{N})\} = \sum_{\{\sigma\}} w(\sigma, \tilde{\sigma}) \exp\{-\hat{H}'_2(\sigma; \zeta_0, \zeta, N)\}, \tag{73}$$

where the *weight function*  $w(\sigma, \tilde{\sigma})$  satisfies

$$\sum_{\{\tilde{\sigma}\}} w(\sigma, \tilde{\sigma}) = 1. \tag{74}$$

Running over the set of states  $\tilde{\sigma}$  in (73) will, *in principle*, produce *recurrence relationships*<sup>54</sup>

$$\tilde{\zeta}_j = \mathcal{K}_j(\zeta), \quad j = 1, 2, \dots, n, \tag{75}$$

and for  $\tilde{\zeta}_0$  a recurrence relationship which we choose, for convenience to express in the form

$$\tilde{\zeta}_0 = \lambda^d [\zeta_0 + \mathcal{K}_0(\zeta)]. \tag{76}$$

The ‘in principle’ caveat entered here is important. As we shall see it is rarely possible to implement this programme and to choose a weight function without some kind of approximation being applied. And it is frequently the case that consistency can be achieved only by increasing the value of  $n$  from its initial value. When this happens it is necessary, in order to apply repeated iterations, to back-track and for the extra couplings to be included from the start.

The importance of (76) is that it can be used, together with (71) and (72) to obtain the relationship

$$\phi_2(\tilde{\zeta}) = \lambda^d \phi_2(\zeta) - \lambda^d \mathcal{K}_0(\zeta), \tag{77}$$

between the free-energy densities per lattice site at  $\zeta$  and  $\tilde{\zeta}$ . Then, given that (75) can be iterated to produce a sequence of points  $\zeta^{(0)} \rightarrow \zeta^{(1)} \rightarrow \zeta^{(3)} \rightarrow \dots$  in  $\Xi_2$ ,

$$\phi(\zeta^{(0)}) = \sum_{s=0}^{\infty} \frac{1}{\lambda^{sd}} \mathcal{K}_0(\zeta^{(s)}), \tag{78}$$

is the free-energy density at an initial point  $\zeta^{(0)}$ . Although this result seems to imply the need for an infinite number of iterations, this is clearly not possible in practical computations. It is, therefore, fortunate that it is usually found that this series converges after a very few iterations, allowing densities and response functions to be calculated (see the discussion Sect. 3.5.1).

<sup>54</sup> To be precise, the recurrence relationships are derived from (73) as relationships between the Boltzmann factors  $\exp(\tilde{\zeta}_j)$  and  $\exp(\zeta_j)$ ,  $j = 0, 1, \dots, n$  of the couplings.

A fixed point  $\zeta^*$  of (75) is associated with either a single-phase region or a critical region  $\mathcal{C}$  in  $\mathcal{E}_2$ . To analyze its nature we linearize with  $[\Delta\zeta^{(s)}]^\top := \zeta^{(s)} - \zeta^*$  to give  $\Delta\zeta^{(s+1)} \simeq L^* \Delta\zeta^{(s)}$ , where  $L^*$  is the fixed-point value of the matrix  $L$  with elements  $L_{ij} := \partial \mathcal{K}_i / \partial \zeta_j$ . In general  $L^*$  is not symmetric, with different left and right eigenvectors  $w_j$  and  $x_j$  for the eigenvalue  $\Lambda_j$ . It can then be shown<sup>55</sup> that in a neighbourhood of the fixed point there exist scaling fields  $\theta_j = \theta_j(\Delta\zeta)$ ,  $j = 1, 2, \dots, n$  which are smooth functions of the couplings with

$$\theta_j(0) = 0, \quad \theta_j^{(s+1)} = \Lambda_j \theta_j^{(s)}, \tag{79}$$

$$\theta_j \simeq w_j \cdot \Delta\zeta, \quad \Delta\zeta \simeq \sum_{j=1}^n x_j \theta_j, \tag{80}$$

which is a realization of the relationship between scaling fields and couplings described in Sect. 2.4.

From (79)  $\theta_j^{(s+k)} = \Lambda_j^{k+s} \theta_j^{(0)}$  and the semi-group character of this transformation implies that  $\Lambda_j = \lambda^y$ , for  $j = 1, 2, \dots, n$  and a set of exponents  $y_1, y_2, \dots, y_n$ . Then, in a neighbourhood of the fixed point  $\zeta^*$  the couplings  $\zeta_j$  and  $\tilde{\zeta}_j$  in (77) can be expressed as

$$\zeta_j = \zeta_j^* + \sum_{i=1}^n x_i^{(j)} \theta_i, \quad \tilde{\zeta}_j = \zeta_j^* + \sum_{i=1}^n x_i^{(j)} \lambda^{y_i} \theta_i, \tag{81}$$

where  $x_i := (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(n)})$ . In (77) the function  $\mathcal{K}_0(\zeta)$  is regular. So in a region around  $\zeta^*$  the singular part  $\phi_{\text{sing}}(\Delta\zeta)$  of  $\phi_2(\zeta)$ , with  $\phi_{\text{sing}}(0) = 0$ , can be re-expressed in terms of the scaling fields to give

$$\phi_{\text{sing}}(\lambda^{y_1} \theta_1, \dots, \lambda^{y_n} \theta_n) = \lambda^d \phi_{\text{sing}}(\theta_1, \dots, \theta_n), \tag{82}$$

which is a substantiation of (25).

### 3.4.2 Finite-Size Systems

This treatment of criticality, which plays an important role in our understanding of PTCP in real systems (see Sect. 4), was initiated by Fisher [33] and Fisher and Barber [34].<sup>56</sup> For simplicity we suppose, as in Sect. 3.3, that the system under consideration consists of  $N$  identical microsystems on the sites of a  $d$ -dimensional hypercubic lattice  $\mathcal{N}_d$  with  $N_k$  sites in the  $k$ -direction and  $N_1 N_2 \dots N_d = N$ . A partially-infinite system of thickness  $\mathfrak{N} := [N^{(\mathfrak{d})}]^{1/(d-\mathfrak{d})}$ , where  $N^{(\mathfrak{d})} := N_{\mathfrak{d}+1} N_{\mathfrak{d}+2} \dots N_d$ , is obtained if the thermodynamic limit  $N_k \rightarrow \infty$  is taken only for  $k = 1, 2, \dots, \mathfrak{d} < d$ . In a fully-finite system  $\mathfrak{d} = 0$  and  $N^{(\mathfrak{d})} = N$ . We denote the critical region in the partially-infinite system, when  $\mathfrak{d} > d_{\text{LC}}$ , by  $\mathcal{C}(\mathfrak{d}; \mathfrak{N})$ , with  $\mathcal{C}(d; \infty) = \mathcal{C}$ . Finite-size scaling

<sup>55</sup> Assuming that  $L^*$  is a simple matrix.

<sup>56</sup> For a review see Barber [3] and, for a collection of papers on finite-size scaling, Cardy [24].

theory can be applied both to a partially-infinite system, where there is the possibility of a critical region consisting of some kind of singular behaviour, and a fully-finite system where there is not. In a fully-finite system or a partially infinite system with  $\mathfrak{d} \leq d_{LC}$  the critical region is replaced by:

**Definition 1** For a fully-finite, or partially-infinite system with  $\mathfrak{d} \leq d_{LC}$ , a region  $\mathcal{IS}(\mathfrak{d}; \aleph)$  in the space of couplings is one of *incipient singularity*,<sup>57</sup> if in the limit  $\aleph \rightarrow \infty$ , it maps into a critical region  $\mathcal{C}$  of the infinite system.

Expressed in a slightly different way a system has an incipient singularity at certain size-dependent values of its couplings if, as the system size  $\aleph$  is increased, those values converge to ones where thermodynamic functions exhibit *properties that have no finite limits*.

The basic assertion of finite-size scaling is that  $\theta_{\aleph} := 1/\aleph$ , which is a measure of the inverse of finite linear extent of the system measured in units of lattice spacing, can be treated as another scaling field with  $y_{\aleph} = 1$ , meaning that  $\theta_{\aleph}$  is a relevant scaling field, and  $\theta_{\aleph} = 0$  for the infinite system. The only condition required for this is that the system is sufficiently large for the renormalization group transformation in the space of all the other couplings to be unmodified by the finite size of the system. That is to say, that the renormalized couplings can be represented in the system. For simplicity we confine our attention to the simple magnetic system used in Sect. 2.4. The critical region for the infinite system is just a critical point  $T = T_c$ ,  $\mathcal{H} = 0$  with scaling fields  $\theta_T$  and  $\theta_{\mathcal{H}}$ , given by (27), measuring departures from this point. When the system has finite thickness ( $\theta_{\aleph} \neq 0$ ), the incipient singularity is at a different temperature, but because of the symmetry of the system still with  $\mathcal{H} = 0$ . Again, for simplicity, attention will be restricted to the zero-field axis where two temperatures come into play:

- (i) For a system of finite thickness  $\aleph$ ,  $\tilde{T}(\aleph)$  is the *shift temperature* such that, as  $\aleph \rightarrow \infty$ ,  $\tilde{T}(\aleph) \rightarrow T_c$ , the temperature at which *the infinite system has a singularity*. If  $\mathfrak{d} > d_{LC}$  then  $\tilde{T}(\aleph)$  is also a critical temperature, but for the system of finite thickness. If  $\mathfrak{d} \leq d_{LC}$ , and in particular when  $\mathfrak{d} = 0$  and the system is fully-finite,  $T = \tilde{T}(\aleph)$  is a *quasicritical temperature* [34] which is exhibited by a maximum in the susceptibility.<sup>58</sup> This temperature is an example of an *incipient singularity*. In keeping with the other assumptions of scaling theory it is assumed that this convergence is algebraic, so, with scaling field

$$\tilde{\theta}_T(T, \aleph) := \epsilon \left( \frac{1}{T} - \frac{1}{\tilde{T}(\aleph)} \right), \tag{83}$$

<sup>57</sup> It should be noted that this is a slightly different usage from that in Lavis [67, Chap. 11], where such occurrences are called ‘incipient phase transitions’.

<sup>58</sup> Another response function like the heat capacity can replace the susceptibility, with a slightly different quasicritical temperature.

the condition

$$\begin{aligned} \tilde{\Delta}(\aleph) &:= \theta_T(T) - \tilde{\theta}_T(T, \aleph) = \varepsilon \left( \frac{1}{\tilde{T}(\aleph)} - \frac{1}{T_c} \right) \\ &= \theta_T(\tilde{T}(\aleph)) = -\tilde{\theta}_T(T_c, \aleph) \simeq C_s \aleph^{-\chi} \quad \text{as } \aleph \rightarrow \infty, \end{aligned} \tag{84}$$

where  $\chi > 0$  is the *shift exponent*, is sufficient to ensure convergence.

- (ii)  $\mathring{T}(\aleph)$ , called the *rounding temperature* is an important, but rather more elusive, property of the system. It is the temperature at which the susceptibility first shows significant deviation from that of the fully-infinite system. With

$$\mathring{\theta}_T(T, \aleph) := \varepsilon \left( \frac{1}{T} - \frac{1}{\mathring{T}(\aleph)} \right), \tag{85}$$

it is supposed that

$$\begin{aligned} \mathring{\Delta}(\aleph) &:= \tilde{\theta}_T(T, \aleph) - \mathring{\theta}_T(T, \aleph) = \varepsilon \left( \frac{1}{\tilde{T}(\aleph)} - \frac{1}{\mathring{T}(\aleph)} \right) \\ &= \tilde{\theta}_T(\mathring{T}(\aleph), \aleph) = -\mathring{\theta}_T(\tilde{T}(\aleph), \aleph) \simeq C_r \aleph^{-\tau}, \quad \text{as } \aleph \rightarrow \infty, \end{aligned} \tag{86}$$

where  $\tau > 0$  is the *rounding exponent*.

Scaling around the infinite system critical point is shown in Fig. 6. Our interest in this work is in the occurrence of an incipient singularity; so henceforth the assumption is that  $\mathfrak{d} \leq d_{LC}$ .<sup>59</sup> Thus we have three relevant scaling fields with the critical region of the infinite system at the origin  $(\theta_T, \theta_{\mathcal{H}}, \theta_{\aleph}) = (0, 0, 0)$ . However, this is not the complete picture; in general there will be a number of irrelevant scaling fields, which parametrize the critical region and affect its asymptotic properties. For the sake of simplicity we just include the most nearly relevant.<sup>60</sup> of these designated as  $\theta_{\star}$ , with exponent  $y_{\star} < 0$ . Then on the zero-field axis (82) is replaced by

$$\phi_{\text{sing}}(\lambda^{y_T} \theta_T, \lambda^{y_{\star}} \theta_{\star}, \lambda \theta_{\aleph}) = \lambda^d \phi_{\text{sing}}(\theta_T, \theta_{\star}, \theta_{\aleph}). \tag{87}$$

As we have already seen, singular parts of thermodynamic functions like densities and response functions are obtained by differentiations with respect to the scaling fields. In particular, for the susceptibility  $\varphi_T$ , given by (108),

$$\varphi_T(\theta_T, \theta_{\star}, \theta_{\aleph}) = \lambda^{\omega} \varphi_T(\lambda^{y_T} \theta_T, \lambda^{y_{\star}} \theta_{\star}, \lambda \theta_{\aleph}), \tag{88}$$

with  $\omega := 2y_{\mathcal{H}} - d = \gamma/\nu$ , where  $\gamma$  is given by (34) and  $\nu := 1/y_T$ , given in (57), is the critical exponent of the correlation length. Asymptotic behavior in a

<sup>59</sup> In Sect. 3.3 on transfer matrix methods, and Sect. 3.4.3(c) on phenomenological renormalization,  $\mathfrak{d} = d_{LC} = 1$  and  $d = 2$ . Our later discussion in Sect. 4 is concerned with phase transitions in fully-finite systems where  $\mathfrak{d} = 0$ .

<sup>60</sup> That is  $|y_{\star}| := \min_{j \in s+1, \dots, n} |y_j|$ .

neighbourhood of the critical point, that is when  $|\theta_T| \ll 1$ , is then as usual exposed by choosing the scale parameter  $\lambda = |\theta_T|^{-1/y_T}$ , giving

$$\varphi_T(\theta_T, \theta_*, \theta_{\aleph}) = |\theta_T|^{-\gamma} \varphi_T(\pm 1, \mathfrak{X}_*, \mathfrak{X}_{\aleph}), \tag{89}$$

where the  $\pm 1$  branches of  $\varphi_T(\pm 1, \mathfrak{X}_*, \mathfrak{X}_{\aleph})$  apply to the cases  $\theta_T > 0$  and  $\theta_T < 0$ , respectively, and  $\mathfrak{X}_*(T, \aleph) := |\theta_T(T)|^{-y_*\nu} \theta_*$ ,  $\mathfrak{X}_{\aleph}(T, \aleph) := |\theta_T(T)|^{-\nu} \aleph^{-1}$  are scaling functions. In a similar way, with  $\lambda = \aleph$ ,

$$\varphi_T(\theta_T, \theta_*, \theta_{\aleph}) = \aleph^{\omega} \varphi_T(\mathfrak{X}_{\aleph}^{-1/\nu}, \mathfrak{X}_*^{1/y_*}, 1). \tag{90}$$

In the thermodynamic limit  $\aleph \rightarrow \infty$ , it follows from (89) that the susceptibility has the form

$$\varphi_T(\theta_T, \theta_*, 0) = A_T^{(\pm)}(\mathfrak{X}_*) |\theta_T|^{-\gamma}, \tag{91}$$

where the amplitudes

$$A_T^{(\pm)}(\mathfrak{X}_*) := \varphi_T(\pm 1, \mathfrak{X}_*, 0), \tag{92}$$

which are, in general, different for  $\theta_T > 0$  and  $\theta_T < 0$ , are dependent on  $\theta_T$  by virtue of the presence of the irrelevant scaling field  $\theta_*$ . This contribution will become small, as  $|\theta_T|^{-y_*\nu} \rightarrow 0$  for  $|\theta_T| \rightarrow 0$ , eventually becoming negligible for sufficiently small  $|\theta_T|$ . The susceptibility will then display an asymptotic algebraic singularity of the form

$$\varphi_T \simeq A_T^{(\pm)}(0) |\theta_T|^{-\gamma}, \quad \text{as } |\theta_T| \rightarrow 0. \tag{93}$$

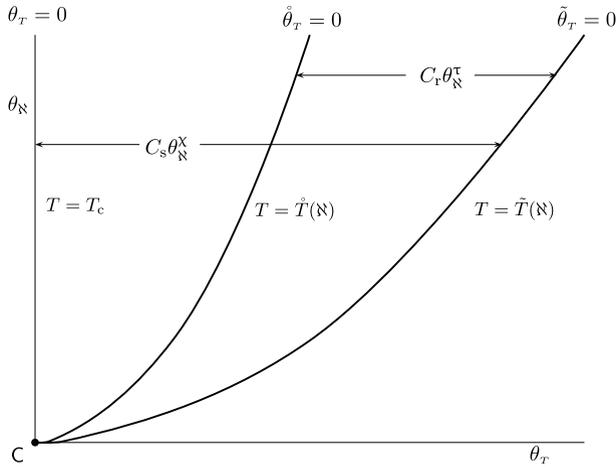
The singularity is a divergence, if  $\gamma > 0$ , which is generally the case for response functions.

Given that both (89) and (90) are valid, and that a finite statistical mechanical system cannot exhibit non-analytic behaviour, whereas singular behaviour does occur at critical points in the limit of infinite system size, the scaling function  $\varphi_T(\pm 1, \mathfrak{X}_*, \mathfrak{X}_{\aleph})$  in (89) must exhibit asymptotic behaviour of the form

$$\varphi_T(\pm 1, \mathfrak{X}_*, \mathfrak{X}_{\aleph}) \simeq B_T^{(\pm)}(\mathfrak{X}_*) \mathfrak{X}_{\aleph}^{-\omega}. \tag{94}$$

Since the susceptibility has maxima along the curve  $T = \tilde{T}(\aleph)$  of shift temperatures in Fig. 6 these maxima will be in one of the branches of  $B_T^{(\pm)}(\mathfrak{X}_*)$  with the other branch being a monotonically decreasing function of  $\mathfrak{X}_*$  in the vicinity of  $\mathfrak{X}_* = 0$ . Along the curve of shift temperatures, from (84),  $\mathfrak{X}_*(\tilde{T}(\aleph), \aleph) \simeq C_s^{-\nu y_*} \aleph^{\nu y_*} \theta_*$  and  $\mathfrak{X}_{\aleph}(\tilde{T}(\aleph), \aleph) \simeq C_s^{-\nu} \aleph^{\nu-1}$ . On this curve  $\theta_* \neq 0$ , and if it is supposed that the two shift functions have the same asymptotic dependence on  $\aleph$ , the shift exponent will be related to  $y_T = 1/\nu$  and  $y_* < 0$  by  $\chi = y_T/(1 - y_*)$  with the shift amplitude  $C_s \simeq [\mathfrak{X}_*(\tilde{T}(\aleph), \aleph)/\mathfrak{X}_{\aleph}(\tilde{T}(\aleph), \aleph)\theta_*]^\chi$ .

As already indicated finite-size corrections to the pure power-law behaviour of  $\varphi_T$ , as described by (93), will begin to be observed whenever the system is finite (with  $\theta_{\aleph} := \aleph^{-1} \neq 0$ ) at the rounding temperature  $\tilde{T}(\aleph)$ . It has been argued [31] that



**Fig. 6** Scaling around the critical point  $C$ , showing the curves  $\dot{\theta}(T, \aleph) = 0$  and  $\tilde{\theta}(T, \aleph) = 0$  of rounding and shift temperatures

this is the temperature at which the size  $\aleph$  of the system is of the same order as the correlation length  $\xi(T)$ .<sup>61</sup> It follows from (111) that  $|\dot{\theta}_T(\dot{T}(\aleph), \aleph)|^{-\nu} \aleph^{-1} \simeq C$ , where  $C$  is a constant, which establishes, from (86), that  $C = C_r$  and the rounding exponent  $\tau = 1/\nu = y_T$  with  $\omega = \gamma\tau$ . Thus on the basis of some plausible assumptions we have the condition  $\chi < \tau$ , which, for large systems, motivates the disposition of the curves in Fig. 6.

### 3.4.3 Renormalization Schemes

The practical implementation of the renormalization group procedure in Sect. 3.4.1 involves the choice of a weight function and leads to recurrence relationships between systems related by a size parameter  $\lambda$ , together with a method for calculating the free-energy density which satisfies the scaling relationship. In (a) and (b) in this section we give examples of the implementation of two of the most commonly used weight functions and in (c) we briefly outline a different scheme which, using transfer matrix methods, relates the correlation lengths of systems of different sizes.

For  $d$ -dimensional lattices, most weight functions are based on a division of the lattice  $\mathcal{N}$  into equal blocks of  $\lambda^d$  sites. The mapping from  $\mathcal{N}$  to  $\tilde{\mathcal{N}}$  is given by associating each lattice site  $\tilde{\mathbf{r}} \in \tilde{\mathcal{N}}$  with a blocks of sites in  $\mathcal{N}$  denoted by  $\mathcal{B}(\tilde{\mathbf{r}})$ .

- (a) The decimation weight function. For this weight function the sites of  $\tilde{\mathcal{N}}$  consist of a subset of the sites of  $\mathcal{N}$ , chosen so that  $\tilde{\mathcal{N}}$  forms a lattice which is isomorphic to  $\mathcal{N}$ . So we can take  $\tilde{\mathbf{r}} \in \mathcal{B}(\tilde{\mathbf{r}})$  with

<sup>61</sup> This can quite easily be established explicitly for a one-dimensional Ising model on a ring of  $N$  sites, where the magnetization density  $m(T, \mathcal{H}, N)$  is given by  $m(T, \mathcal{H}, N) = \tanh[N/2\xi(T)] \times m(T, \mathcal{H}, \infty)$ .

$$w(\sigma, \tilde{\sigma}) := \prod_{\{\tilde{r}\}} \delta^{\text{kr}}(\tilde{\sigma}(\tilde{r}) - \sigma(\tilde{r})). \tag{95}$$

The effect of this is that the summation on the right-hand side of (73) is a partial sum over all the sites of the lattice  $\mathcal{N}$  except those of  $\tilde{\mathcal{N}}$ . For a range of one-dimensional models (including the Ising and Potts models), which can be solved exactly using transfer matrix methods, exact RSRG decimation transformations can also be obtained. For the one-dimensional ferromagnetic case of the Ising model it can be shown [67, 87] that the most convenient variables are not those given in Appendix 2 but rather  $\zeta_1 := \tanh([2J + H]/2T)$ ,  $\zeta_2 := \exp(-2H/T)$ , and for  $\lambda = 2$ , with the partial summation in (73) over alternate sites, the recurrence relationships take the form

$$\tilde{\zeta}_1 = \frac{4\zeta_1^2 - (1 - \zeta_2)(\zeta_1^2 - 1)}{4 + (1 - \zeta_2)(\zeta_1^2 - 1)}, \quad \tilde{\zeta}_2 = \frac{\zeta_2^2(1 + \zeta_1)^2 + (1 - \zeta_1)^2}{2(1 + \zeta_1^2)}. \tag{96}$$

It is then not difficult to show that there is a fixed point  $\zeta_1 = \zeta_2 = 1$  ( $T = H = 0$ ), with both scaling exponents equal to  $d = 1$ . As we saw in the discussion of scaling theory in Sect. 2.4, an exponent equal to the dimension of the system is indicative of the possibility of a first-order transition. In this case the critical point is at zero temperature on the zero field line, meaning that the first-order coexistence curve has contracted to a point coinciding with the critical point at zero-temperature. At this point there is a first-order transition across the zero-field axis with a change of sign of the magnetization. It can also be shown that the curve

$$\zeta_2 = \left( \frac{1 - \zeta_1}{1 + \zeta_1} \right)^2, \tag{97}$$

which corresponds to the interaction  $J$  between microsystems being set to zero, is invariant under (96). At every point it has exponents 0 and  $-\infty$ ; the first of these is marginal, which indicates that the line consists of fixed points, and the latter that it is ‘infinitely attractive’ to points not on the line. The end points of the line  $\zeta_1 = 0$ ,  $\zeta_2 = 1$  ( $H = \infty, T = 0$ ) and  $\zeta_1 = 1$ ,  $\zeta_2 = 0$  ( $T = \infty, H = 0$ ) are fixed points in their own right in the invariant subspaces  $T = 0$  and  $H = 0$  respectively. The phase diagram is shown in Fig. 7. Of course, for reasons just explained, the one-dimensional Ising model is less interesting than the two-dimensional model where the ferromagnetic critical point is not at zero temperature. So, suppose that we try to carry out the same procedure in that case. A possibility is to choose blocks of two sites as shown in Fig. 8. The lattice  $\tilde{\mathcal{N}}$  consists of the black sites and the partial summation in (73) is over the spin states on white sites. But this will create an interaction between the *four* sites surrounding each white site. So we would need to back-track and increase  $n$  from two to three, inserting this interaction from the beginning. But this would in turn generate an interaction between nine sites. And so on. This proliferation of interactions is typical of the problems encountered with decimation. The usual trick is to cut off the proliferation at a certain level. Such an approximation for

this model was investigated by Wilson [128] with a rather poor outcome compared to the known exact results.

- (b) The majority-rule weight function. This weight function was introduced by Niemeijer and van Leeuwen [90, 91]. The first step in assigning  $\tilde{\sigma}(\tilde{\mathbf{r}})$  for the block  $\mathcal{B}(\tilde{\mathbf{r}})$  can be described in terms of the ‘winner takes all’ voting procedure used in some democracies. Given that each microsystem has  $\nu$  states and that among the sites of  $\mathcal{B}(\tilde{\mathbf{r}})$  one of the  $\nu$  state occurs more than any other,  $\tilde{\sigma}(\tilde{\mathbf{r}})$  is assigned to have this value. If  $\nu := 2$  and the number of sites  $\lambda^d$  in a block is odd this rule works; a case in point being the treatment of the Ising model on the triangular lattice with a block of nine sites ( $\lambda := 3$ ) by Schick et al. [115]. But unless these conditions hold it is clear that the simple majority rule is not sufficient to determine  $\sigma(\tilde{\mathbf{r}})$  for every configuration of the block. A ‘tie’ can occur in the voting procedure and a strategy must be adopted to deal with such cases. One possibility is to assign to  $\sigma(\tilde{\mathbf{r}})$  one of these predominating values on the basis of equal probabilities. In some cases this may not, however, be the most appropriate choice. In their work on the Ising model using a square first-neighbour block ( $\lambda = 2$ ) Nauenberg and Nienhuis [86] divided the configurations with equal numbers of up and down spins between block spins up and down with equal probabilities. The rule (one of four) which they chose ensured that the reversal of all the spins in the block reversed the block spin.
- (c) Phenomenological renormalization. The idea of finite-size scaling, introduced in Sect. 3.4.2, leads quite naturally [3, Sect. IV] to the RSRG method developed by Nightingale [93]. The essential feature of finite-size scaling is that, for a  $d$ -dimensional system, infinite in  $\mathfrak{d}$  dimensions and of thickness  $\aleph$ , the quantity  $1/\aleph$  is treated as an additional scaling field  $\theta_{\aleph}$ . If attention is restricted to the simple magnetic system with the two other scaling fields  $\theta_T$  and  $\theta_{\mathcal{H}}$ , the response function  $\varphi_T$  satisfies the scaling relationship (88). A similar inclusion of  $\theta_{\aleph}$  in the scaling relationship (56) for the correlation length gives

$$\xi(\lambda^{\nu_T} \theta_T, \lambda^{\nu_{\mathcal{H}}} \theta_{\mathcal{H}}, \lambda \theta_{\aleph}) = \lambda^{-1} \xi(\theta_T, \theta_{\mathcal{H}}, \theta_{\aleph}). \tag{98}$$

With the slight change of notation  $\xi^{(\aleph)}(\theta_T, \theta_{\mathcal{H}}) := \xi(\theta_T, \theta_{\mathcal{H}}, \theta_{\aleph})$ , (98) can be regarded as relating the correlation lengths of two similar systems denoted by  $\mathcal{L}_{\mathfrak{d}}(\aleph)$  and  $\mathcal{L}_{\mathfrak{d}}(\tilde{\aleph})$  with couplings  $\zeta_T, \zeta_{\mathcal{H}}$  and  $\tilde{\zeta}_T, \tilde{\zeta}_{\mathcal{H}}$  and thicknesses  $\aleph$  and  $\tilde{\aleph} := \aleph/\lambda, \lambda > 1$ , respectively. The relationship (98) can be reexpressed as

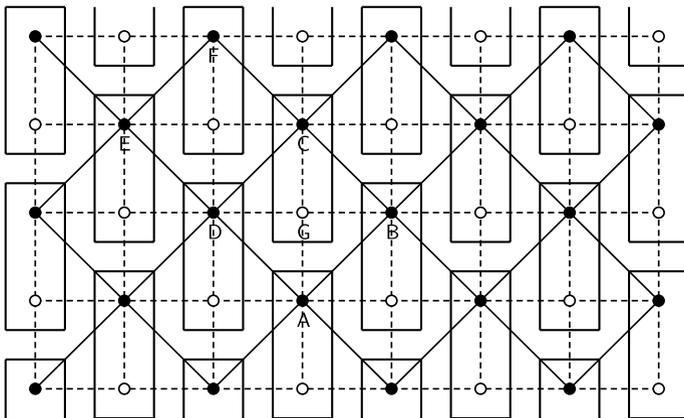
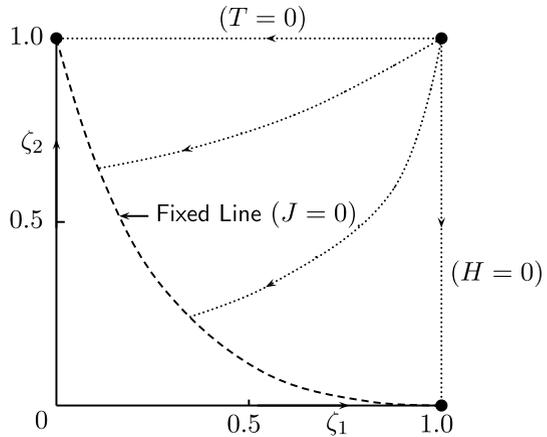
$$\xi^{(\aleph)}(\theta_T, \theta_{\mathcal{H}}) = \lambda \xi^{(\tilde{\aleph})}(\tilde{\theta}_T, \tilde{\theta}_{\mathcal{H}}), \tag{99}$$

where

$$\tilde{\theta}_T = \theta_T(\tilde{\zeta}_T, \tilde{\zeta}_{\mathcal{H}}) = \lambda^{\nu_T} \theta_T(\zeta_T, \zeta_{\mathcal{H}}), \quad \tilde{\theta}_{\mathcal{H}} = \theta_{\mathcal{H}}(\tilde{\zeta}_T, \tilde{\zeta}_{\mathcal{H}}) = \lambda^{\nu_{\mathcal{H}}} \theta_{\mathcal{H}}(\zeta_T, \zeta_{\mathcal{H}}) \tag{100}$$

relate the scaling fields for  $\mathcal{L}_{\mathfrak{d}}(\aleph)$  and  $\mathcal{L}_{\mathfrak{d}}(\tilde{\aleph})$ . These relationships form the basis of Nightingale’s phenomenological renormalization method, where the correlation lengths for systems of the two widths are obtained from transfer matrix calculations using (66). In the case of one scaling field ( $\mathcal{H} = 0$ ) the method yields the critical temperature fixed point  $\theta_T^* := \theta_T = \theta_{\tilde{T}}$  and the thermal exponent for a number of

**Fig. 7** The trajectory flows for the renormalization group transformation of the one-dimensional Ising model



**Fig. 8** Two site blocks for the first-neighbour Ising model on a square lattice. The lattice  $\mathcal{N}$  consists of both white and black sites and  $\tilde{\mathcal{N}}$  of only black sites

different models [60, 93, 94, 118], which in the case where exact results are known are at a high level of accuracy.<sup>62</sup>

### 3.5 The Thermodynamic Limit

In the development of statistical mechanics represented by the right-hand column in Fig. 1 system size appears twice. Firstly in the passage for SM1 to SM2, where

<sup>62</sup> However, if there is more than one coupling ( $\mathcal{H} \neq 0$ ) then substituting  $\theta_T^* := \theta_T = \bar{\theta}_T$  and  $\theta_{\mathcal{H}}^* := \theta_{\mathcal{H}} = \bar{\theta}_{\mathcal{H}}$  in (99), does not completely define the critical point. A number of methods for dealing with this case have been developed, [60, 118]

the system becomes large yielding approximate extensivity. This is needed for the discussion of finite-size phase transitions represented by SM5. Secondly in the other branch from SM2, via the thermodynamic limit, to an infinite system represented by SM3. This entails the identification of the infinite statistical mechanical system SM3 with thermodynamics, or at least the version, labelled TD3 in Fig. 1, of thermodynamics with some PTCP defined. But are SM3 and TD3 actually identical? The answer is clearly ‘no’. TD3 is the result of a development in the left-hand column of Fig. 1, from the basic structure through the assumption of extensivity to a grafting on of a picture of PTCP, in the manner of Pippard [105] or Buckingham [18]. On the other hand, as we have just indicated, SM3 is the result of a statistical mechanical development in the right-hand column in Fig. 1. It retains its microstructure with a probability distribution, and in most cases it is the result of the implementation of the thermodynamic limit for a particular model, the most well-known examples being the two-dimensional zero-field Ising model and the eight-vertex model. Thus it should be recognised for later reference (see Sect. 5.1) that this way of understanding the relation between thermodynamics and statistical mechanics involves the unwarranted *conflation of two quite different pictures*. Although one can argue that SM3 is an enrichment of TD3, since the former has all the features of the latter together with the extra ones provided by microstructure and precise results concerning critical values and exponents. That having been said, one may still question whether the thermodynamic limit is:<sup>63</sup>

- (1) Necessary, in principle, because statistical mechanics is not complete without it.
- (2) Useful because calculations become much simpler in the thermodynamic limit and the relationship FSM–3 of SM3 to TD3 makes it easier to identify the order of phase transitions.

Although both of these possibilities deserve consideration it is the the first which has received the most attention, principally because of the role of the thermodynamic limit in the understanding of PTCP; this will be discussed in detail in Sect. 3.5.1.

In this work we propose, in Sect. 4, a particular view of the usefulness of the thermodynamic limit in the context of phase transitions in finite systems. However, it is pertinent to note the range of possible circumstances calling for the use of the thermodynamic limit. In particular one might suppose an additional kind of necessity interposed between the two items in our list:

(1a) Necessary in practice, because calculations for particular models are not tractable without its use.

However, of course, tractability, and hence necessity in practice, is ephemeral, evolving (one might hope) with an increase in computing power and technical ingenuity into mere usefulness.

<sup>63</sup> A third possibility that the thermodynamic limit is neither necessary nor useful can surely be discounted, with respect to usefulness, after a cursory survey of the corpus of work on statistical mechanics.

### 3.5.1 Phase Transitions in Infinite Systems

The argument for the necessity in principle of the thermodynamic limit for PTCP effectively involves asserting the truth of the contradictory set of propositions:

- P-IA** PTCP occur in nature.
- P-IB** PTCP occur in nature as discontinuities in densities (first-order transitions) and as singularities in response functions (higher-order transitions).<sup>64</sup>
- P-IIA** PTCP in thermodynamics are defined by singularities in derivatives of first or higher order in the free energies and are treated as such using scaling theory.
- P-IIB** PTCP must necessarily be represented in thermodynamics by singularities.
- P-IIIA** PTCP should be able to be modelled in statistical mechanics.
- P-IIIB** PTCP should be modelled in statistical mechanics in the same way that they are in thermodynamics.
- P-IV** Real systems are of finite size.
- P-V** Thermodynamic functions for finite systems in statistical mechanics are regular functions.
- P-VI** Thermodynamic functions for infinite systems in statistical mechanics can show singularities.

For later use it is relevant to compare this list with that of Callender [22, p. 589] (repeated by Mainwood [80, pp. 13–14]):

- CP-I** Real systems have finite [size].
- CP-II** Real systems display phase transitions.
- CP-III** Phase transitions occur when the partition function has a singularity.<sup>65</sup>
- CP-IV** Phase transitions are governed/described by classical or quantum statistical mechanics (through [the partition function]).

A number of items in our list are indisputable and are not included in Callender's list:

- That PTCP are defined in thermodynamics by singularities, can be confirmed by a visit to the thermodynamics section of any academic library (**P-IIA** is true). Whether it is necessary for thermodynamics to be formulated in this way (that **P-IIB** should be accepted), given a possible denial that PTCP occur in nature as singularities (that **P-IB** is true) is a different question.
- The joint assertions that thermodynamic functions are regular for finite systems but can have singularities for infinite systems (included in our list as **P-V** and **P-VI**, respectively, but not contained in Callender's list) are facts about the math-

<sup>64</sup> This latter group also includes other sorts of weaker singular behaviour.

<sup>65</sup> In relation to this statement, see footnote 76.

ematical structure of statistical mechanics which cause the total list to be contradictory.

And on Callender's list:

- It is difficult to argue that phase transitions do not occur in real systems (that **P-IA (CP-II)** is false), although it is plausible to deny that they arise as some kind of singularities (to argue that **P-IB** (not in Callender's list) is false), on the grounds that a first-order transition (say that between liquid water and water vapour) may look like a sudden change of density, but on closer observation would turn out to be a very steep continuous change. Likewise, apparent singularities in compressibility in fluids and susceptibility in magnets may just be very steep maxima.
- It is also difficult to argue that real systems are not finite (that **P-IV (CP-I)** is false), given that no system takes up the whole of the universe.<sup>66</sup> A sort of argument could be constructed on the basis that no system is completely isolated, but this would mean accepting the need for computation, not with an infinite system as envisaged here, but with a system joined to a complicated and largely undetermined environment.
- If the ability to model PTCP were not deemed to be a necessary part of statistical mechanics (**P-IIIA (CP-IV)** is rejected), then most of the work on statistical mechanics in the last half century and more would be pointless. It is, however, relevant here to mention the work of the late Ilya Prigogine (in particular, [106]). Although, in a sense he accepts **P-IIIA**, it is a radically different form of statistical mechanics that he has in mind. From the assertion that "[a]s long as we consider merely a few particles, we cannot say if they form a liquid or gas" (ibid, p. 45) he concludes that "[s]tates of matter as well as phase transitions are ultimately defined by the thermodynamic limit. ... Phase transitions correspond to emerging properties. They are meaningful only at the level of populations and not of single particles" (op. cit.). This entails for him the reformulation of statistical mechanics so that the underlying dynamics is not that of trajectories but of measure.<sup>67</sup>

There remain **P-IB** and **P-IIIB**, which together with **P-IIB** is equivalent to **CP-III**, and we now consider the consequences of denying one or both of them.

- (i) If **P-IB** is accepted, that is that PTCP in nature do occur as singularities, then it is clearly necessary for thermodynamics to represent them in this way; **P-IIB** must be accepted. Then we seem to be driven toward the conclusion that statistical mechanics should model them in the same way (that is the acceptance of **P-IIIB**) which leads back to the contradiction. This is avoided by

<sup>66</sup> Which, in any event, may be finite.

<sup>67</sup> This being the approach that he and his Brussels group also used to resolve the problem of irreversibility (see, for example, [107]).

denying **P–IIIB**. Then PTCP can be modelled in statistical mechanics without singularities, by, for example, transfer matrix methods, while at the same time admitting that this is not the situation in reality.

- (ii) If **P–IB** is denied then it can be argued either:
- (a) That it is not necessary for thermodynamics to model PTCP as singularities (**P–IIB** is false). In this case **P–IIIB** can be accepted, with PTCP modelled without singularities in statistical mechanics, with thermodynamics reformulated to do the same.  
or
  - (b) That in statistical mechanics PTCP should be modelled without singularities, but because for large systems steep maxima in response functions and steep changes in densities look very much like singularities and discontinuities, it is still necessary (on the grounds of tractability and simplicity) to model PTCP in thermodynamics as singular behaviour; **P–IIB** is accepted and **P–IIIB** is rejected.<sup>68</sup>

So given that all of **P–I** to **P–VI** are accepted is there any way out of the paradox? One radical approach, which has already been noted, is that due to Prigogine, where statistical mechanics is reformulated to ‘build in’ the thermodynamic limit.<sup>69</sup> Somewhat similar, but less radical, is the approach of Robert Batterman, a philosopher of physics who has written extensively on questions related to phase transitions, the renormalization group and the thermodynamic limit [4–9]. Rather than formulating a novel form of the mechanics underlying statistical mechanics, his argument, following the lead of Kadanoff [57], is that the renormalization group is itself a novel approach, revolutionary in the sense of Kuhn [62], which has the thermodynamic limit built in. His starting point is that thermodynamics<sup>70</sup>

is correct to represent [phase transitions] mathematically as singularities. (**A**: [5, p.234].)

And:

Further, without the thermodynamic limit, statistical mechanics would completely fail to capture a genuine feature of the world. Without the thermodynamic limit, in fact, *statistical mechanics is incapable even of establishing the existence of distinct phases of systems*. (**B**: op. cit.)

If there is any doubt about his view of real systems, this is dispelled by his forthright assertion that he wants

to champion the manifestly outlandish proposal that despite the fact that real systems are finite, our understanding of them and their behaviour requires, in a

<sup>68</sup> To preview Sect. 4, this is the position we shall defend.

<sup>69</sup> This involves an extension of the Koopman [61] formulation to a space beyond the Hilbert space in which it is set.

<sup>70</sup> For reference in the summary (a)–(f) of his position on the renormalization group in Sect. 3.5.2 the quotations from Batterman’s work are given labels **A–F**.

very strong sense, the idealization of infinite systems and the thermodynamic limit. (C: *ibid*, p. 231.)

‘Outlandish’ or not his position is one which would appear, in our experience, to be that adopted implicitly or explicitly by many working physicists, including, albeit in a radical sense as indicated above, by Prigogine, and Kadanoff [55, p. 238], who asserts that the “existence of a phase transition requires an infinite system. No phase transitions occur in systems with a finite number of degrees of freedom”. Kadanoff calls this the “extended singularity theorem” [57, pp. 154–156] because “these singularities have effects that are spread out over large regions of space” [58, p. 24]. Having asserted that

the idea that we can find analytic partition functions that “approximate” singularities is mistaken, because the very notion of approximation required fails to make sense when the limit is singular, [which it is in this case because the] behaviour at the limit (the physical discontinuity, the phase transition) is *qualitatively* different from the limiting behaviour as that limit is approached. (D: *ibid*, p. 236)

Batterman’s proposal for resolving the puzzle is to resort to the renormalization group. In the next section this possibility is examined.

### 3.5.2 Infinite Systems and the Renormalization Group

‘Infinity’ as it arises in accounts of renormalization group methods consists not so much in the limiting process, evident in, say, Onsager’s solution of the zero-field two-dimensional Ising model, whereby the dimensions of the system are taken to infinity, but rather in the perception that to make the method intelligible one must be working with a system which is already infinite [101, 102]. To spell this out, a renormalization group scheme consists of the following:

- (i) In the space of couplings (or of functions thereof) a semigroup of transformations is derived which generates recurrence relationships under which any critical regions are invariant.
- (ii) In this ‘dynamic system’ the critical regions are the basins of attraction of critical fixed points. And there are sinks associated with non-critical regions (phases) of the system.
- (iii) A critical fixed point determines the universality class of the system at each point in its basin of attraction, with an associated set of critical exponents.
- (iv) In general a system may be able to be in more than one universality class determined by the symmetry group of the Hamiltonian when there is a particular relationship between the couplings.<sup>71</sup>

<sup>71</sup> For example, a renormalization scheme for the spin-1 Ising model will have fixed points for both the spin- $\frac{1}{2}$  Ising model and the 3-state Potts model.

It is clear that this way to do statistical mechanics is very different from the standard procedures (mean-field and other classical approximations, series expansions and exact solutions). So much so that, as we have already indicated, it is characterized by Kadanoff [57] as a Kuhn-type revolution, a view endorsed by Batterman [8]. The argument presented by Batterman concerning the whole question of singularities/real singular systems/the thermodynamic limit needs to be carefully rehearsed and for this his [8] tribute to Leo Kadanoff provides the clearest account.

He presents his view in contradistinction to that of Jeremy Butterfield who contends [21, p. 1077] that: “The use of the infinite limit ... is justified, despite  $N$  being actually finite, by its being mathematically convenient and empirically correct (up to the required accuracy)”. For an understanding of Batterman’s view two quotes are particularly useful. In the first he asserts that:

the RG is not just a theory of the critical point, but rather it is a *theory of the critical region*. And this covers large but finite systems. So contrary to the line of reasoning presented [by Butterfield] the explanation of the behaviour of real finite systems requires the use of mathematical infinities, but does not require there to *be* infinite real systems. (E: [8, p.571].)

At this point we have cause to be grateful to a referee of his paper, who objected that this quote was actually in line with “the claims of those supporting the idea that real phase transitions aren’t sharp”. In response to this Batterman added a footnote in which he clarified his position in the following way:

It seems to me that if one is going to hold that the use of the infinite limits is a convenience, then one should be able to say how (even if inconveniently) one might go about finding a fixed point of *the RG transformation* without infinite iterations. I have not seen any sketch of how that is to be done. The point is that the fixed point, as just noted, determines the behaviour of the flow in its neighbourhood. If we want to explain the universal behaviour of finite large systems using the RG, then we need to find a fixed point and, to my knowledge, this requires an infinite system. (F: op. cit.)

So to summarize his view (using the labelled quotes A–F, given above):

- (a) Phase transitions are real discontinuities in experimental systems (A). [An acceptance of P–IA,B and P–IIA,B].
- (b) The thermodynamic limit is needed in statistical mechanics to exhibit phase transitions (B). [An implicit acceptance of P–V and P–VI and an endorsement of P–IIIA,B.]
- (c) Real systems are finite but in order to understand them we need the idealization of infinite systems and the thermodynamic limit (C). [An endorsement of P–IV, and more.]
- (d) The idea that the study of large systems can play a role here is wrong because the properties of large systems and infinite systems are qualitatively different (D).
- (e) To represent the situation correctly we need to engage with mathematical singularities but not real infinite systems (E).

- (f) We need infinite iteration (of the RG transformation) to obtain fixed points (and all the information they provide) (**F**).

And to summarize the summary of Batterman's position:

Although phase transitions in real systems are accompanied by singular behaviour, and in statistical mechanical models this singular behaviour is exhibited only by infinite systems, we don't need infinite systems, just the use of mathematical singularities, these being required to derive the fixed points in renormalization group calculations.

At this point we wish to challenge the last part of this statement by providing<sup>72</sup> the 'sketch' that Batterman (quote **F**) requires of the means of the determination of renormalization group fixed points.

The first thing to note is that the recurrence relationships (75) and (76) are derived (almost always with some approximations involved) between the couplings of two *finite* systems with sizes  $N$  and  $\tilde{N}$  with  $N/\tilde{N} = \lambda^d > 1$ . Once this is done no point in the space of couplings is intrinsically associated with a system of a particular size and, by the same token, fixed points, obtained from (75) with  $\tilde{\zeta}_j = \zeta_j$ , are not associated with infinite systems. However, *if we were to choose* to associate a particular system-size  $N$  with the first point of a trajectory, it would be necessary to assume only that we are working with a system large enough to allow the required number of iterations.<sup>73</sup> (Hence the inclusion of SM2 in the path from SM1 to SM4 in Fig. 1.) As Norton [97, p. 222] says, fixed points are the "limit points" of the sequences generated by the recurrence relationships; the "mathematical pegs on which to hang limit properties" which are never reached in a finite number of iterations. They do not arise from an investigation of the properties of infinite limit systems, and, although they are properties of the transformation, iteration is not always needed for their determination. In some simple cases, like the one-dimensional Ising model described in Sect. 3.4.3, the fixed points can be extracted by direct analytic solution of the fixed point equations. But, in more complicated cases numerical computation comes into play. Although in principle iteration of the recurrence relationships starting from a point in the basin of attraction of a fixed point will generate a sequence of points approaching the fixed point, this is not usually a viable strategy for their determination. Since those of greatest interest, associated with critical regions, have both irrelevant directions of attraction within the critical region (the basin of attraction) and relevant directions along which the trajectory is driven away from the critical region. Except in special cases it is difficult to start a trajectory in a critical region, but nearby points are useful and possible. Then the trajectory will hover near the critical fixed point before it moves away to the sink associated with the phase containing the trajectory. These 'hover points' can be spotted by inspection of the computer output and used as initial guesses for a numerical solution of the fixed point equations. These kinds of numerical techniques, used also to map the

<sup>72</sup> Based on practical experience [70, 71, 119, 120, 129].

<sup>73</sup> Here we are agreeing with Batterman that an infinite system is not required.

critical regions themselves, provide a good picture of the whole phase diagram. And linearization of the recurrence relationships about the fixed points allows the critical exponents to be obtained.

### 3.6 Phase Transitions in Finite Systems: Mainwood’s Proposal

Given, as we have concluded in the previous section, that the thermodynamic limit is not necessary to enable renormalization group calculations to provide the PTCP structure, is it still useful in other statistical mechanical treatments of PTCP? An assessment of usefulness, as distinct from necessity, is obviously heavily influenced by the position adopted with respect to whether PTCP occur in nature as singularities (**P–IB**). If it is false and real systems, by virtue of their size ( $\sim 10^{23}$  microsystems) exhibit behaviour approximating to singular behaviour, in the sense, say, that the maximum in the compressibility of a fluid is experimentally indistinguishable from a singularity, then we have the means to remove the contradiction in the set of statements at beginning of Sect. 3.5.1. One way would be to deem it unnecessary for PTCP to be treated as singularities in thermodynamics (a denial of **P–IIB**). Although this would allow thermodynamics and statistical mechanics to be modelled in the same way (for **P–IIIB** to be accepted) we would argue, for the reasons given in Sect. 4, that it is not a tenable possibility.

The alternative, which is the one discussed in this section, and which is favoured by ourselves, is to accept that thermodynamics must represent PTCP in terms of singularities (**P–IIB**) on the basis that this is an appropriate approximation to real systems. Thus rejecting the assertion that thermodynamics and statistical mechanics must model PTCP in the same way (**P–IIIB**), since statistical mechanics models phase transitions in finite systems. Given that real systems are very large (in terms of the number of microsystems) and finite, with phase transition giving the appearance, but not the exact reality of singularities, can calculations avoid using the thermodynamic limit? Or, more generally can recourse to a system where PTCP occur as singularities be avoided? Here we examine a proposal of Mainwood [80] which definitely answers the question in the negative and in the next section we propose an answer which is more nuanced.

The definition of a phase transition provided by Mainwood (ibid, p. 28) can<sup>74</sup> be described in the following way. For a statistical mechanical system  $\mathfrak{S}_N$  of size  $N$  with partition function  $Z_2(\zeta_1, \zeta_2, N)$ , the free energy  $\Phi_2(\zeta_1, \zeta_2, N)$  is given by (40) and satisfies (5) and (6).<sup>75</sup> Suppose that the thermodynamic limit

$$\lim_{N \rightarrow \infty} \frac{\Phi_2(\zeta_1, \zeta_2, N)}{N} = \phi_2(\zeta_1, \zeta_2) \tag{101}$$

exists, with  $\phi_2(\zeta_1, \zeta_2)$  the free-energy density of the system  $\mathfrak{S}_\infty$ . Then:

<sup>74</sup> With some changes of notation to give conformity with our usage.

<sup>75</sup> We have chosen the system with two independent couplings for convenience.

**Definition 2**  $(\zeta_1, \zeta_2)$  is a point with a particular criticality for  $\mathfrak{S}_N$  iff  $(\zeta_1, \zeta_2)$  is a point where  $\mathfrak{S}_\infty$  has a singularity associated with this same criticality.

And Mainwood (ibid, p. 29) asserts that:<sup>76</sup>

Rather surprisingly, using this definition it is possible to hold on to all of Callender’s four statements [(given above as **CP–I** to **CP–IV**)] without contradiction; though only in a Pickwickian sense—it is a “trick” possible only due to his choice of wording. Namely, the singularity referred to in [**CP–III**] is one not in the partition function [of  $\mathfrak{S}_N$ ] but in [the partition function of  $\mathfrak{S}_\infty$ ].

If this is regarded as a positive point in favour of Mainwood’s definition, the overall conclusion seems to be more mixed. Mainwood ‘worries’ that:<sup>77</sup>

- (1) The definition means that a phase transition can be predicted in a finite system, however small it might be (ibid, p. 32).
- (2) “While there exist standard procedures for taking the thermodynamic limit, ... these procedures are human inventions, and choices could have been made differently. ... The definition of a phase transition thus seems arbitrary in a disastrous sense: we can choose whether one is occurring or not by modelling it differently, or taking the limit according to a different scheme” (ibid, p. 31).
- (3) “[T]he facts we need to decide whether or not [a physical system] is undergoing a phase transition should be physical facts, about actual states of affairs ... They should not exist only in an idealized model on a theoretician’s blackboard” (ibid, p. 29).

Although Mainwood adds (1) as a final difficulty it is probably the one which would first spring to mind, since the definition would imply a phase transition in an Ising model of four spins in a square at the critical temperature given by Onsager’s solution. Mainwood thinks that “this bullet can and should be bitten” (ibid, p. 32), but the consequences are not, we think, ones which would recommend themselves to any working physicists; not to put too fine a point on it, they would bring chaos to discussions of critical phenomena. The tractable alternative, also suggested by

<sup>76</sup> In relation to both this assertion and CP–III, the following quibble might not be out of place. The thermodynamic limit is taken for *thermodynamic functions* which are approximately extensive for large systems and become extensive in the thermodynamic limit [41]. The partition function is not of this sort, as one can see by using a little ‘reverse engineering’ to define the partition function of  $\mathfrak{S}_\infty$  as  $Z_\infty(\zeta_1, \zeta_2, N) := \exp\{-N\phi_2(\zeta_1, \zeta_2)\}$ . Apart from the retained dependence on  $N$ , a singularity, which is an infinity of  $\phi_2(\zeta_1, \zeta_2)$  would be a zero of  $Z_\infty(\zeta_1, \zeta_2, N)$ . In fact this brings to the fore a problem with CP–III. Phase transitions do not correspond to points “when the partition function has a singularity”. For (say a lattice system) the partition function is, for a finite system, a polynomial whose zeros give singularities of the free energy, none of which lies on the positive real axis. In the thermodynamic limit a phase transition corresponds to a point of accumulation of zeros on the real axis. The quibble is resolved by replacing ‘partition function’ by ‘free energy’ in CP–III and Mainwood’s assertion.

<sup>77</sup> It is convenient to take his worries in reverse order.

Mainwood, is to restrict the definition to large systems.<sup>78</sup> This would seem to us to be an inevitable step, but it also has consequences which we discuss in more detail below.

At one level both (2) and (3) are examples of the standard concern with respect to modelling, namely that we may not have a very good model which is not giving results which agree with experiment. And Mainwood's response to this is, as would be expected, that we should find a better model. But worry (3) also contains a second element, namely that his definition contains the use of a counterfactual, an idealized infinite model. His argument here is more complex and draws on a strong parallel with Lewis's [72] analysis of counterfactuals. On this basis he argues that

it is the character of [the real finite system] that determines the nature of the infinite system that we then consider. When we draw conclusions about the nature of the phase transitions, they are conclusions about the character of [the real finite system], but by reference to the infinite model we can express them in a concise and illuminating form (ibid, p. 30).

However we have worries of our own which do not seem to concern Mainwood. These can best be described by considering the transfer matrix treatment in Sect. 3.3, where, if we restrict attention to the two-dimensional square-lattice spin- $\frac{1}{2}$  Ising model in zero field, the exact critical temperature is known for the model on an infinite lattice (see Appendix 2). To apply the transfer matrix method (see Sect. 3.3), the square lattice is taken to have  $N_H$  sites in the horizontal direction and  $N_V$  sites in the vertical direction, so that  $N = N_H N_V$ . Periodic boundary conditions are applied so that the lattice forms a torus with horizontal rings of  $N_H$  sites and rings in a vertical plane of  $N_V$  sites. It is assumed that the system is large in the horizontal direction, so that, parameterized by  $N_V$ , we have a sequence of one-dimensional models of increasing complexity. Each exhibits a maximum in the heat capacity, including the simplest case  $N_V = 1$  [27, p. 166].<sup>79</sup> These maxima (although they will differ slightly for all  $N_V$  however large and finite) are taken as incipient singularities<sup>80</sup> and for increasing  $N_V$  show good agreement with the Onsager result, which is the case  $N_V = \infty$ .

However, the prescription to be applied by the Mainwood proposal is that their critical temperatures, *for all*  $N_V$ , are the Onsager value. This would seem to us to reverse the order of the way of working of physicists. We think it is probably true

<sup>78</sup> This is also discussed by Ardourel [2], who proposes to use the Lee–Yang formulation of phase transitions in terms of the zeros of the partition function to describe, and to understand, the emergence of anomalies of thermodynamic functions in terms of accumulations of Lee–Yang zeros in the vicinity of the critical temperature on the real axis in the complex temperature plane. While this does indeed provide a useful intuition, it is not substantially different from exactly solving the statistical mechanics of finite systems, and does not by itself allow us to predict the way in which anomalies approach singularities of the infinite system, when the system size is increased.

<sup>79</sup> Although we have shown by an exact renormalization group method in Sect. 3.4.3 that when  $N_V = 1$  the critical fixed point is at zero temperature.

<sup>80</sup> As we have indicated in Sect. 3.3 maxima in other response functions and also the behaviour of the ratio of the two largest eigenvalues of the transfer matrix can also be used as identifiers of incipient singularities.

to say that, with notable exceptions like Kadanoff [56–58], physicists involved in model calculations do not consider whether their interest is in very large systems or infinite systems. Their concern is whether a phase transition occurs. If they suppose that it does, one tool<sup>81</sup> to determine its location is to use transfer matrix calculations [12, 65, 99, 100, 111, 112]. The method is to determine incipient singularities for as large a vertical width of system as possible as an estimate for the transition temperature for a very large/infinite width. Here one cannot use Mainwood’s prescription to assign the infinite-width result to the finite-width systems, since the former is not known.<sup>82</sup> When, as in the case of the zero-field spin- $\frac{1}{2}$  Ising model, the infinite-width result is known exactly or has been determined to a good approximation by series methods, the motivation for determining finite-width results is to test the efficacy of the method, or to cross-check with other results.

In his discussion of Mainwood’s proposal Butterfield [21, p. 1130] states it in a more restricted form. Again using our notation this is:

**Definition 3** A phase transition occurs in  $\mathfrak{E}_N$  iff  $\mathfrak{E}_\infty$  has non-analyticities.

This Mainwood–Butterfield proposal has the advantage that it doesn’t project a result from the infinite system onto finite systems of any size (or maybe onto just large-size systems). However, given that it asserts the existence of a phase transition in a finite system of any size  $N$ , where does this occur? At the maximum of one of the response functions (heat capacity or susceptibility/compressibility), or by extraction from the behaviour of the ratio of the two largest eigenvalue of the transfer matrix? These will all give different results, as will also the results of taking the limits in different ways and for differing numbers of dimensions, all of which in turn will differ with  $N$ . If all these values are taken to be estimates of some ‘true’ value will this be  $N$ -dependent or the same for all  $N$ , including presumably  $N = \infty$ , when we would be back with the problems of Mainwood’s original proposal?

## 4 Phase Transitions in Large Systems: Our Proposal

As we shall see, our discussion in previous sections of the structure of thermodynamics and of statistical mechanics in general, and of PTCP in particular, will allow us to paint a more nuanced and quantitative picture of their relationship than that provided by previous approaches. In particular we are concerned with the role in that relationship played by large finite systems. Mainwood suggests that we ‘bite the bullet’ by countenancing the possibility of phase transitions in small systems. However, we suggest that he is proposing to bite the wrong bullet. The one which should be bitten is the need for a criterion giving a demarkation in system size between

<sup>81</sup> Among others, including duality transformations and series expansions.

<sup>82</sup> Although one can, in some cases, prove the existence of a phase transition, even if the transition temperature is not known [45–47, 103].

small systems and large systems, and our proposal, which uses the discussion of finite-size scaling in Sect. 3.4.2, is intended to encompass this need.

Thermodynamics, on the one hand, characterises PTCP in terms of singularities of thermodynamic functions, which may occur at special values of externally controllable parameters. This characterisation appears, *at first sight*, to be warranted by the phenomenology of phase transitions as they are observed in nature—apparent discontinuities of thermodynamic functions at first-order phase transitions, and apparent algebraic singularities of thermodynamic functions including divergent response functions at second-order phase transitions. In statistical mechanics, on the other hand, singularities of thermodynamic functions can emerge *only* in the limit of infinite system size. As realistic systems are clearly of finite size, this creates an internal inconsistency in the list **P–I** to **P–VI** of propositions given above, if indeed the characterisation of PTCP *as they occur in nature* in terms of singularities (that is proposition **P–IB**) is accepted.

Our aim now is to present an argument, based on the account of finite-size scaling in Sect. 3.4.2, which shows that this inconsistency can be resolved within statistical mechanics *and in a fully quantitative manner*. In Sect. 3.4.2, and also here, discussion is restricted to a system with a thermal coupling  $\theta_T$  and a magnetic coupling  $\theta_{\mathcal{H}}$ , in the cases where (i) it is fully-finite with thickness  $\aleph$  and (ii) it is fully-infinite with  $\aleph = \infty$ . In case (ii) on the zero-field axis  $\mathcal{H} = 0$ ,  $\theta_{\mathcal{H}} = 0$  there is a critical temperature  $T = T_c$  with  $\theta_T = 0$  where response functions are singular. There is no singularity in the finite system but maxima appear in the response functions. We now summarize the relevant conclusions of finite-size scaling:

- FSS–I** In the thermodynamic limit  $\aleph \rightarrow \infty$  when  $\theta_T$  is small, but not infinitesimal, the asymptotic form for the susceptibility at  $T = T_c$ , given by (91), has a singular component with exponent  $\gamma$ , but amplitudes which, by virtue of the presence of an irrelevant field  $\theta_*$ , are dependent on  $\theta_T$ .
- FSS–II** As  $\theta_T \rightarrow 0$ , the influence of  $\theta_*$  becomes negligible and the susceptibility exhibits a pure power-law singularity at  $T = T_c$  as described by (93).
- FSS–III** When  $\aleph$  is finite there is no singular behavior and two temperatures are defined:<sup>83</sup> the shift temperature  $\hat{T}(\aleph)$  where the susceptibility has a maximum and the rounding temperature  $\hat{T}(\aleph)$  at which the profile of the susceptibility in the finite system begins to diverge from that in the infinite system.
- FSS–IV** Assuming, as in (85) and (86), that  $|T_c - \hat{T}(\aleph)| \sim \mathcal{O}(\aleph^{-\chi})$  and  $|\hat{T}(\aleph) - \tilde{T}(\aleph)| \sim \mathcal{O}(\aleph^{-\tau})$ , it can be shown that the shift exponent  $\chi = [\nu(1 - y_*)]^{-1}$  and the rounding exponent  $\tau = \nu^{-1}$ ; that is that the rate of convergence of both the incipient singularity and the range of influence of finite-size effects around the incipient singularity are determined by exponents present *in the infinite system*.

<sup>83</sup> Each will, of course, depend of the particular response function under consideration.

This renormalization group scaling approach to the description of critical phenomena thus explains in a quantitative way, how singularities that might occur in infinite systems are smoothed out by finite-size effects. This, being fully in line with the fundamental observation that statistical mechanical systems of finite size cannot exhibit any singularities, resolves the inconsistency in the list of propositions **P–I** to **P–VI**. In particular **FSS–IV** gives a *quantitative measure* of the deviations of critical phenomena, as observed in finite systems, from the behaviour expected for infinite system size. From (89), deviations from critical behaviour characteristic of the infinite system will be observable in a narrow region around the infinite system critical point. This, however, is precisely the region, where one would stand the chance of observing *asymptotic singular behaviour*, as only in this region is the influence of irrelevant scaling fields on PCTP expected to be sufficiently small. In order to observe asymptotic critical singularities it is thus required that  $|\theta_T|$  be *sufficiently small* to keep corrections to asymptotic critical singularities due to irrelevant scaling fields under control, *but also not too small*, in order to prevent finite-size corrections from becoming significant. As the range of  $\theta_T$  within which finite-size corrections dominate critical behaviour shrinks with system size  $\aleph$  like  $\aleph^{-1/\nu}$ , one has to choose systems *sufficiently large in a quantitatively well-defined sense* in order to be able to observe asymptotic critical singularities characteristic of the respective universality class of a system.

In the context of the list **P–I** to **P–VI** of propositions, it is important to realise that the characterisation of PTCP in terms of singularities of thermodynamic functions constitutes an *extrapolation of empirical observations*, as properly establishing the existence of a discontinuity of a thermodynamic function would require experimental control of infinite precision, while establishing a divergence of a response function would require an actual measurement of an infinite quantity. Neither requirement can conceivably be met in any realistic experiment. Given that realistic systems contain  $\mathcal{O}(10^{23})$  constituents, the linear dimension  $\aleph$  of such systems, measured in terms of atomic distances, is very large and the temperature range over which finite-size corrections to singular behaviour would manifest themselves, will be very small. It is thus understandable that such effects have been beyond experimental resolution.<sup>84</sup> On the other hand, in computer simulations of statistical mechanical systems, one can handle only relatively small systems, and finite-size roundings of critical singularities are therefore quite prominent. In such situations such roundings, as predicted (and captured) by finite-size scaling are indeed observed and routinely used to extract asymptotic critical exponents from finite-size data [14]. The renormalization group and its formulation of finite-size scaling theory thus predicts in a quantitative way, both, the emergence of critical singularities, described as pure power-law singularities sufficiently close to an infinite system critical point,<sup>85</sup> and their shifting and rounding in systems of finite size.

<sup>84</sup> Except for fairly recently in thin films [73, 124].

<sup>85</sup> For the sake of completeness, it should be mentioned that under certain well understood conditions, logarithmic corrections to pure power laws can occur [125]. They translate into analogous logarithmic corrections in finite-size scaling relations [82].

According to our definition of an incipient singularity (Definition 1, above) such will occur in a finite system at certain values of their external parameters, if at those values thermodynamic functions exhibit properties that have no finite limits as the system size  $\aleph$  is increased. This could be a steep increase in the the slope of magnetization as a function of the external field across the zero-field axis at low temperatures, as shown in Fig. 9, which is indicative of the possibility of a first-order transition in the infinite system. Or it could be the size-dependent height of the maximum of a response function as shown in (90) with  $\omega > 0$ , which is indicative of the possibility of a second-order transition in the infinite system. However it is important to note that an assertion of the occurrence of a incipient singularity in a finite system can never be made with absolute certainty by looking at the behaviour of a single system of any fixed finite size, but *only by comparing the behaviour of systems of different sizes*. That said, our investigations have now provided us with a well-defined notion of a large system:

**Definition 4** *For a system to be counted as large* it must be big enough to exhibit a range of values of a thermodynamic variable (for example, the temperature) within which the following two phenomena can both be avoided:

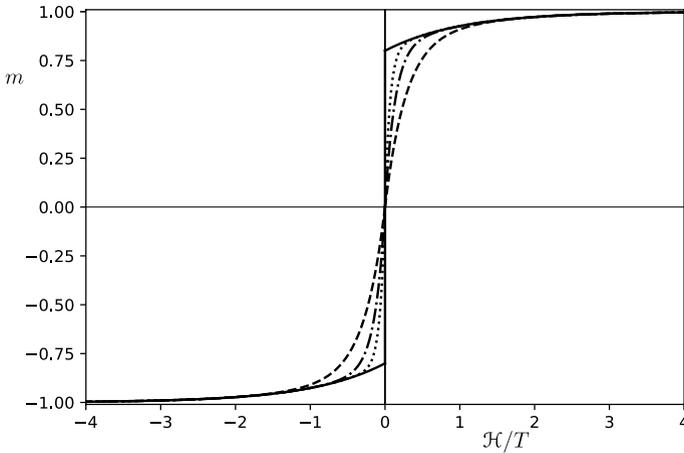
- (i) the corrections to scaling (due to the existence of non-zero irrelevant scaling fields) which require the system to be close to an incipient singularity,
- (ii) the noticeable finite-size corrections in a close neighbourhood of an incipient singularity (due to a finite value of  $\aleph$ ), which requires the system to be sufficiently far away from an incipient singularity.

Although, as we saw above, these two conditions pull in opposite directions this tension will become less acute as the system size increases. For such systems incipient singularities will be observable in a range of temperatures (or couplings), which are described by the asymptotic critical exponents of infinite systems. These exponents describe incipient singularities which will never fully materialize in a system of finite extent. They do, however, provide an economy of description, and lead to a classification of systems according to their universality class, as described earlier. Quite often the full complexity of the crossover between behaviour described by asymptotic critical exponents and finite-size rounding of thermodynamic functions is far beyond the capabilities of available analytic tools. Taking the thermodynamic limit in a statistical mechanical analysis of a system is also often,<sup>86</sup> the only way to carry the calculation through to its end.

The renormalization group approach to PTCP actually plays a dual role in the analysis of critical phenomena.<sup>87</sup> On the one hand it provides micro-reductive methods, firmly embedded in the arsenal of techniques of statistical mechanics, to evaluate critical exponents for given statistical mechanical systems, albeit in most cases only approximately. On the other hand it embodies a new way of looking at such

<sup>86</sup> As in the case of the Onsager [98] solution of the zero-field two-dimensional Ising model.

<sup>87</sup> This aspect is also highlighted in Hüttemann et al. [50].



**Fig. 9** Isothermal curves of magnetization density plotted against the field coupling. System size increases from the broken to the chain to the dotted curves with the infinite system represented by the continuous line

systems, by describing statistical properties of systems at different length scales. It is this radically new way of analysing systems which allows it to put *systems with different microscopic properties into a common context*, which in turn leads to the identification of fixed points and their basins of attraction as universality classes, thereby revolutionizing the analysis of critical phenomena.

It is perhaps appropriate to add a final twist. Asymptotic critical exponents characterising singularities at phase transitions as they would occur in infinite systems, including exponents that describe corrections to scaling due to irrelevant scaling fields, are obtained from the eigenvalues of a renormalisation group transformation that is linearized in the vicinity of (one of) its fixed points. They are thus obtainable *without ever touching or contemplating systems of infinite size!* As we have seen in our discussion above, these critical exponents also govern the way in which finite-size corrections to critical phenomena manifest themselves. In some sense, therefore, it would be fair to say that *critical exponents are bona-fide properties of finite systems*—rather than, as mostly discussed, simply properties of potentially infinite systems.

The aim of our analysis has been to eliminate some of the confusion that has characterised much of the discussion surrounding PTCP in the philosophical (and physics) literature. To summarize our position:

- It cannot be denied that phase transitions occur in nature. (**P-IA** is accepted).
- The assertion that they are characterized by singularities is an unwarranted extrapolation of empirical findings. (**P-IB** is rejected). (Asserting the existence of a singularity in an experimental result requires infinitely precise experimental control, or an actual ‘measurement of the infinite’, which is clearly infeasible.)
- Within thermodynamics, there is *no choice* but to describe phase transitions in terms of singularities. (That is, **P-IIA** and **P-IIB** are valid statements about the

structure of thermodynamics). Equations of state either have unique solutions – in which case there is no phase transition – or they may exhibit bifurcations in their solution manifolds, in which case singularities and discontinuities arise.

- Phase transitions, as they occur in nature, are correctly described by statistical mechanics, the renormalization group and finite-size scaling. Thermodynamics, on the contrary, is fundamentally incapable of an adequate description as it is, from the outset, conceived as a theory of infinitely large systems. (**P-III A** is accepted but **P-III B** is rejected).
- Investigating systems in the limit of infinite system size provides added value in that it allows one to (i) identify exact asymptotic power laws, which the incipient singularities would follow if system sizes could be taken arbitrarily large, (ii) provide a classification of systems according to their universality class.

## 5 After-Thoughts on Reduction and Emergence

Figure 1 is a diagrammatic attempt to encapsulate the relationship between thermodynamics including scaling theory, and the Gibbsian version of statistical mechanics including various approaches to PTCP: the use of the thermodynamic limit, the renormalization group and phase transitions in finite systems. Apart from the formal links spelled out as messages FSM-1, ..., FSM-4 from statistical mechanics to thermodynamics and the connecting relationships FTD-1, ..., FTD-3, provided by thermodynamics to statistical mechanics, there is another element of collaborative interaction, as shown in Fig. 1, in the direction from statistical mechanics to thermodynamics; specifically from the renormalization group to scaling theory. This has two aspects *substantiation and enrichment*:

- The Kadanoff scaling relationship (25) is introduced as a hypothesis, which is *substantiated* as (82) in renormalization group theory.
- Scaling about an arbitrary origin in a critical region with relevant and irrelevant directions is a consequence of scaling theory. This picture is *enriched* in renormalization group theory, where scaling origins are not arbitrary, but fixed points of the flow determined by the recurrence relationships, and corresponding to different universality classes. Relevant and irrelevant directions correspond to directions in which a fixed point is repulsive and attractive to the flow. Following a trajectory as it approaches one fixed point, but is finally repulsed towards another, is an example of crossover between different types of critical behaviour, that is between different universality classes.

Having spelled out a picture of the anatomy of thermodynamics and statistical mechanics, as well as the relationships between their different parts, we can now ask what consequences this has for our understanding of reduction and emergence as regards PTCP. The literature on reduction and emergence is vast, even when restricted to the specific area of PTCP. So reviewing and discussing this entire

literature is beyond the scope of this work,<sup>88</sup> with a more extensive treatment being the subject of a future paper. Our aim in this section is simply to sketch the main contours of the lie of the land in the light of the picture we have developed, hoping that this will serve as a springboard for further discussions.

To aid our account, we introduce the following terminology. Let  $\mathfrak{T}_C$  and  $\mathfrak{T}_F$  be two theories, where ‘C’ stands for ‘coarse’, meaning less detailed, and ‘F’ stands for ‘finer’, meaning more detailed.<sup>89</sup> Intuitively,  $\mathfrak{T}_C$  is the theory that is supposed to be reduced to  $\mathfrak{T}_F$ . In the terminology that has become standard in the philosophical literature on the topic,  $\mathfrak{T}_F$  is supposed to be the *reducing theory* and  $\mathfrak{T}_C$  is supposed to be the *reduced theory*. We say ‘supposed to be’ because this is what reductionists would expect. The question is whether this expectation bears out, and if so in what sense of reduction.

Accounts of reduction might be divided into two broad families, called *limit reduction* and *deductive reduction*.<sup>90</sup> We now have a look at each in turn and consider whether they can account for the relation between  $\mathfrak{T}_C$  and  $\mathfrak{T}_F$  that emerges from our account.

## 5.1 Limit Reduction

The core idea of limit reduction is that  $\mathfrak{T}_C$  reduces to  $\mathfrak{T}_F$  if the former turns out to be a regular limit of the latter. An example of such a reduction is letting the parameter  $c$ , the speed of light in the special theory of relativity, tend toward infinity and thereby recovering classical Newtonian mechanics [89].<sup>91</sup> In general, let us call the relevant parameter  $\alpha$ ; the limit, denoted as  $\lim_\alpha$ , can be toward any value of  $\alpha$ , the most frequent cases being  $\alpha \rightarrow 0$  and  $\alpha \rightarrow \infty$ . Batterman [10] adds the further requirement that the limit be regular, which means that the relevant formulae in  $\mathfrak{T}_F$  approach the relevant formulae in  $\mathfrak{T}_C$  smoothly as the parameter approach the relevant limit value.<sup>92</sup> Taking these elements together yields the following:

### Definition 5 Limit Reduction

$\mathfrak{T}_C$  *limit-reduces* to  $\mathfrak{T}_F$  iff  $\lim_\alpha \mathfrak{T}_F = \mathfrak{T}_C$  and the limit is *regular*.

<sup>88</sup> For surveys of the various positions in the discussion concerning reduction see Hüttemann and Love [49] and Van Riel and Van Gulick [123]. For surveys of the discussions of emergence see Humphreys [48] and the contributions of Gibb et al. [38]. For an overview of discussions of phase transitions see Shech [116].

<sup>89</sup> Butterfield [20, p. 928] replaces ‘C’ by ‘t’ to stand for top, tangible or tainted and ‘F’ by ‘b’ to stand for bottom, basic or best.

<sup>90</sup> The distinction goes back to Nickles [89]. Batterman [10] calls them the “physicist’s sense of reduction” and the “philosopher’s sense of reduction”. We avoid this terminology because, as we will argue, physicists do present deductive reductions.

<sup>91</sup> We note that the terminology of what reduces to what varies. Nickles says that in taking the limit  $\mathfrak{T}_F$  reduces to  $\mathfrak{T}_C$ ; in keeping with the terminology previously introduced, we say that  $\mathfrak{T}_C$  reduces to  $\mathfrak{T}_F$ . Nothing in what follows depends on this purely terminological matter.

<sup>92</sup> See Berry [13] and for discussions of singular and regular limits see Butterfield [21] and Nguyen and Frigg [88].

This definition plays an important role in the discussion about the reduction of PTCP because  $\mathfrak{T}_C$  is commonly associated with thermodynamics and  $\mathfrak{T}_F$  with statistical mechanics. The failure of the limit to be regular as the number of microsystems tends to infinity is then seen as an indication that reduction fails.

How does this argument play out in our scheme? To answer this question we first need to identify certain elements in Fig. 1 with  $\mathfrak{T}_C$  and  $\mathfrak{T}_F$ . There are two possibilities:

- (a) Work within the renormalization group SM4. In this case, as described in Sect. 3.4.1, the limiting process is implemented by the renormalization transformation which applies a succession of reductions in the number of lattice sites. This reduces the fluctuations (and correlation length) away from critical regions, but leaves the essential statistical mechanical structure intact.
- (b) Apply the infinite system limit SM2  $\rightarrow$  SM3. Away from critical regions this removes fluctuations in the uncontrolled extensive variables, but leaves the microstructure and the probability distribution intact.

However, neither of these is a reduction to a version of thermodynamics. Both (a) and (b) are procedures lying entirely *within* statistical mechanics. That having been said, (b) is probably the closest to the above idea of reduction. However, while it uses the thermodynamic limit, that limit does not take the system to a thermodynamic system, but to an *infinite statistical mechanical system* (SM3). To arrive at thermodynamics it is necessary to *conflate* SM3 with TD3. While SM3 like TD3 contains the singular characteristics deemed necessary (by some) for the occurrence of phase transitions it also has a microstructure which is lacking in TD3.

So there is no part of Fig. 1 which involves the kind of limit that would ground a limit reduction. However, far from being a problem, this is simply irrelevant to the issue of the reduction of PTCP. As we have indicated in Sect. 4 the role of the thermodynamic limit is, in the first instance, to provide a condition for maxima in response functions to be incipient singularities; some finite systems do not show PTCP no matter how large they become. In the second instance it provides the critical exponents that can be regarded as properties of the real system. Limits and renormalization group techniques are classification tools that enable us to separate phase transitions into different universality classes.

## 5.2 Deductive Reduction

This notion of reduction is closely associated with Nagel. The broad idea is that  $\mathfrak{T}_C$  is reduced to  $\mathfrak{T}_F$  if the laws of  $\mathfrak{T}_C$  are deducible from the laws of  $\mathfrak{T}_F$  and some auxiliary assumptions. A mature formulation of this idea, known as the *Generalised Nagel-Schaffner Model of Reduction*, is as follows:<sup>93</sup>

<sup>93</sup> The original reference is the first (1961) edition of Nagel [85], of which Schaffner [114] provides a reformulation. An alternative account by Butterfield [20, 21] uses the notion of a definitional extension. Our presentation here follows that of Dizadji-Bahmani et al. [26].

**Definition 6 Deductive Reduction**

$\mathfrak{T}_C$  reduces to  $\mathfrak{T}_F$  iff there is a *corrected* version  $\mathfrak{T}_C^*$  of  $\mathfrak{T}_C$  such that:

- (i) *Connectability*: If  $\mathfrak{T}_C$  contains terms that do not appear in  $\mathfrak{T}_F$ , then for every such term there is a bridge law connecting it to a term in  $\mathfrak{T}_F$ .
- (ii) *Derivability*: Given the associations in (i),  $\mathfrak{T}_C^*$  is derivable from  $\mathfrak{T}_F$  plus bridge laws and, possibly, some auxiliary assumptions.
- (iii) *Strong analogy*:  $\mathfrak{T}_C$  and  $\mathfrak{T}_C^*$  are *strongly analogous* to one another.

As a simple example, consider the derivation of the perfect gas law  $PV = NT$  (given as the second of equations (8)) from the kinetic theory of gases. Here the perfect gas law is  $\mathfrak{T}_C$  and the kinetic theory is  $\mathfrak{T}_F$ .  $\mathfrak{T}_C$  contains the term ‘temperature’, which is not in  $\mathfrak{T}_F$ . The bridge law  $T := 2U\epsilon/(3N)$  (which is the first of equations (8)) with the internal energy  $U$  identified as the expectation value of the kinetic energy of the gas, connects this term to  $\mathfrak{T}_F$ .  $\mathfrak{T}_C^*$  is the version of the perfect gas law in which, subject to the physical constraints on the system,  $P$ ,  $V$ , and  $T$  are variables that can fluctuate (something they cannot do in  $\mathfrak{T}_C$ ).  $\mathfrak{T}_C^*$  and  $\mathfrak{T}_C$  are strongly analogous in that fluctuations are small (to the point of being negligible) in contexts in which  $\mathfrak{T}_C$  is applied.

The introduction of  $\mathfrak{T}_C^*$  is a concession to practice. Ideally one would be able to derive  $\mathfrak{T}_C$  from  $\mathfrak{T}_F$ , but that is usually not possible. So one rests content with deriving a theory  $\mathfrak{T}_C^*$  that is not identical with, but strongly analogous to,  $\mathfrak{T}_C$ . What does it mean to be strongly analogous? Schaffner [114] blocks the worry that an appeal to strong analogy is an entry ticket to ‘anything goes’ by imposing the following two conditions:

- (a)  $\mathfrak{T}_C^*$  corrects  $\mathfrak{T}_C$  in that  $\mathfrak{T}_C^*$  makes *more accurate predictions* than  $\mathfrak{T}_C$ .
- (b)  $\mathfrak{T}_C$  is *explained* by  $\mathfrak{T}_F$  through  $\mathfrak{T}_C^*$  being a deductive consequence of  $\mathfrak{T}_F$  and  $\mathfrak{T}_C^*$  being strongly analogous to  $\mathfrak{T}_C$ .

With this in place, we can now ask whether the above schema indicates that a deductive reduction is taking place. For this we first need to know which theories are in play: what is reduced to what? Since we are interested in a reduction of PTCP, we should focus on a version of thermodynamics with PTCP in it. So we set  $\mathfrak{T}_C := \text{TD3}$ . Then it might seem tempting to choose **SM3** as the reducing theory because, in Fig. 1, **TD3** ‘communicates’ with **SM3**. This, however, would be the wrong choice. What we are interested in is the reduction of thermodynamics to a fundamental theory of large systems, and this is **SM2**. This is because **SM2** contains the fundamental principles of statistical mechanics with the only added assumption being that systems are large; so  $\mathfrak{T}_F := \text{SM2}$  is appropriate. **SM3**, by contrast, contains a limit assumption which does not belong in the fundamental theory. So the task we set ourselves here is to check whether the reduction of **TD3** to **SM2** fits the mould of deductive reduction. We shall argue that it does and, to this end, we now consider this contention in relation how to the elements (i)–(iii) of Def. 6 play out in Fig. 1.

- For (i)** connectivity requires a number of bridge laws. We have avoided this designation for the relationships FTD–1, FTD–2 and FTD–3 in Fig. 1, preferring to call them ‘inter-theory connections’. However, now we shall consider the possibility that they can assume the role of bridge laws as required in the present context. The paradigmatic example of a bridge law in the philosophical literature is provided, as indicated above, by the perfect gas. There the bridge law identifies the temperature in statistical mechanics using the underlying identification of the expectation value of kinetic energy of the gas with its internal energy. But, on closer examination, this example glosses over two other identifications, of volume and pressure.<sup>94</sup> In a perfect gas contained in a cylinder closed by a movable piston, the piston position will fluctuate; that is to say, from a statistical mechanical point of view, the volume of the gas is a fluctuating quantity. So, just as the internal energy must be identified with the expectation value of the kinetic energy, the thermodynamic volume must be identified with the expectation value of the statistical mechanical volume. Other instances of the same kind are provided by other systems and they are all covered by FTD–3, which in the current context plays the role of a bridge law. In the case of the perfect fluid the identification of internal energy and the expectation value of the kinetic energy and of the thermodynamic volume with the expectation value of the statistical mechanical volume is sufficient to provide a bridge for temperature, pressure and for entropy via the Sackur-Tetrode formula and consequentially for all other thermodynamic variables, as described by the connecting relationships FTD–1 and FTD–2. These could, therefore, be regarded as consequences of the underlying bridge law FTD–3, rather than as bridge laws in their own right. In more complicated situations, where there is a need to connect a larger set of thermodynamic and statistical mechanical variables, it is a reasonable economy to regard them, together with FTD–3 as comprising an exhaustive set of bridge laws.
- For (ii)** by definition  $\mathfrak{T}_C^*$  is a corrected version of  $\mathfrak{T}_C$  that can be derived from  $\mathfrak{T}_F$  plus bridge laws. In the current context  $\mathfrak{T}_C^*$  is a version of TD3 in which the relevant quantities are allowed to fluctuate, and the fluctuations show roughly the pattern given in SM2 (but without  $\mathfrak{T}_C^*$  containing any of the microstructure of matter specified in statistical mechanics). It is obvious that  $\mathfrak{T}_C^*$  thus defined is a deductive consequence of SM2: it is obtained simply by applying the bridge laws to SM2.<sup>95</sup>
- For (iii)** we need to show that  $\mathfrak{T}_C^*$  and  $\mathfrak{T}_C$  stand in the proper strong analogy relationship. In effect the derivation of SM3 from SM2 through the

<sup>94</sup> And also of the number of particles of the gas if it is enclosed in a permeable container.

<sup>95</sup> Terminological note: the term ‘corrected’, which is customary in the discussion of reduction, is somewhat ill-chosen, because it might suggest that that TD3 is in some way faulty, which it is not. It is in fact one of the most successful and enduring models in physics. The term ‘corrected’ here should be read in a unemphatic (and non-pejorative) way, as simply indicating that conditions (a) and (b), listed above, are met.

thermodynamic limit and the fact that SM3 corresponds to TD3 amounts to saying that there is a strong analogy between SM2 and TD3. However, a more detailed analysis is useful and for this we check whether Schafner's two criteria are satisfied:

- For (a) the messages FSM–1 and FSM–2 are relevant. FSM–1 asserts that uncontrolled thermodynamic variables fluctuate with variances of  $\mathcal{O}(N)$  related to response functions. This means that the variances of the corresponding density variables are  $\mathcal{O}(1/N)$ . That these fluctuations are small for large systems is related to, but not exactly equivalent to the fact, asserted in FSM–2, that extensivity is an approximate property of large systems. So  $\mathfrak{Z}_C^*$  modifies  $\mathfrak{Z}_C$  by replacing equality in the basic relationship with approximate equality, valid when the system is large. It also contains fluctuation–response function relationships between fluctuations, which are recognised in  $\mathfrak{Z}_C^*$  but not in  $\mathfrak{Z}_C$ , and response functions which appear in both. Thus  $\mathfrak{Z}_C^*$  makes more adequate predictions than  $\mathfrak{Z}_C$  because real systems *do* show fluctuations.
- For (b) the way that  $\mathfrak{Z}_C := \text{TD3}$  is explained by  $\mathfrak{Z}_F := \text{SM2}$  follows straightforwardly from Sect. 4 once the bridge laws are accepted and we have in place the definition of an incipient singularity (Def. 1). Maxima in response functions are identified as incipient singularities if they map into real singularities in the thermodynamic limit, which is the step from SM2 to SM3. And, as we have already noted, TD3 communicates with SM3 in the sense that it communicates its understanding of the singularities in SM3 to TD3.

From the above we conclude, that TD3 reduces to SM2 in the sense of deductive reduction. However, the structure of Fig. 1 prompts a consideration of the possibility of further reductive relationships higher in the figure. In particular does  $\mathfrak{Z}_C := \text{TD4}$  and  $\mathfrak{Z}_F := \text{SM4}$  satisfy the required conditions?<sup>96</sup> It is straightforward to see that connectability and derivability, where  $\mathfrak{Z}_C^*$  is a version of TD4 that has certain of the features of SM4 built into it, are satisfied as before. Scaling in TD4 is a phenomenological means of capturing the structure of the way thermodynamic functions in critical regions depend on variables (in the form of homogeneous functions of controllable variables). It can in a sense be regarded as being built from renormalization group theory with the scaffolding removed. This is what we referred to above as the substantiation of scaling theory by the renormalization group. On the other hand the *values* of critical exponents and the interpretation of the origin of scaling as the fixed point of a semi-group transformation is absent from TD4 but present in SM4. In that sense the later provides an *explanation* or *enrichment* of the former.

<sup>96</sup> Replacing SM4 with SM5, would rather complicate the situation, since TD4 has extensivity in all  $d$  dimensions, whereas this is the case for only  $\mathfrak{d}$  dimensions (which includes the fully-finite case  $\mathfrak{d} = 0$ ) in SM5.

### 5.3 Emergence

In the case of emergence things are even more difficult than with reduction. As Humphreys notes in a recent review of the field, not only is there no unified framework or account of emergence, there is not even a generally agreed set of core examples of emergent phenomena on which a discussion could build [48]. Our aim here is not, therefore, to comprehensively review the field; we rather discuss some senses of emergence that have played a role in the debate and assess whether, in the light of our analysis, PTCP are emergent in these senses.

For Butterfield, whose view of reduction is essentially Nagelian, there is no conflict between reduction and emergence. The view that reduction and emergence are compatible is based on an understanding of emergence as there being “properties or behaviour of a system which are *novel* and *robust* relative to some appropriate comparison class” ([20, p. 921], orig. emph.). He adds the comment that this is intended to cover the case where a system consists of parts, where the idea is that a composite system’s “properties and behaviour are novel and robust compared to those of its component systems, especially its microscopic or even atomic component” (op. cit.). We agree that thus understood, there is emergence in the large but finite systems we are studying and PTCP can be regarded as both emergent and reduced. Illustrative of this is the transfer matrix approach where maxima in response functions and the correlation length (or critical properties if  $\mathfrak{d} > d_{LC}$ ), calculated for a lattice which is infinite in  $\mathfrak{d}$  dimensions, converge towards the critical properties of a  $(\mathfrak{d} + 1)$ -dimensional system as the size in that dimension is increased. This account, affords a understanding of dimensional crossover between universality classes, with the ‘gradual emergence’ of critical behavior.

Humphreys [48] introduces the triplet of conceptual emergence, ontological emergence and epistemological emergence, which we now consider:

- (1) We have *conceptual emergence* “when a reconceptualization of the objects and properties of some domain is required in order for effective representation, prediction, and explanation to take place” (op. cit. p. 762). This is close to Butterfield’s notion of reduction, and there is emergence in this sense because various notions that are not native to statistical mechanics, have been introduced into the theory in order to deal with PTCP, both through inputs from thermodynamics (FTD-1, FTD-2 and FTD-3) and through the introduction of the notion of a large system at level SM2. As we have argued in Sect. 4 and in our discussion of transfer matrix methods, it is precisely in such large systems that PTCP are manifested in the form of incipient singularities.
- (2) *Ontological emergence* amounts to the following: “A ontologically emerges from B when the totality of objects, properties, and laws present in B are insufficient to determine A” (op. cit. p. 762). As we have seen in Sect. 4, the properties of a system’s micro-constituents together with the laws that govern them are sufficient to determine PTCP; in fact they can be shown to happen in finite systems. So PTCPs are not ontologically emergent.
- (3) *Epistemological emergence* is present when the limitations in our knowledge prevent us from predicting the relevant phenomenon. As Humphreys puts it, A

epistemically emerges from  $B$  “when full knowledge of the domain to which  $B$  belongs is insufficient to allow a prediction of  $A$  at the time associated with  $B$ ” (op. cit. p. 762). This is also the notion of emergence that Morrison appeals to when she notes that “what is truly significant about emergent phenomena is that we cannot appeal to microstructures in explaining or predicting these phenomena, even though they are constituted by them” [84, p. 143].<sup>97</sup> We submit that PTCP are not epistemically emergent because, as we have seen in Sect. 4, they in fact can be deduced and predicted from the underlying micro-theory. What is important here is PTCP appear in *finite* systems.

Batterman’s account of emergence [7], centres around the application of the renormalization group. As we have seen in Sect. 3.5.2 he (and Kadanoff) regard the use of renormalization group as a wholly different type of approach to PTCP from which novel properties emerge. In particular the fixed points of the renormalization transformation which allocate the universality classes. We agree with this except for two reservations:

- (i) Batterman takes the thermodynamic limit as an essential feature of this method. As we have indicated in Sect. 3.5.2 we do not regard this as being necessary.
- (ii) There is nothing automatic about setting up a renormalization group analysis of a system. It does not arise in a straightforward algorithmic way from the basic structure of statistical mechanics. Indeed *physical insight* is required both in the choice of the lattice scaling  $\mathcal{N} \rightarrow \tilde{\mathcal{N}}$  and of the weight function. These must be compatible with the nature of the ordered state and the critical phenomena to be explored. The recurrence relationships are determined by these choices, and the fixed points ‘emerge’ as properties of the recurrence relationships. These in turn have exponents which give the universality classes of the various critical regions. As we have already indicated, most renormalization schemes involve some degree of approximation, with a consequent variation in fixed points and their exponents.<sup>98</sup> However, weight-function dependent variations can also occur even when no approximation is involved. An example of this is the one-dimensional Ising model with the scheme described in Sect. 3.4.3 with  $\lambda = 2$ , but with  $J < 0$ , that is the antiferromagnetic case. In principle one expects a fixed point associated with antiferromagnetism, but, although the free-energy density is correctly computed the fixed point is missing. For this to appear, as is shown by Nelson and Fisher [87], one needs to take  $\lambda = 3$ ; that is blocks of three sites. That, in general, different fixed points and hence different universality classes emerge from different choices

<sup>97</sup> For detailed discussion of Morrison’s position see Hüttemann et al. [50].

<sup>98</sup> An example of such variations, in the case of the Ising model on a triangular lattice, for the exponents  $y_T$  and  $y_H$  and for the location of the fixed point (the Curie temperature) is provided by Table IV on page 482 of Niemeijer and van Leeuwen [92].

of lattice scaling and weight function *for the same system* means that this is a qualified type of emergence.

Finally, emergence is often characterised as the failure of reduction [59, p. 21]. That is, reduction and emergence are taken to be mutually exclusive and a property is emergent only if it fails to be reducible. PTCP are not emergent in this sense because, as we have seen above, they are reducible in the sense of a deductive reduction.

## 6 Conclusions

We have presented a picture of the way that thermodynamics and statistical mechanics coexist and collaborate within the envelope of thermal physics. We showed that the relationship between the two developments, represented by the columns in Fig. 1 depends, on the one hand, on inter-theory connecting relationships from thermodynamics to statistical mechanics, one of which, FTD-3, can, in the context of deductive reduction be regarded as a bridge law, with the remaining two, FTD-1 and FTD-2, being consequences of FTD-3. On the other hand, from statistical mechanics to thermodynamics, there is also a sequence of ‘messages’ that are effectively warnings about the idealized nature of thermodynamics.

We address the problem that real systems are finite, and singular behaviour associated with PTCP can occur only in infinite systems, using finite-size scaling and a clear specification of a large system. This enables us to develop a picture of the way that PTCP in finite systems can be defined in terms of incipient singularities. Within this picture the role of the infinite system is threefold: (a) the existence of a critical region in the thermodynamic limit is a necessary condition for there to be a region of incipient singularity in the real finite system, (b) as one (but not the only) way to determine quantitative properties like the value of critical exponents of the real system (c) to simplify calculations. In these senses the infinite system is an indispensable, idealized approximation to the real finite system.

The usual arguments for limit reduction are based on an unwarranted conflation between a thermodynamic system with critical behaviour (TD3) and an infinite statistical mechanical system (SM3). On the other hand, the arguments for the deductive reduction of TD3 to the statistical mechanics of a large system (SM2) are valid. Next we argue that PTCP are neither ontologically or epistemologically emergent, but they are conceptually emergent. Rather less frequently remarked upon are the ways that statistical mechanics both substantiates and enriches the picture of PTCP in thermodynamics.

## Appendix 1

### Response Functions and Critical Exponents

In terms of densities and fields the response functions are

$$c_x := k_B T \left( \frac{\partial s}{\partial T} \right)_x, \quad c_\xi := k_B T \left( \frac{\partial s}{\partial T} \right)_\xi, \quad \varphi_T := T \left( \frac{\partial x}{\partial \xi} \right)_T, \quad \alpha_\xi := T \left( \frac{\partial x}{\partial T} \right)_\xi. \tag{102}$$

And in the coupling–density representation the densities are given by

$$u = \frac{\partial \phi_2}{\partial \zeta_1}, \quad x = -\frac{\partial \phi_2}{\partial \zeta_2} \tag{103}$$

with the response functions by

$$\begin{aligned} c_\xi &= -k_B \left( \zeta_1^2 \frac{\partial^2 \phi_2}{\partial \zeta_1^2} + 2\zeta_1 \zeta_2 \frac{\partial^2 \phi_2}{\partial \zeta_1 \partial \zeta_2} + \zeta_2^2 \frac{\partial^2 \phi_2}{\partial \zeta_2^2} \right), \\ \varphi_T &= -\frac{\partial^2 \phi_2}{\partial \zeta_2^2}, \quad \alpha_\xi = \zeta_1 \frac{\partial^2 \phi_2}{\partial \zeta_2 \partial \zeta_1} + \zeta_2 \frac{\partial^2 \phi_2}{\partial \zeta_2^2}. \end{aligned} \tag{104}$$

The simplest way to treat  $c_x$  is to use the standard formula

$$c_\xi - c_x = k_B \alpha_\xi^2 / \varphi_T. \tag{105}$$

In the standardized notation of PTCP [18] the critical exponents  $\alpha, \alpha', \beta, \delta, \gamma$  and  $\gamma'$  are defined by

$$c_x \sim \begin{cases} (T - T_c)^{-\alpha}, & \text{along the critical isochore, } T > T_c, \\ (T_c - T)^{-\alpha'}, & \text{along the coexistence curve, } T < T_c, \end{cases} \tag{106}$$

$$x - x_c \sim (T_c - T)^\beta, \text{ along the coexistence curve, } T < T_c, \tag{107}$$

$$\varphi_T \sim \begin{cases} (T - T_c)^{-\gamma}, & \text{along the critical isochore, } T > T_c, \\ (T_c - T)^{-\gamma'}, & \text{along the coexistence curve, } T < T_c, \end{cases} \tag{108}$$

$$\xi - \xi_c \sim (x - x_c) |x - x_c|^{\delta-1}, \text{ along the critical isotherm.} \tag{109}$$

It also convenient to define the exponents  $\sigma$  and  $\sigma'$  according to

$$c_\xi \sim \begin{cases} (T - T_c)^{-\sigma}, & \text{along the critical isochore, } T > T_c, \\ (T_c - T)^{-\sigma'}, & \text{along the coexistence curve, } T < T_c. \end{cases} \tag{110}$$

In addition to these purely thermodynamic critical exponents three more exponents  $\nu, \nu'$  and  $\eta$  arise, from statistical mechanics, for the pair correlation function and correlation length defined in Sect. 3.2. For the correlation length the exponents  $\nu$  and  $\nu'$  are given by

$$\xi \sim \begin{cases} (T - T_c)^{-\nu}, & \text{along the critical isochore, } T > T_c, \\ (T_c - T)^{-\nu'}, & \text{along the coexistence curve, } T < T_c. \end{cases} \tag{111}$$

which encapsulates the asymptotic behaviour of the correlation length in a neighbourhood of that critical point. The situation for the correlation function is rather more complicated since we are concerned not only with its dependence on the couplings near to a critical region but also on its asymptotic form for a pair of widely separated lattice sites. However, the result

$$\Gamma(\vec{r}; \zeta_1, \zeta_2) = \frac{f_d(|\vec{r}|/\xi)}{|\vec{r}|^{d-2-\eta}}, \tag{112}$$

from Ginzburg–Landau theory (see, for example, [67], Chap. 5) in which dependence on the couplings is mediated through the correlation length is believed to have wide applicability.

## Appendix 2

### The Ising Model

For simplicity we consider a  $d$ -dimensional hypercubic lattice  $\mathcal{N}_d$  of  $N$  sites with periodic boundary conditions. At  $\mathbf{r} \in \mathcal{N}_d$  there is a microsystem  $\sigma(\mathbf{r})$  with values  $\pm 1$ <sup>99</sup> so that the microstate of the system is  $\boldsymbol{\sigma} := \{\sigma(\mathbf{r})\}$  and the Hamiltonian is

$$\hat{H}(\zeta_T, \zeta_{\mathcal{H}}; \boldsymbol{\sigma}) := -\zeta_T \sum_{\{\mathbf{r}, \mathbf{r}'\}}^{(1)} \sigma(\mathbf{r})\sigma(\mathbf{r}') - \zeta_{\mathcal{H}} \sum_{\{\mathbf{r}\}} \sigma(\mathbf{r}), \tag{113}$$

where the first summation is over all first-neighbour pairs of lattice sites and the thermal and field couplings are

$$\zeta_T := J/T, \quad \zeta_{\mathcal{H}} := \mathcal{H}/T, \tag{114}$$

respectively,  $J$  being an energy parameter, so that  $J > 0$  corresponds to ferromagnetic behavior, where alignment of the states of all first-neighbour pairs of sites is favoured, and  $J < 0$  corresponds to antiferromagnetic behaviour, where anti-alignment of the states of all first-neighbour pairs of sites is favoured;<sup>100</sup>  $\mathcal{H}$  is a magnetic field. It will be noted that these are the couplings introduced for the simple magnetic system in Sect. 2.4 except that there we used  $\varepsilon$  instead of  $J$  and it was assumed that  $\varepsilon > 0$ . This two-state model which was first solved for  $d = 1$  by Ising [51],<sup>101</sup> is usually now called the *spin- $\frac{1}{2}$  Ising model*.<sup>102</sup> On the basis of this solution, and a wealth of other results for larger dimensions (including the exact zero-field solution for a

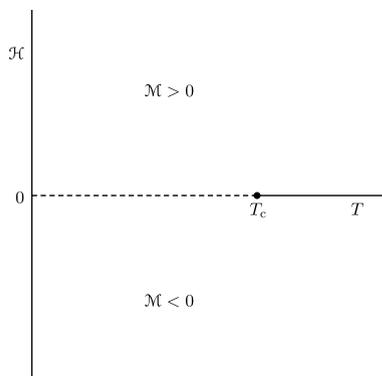
<sup>99</sup> These states are usually thought of as spin directions up (+ 1) and down (− 1).

<sup>100</sup> A perfect arrangement like this is possible for hypercubic lattices, but not for so-called ‘close-packed’ lattices like the plane triangular lattice and the face-centred cubic lattice.

<sup>101</sup> In fact it was suggested to Ising by his research director Wilhelm Lenz and historians of science like Brush [16] and Niss [96] often call it the *Lenz-Ising model*.

<sup>102</sup> The corresponding three-state model with states 0, ±1 being the spin-1 Ising model.

**Fig. 10** The phase diagram for the Ising model. The first-order transition is shown as a broken line



square lattice by Onsager [98]) the ferromagnetic phase diagram in the space of the temperature  $T$  and the field  $\mathcal{H}$  is known to have the form shown in Fig. 10, where, for  $d = 1$ ,  $T_c = 0$ , for  $d = 2$ ,  $T_c = 2.2692J$  and, for  $d = 3$ ,  $T_c = 4.5108J$ . Apart from in the one-dimensional case where the critical point is at zero temperature (see Sect. 3.4.3(a)), there is a line of first-order transitions along the interval  $[0, T_c)$  of the zero-field axis across which the magnetization  $\mathcal{M}$  changes between equal and opposite values. The universality class of the second-order transition at the critical point depends on the dimension of the system. The critical exponents for  $d = 2$  are  $\alpha = 0$  (a logarithmic singularity),  $\beta = \frac{1}{8}$ ,  $\gamma = \frac{7}{4}$ ,  $\delta = 15$ ,  $\nu = 1$ ,  $\eta = \frac{1}{4}$ . For  $d = 3$  the exponents are obtained to a high level of accuracy by series methods with  $\alpha \simeq 0.11008$ ,  $\beta \simeq 0.326419$ ,  $\gamma \simeq 1.237075$ ,  $\delta = 4.78984$ ,  $\nu \simeq 0.629971$ ,  $\eta \simeq 0.036298$ . When  $d \geq d_{UC} = 4$  the critical exponents take their classical values  $\alpha = 0$ ,  $\beta = \frac{1}{2}$ ,  $\gamma = 1$  and  $\delta = 3$ .<sup>103</sup>

**Acknowledgements** We are grateful to the reviewers of this work for their detailed comments and helpful criticisms.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Amit, D.J.: Field Theory, the Renormalization Group and Critical Phenomena. McGraw-Hill, New York (1978)

<sup>103</sup> For a definition of the upper-critical dimension see footnote 42.

2. Ardourel, V.: The infinite limit as an eliminable approximation for phase transitions. *Stud. Hist. Philos. Mod. Phys.* **62**, 71–84 (2018)
3. Barber, M.N.: Finite-size scaling. In: Domb, C., Lebowitz, J.L. (eds.) *Phase Transitions and Critical Phenomena*, vol. 8, pp. 145–266. Academic Press, London (1983)
4. Batterman, R.W.: *The Devil in the Details*. Oxford University Press, Oxford (2002)
5. Batterman, R.W.: Critical phenomena and breaking drops: infinite idealizations in physics. *Stud. Hist. Philos. Mod. Phys.* **36**, 225–244 (2005)
6. Batterman, R.W.: Reduction and renormalization. In: Ernst, G., Hüttemann, A. (eds.) *Time, Chance and Reduction*, pp. 159–179. Cambridge University Press, Cambridge (2010)
7. Batterman, R.W.: Emergence, singularities and symmetry breaking. *Found. Phys.* **41**, 1031–1050 (2011)
8. Batterman, R.W.: Philosophical implications of Kadanoff's work on the renormalization group. *J. Stat. Phys.* **167**, 559–574 (2017)
9. Batterman, R.W.: Universality and RG explanations. *Perspect. Sci.* **27**, 26–47 (2019)
10. Batterman, R.W.: *Intertheory Relations in Physics*. Stanford Encyclopedia of Philosophy. Stanford University Press, Stanford (2020)
11. Baxter, R.J.: Partition function of the eight-vertex lattice model. *Ann. Phys.* **70**, 193–228 (1972)
12. Bellemans, A., Nigam, R.K.: Phase transitions in two-dimensional lattice gases of hard-square molecules. *J. Chem. Phys.* **46**, 2922–2935 (1967)
13. Berry, M.: Singular limits. *Phys. Today* **55**(5), 10–11 (2002)
14. Binder, K.: Finite size scaling analysis of Ising model block distribution functions. *Z. Phys. B* **43**, 119–140 (1981)
15. Binney, J.J., Dowrick, N.J., Fisher, A.J., Newman, M.E.J.: *The Theory of Critical Phenomena*. Oxford University Press, Oxford (1993)
16. Brush, S.G.: History of the Lenz-Ising model. *Rev. Mod. Phys.* **39**, 883–893 (1967)
17. Buchdahl, H.A.: *The Concepts of Classical Thermodynamics*. Cambridge University Press, Cambridge (1966)
18. Buckingham, M.J.: Thermodynamics. In: Domb, C., Green, M.S. (eds.) *Phase Transitions and Critical Phenomena*, vol. 2, pp. 1–38. Academic Press, London (1972)
19. Burkhardt, T.W., van Leeuwen, J.M.J.: *Real-Space Renormalization*. Topics in Current Physics, vol. 30. Springer, Berlin (1982)
20. Butterfield, J.: Emergence, reduction and supervenience: a varied landscape. *Found. Phys.* **41**, 920–959 (2011a)
21. Butterfield, J.: Less is different: emergence and reduction reconciled. *Found. Phys.* **41**, 1065–1135 (2011b)
22. Callender, C.: Taking thermodynamics too seriously. *Stud. Hist. Philos. Mod. Phys.* **32**, 539–553 (2001)
23. Carathéodory, C.: Untersuchungen über die Grundlagen der Thermodynamik, *Mathematische Annalen* **67**: 355–386. Page references given to the English translation: Investigation into the Foundations of Thermodynamics, by J. Kestin in *The Second Law of Thermodynamics*, Benchmark Papers on Energy, Vol. 5, (Ed. J. Kestin), Dowden, Hutchinson and Ross, 1976, pp. 229–256. (1909)
24. Cardy, J.L.: *Finite-Size Scaling*. Current Physics—Sources and Comments, vol. 2. North-Holland, Amsterdam (1988)
25. Cardy, J.L.: *Scaling and Renormalization in Statistical Physics*. Cambridge University Press, Cambridge (1996)
26. Dizadji-Bahmani, F., Frigg, R., Hartmann, S.: Who's afraid of Nagelian reduction. *Erkenntnis* **73**, 393–412 (2010)
27. Domb, C.: On the theory of cooperative phenomena in crystals. *Philos. Mag. Supp.* **9**, 149–361 (1960)
28. Domb, C.: Critical phenomena: a brief historical survey. *Contemp. Phys.* **26**, 49–72 (1985)
29. Domb, C., Green, M.S.: *Phase Transitions and Critical Phenomena*, vol. 6. Academic Press, London (1976)
30. Essam, J.W., Fisher, M.E.: Padé approximant studies of the lattice gas and Ising ferromagnet below the critical point. *J. Chem. Phys.* **38**, 802–812 (1963)
31. Ferdinand, A.E., Fisher, M.E.: Bounded and inhomogeneous Ising models. I Specific heat anomaly on a finite lattice. *Phys. Rev.* **185**, 832–846 (1969)

32. Fisher, M.E.: Rigorous inequalities for critical-point correlation exponents. *Phys. Rev.* **180**, 594–600 (1969)
33. Fisher, M.E.: The theory of critical point singularities. In: Green, M. (ed.) *Critical Phenomena, Proceedings of the 51st Enrico Fermi Summer School, Varenna, Italy*, pp. 1–99. Academic Press, London (1971)
34. Fisher, M.E., Barber, M.N.: Scaling theory for finite-size effects in the critical region. *Phys. Rev. Lett.* **28**, 1516–1519 (1972)
35. Frigg, R.: A field guide to recent work on the foundations of statistical mechanics. In: Rickles, D. (ed.) *The Ashgate Companion to Contemporary Philosophy of Physics*, pp. 99–196. Ashgate, London (2008)
36. Frigg, R., Werndl, C.: Statistical mechanics: a tale of two theories. *The Monist* **102**, 424–438 (2019)
37. Gantmacher, F.R.: *Applications of the Theory of Matrices*. Interscience, New York (1959)
38. Gibb, S., Hendry, R.F., Lancater, T.E.: *The Routledge Handbook of Emergence*. Routledge, London (2019)
39. Giles, R.: *Mathematical Foundations of Thermodynamics*. Pergamon Press, Oxford (1964)
40. Goldenfeld, N.: *Lectures on Phase Transitions and the Renormalization Group*. Addison-Wesley, Reading (1992)
41. Griffiths, R.B.: A proof that the free energy of a spin system is extensive. *J. Math. Phys.* **5**, 1215–1222 (1964)
42. Griffiths, R.B.: Ferromagnets and simple fluids near the critical point: some thermodynamic inequalities. *J. Chem. Phys.* **43**, 1958–1968 (1965)
43. Griffiths, R.B.: Rigorous results and theorems. In: Domb, C., Green, M.S. (eds.) *Phase Transitions and Critical Phenomena*, vol. 1, pp. 7–109. Academic Press, London (1972)
44. Hankey, A., Stanley, H.E.: Systematic application of generalized homogeneous functions to static scaling, dynamic scaling and universality. *Phys. Rev. B* **6**, 3515–3542 (1972)
45. Heilmann, O.J.: Existence of phase transitions in certain lattice gases with repulsive potential. *Lett. Nuovo Cim.* **3**, 95–98 (1972)
46. Heilmann, O.J.: Existence of an ordered phase for the repulsive lattice gas on the fcc lattice. *J. Phys. A Math. Gen.* **13**, 1803–1810 (1980)
47. Heilmann, O.J., Huckaby, D.A.: Phase transitions in lattice gas models for water. *J. Stat. Phys.* **20**, 371–383 (1979)
48. Humphreys, P.W.: Emergence. In: Humphreys, P. (ed.) *The Oxford Handbook of the Philosophy of Science*, pp. 759–778. Oxford University Press, Oxford (2015)
49. Hüttemann, A., Love, A.: Reduction. *The Oxford Handbook of the Philosophy of Science*, pp. 460–484. Oxford University Press, Oxford (2015)
50. Hüttemann, A., Kühn, R., Terzidis, O.: Stability, Emergence and Part-Whole Reduction, Why More is Different: Philosophical Issues in Condensed Matter Physics and Complex Systems, pp. 169–200. Springer, Berlin (2015)
51. Ising, E.: Beitrag zur theorie des ferromagnetismus. *Z. Phys.* **31**, 253–258 (1925)
52. Josephson, B.D.: Inequality for the specific heat. *Proc. Phys. Soc.* **92**, 269–275, 276–284 (1967)
53. Kadanoff, L.P.: Scaling laws for Ising models near  $T_c$ . *Physics* **2**, 263–272 (1966)
54. Kadanoff, L.P.: Scaling, universality and operator algebras. In: Domb, C., Green, M.S. (eds.) *Phase Transitions and Critical Phenomena*, vol. 5a, pp. 1–34. Academic Press, London (1976)
55. Kadanoff, L.P.: *Statistical Physics: Statics, Dynamics, and Renormalization*. World Scientific, Singapore (2000)
56. Kadanoff, L.P.: More is the same; phase transitions and mean field theories. *J. Stat. Phys.* **137**, 777–797 (2009)
57. Kadanoff, L.P.: Theories of matter: infinities and renormalization. In: Batterman, R.W. (ed.) *The Oxford Handbook of Philosophy of Physics*. Oxford University Press, Oxford (2013a)
58. Kadanoff, L.P.: Relating theories via renormalization. *Stud. Hist. Philos. Mod. Phys.* **44**, 22–39 (2013b)
59. Kim, J.: Making sense of emergence. *Philos. Stud.* **95**, 3–36 (1999)
60. Kinzel, W., Schick, M.: Phenomenological scaling approach to the triangular Ising antiferromagnet. *Phys. Rev. B* **23**, 3435–3441 (1981)
61. Koopman, B.O.: Hamiltonian systems and transformations in Hilbert space. *Proc. Nat. Acad. Sci.* **17**, 315–318 (1931)
62. Kuhn, T.S.: *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago (1963)

63. Lancaster, P., Tisamenetsky, M.: *The Theory of Matrices*, 2nd edn. Academic Press, London (1985)
64. Landsberg, P.T.: Heat engines and heat pumps at positive and negative absolute temperatures. *J. Phys. A* **10**, 1773–1780 (1977)
65. Lavis, D.A.: An exact matrix calculation for a two-dimensional model of the steam-water-ice system: bulk. *J. Phys. A Math. Gen.* **9**, 2077–2095 (1976)
66. Lavis, D.A.: Boltzmann and Gibbs: an attempted reconciliation. *Stud. Hist. Philos. Mod. Phys.* **36**, 245–273 (2005)
67. Lavis, D.A.: *Equilibrium Statistical Mechanics of Lattice Systems*. Springer, Berlin (2015)
68. Lavis, D.A.: The problem of equilibrium processes in thermodynamics. *Stud. Hist. Philos. Mod. Phys.* **62**, 136–144 (2018)
69. Lavis, D.A.: The question of negative temperatures in thermodynamics and statistical mechanics. *Stud. Hist. Philos. Mod. Phys.* **67**, 26–63 (2019)
70. Lavis, D.A., Southern, B.W.: Renormalization group study of a three-dimensional lattice model with directional bonding. *J. Stat. Phys.* **35**, 489–506 (1984)
71. Lavis, D.A., Southern, B.W., Bell, G.M.: Phase transitions in monolayers at air/water and oil/water interfaces. *J. Phys. C* **15**, 1077–1088 (1982)
72. Lewis, D.K.: *On the Plurality of Worlds*. Blackwell, Oxford (1986)
73. Li, Y., Baberschke, K.: Dimensional crossover in ultrafine Ni(111) films on W(110). *Phys. Rev. Lett.* **68**, 1208–1211 (1992)
74. Lieb, E.H., Yngvason, J.: A guide to entropy and the second law of thermodynamics. *Am. Math. Soc.* **45**, 571–581 (1998)
75. Lieb, E.H., Yngvason, J.: The physics and mathematics of the second law of thermodynamics, *Phys. Rep.* **310**: 1–96. For Fig. 8 see *ibid.* (1999), **314**, 699 (1999)
76. Lieb, E.H., Yngvason, J.: A fresh look at entropy and the second law of thermodynamics. *Phys. Today* **53**, 32–37 (2000)
77. Lieb, E.H., Yngvason, J.: The entropy concept for non-equilibrium states. *Proc. R. Soc. A* **469**, 20130408 (2013)
78. Ma, S.K.: *Modern Theory of Critical Phenomena*. Benjamin, Reading (1976)
79. MacDonald, D.K.C.: *Noise and Fluctuations: An Introduction*. Dover Publications, New York (2013)
80. Mainwood, P.: Phase transitions in finite systems. Unpublished MS (corresponds to Chap. 4 of: *Is More Different? Emergent Properties in Physics*, D. Phil. thesis, University of Oxford, 2006). <http://philsci-archive.pitt.edu/8340/>
81. Maxwell, J.C.: On the dynamical evidence of the molecular constitution of bodies. *Nature* **11**, 357–374 (1875)
82. Mazzeo, G., Kühn, R.: Critical behaviour of the two-dimensional spin-diluted Ising model via the equilibrium ensemble approach. *Phys. Rev. E* **60**, 3823–3836 (1999)
83. Menon, T., Callender, C.: Turn and face the strange ... ch-ch-changes: philosophical questions raised by phase transitions. In: Batterman, R. (ed.) *Oxford Handbook of the Philosophy of Physics*, pp. 189–223. Oxford University Press, Oxford (2013)
84. Morrison, M.: Emergence physics and micro-ontology. *Philos. Phys.* **79**, 141–166 (2012)
85. Nagel, E.: *The Structure of Science: Problems in the Logic of Scientific Explanation*, 2nd edn. Hackett Publishing Co, Cambridge (1979)
86. Nauenberg, M., Nienhuis, B.: Critical surface for square Ising spin lattice. *Phys. Rev. Lett.* **33**, 944–946 (1974)
87. Nelson, D.R., Fisher, M.E.: Soluble renormalization groups and scaling fields for low-dimensional Ising systems. *Ann. Phys.* **91**, 226–274 (1975)
88. Nguyen, J., Frigg, R.: Unlocking limits. *Argumenta* **6**, 31–45 (2020)
89. Nickles, T.: Two concepts of intertheoretic reduction. *J. Philos.* **70**, 181–201 (1973)
90. Niemeijer, T., van Leeuwen, J.M.J.: Wilson theory for spin systems on a triangular lattice. *Phys. Rev. Lett.* **31**, 1411–1414 (1973)
91. Niemeijer, T., van Leeuwen, J.M.J.: Wilson theory for two-dimensional Ising spin systems. *Physica* **71**, 17–40 (1974)
92. Niemeijer, T., van Leeuwen, J.M.J.: Renormalization theory for Ising-like spin systems. In: Domb, C., Green, M.S. (eds.) *Phase Transitions and Critical Phenomena*, vol. 6, pp. 425–505. Academic Press, London (1976)
93. Nightingale, M.P.: Scaling theory and finite systems. *Physica A* **83**, 561–572 (1976)
94. Nightingale, M.P.: Non-universality for Ising-like spin systems. *Phys. Lett. A* **59**, 486–488 (1977)

95. Nightingale, M.P., THoof, A.H.: Scaling theory and logarithmic singularities. *Physica* **77**, 390–402 (1974)
96. Niss, M.: History of the Lenz-Ising model 1920–1950: from ferromagnetic to cooperative phenomena. *Arch. Hist. Exact. Sci.* **59**, 267–318 (2005)
97. Norton, J.D.: Approximation and idealization: why the difference matters. *Philos. Sci.* **79**, 207–232 (2012)
98. Onsager, L.: Crystal statistics. I Two-dimensional model with order-disorder transition. *Phys. Rev.* **65**, 117–149 (1944)
99. Orban, J., Bellemans, A.: Phase transitions in two-dimensional lattice gases of hard-core molecules: the triangular lattice. *J. Chem. Phys.* **49**, 363–370 (1968)
100. Orban, J., Van Craen, J., Bellemans, A.: Lattice models of hard-core molecules with attractions: the phase diagram. *J. Chem. Phys.* **49**, 1778–1783 (1968)
101. Palacios, P.: Had we but world enough, and time  $\dots$  but we don't!. *Found. Phys.* **48**, 526–541 (2018)
102. Palacios, P.: Phase transitions: a challenge for intertheoretic reduction? *Philos. Sci.* **86**, 612–640 (2019)
103. Peierls, R.: On Ising's model of ferromagnetism. *Proc. Camb. Philos. Soc.* **32**, 477–481 (1936)
104. Perron, O.: Grundlagen für eine Theorie des Jacobischen Kettenbruchalgorithmus. *Math. Ann.* **64**, 11–76 (1907)
105. Pippard, A.B.: *Elements of Classical Thermodynamics*. Cambridge University Press, Cambridge (1957)
106. Prigogine, I.: *The End of Certainty*. The Free Press, New York (1996)
107. Prigogine, I.: Laws of nature, probability and time symmetry breaking. *Physica A* **263**, 528–539 (1999)
108. Ramsey, N.F.: Thermodynamics and statistical mechanics at negative absolute temperatures. *Phys. Rev.* **103**, 20–28 (1956)
109. Rowlinson, J.S.: English Translation (with Introductory Essay) of J.D. van der Waals: *On the Continuity of the Gaseous and Liquid States of Matter*, *Studies in Statistical Mechanics*, vol. 14. North-Holland, Amsterdam (1988)
110. Ruelle, D.: *Statistical Mechanics: Rigorous Results*, 4th edn. Imperial College Press and World Scientific, London and Singapore (1999)
111. Runnels, L.K., Combs, L.L.: Exact finite method of lattice statistics. I. Square and triangular lattice gases of hard molecules. *J. Chem. Phys.* **45**, 2482–2492 (1966)
112. Runnels, L.K., Combs, L.L., Salvant, J.P.: Exact finite method of lattice statistics. II. Honeycomb-lattice of hard molecules. *J. Chem. Phys.* **47**, 4015–4020 (1967)
113. Rushbrooke, G.S.: On the thermodynamics of the critical region for the Ising model. *J. Chem. Phys.* **39**, 842–843 (1963)
114. Schaffner, K.F.: Reductionism in biology: prospects and problems. *Philos. Sci.* **1974**, 613–632 (1976)
115. Schick, M.A., Walker, J.S., Wortis, M.: Antiferromagnetic triangular Ising model. *Phys. Lett. A* **58**, 479–480 (1976)
116. Shech, E.: Infinite idealizations in physics. *Synthese* **13**, e12514 (2018)
117. Sklar, L.: The reduction (?) of thermodynamics to statistical mechanics. *Philos. Stud.* **95**, 187–202 (1999)
118. Sneddon, L.: Critical properties of two-dimensional spin  $s$  Ising systems with lattice anisotropy. *J. Phys. C Solid State Phys.* **11**, 2823–2828 (1978)
119. Southern, B.W., Lavis, D.A.: A model for absorbed monolayers of orientable molecules. *J. Phys. C* **12**, 533–543 (1979)
120. Southern, B.W., Lavis, D.A.: Renormalization group study of a two-dimensional lattice model with directional bonding. *J. Phys. A* **13**, 251–262 (1980)
121. Uffink, J.: Compendium of the foundations of statistical physics. In: Butterfield, J., Earman, J. (eds.) *Handbook of the Philosophy of Science: Philosophy of Physics*, pp. 923–1074. North-Holland, Amsterdam (2007)
122. van der Waals, J.D.: *Over de Continuïteit van der Gas - en Vloeïstoftoestand*, PhD thesis, Leiden. For an English translation see Rowlinson [109] (1873)
123. Van Riel, R., Van Gulick, R.: Scientific reduction. In: Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy* (Spring, 2019th edn. Stanford University Press, Stanford (2019))

124. Wang, J., Wu, W., Zhao, F., Zhao, G.: Curie temperature reduction in SiO<sub>2</sub>-coated ultrafine Fe<sub>3</sub>O<sub>4</sub> nanoparticles: quantitative agreement with a finite-size scaling law. *Appl. Phys. Lett.* **98**, 083107 (2011)
125. Wegner, F.: Corrections to scaling laws. *Phys. Rev. B* **5**, 4529–4536 (1972)
126. Widom, B.: Degree of the critical isotherm. *J. Chem. Phys.* **41**, 1633–1634 (1964)
127. Widom, B.: Equation of state in the neighbourhood of the critical point. *J. Chem. Phys.* **43**, 3898–3905 (1965)
128. Wilson, K.G.: The renormalization group: critical phenomena and the Kondo problem. *Rev. Mod. Phys.* **47**, 773–840 (1975)
129. Young, A.P., Lavis, D.A.: Critical behaviour of a two-dimensional bonded lattice fluid. *J. Phys. A* **12**, 229–243 (1979)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.