Chapter 8

An overview of the renormalization "group"

8.1 Relating small to large

In physics, we're quite accustomed to understanding macroscopic phenomena by first analysing the underlying microscopic behavior. Many, both inside and outside the field, sometimes forget how remarkable this is. Interactions between electrons and ions in a material are governed by electromagnetism, which has been well understood for many decades. It looks completely impossible to extract anything useful from the Hamltonian for the Coulomb interactions between 10^{23} charged particles. Nonetheless, in many cases we compute, often to stunning accuracy, properties of many-body systems. One reason is that people have been clever enough to write down effective theories, governing collective properties of the system. A key part of this is to choose very carefully which quantities are to be computed.

Despite many triumphs, condensed-matter physics in the present day exhibits many different kinds of behavior that cannot be understood by 19th and early-20th century methods. One example of such exotic phenomena is the appearance of fractional charge. There are experimentally realized systems comprised entirely of electrons which the effective "quasiparticle" excitations carry a charge a fraction of e. Quite obviously, doing the statistical mechanics of weakly interacting electrons will never explain such phenomena – it requires understanding *strong interactions*.

And this is not an issue only condensed-matter and many-body physicists need to worry about. In the energy ranges of interest to particle physicists, $E = mc^2$ means that particle production occurs regularly, so effectively can make a system many-body (e.g., "jet" processes at accelerators). Even more worrying is that when one analyzes particle physics via field theory, one encounters large quantities at higher orders in perturbation theory even when the expansion parameter is small. Frequently there appear divergent integrals apparently due to the creation of "virtual" particles. It is still heard today that these theories have "infinities" which must be "swept under the rug".

It is understandable why physicists thought this at the time, but it's now fairly embarrassing for an otherwise-respectable theorist to say such things. Nearly a half-century ago, the framework for understanding how to approach such problems was developed.

The key is to understand how physics on different length scales affects each other, how to continuously connect the micro with the macro.

This framework is called the *renormalization group*. One prominent feature is shows how to find effective theories that describe exactly long-distance behavior of a theory defined microscopically. For example, in many cases, it allows one to find a field theory, where the degrees of freedom are present at every point in space, that describes effectively a lattice statistical-mechanical model. Of course, the effective theory does not describe every detail of the microscopic theory. However, one of the many remarkable features of the renormalisation-group framework is that not only does it yield the effective theory, but it shows precisely which quantities the effective theory gives exactly, and which it does not. Such quantities are called *universal*, and they play a central role in what follows.

The renormalization group is much more "meta" than most subjects in theoretical physics. It does not merely give a method of computing experimentally relevant quantities, by starting with some Hamiltonian or Lagrangian exactly or approximately describing the physical system of interest. As has been noted by many, it is a "theory about theories" – it explains how different theories exhibit not just qualitatively similar phenomena, but quantitatively exact correspondences. It gives a precise method for relating such theories and for determining the universal quantities. It gives a precise notion of the "simplest" theories exhibiting the appropriate universal behavior.

Indeed, in the field-theory approach to particle physics, the renormalization group is necessary just to define the theory itself. An improperly defined theory indeed will result in nonsensical calculations, the infinities that plague a naive approach to quantum field theory. Appropriately taking into account the microscopic theory (what is typically called in the field theory literature "regulating" the theory), enables the theory to be defined so that physically measurable quantities are indeed sensible and finite. So-defined theories are typically called "renormalizable". However, it is worth noting that effective theories dubbed "non-renormalizable" are often quite useful, because the renormalization-group framework explains on which length scales such effective theories still describe the physics quantitatively.

I'm not sure who coined the name "renormalization group", but it's half-good, half-terrible. The word renormalization does accurately capture a key part of the analysis, that certain carefully defined quantities behave nicely under scaling. For example, in an effective theory, the couplings are related at different length scales by simply multiplying by an appropriate function of the scale, and hence are "re-normalized". So whereas that's only a part of the story, it's a central one, and it gives a good flavor of what's going on. The "group" part is simply wrong. The name presumably arose because as part of the procedure relating different theories, one does transformations e.g. of the couplings of the different theories. This does seem quite analogous to a symmetry transformation, where seemingly different theories are show to be equivalent by an appropriate redefinition of the couplings. Such symmetry transformations are given by elements of a group (e.g. in quantum mechanics implemented by a unitary operator acting on the Hilbert space). However, a key property of a group is that it is invertible – any transformation can be undone. However, when we do a "renormalization-group transformation" to relate a microscopic theory to a macroscopic one, the microscopic details are are forgotten – the

procedure is *not* invertible. Only the universal quantities are the same in the micro and macro theories.

The renormalization group is thus not a group! I thus will for the most part avoid using the word group; if it would really be awkward to not use it, I will just say RG and let you decided what the G stands for. I note that many of the original key insights were made by Kenneth G. Wilson, and so I nominate his middle name (Geddes).

8.2 Analyticity Analyticity Analyticity

Underlying not only dimensional analysis but most of theoretical physics is *analyticity*. The idea is simple: physical quantities typically analytically on the parameters and variables. This is the only reason calculations work! We can then do perturbation theory around a simple limit, i.e. Taylor-expand some complicated expression and compute the corrections term by term. Or even simpler, infer physical laws by taking the first corrections. Many physical laws such as Ohm's law are simply keeping the linear expansion in a Taylor expansion.

One major consequence of the analyticity of most expressions in theoretical physics is the fact that dimensional analysis works. There are only a few "fundamental" dimensionful constants describing much of physics, such as $\hbar$, c, and G. In a given problem, of course there are the parameters describing the degrees of freedom: the charge e of the electron, the electroweak symmetry-breaking scale, the band gap of a semiconductor. If you want to call some of the latter fundamental, and not others, that's a question of taste. I won't.

Thus when doing dimensional analysis, it is customary for theorists to start out by listing all the ones believed to be relevant to this problem (here I use the word "relevant" loosely – later we will see that there is a very precise physical definition). For example, gravity is a negligible effect in the how current is conducted in most materials, so it is reasonable to not include G in the list of such constants in this problem

A simple example of the power of dimensional analysis is the ideal gas law. Here you assume the pressure P of a gas depends only on the number density N/V (the number of molecules per volume), the temperature T of the gas, and Boltzmann's constant k_B that relates a temperature scale to an energy scale (i.e. has units energy/(absolute temperature). The pressure has units Force/area = Energy/(length)^3, so then

$$P \propto \frac{N}{V} k_B T \ .$$

The proportionality constant of course turns out to be 1 in this example, but showing this requires more assumptions about what is going on (i.e. a proper definition of what temperature is).

Of course, if there are more "fundamental couplings", dimensional analysis does not give such a simple answer. It does still give a useful simplification – a functional form involving less variables than unknowns. For example, say in the ideal gas example, we wish also to take into account the fact that the molecules occupy a non-zero volume Nb of space. Since this also has dimensions of length, dimensional analysis constrains the

form only:

$$P \propto \frac{N}{V} f(b/V) k_B T \ ,$$

where $f(b/V)$ is a dimensionless function of the dimensionless quantity b/V. In the Van der Waals approximation, we take this function to be $1+bN/V$. This is an approximation born of analyticity: it is assumed that as long as b is small enough, we can just expand $f(b/V)$ in a Taylor series around the ideal gas case $b = 0$.

It is important to understand that analyticity applies to *regions* of couplings. An illustration of this comes from Kepler's third law, which relates the radius R of a planet's orbit to its orbital period $\mathcal{T}$. It follows directly by assuming that the only "fundamental" constants relevant to the problem are Newton's gravitational constant G and the mass of the sun M_s. The only constant other than R to have a length scale involved is G, which has dimensions of $(\text{length})^3/(\text{mass} \times \text{ time})$. Thus $R \propto G^{1/3}$. Since the dimensions of R is a pure length, it must therefore be proportional to the combination

$$R \propto G^{1/3} M_s^{1/3} \mathcal{T}^{2/3} \ .$$

This is indeed Kepler's third law, which applies to all objects orbiting the sun. The fractional powers make it looks like the physics is not analytic. However, the only non-analyticity is at the origin $R = 0$, inside the sun. Obviously, Kepler's laws apply only outside the sun, and there they are perfectly analytic.

However, this example does bring up an interesting question: are there any examples where the physics is genuinely non-analytic as a function of the parameters and/or variables? Can the analyticity be even stronger, where exponents are irrational? The answer to both turns out to be yes, and understanding how and when turns out to provide the key to understanding renormalization.

8.3 Phase transitions and critical phenomena

I described in earlier parts of this book different phases that arise in classical lattice models and quantum spin chains. A general taxonomy is given in chapter 1. The RG gives valuable information about systems of all types. However, it turns out understanding *critical phenomena* is absolutely central to the story.

A key concept is that of the correlation length, as defined in chapter 1 (see also section 7.5). Qualitatively, it gives a notion of how far any local perturbation of the system will propagate. This is pretty obviously quite important if one is trying to construct an effective theory. For example, in the block spin approach pioneered by Kadanoff, one replaces a group of spins with a single effective spin. Clearly one this must effect the correlation length in some non-trivial way, since replacing spins changes the fundamental length scale. (I will someday write my own version of the block spin procedure, but there is a very nice explanation of this procedure in chapter 13 of Pathria's stat mech book.)

The phase diagram of magnet includes two kinds of transitions, which in the old but still useful language are described as first-order and second-order. The name tells you how analyticity is violated: at a first-order transition, the free energy is continuous, but the first derivative (with respect to the parameter whose variation is driving the

transition) is not. For example, in the Ising model in two dimensions and more, at any $T < T_c$ there is a first-order transition caused at $h = 0$ by varying h, i.e. $\partial F/\partial h$ is discontinuous at $h = 0$. Physically, this means that energy is required to flip the spins from ordered up to ordered down, or vice versa. At a first-order phase transition, the correlation length may be discontinuous, but it remains finite.

Phase transitions of higher than first order are called *critical*. At a critical point, *the correlation length diverges*. Most examples we'll encounter are second-order, where the correlation length diverges algebraically, e.g. $\xi \sim |T - T_c|^{-\nu}$ for some critical exponent ν. However, there is a very important infinite-order transition in 2d classical/1d quantum systems: the Kosterlitz-Thouless transition. This transition is not a conventional order-disorder transition, because there is no non-vanishing local order parameter on the two sides. The free energy is still singular, but it is only an essential singularity, i.e. the correlation length diverges not as a power law, but as $e^{-1/|T-T_c|}$. For now, I will stick with the more common second-order critical points.

A critical theory is *scale invariant*. As long as one is studying physics at length scales much longer than the fundamental length of the system (e.g. lattice spacing or average distance between particles), and much smaller than the total size of the system, the physics will look the *same* on different length scales. For example, in water at the critical point, there are dense and not-dense (i.e. liquid and gas) regions of all sizes. There *appear* to be no *inherent* length scales in a critical theory.

We will soon make this notion precise. What we will see is that the reason for the breakdown of analyticity at a critical point is that there are "hidden" *physical* parameters that must be taking into account at and near a critical point. In short, the short-distance physics does still qualitatively and quantitatively affect the long distance behavior. Dimensional analysis still of course is valid, but with an additional parameter, all bets are off. The key insight of Wilson was to showed precisely how to take this into account.

8.4 Universality

The remarkable fact that very carefully chosen quantities in different theories are *exactly* the same. Such quantities are called *universal*. The use of the word is in the same spirit as Newton's, but is much broader. Obviously, if seemingly different theories share exactly the same quantities, there must be some way of forgetting "irrelevant" details of the systems. The renormalization group provides such a method.

A simple example of why this isn't crazy: the mapping of fluid onto magnet.

8.5 The RG idea

Mean field theory clearly fails quantitatively in dimensions below the critical one. Nevertheless, it not only often gives the correct qualitative picture in many examples, but provides a way of thinking that guides all that follows. Moreover, the idea of Landau's to replace the microscopic theory with an effective one valid for the long-distance physics is central to the RG analysis that follows. Indeed, Wilson and Kogut say

The renormalization group approach is best seen as a more sophisticated realization of Landau's ideas.

The RG is a way of obtaining the long-distance physics from the short. It typically involves "forgetting" microscopic details when understanding the long-distance physics, and so gives a framework for explaining how different models can give the same universal behaviour. It is not a group, as the transformation is not invertible. Presumably the naming came from the similarity (in some ways) of an RG transformation with that of a symmetry group. A symmetry group typically describes transformations of the degrees of freedom in some fashion that leaves the action or Hamiltonian invariant. It is a group because not only can one repeat the transformation, it can be inverted. The RG transforms both the degrees of freedom *and* the couplings themselves to give a model with the same (or approximately the same) partition function. It is a *relation between different theories*. This *RG transformation* typically yields a system with *less* degrees of freedom. Thus although it can be repeated as with a group transformation, it cannot be inverted.

Namely, the partition function for an N-site system is labeled as $Z_N(\{K\})$, where the dimensionless couplings of the theory are $\{K\} = K_1, K_2 \ldots$. For Ising, for example, $K_1 = J/T$ and $K_2 = h/T$. The RG transformation on the couplings is then labeled as

$$\{K\} \to \{K'\} \equiv R(\{K\}) \ , \tag{8.1}$$

such that

$$Z_N(\{K\}) \approx Z_{N'}(\{K'\}) \ . \tag{8.2}$$

The use of $\approx$ here needs some commentary. Ideally, there would be a RG transformation where instead of being approximately equal, the two partition functions would be identical. In many 1d (classical) examples, making (8.2) an equality is possible, and this is illustrated in section 8.6 in the Ising case. In general, however, it is impossible to make the RG transformation exact, so one can only hope for an approximate equality. However, the idea is find a RG transformation so that as it is repeated again and again, the $\approx$ gets closer and closer to equality. A key issue then to address is which couplings need to be kept in $\{K\}$, and which can be forgotten. The RG framework gives a precise way of finding which couplings are relevant and which are irrelevant.

In an RG transformation it is convenient to define the dimensionless *scale factor* b, which in a lattice model relates the numbers of sites of N and N'. It is defined by

$$N' = b^{-d} N \tag{8.3}$$

for a system in d dimensions. The scale factor is defined so that $b > 1$, so that $N' < N$. In order to define the scale factor to make applicable to any system near criticality, not only lattice models, let a be some fundamental length scale in the system. For example, in a lattice model it might be taken to be the distance between nearest neighbors, or the length of the unit cell. A common choice in field theory is to take $a = 1/\Lambda$, where the momenta of any degrees of freedom are constrained to be $1/\Lambda$. Thus when evaluting

a Feynman diagram with a loop, the integral over any internal momenta is cutoff at Λ. Whatever your choice is, under a RG transformation

$$a \to ba \tag{8.4}$$

Since $b > 1$, one can think of an RG transformation as increasing the fundamental length scale in the system. In momentum space, this then amounts to decreasing Λ, or effectively integrating out high-momentum (and so high-energy) degrees of freedom.[1] Physical quantities such as the correlation length are defined to be invariant under RG transformations.

8.6 The RG equations for the 1d Ising model

8.7 The block-spin RG equations for the 2d Ising model

8.8 Fixed points and relevance/irrelevance

The RG transformations give a way of varying the couplings that leaves invariant the partition function and the ensuing physical quantities derived from it. This varying of couplings is called a *flow*. A *fixed point* of the flow is one where *all* the couplings remain the same under RG. Since an RG transformation amounts to a change of the fundamental length scale of the system, at a fixed point the system must behave the same at all length scales. The system thus has no inherent length scale like a correlation length at an RG fixed point. Therefore as the couplings are tuned to their fixed-point values, the correlation length must increase either to zero or infinity.

Typically, $\xi = 0$ fixed points are easy to understand. Because the degrees of freedom are effectively not interacting, the model often can be solved exactly in this limit. For example, the partition function of the Ising magnet is easily computed at either $T = 0$ or at $T = \infty$, where it has $\xi = 0$. Thus completely ordered models and completely disordered models typically have $\xi = 0$. However, some $\xi = 0$ models have much subtler (although still computable) behavior. One famous example in a quantum model is the Levin-Wen "string-net" lattice Hamiltonian for a 2+1-dimensional topological phase. There the Hamiltonian can be written as a sum over mutually commuting local operators, and so each can be diagonalised independently. One finds, however, that there are different ground states not related by any symmetry. Moreover, the number of such ground states depends on the genus of the two-dimensional surface the model is defined on. Such behavior is characteristic of topological order, and so the $\xi = 0$ fixed point provides an exceptionally valuable way of studying this physics.

In this chapter I focus mainly on the much subtler case $\xi \to \infty$, a critical point. Thus the RG way provides a framework to investigate physical behavior at and near

[1] A beautiful explanation of how Wilson's RG works in the field-theory language used by particle theorists was given by Polchinski, Nucl. Phys. B231 (1984) 269; see also his TASI lectures at `https://arxiv.org/abs/hep-th/9210046`.

critical points. In other words, it provides a way to go beyond, both qualitatively and quantitatively, mean-field theory.

8.9 Ginzburg revisited

Near a fixed point, the RG equations can be linearized, and then couplings defined so that each scales multiplicatively. In section 8.7 the corresponding variables and how they scale was found for the 2d Ising model. That calculation, although totally straightforward, was a bit tedious. So here I head back to both a simpler and more general setting, the Gaussian field theory that is at the core of mean field theory.

Here one can find a critical point almost by inspection: the *Gaussian fixed point*, i.e. the fixed point that comes from mean field theory. Recall that in MFT, all terms beyond second order in ϕ are neglected (in the low-T phase, this neglecting is done after the field is shifted so that $\phi = 0$ at the mininum). At the critical point, the coefficient of the ϕ^2 term vanishes also, leaving only the $(\nabla\phi)^2$ term. The reason for the name is thus obvious: the partition function is then simply a set of Gaussian integrals, as detailed in the previous chapter.

The analysis of this fixed point gives a way of understanding the Ginzburg criterion simply by power counting. To do this, it is convenient to keep track of explicit factors of a in the free energy. These arise when going from the sum over lattice sites to the integral over all space:

$$\sum_j \sim \frac{1}{a^d} \int d^d x$$

where the integral is now over the physical space x, so that it is invariant under the RG like all physical quantities. Thus the RG transformation $a \to ba$ scales this by a factor of b^{-d}, which is a fancy way of saying that the transformation reduces the number of points in the system by a factor of b^d. Similarly, a lattice difference turns into a derivative via

$$\phi_i - \phi_j \ \sim \ a\partial\phi$$

for nearest-neighbor sites i and j. The original definition of the stiffness κ absorbed these factors of a^{d-2} so that κ had dimension (length)$^{2-d}$, while ϕ was dimensionless. Unabsorbing these factors gives a dimensionless version of the stiffeness defined as $\tilde{\kappa} = \kappa a^{d-2}$

$$\beta F_{\rm GFP} = \kappa \int d^d x (\nabla\phi)^2 = \frac{\tilde{\kappa}}{a^{d-2}} \int d^d x (\nabla\phi)^2 \tag{8.5}$$

where GFP stands for "Gaussian fixed point".

As always, the partition function is invariant under RG transformations. Moreover, at a fixed point the couplings do not change under RG transformations. Thus to make a scale invariant theory, the field ϕ must therefore scale under the RG as

$$\phi \to b^{(d-2)/2}\phi$$

at the Gaussian fixed point. Since the two-point correlation function is a physically measurable quantity, it must be independent of the RG. Therefore it must obey

$$\langle \phi(\vec{r})\phi(0)\rangle = \mathcal{A}\left(\frac{a}{r}\right)^{d-2} \tag{8.6}$$

for some dimensionless quantity $\mathcal{A}$; the powers of b on the right- and left-hand sides indeed cancel. The *scaling dimension* x_ϕ of the field ϕ at the Gaussian fixed point is thus $x_\phi = (d-2)/2$.

It is now convenient to follow particle-physicist conventions and define a *renormalized* field $\tilde{\phi} = a^{(2-d)/2}\phi$ that absorbs the powers of a at the Gaussian fixed point, so that

$$\beta F_{\text{gaussian}} = \tilde{\kappa} \int d^d x (\nabla \tilde{\phi})^2 \tag{8.7}$$

The renormalized field and $\tilde{\kappa}$ are independent under the RG, and its two-point function is independent of the lattice scale:

$$\langle \tilde{\phi}(\vec{r}) \tilde{\phi}(0) \rangle = \mathcal{A}\, r^{2-d} \ . \tag{8.8}$$

The renormalized field $\tilde{\phi}$ is no longer dimensionless. Instead, the physical dimension (as measured by the powers of energy) of the renormalized field is simply the scaling dimension of the "bare" field.

The Ginzburg criterion for the Gaussian fixed point is now simple to see. Reinstating the ϕ^4 term and explicitly including the powers of a, the free energy near the critical point is of the form

$$\beta F = \int \frac{d^d x}{a^d} \left[a^2 \tilde{\kappa} (\nabla \phi)^2 + B t \phi^2 + C \phi^4 \right]$$

where the constants B and C are independent of the reduced temperature $t = (T - T_c)/T_c$. It is illuminating to rewrite this in terms of the renormalized field:

$$\beta F = \int d^d x \left[\tilde{\kappa} (\nabla \tilde{\phi})^2 + B t a^{-2} \tilde{\phi}^2 + C a^{d-4} \tilde{\phi}^4 \right] \tag{8.9}$$

The scaling of the couplings $\tilde{B}$ and $\tilde{C}$ then follows from the fundamental requirement that the partition function and hence the free energy remain invariant the RG. Since $\tilde{\phi}$ is invariant under $a \to ba$, the couplings must scale as

$$B t \to b^2 B t \ , \qquad C \to b^{4-d} C \tag{8.10}$$

In other words $y_{\phi^2} = 2$, while $y_{\phi^4} = 4 - d$.

The scaling under the RG from (8.10) shows that Bt grows in any dimension. The only way to have a fixed point is to set $t = 0$. This indeed is apparent from the computation of the two-point function in the previous chapter, where any non-zero B leads to a non-vanishing correlation length (when $C = 0$). However, the behaviour of C depends on the dimensionality: under the RG it grows for $d < 4$, shrinks to zero for $d > 4$. This is the same conclusion as that coming from Ginzburg criterion! One cannot neglect the ϕ^4 term for $d < 4$. In $d > 4$ (and it turns out $d = 4$ as well, this analysis goes beyond Ginzburg to strongly indicate that not only is mean field theory consistent, but actually works, at least close enough to the critical point. The catch is that $d = 3$, our world, one needs to understand how to deal with a *relevant* ϕ^4 term.

The remarkable consequence of this analysis is that the Ginzburg criterion follows from simple dimensional analysis. The powers of a in (8.9) and hence the Ginzburg

criterion simply follow from rescaling the field so that derivative term has no explicit powers of a in it. Since a is the only dimensionful quantity at the critical point, the powers in the other terms follow automatically. The scaling behaviour at and near the Gaussian fixed point in $d > 2$ is thus easy to understand. (The reason for the restriction $d > 2$ is apparent from (8.8).) Though although the derivation of the Ginzburg criterion in the last chapter was nice in that it made direct contact with the original lattice quantities, using the field-theory approach with the RG is much simpler. The moral is that once you understand the RG, life gets much easier.

8.10 Scaling functions

The critical exponents are directly related to quantities arising in the RG, namely the eigenvalues of the linearized RG transformations valid near a critical point. The easiest case to understand is the correlation length. Since it has dimension length, it can be parametrised as a function of the dimenionless couplings $\{u_i\}$ as

$$\xi = af(u_1, u_2, \dots) \ . \tag{8.11}$$

Since ξ is a physical quantity, it does not change under the RG transformations. Thus for a transformation that changes $a \to ba$ and $u_j \to u'_j$,

$$f(u_1, u_2, \dots) = bf(u'_1, u'_2, \dots) \tag{8.12}$$

When the $\{u_i\}$ are scaling variables, this functional relation simplifies to

$$f(u_1, u_2, \dots) = bf(\lambda_1 u_1, \lambda_a u_2, \dots) \ , \tag{8.13}$$

where the λ_i are the corresponding eigenvalues found by linearizing the RG equations around a critical point. Note that since by definition all $u_i = 0$ at a fixed point, the only solutions to this equation in this case are indeed $f(0, 0, \dots) = 0$ or $\to \infty$. Consider now the latter case of a critical point, and tuning all but one of the u_i to be zero, leaving say only $u_j \neq 0$. Assuming the correlation length diverges by some power law, the critical exponent y_j is then defined by setting

$$f(0, 0, u_j, 0, 0 \dots) = C_j |u_j|^{-1/y_j} \qquad \text{for } u_j \to 0 \tag{8.14}$$

for some constant C_j. Using this form in (8.13) gives

$$1 = b\lambda_j^{-1/y_j} \quad \Rightarrow \quad y_j = \frac{\ln \lambda_j}{\ln b} \ .$$

It is possible for different ν_j to occur for different signs of u_j, but this rarely (never?) happens in practice.

If, as in the two-dimensional Ising model $h = 0$, there is only a single relevant coupling, $u_1 \propto (T - T_c)/T_c$. The scaling form (8.14) then reduces to the usual definition of ν:

$$\xi \propto |T - T_c|^{\nu}, \qquad \nu = 1/y_1 \ . \tag{8.15}$$

For models where there are multiple relevant couplings, then a $\nu_j = 1/y_j$ can be for each coupling. The most relevant coupling corresponds to the largest value of λ, and so the largest (positive) y_j present.

Thus the critical exponent ν is simply related to the dimension of the relevant coupling that drives the system away from critical point. One of many remarkable features of the RG is that it gives many more relations between among critical exponents. In fact, it shows man many of the critical exponents depend just the dimensions y_i of the relevant couplings, along with the dimension of space itself (in quantum theories, space-time).

The key to deriving these relations is to find the *scaling function* for the free energy. The scaling function arising from the RG a more sophisticated version of dimensional analysis. Namely, one of the key characteristics of dimensional analysis is that it constrains expressions for physical quantities to depend on the couplings in a particular way, to be consistent dimensionally. While of course dimensional analysis remains valid when applying the RG, a short-distance dimensionful parameter, the length scale a, must also be included. Including it seems to make matters worse, effectively increasing the number of parameters a physical quantity can depend on.

However, while the RG shows why the short-distance behavior can't be forgotten, it also shows how to deal with it. The RG transformation leaves all physical quantities invariant, and so provides a constraint on the scaling functions in the same spirit as dimensional analysis. Indeed, the fact that the correlation length cannot depend on b constrains the form of its scaling function (8.11), allowing ν be related to the most relevant y. An analogous relation is true for all scaling functions, effectively reducing the number of variables by one.

Consider a situation where there are two relevant perturbations of some fixed point, labeling the corresponding (dimensionless) scaling couplings u_t and u_h. In the example of a magnet $u_t \propto (T - T_c)/T_c$, $u_h \propto h$. At the critical point $u_t = u_h = 0$, while near it the couplings scale under the RG as

$$u_t \to b^{y_t} u_t, \qquad u_h \to b^{y_h} u_h, \ . \tag{8.16}$$

For the Gaussian fixed point, (8.10) gives $y_t = 2$, but that fact is unimportant here; the following scaling functions apply to any fixed point. By construction of the RG,

$$Z_N(u_t,\, u_h) \approx e^{NK_0(u_t,u_h)} Z_{N/b^d}(b^{y_t} u_t,\, b^{y_h} u_j) \ ,$$

where K_0 is the constant shift in the energy that arises when doing the RG transformation. It is typically analytic in the couplings at the transition and so only gives rise to the non-universal constant terms in the free energy. Thus in the (reduced) free energy density $f \equiv \beta F/N = -\ln Z_N/N$, it is convenient to separate it out:

$$f(u_t, u_h) = K_0 + f_s(u_t, u_h) \ . \tag{8.17}$$

The subscript on the latter term stands for *singular*. All the non-analyticities in the free energy are contained in this part.

The RG gives a way of rescaling f_s so that it can be written as a *scaling function*. Namely, the invariance of the physics under an RG transformation results in a one-parameter constraint on a two-parameter function. Thus the scaling function here only

depends on a *single variable*, a very special combination of u_t and u_h. The trick (following Cardy's book) to understand this scaling function is to do the RG transformation many times, while still keeping the rescaled couplings small enough so that the scaling behavior (8.16) still holds. In other words, the couplings still remain close to the critical point. Doing the RG transformation once on the singular part of the free energy gives

$$f_s(u_t, u_h) = b^{-d} f_s(u_t b^{y_t}, u_h b^{y_h}) \ . \tag{8.18}$$

while doing it n times gives

$$f_s(u_t, u_h) = b^{-nd} f_s(u_t b^{ny_t}, u_h b^{ny_h}) \ . \tag{8.19}$$

Stopping after n times gives defines a rescaled coupling u_{t_0}

$$u_{t_0} \equiv |u_t b^{ny_t}| \ ,$$

which still must be small to keep the linearized form of the RG equations valid. It is then convenient to get rid of n in the equation by rewriting in terms of u_{t_0}, giving

$$f_s(u_t, u_h) = \left|\frac{u_t}{u_{t_0}}\right|^{-d/y_t} f_s\Big(\pm u_{t_0}, u_h \left|\frac{u_t}{u_{t_0}}\right|^{-y_h/y_t}\Big)$$

The left-hand-side of course does not depend on n or u_{t_0}. Thus the right-hand-side doesn't either! Moreover, the unknown function on the right-hand-side depends only on u_t and u_h through the combination $u_h|u_t|^{-y_h/y_t}$. The free energy thus depends on this variable via the functional form

$$f_s(u_t, u_h) = |u_t|^{-d/y_t} \Phi(u_h |u_t|^{-y_h/y_t}) \ . \tag{8.20}$$

The RG does not tell you what the function Φ is, but it does tell you the extremely important fact that it only depends on one variable, and moreover, precisely how this variable depends on u_t and u_h. This is an experimentally verifiable prediction. In fact, it is by far the best way to experimentally determine the critical exponents, by taking data for various u_t and u_h, and then seeing if it collapses on to a single-variable curve for an appropriate choice of y_t and y_h.