

The renormalization group: Critical phenomena and the Kondo problem*†

Kenneth G. Wilson

Laboratory of Nuclear Studies, Cornell University, Ithaca, New York 14850

This review covers several topics involving renormalization group ideas. The solution of the s -wave Kondo Hamiltonian, describing a single magnetic impurity in a nonmagnetic metal, is explained in detail. See Secs. VII–IX. “Block spin” methods, applied to the two dimensional Ising model, are explained in Sec. VI. The first three sections give a relatively short review of basic renormalization group ideas, mainly in the context of critical phenomena. The relationship of the modern renormalization group to the older problems of divergences in statistical mechanics and field theory and field theoretic renormalization is discussed in Sec. IV. In Sec. V the special case of “marginal variables” is discussed in detail, along with the relationship of the modern renormalization group to its original formulation by Gell-Mann and Low and others.

CONTENTS

Introduction	773
I. Definition of a Renormalization Group Transformation	777
II. Fixed Points in Perturbation Theory	780
III. Linearized Behavior Near Fixed Points	783
IV. Divergences in Field Theory and Statistical Mechanics, the Triangle of Renormalization	787
V. Marginal Operators and Second Order Perturbations About a Fixed Point	791
VI. Block Spin Methods: The Two Dimensional Ising Model	797
VII. The Kondo Problem: Introduction and Definition of Basis	805
VIII. Renormalization Group Transformation for the Kondo Hamiltonian	816
IX. Impurity Susceptibility and Specific Heat for the Kondo Hamiltonian	825
Appendix: Perturbation Expansions for the Kondo Hamiltonian	837

INTRODUCTION

One of the most basic themes in theoretical physics is the idea that nature is described locally. The basic equations of all physics are local. For example, Maxwell's equations specify the behavior of electric and magnetic fields in an infinitesimal neighborhood of a point x . In order to be able to specify local equations it is necessary to define continuum limits, namely the limits which define derivatives. The idea of the derivative and the idea of a continuum limit that underlies the derivative is therefore of great importance in all of physics.

It is now becoming clear that there is a second form of continuum limit, called the statistical continuum limit, which also has a very broad range of applicability throughout physics. In the statistical continuum limit functions of a continuous variable are themselves *independent* variables. For example, the electric and magnetic fields throughout space can be the independent variables in a statistical continuum limit. This happens in statistical or quantum mechanical problems where there are field fluctuations, so that one has to compute averages over an ensemble of fields. In statistical calculations one does not compute the field at a

point x . Instead one computes correlation functions; that is, expectation values of products of fields such as $\langle E(x,t)E(y,t') \rangle$. In quantum mechanical problems the correlation functions are sometimes replaced by vacuum expectation values of products of fields. In the simplest cases a field average determining a correlation function can be written formally as a functional integral. In the functional integral the fields are the independent variables of integration.

There are two ways in which a statistical continuum limit can arise. The obvious way is when the independent field variables are defined on a continuous space; the case of statistical or quantum fluctuations of the electromagnetic field is an example. If one were to replace the continuum by a discrete lattice of points, the field averages would consist of integrals over the value of the field E at each lattice site n . Thus for the discrete lattice case one has a multiple integration, $\prod_n \int dE_n$, the variables of integration being the fields E_n . In the continuum limit one has infinitely many integration variables E_n . Problems with infinitely many variables can be very difficult to solve.

The second source of statistical continuum limits is the situation where one has a lattice with a fixed lattice spacing, usually an atomic lattice. The number of independent variables (i.e., independent degrees of freedom) at each lattice site is fixed and finite. The continuum limit arises when one considers large size regions containing very many lattice sites. When the lattice is viewed on a macroscopic scale one normally expects the lattice structure to be invisible. That is, large scale effects should be describable by a continuum picture making no reference to the lattice spacing.

Consider, for example, critical phenomena in a magnet. A magnet is built of atoms and the atomic spacing provides a fixed shortest length which does not go to zero. At the critical point (the critical point occurs at the Curie temperature) there are long wavelength fluctuations of the magnetization signalling the onset of spontaneous magnetization. The maximum wavelength of the fluctuations is the corre-

* Supported in part by the National Science Foundation.

† This paper is a compilation of material presented as a series of nine lectures at Cargèse in Summer 1973.

lation length ξ ; precisely at the critical temperature, ξ is infinite.

The continuum limit comes into the problem when one tries to formulate a hydrodynamic picture of the fluctuations on the scale of the correlation length. Macroscopic waves in a fluid are described by the continuum equations of hydrodynamics which make no reference to the atomic structure of the fluid. Similarly one would hope to achieve a theory of the critical fluctuations in a magnet which makes no reference to the atomic spacing.

The difference between a hydrodynamic wave and the critical fluctuations in a magnet illustrates the difference between the classical continuum limit and the statistical continuum limit. A classical hydrodynamic wave is characterized by a definite wavelength, and very little motion of the fluid occurs at much shorter wavelengths. It is therefore a relatively trivial matter to introduce continuum forms of density, pressure, etc. for a hydrodynamic wave. However, the critical fluctuations in a magnet for very long wavelengths are not the dominant fluctuations. Instead, fluctuations occur on *all* wavelength scales from the correlation length to the atomic spacing and all these intermediate wavelengths are crucial to the physics of critical phenomena. In particular there is no gap in wavelengths between the wavelengths of fluctuations and the atomic wavelengths. This means it is difficult to determine which wavelengths of fluctuations to include in a continuum description and which to exclude.

The statistical continuum limit is most difficult to achieve when one tries to maintain locality. The problem is that in a local system, very short wavelength fluctuations in a field are as important as fluctuations at a fixed wavelength. For example, the quantum fluctuations of the electromagnetic field are predominantly at wavelengths much smaller than the electron Compton wavelength. The reason for this is that there is very little difference, locally, between a short wavelength fluctuation and a long wavelength fluctuation. For example, a fluctuation with wavelength 10^{-16} cm and a fluctuation with wavelength 10^{-11} cm both look like long wavelength waves if one looks only at a region of size 10^{-20} cm. In a local theory one specifies initially the behavior of the system only in a region even smaller than 10^{-20} cm. In consequence, in a local theory, it is hard to suppress the 10^{-16} cm wavelength fluctuations more than the 10^{-11} cm fluctuations, and the 10^{-16} cm wavelengths then are more important because there are more of them.

A procedure is now being developed to understand the statistical continuum limit. The procedure is called the renormalization group. It is the tool that one uses to study the statistical continuum limit in the same way that the derivative is the basic procedure for studying the ordinary continuum limit. However, the problems that one studies with the renormalization group are rarely formulated explicitly in terms of continuum limits. Because of this the very general nature of the renormalization group has been less apparent than the general nature of the derivative.

The renormalization group is at a much more primitive stage than the derivative. There is only a small subset of

problems involving the statistical continuum limit that have been solved so far, and to solve these problems a large amount of labor and theoretical artifice is required. One is still a long way from the simple and yet explicit nature as the derivative. Nevertheless, the renormalization group is the only method at present which is explicitly designed to investigate statistical continuum limit problems, and is likely to remain so. Also there are excellent prospects for the renormalization group to become much more powerful in the future than it is at present.

The crucial feature of the statistical continuum limit is the absence of characteristic length or energy or time scales. In the case of critical phenomena, the dominant fluctuations are neither the fluctuations with wavelengths ξ nor the fluctuations with wavelengths of order of the atomic spacing. It is the fluctuations *between* these two wavelengths that dominate.

In contrast, in classical (as opposed to statistical) problems the dominant length scale is determined by a length parameter in the problem. For example, in the hydrogen atom the Bohr radius provides the characteristic length scale. This length scale is determined by simple dimensional analysis. Any length of importance in the hydrogen atom is proportional to the Bohr radius. Dimensional analysis is completely taken for granted now. Physicists find it hard to work with systems for which dimensional analysis is irrelevant.

The absence of a characteristic length has profound consequences for the statistical continuum limit. This originally became apparent in quantum electrodynamics, where short wavelength fluctuations led to divergences.

In quantum electrodynamics, and elementary particle physics generally, the most noticeable missing scale is the lack of an energy scale. There is sometimes a minimum energy scale associated with a particle rest energy mc^2 (m may be the electron rest mass or the π meson rest mass, for example). All energy scales above this minimum occur; for example, in high energy scattering experiments, new particles are created with all energies from the rest energy up to the maximum energy available.

The lack of an energy scale becomes even more apparent in quantum field theories of elementary particles. In these theories one has to compute sums over intermediate states containing arbitrarily large energies. For example, in quantum electrodynamics one must consider electron-positron intermediate states where the electron and positron both have arbitrarily large momenta (but in opposite directions so the total 3-momentum of the pair is small). Sums over such states reduce in many simple cases to the logarithmically divergent integral

$$\int_{mc^2}^{\infty} \frac{dE}{E}$$

A logarithmically divergent integral is a typical (but not universal) symptom of a problem lacking a characteristic scale. The contribution of any given order of magnitude range of energies to the integral is finite. For example,

for any $E_0 > mc^2$, the contribution to the integral from energies between E_0 and $2E_0$ is $\ln 2$; this contribution is independent of E_0 . Thus all energy scales above mc^2 make equal contributions to the integral. The sum of these contributions is infinite because there are an infinite number of order of magnitude ranges of energy above mc^2 .

Renormalization theory, due to Bethe, Schwinger, Feynman, Dyson, etc. [see Schwinger (1958)], eliminates the divergences of quantum electrodynamics. Renormalization theory was the first method developed for computing the statistical continuum limit of a local theory. It continues today to be an important tool for investigating the statistical continuum limit. However, the standard renormalization theory applies only to problems which can be solved by a Feynman diagram expansion. Even more restrictive is the requirement that only a few Feynman diagrams be important after renormalization. (There are techniques for summing infinite subclasses of Feynman diagrams but unfortunately these methods are effective only in a few cases.) The worst feature of the standard renormalization procedure is that it is a purely mathematical technique for subtracting out the divergent parts of integrals in the continuum limit. It gives no insight into the physics of the statistical continuum limit. It is possible to solve the renormalization problems of Feynman diagrams by mathematical techniques without new physical insight because the physical aspects of many length or energy scales are least noticeable for a theory dominated by a few Feynman diagrams. The reason for this is that when the expansion parameter for a Feynman diagram expansion is small there is usually only a small coupling between fluctuations of different wavelengths. The coupling between different wavelengths is essential for the nontrivial consequences of many scales. For example, the divergences in quantum field theory are caused by the influence of very high energy intermediate states on low energy phenomena. When the coupling constant (the expansion parameter for Feynman diagrams) is small, this influence is also small. (To be precise, the influence of single factor of 2 range of high energies on low energy phenomena is small—the divergence due to all high energies is not small.)

The basic physical idea underlying the renormalization group approach is that the many length or energy scales are locally coupled. For example, the behavior of fluctuations in a magnet with wavelengths from 1000 to 2000 Å are assumed to be primarily affected by fluctuations with nearby wavelengths, e.g., 500–1000 Å or 2000–4000 Å. Fluctuations with wavelengths much less or much greater than 1000 Å are less important. The result of this assumption is that there is a cascade effect in the whole system: the atomic fluctuations (1–2 Å) influence the 2–4 Å fluctuations. The 2–4 Å fluctuations influence the 4–8 Å fluctuations, etc.

There are two principal features of the cascade picture. The first feature is scaling. The behavior of fluctuations for intermediate wavelengths tends to be identical except for a change of scale, precisely due to the lack of a characteristic length. The scaling fails for fluctuations with wavelengths near a length parameter. For example, in a magnet the fluctuations at atomic wavelengths and at wavelengths of order ξ do not scale in the same way as intermediate wavelength fluctuations.

The second feature of the cascade picture is the existence of amplification and deamplification as the cascade develops. For example, consider a small change in temperature away from the critical temperature in a magnet. A small change in temperature has little effect on the atomic scale fluctuations. But as the cascade develops, from 1 Å to 2 Å to 4 Å to 8 Å wavelengths, etc., the effect of the temperature change is amplified, finally leading to macroscopic changes at very large wavelengths. In particular, if the temperature was initially precisely at the critical temperature then ξ changes from being infinite to some finite value ξ' . This is a macroscopic change for fluctuations with wavelengths greater than ξ' .

Deamplification also takes place in the cascade. For example, two different magnetic materials can have quite different atomic structures. But the effect of the different atomic structures usually decreases with each stage, finally becoming negligible at large wavelengths. This deamplification underlies the hypothesis of universality [for references see Wilson and Kogut (1974)] in critical phenomena. The universality hypothesis is that many different substances have the same critical behavior.

There is an analogue to universality in the case of an ordinary derivative. Namely, there are many different finite difference approximations to a single derivative. That is, many different discrete lattice differences have identical continuum limits. Universality is the corresponding result for the statistical continuum limit.

The first stage in the renormalization group analysis of a system is to find a way to isolate a particular step in the cascade. This amounts to defining a sequence of intermediate steps in the solution of the full problem, one step for each cascade step. For instance, in the case of critical phenomena, one step in the renormalization group calculation may consist of an explicit statistical averaging over fluctuations with a factor of 2 range of wavelengths. The first step might consist of an averaging over all fluctuations with wavelengths less than 2 Å. The second step would then be to average over wavelengths from 2 to 4 Å; the third step to average over wavelengths from 4 to 8 Å, etc. (See Sec. I for more discussion.)

At the end of each step one is left with an effective interaction or Hamiltonian describing the length or energy scales not yet solved. Scaling is achieved when the effective interaction goes into itself (apart from a similarity transformation) after each step. For example, in critical phenomena the effective interaction generated after averaging the 128–256 Å fluctuations is very similar to the interaction generated after averaging out the 256–512 Å fluctuations. When this similarity occurs one formally has a “fixed point” interaction or Hamiltonian.

By carrying out the same cascade iterations for slightly changed initial conditions one can determine a set of amplification and deamplification factors λ_i . The factors λ_i are the eigenvalues of a linearized version of the iteration, linearized about the fixed point. The eigenvalues λ_i are the factors by which small changes in the initial interaction are increased or decreased with each step in the renormalization group calculation. If a particular parameter in the

initial Hamiltonian causes amplification (i.e., λ_i is greater than 1), it is called a "relevant variable." A parameter whose effect is deamplified is called "irrelevant." The number of relevant variables is equal to the number of amplification factors λ_i greater than 1.

Renormalization group theory divides, very roughly, into four parts. There is the formal theory of fixed points and linear (and nonlinear) behavior near fixed points. Much of this theory is due to Wegner (1972); reviews of this theory are provided in Wilson and Kogut (1974), Ma (1973), and Fisher (1974). The second part is a diagrammatic formulation of the renormalization group for critical phenomena. This formulation [discussed in Wilson and Kogut (1974)] can be solved explicitly for the case of space dimension $d = 4 - \epsilon$ with ϵ small. It can be used to do calculations to order ϵ of numerous aspects of critical behavior and illustrates the physics of fluctuations at many wavelengths.

The third aspect of renormalization group theory includes the original Gell-Mann-Low (1954) renormalization group theory and the Callan-Symanzik equations [Callan (1970) and Symanzik (1970)]. These methods are quantum field theoretical methods, originally, and they apply only to Feynman diagram expansions. They are efficient calculational methods (for Feynman diagrams): they are used, for example, to do high order calculations in ϵ . [See Brèzin *et al.* (1973)]. They completely hide the physics of many scales. These methods are hard to follow in detail for physicists without quantum field theoretical training. They are closely related to the original renormalization procedures for Feynman diagrams, which is the reason for the term "renormalization" in "renormalization group."

The fourth aspect of renormalization group theory is the construction of nondiagrammatic renormalization group transformations, which are then solved numerically, usually using a digital computer. This is the most exciting aspect of the renormalization group, the part of the theory that makes it possible to solve problems which are unreachable by Feynman diagrams. The Kondo problem has been solved by a nondiagrammatic computer method. The renormalization group solution of the Kondo problem is explained in detail in this paper: see Sec. VII-IX. The two dimensional Ising model has been solved approximately by several nondiagrammatic ("block spin") renormalization group methods, by Niemeier and Van Leeuwen (1973, 1974, 1975) and others. An example is detailed in Sec. VI. The Ising calculation is only a practice calculation, since the exact solution is known. Recently, Kadanoff (1975) and Kadanoff and Houghton (1975) have developed very powerful block spin methods which have been applied to the three dimensional Ising model, with considerable success.

There is another renormalization group method being developed by Golner and Riedel (1975) which also shows promise for solving three dimensional problems in critical phenomena more reliably than the ϵ expansion.

The development of a nonperturbative renormalization group method which can be solved in practice to a reasonable approximation is difficult. The obstacles to be overcome will be evident from the discussion of the two dimensional

Ising model (Sec. VI) and the Kondo problem (Sec. VII-IX). Nevertheless, the success of Kadanoff and Kadanoff and Houghton for the Ising model and the success of the Kondo calculations described here should encourage the study of other problems in the same spirit.

A principal part of this review is a detailed explanation of the renormalization group solution of the Kondo problem (in Sec. VII-IX). The Kondo problem is the problem of magnetic impurities in a metal. An example is copper with iron impurities, with concentrations of 0.01% or less of iron. At low temperatures (typically a few degrees K) there are strong effects due to the impurities, for example, in the resistivity of the alloy. See Kondo (1969). The especially difficult problem for theorists has been to predict the zero temperature behavior of the system, including the resistivity, the impurity susceptibility, and the impurity specific heat. The problem is caused by spin-spin coupling of the impurity spin to the spins of the conduction electrons. Without the spin-spin coupling, the electrons in the conduction band scatter independently from the impurities and this scattering is easily computed. The spin-spin coupling causes spin-flip scattering which in turn breaks the independence of the electrons (see Sec. VII).

Typically the spin-spin coupling strength J is small and antiferromagnetic. Therefore, one can analyze the effects of the spin-spin coupling by perturbation theory. Unfortunately, the terms of order J^2 involve a logarithm, namely $J^2 \ln(D/kT)$, where D is the maximum electron energy in the conduction band and T is the temperature (k is Boltzmann's constant). When the temperature goes to zero the logarithm becomes infinite and the perturbation series becomes nonsense.

The logarithm in perturbation theory reflects the existence of many energy scales in the conduction band; the logarithm comes from a logarithmic integral $\int dE/E$ with the lower bound on the energy being kT , and the upper bound being D . The low energy scales come from electrons very near the Fermi surface which require very little energy to be excited.

The Kondo problem has fascinated theorists more than is justified by its experimental significance. The reason for this is that the theoretical models involving a single impurity coupled to a free electron band look simple enough so that they ought to be soluble. Much effort has gone into studies of the Kondo problem using a variety of techniques: graph summation methods, Chew-Low theory, etc. [See Kondo (1969).] Perhaps the most successful of the various previous approaches is the analogy developed by Anderson, Yuval, and Hamann (1970a, 1970b) to a one dimensional Coulomb gas, combined with Monte Carlo calculations by Schotte and Schotte (1971). These papers give a picture of the low temperature behavior which is qualitatively in agreement with the precise calculations reported in this review. However, none of the approaches prior to the renormalization group were reliable—they all involved unverifiable approximations.

The simplest models of the Kondo problem are simple enough so that they can be solved with good precision (about 1% accuracy) using renormalization group methods. The details are reported in Sec. VII-IX of this review.

The first part of this review is a review of the essentials of renormalization group theory applied to critical phenomena using the diagrammatic renormalization group formulation. The review is not as detailed as the review of Wilson and Kogut (1974). It should, however, provide sufficient background for the sections on block spin techniques and the Kondo problem. Also, there are some differences in emphasis in the present review from Wilson and Kogut. In Sec. I, the idea of averaging over fluctuations is made precise and it is shown how this leads into the diagrammatic framework. In Sec. II the fixed points of the diagrammatic theory are discussed. In particular, the Gaussian fixed point (see Sec. II for the definition of "Gaussian") and fixed points near the Gaussian will be discussed. In Sec. III the linearized equations about these fixed points and some of the eigenvalues λ_i are computed. The relation of one of these eigenvalues to physics (through a critical exponent) is explained.

Section IV is concerned with the relation of the renormalization group ideas to the problems of divergences in quantum field theory and critical phenomena. There are differences between the limit of zero lattice spacing (relevant to field theory) and the limit of infinite correlation length with a fixed lattice spacing. These differences will be discussed. Then the outline of a theory of the statistical continuum limit (i.e., renormalization) in field theory is presented in the renormalization group framework.

Section V is devoted to a problem which is special and yet very important, namely the theory of marginal variables. A marginal variable is one which is neither amplified nor deamplified by a renormalization group transformation. A marginal variable causes special problems. Suppose it represents a 1% correction in each of more than 10 000 iterations of the renormalization group calculation. Nonlinear effects of the marginal variable, which are of order 0.01% for a single interaction, then may become 100% effects when accumulated over 10 000 iterations. These nonlinear effects will be discussed in Sec. V.

The study of marginal variables is important background for parts of the Kondo calculation. In addition, the field theoretic methods of Gell-Mann and Low and Callan and Symanzik are based *entirely* on the special behavior of marginal variables. The relation to the Gell-Mann-Low theory is also explained in Sec. V.

In Sec. VI the theory of block spin transformations is explained with a particular example applied to the two dimensional Ising model. Near the end of this section, the results of some computer calculations using block spin methods are reported.

Sections VII to IX include the author's calculations for the Kondo problem described in detail. The solution of the Kondo problem involves many aspects of renormalization group theory, from fixed points and marginal operators to nonperturbative methods. There are also many special tricks which help to make the calculation practical; these are also explained.

An extensive list of references on the renormalization group approach has been given in Wilson and Kogut (1974). Further references can be found in Science Citation

Index [through citations of Wilson and Kogut (1974), for example]. The references in this paper are limited to the particular topics discussed.

Further or alternative review papers on the renormalization group can be classified as follows: There are two short, intuitive introductions based on colloquium talks. Wilson (1974b) explains the application to critical phenomena and the ϵ expansion. Wilson (1975) explains the basic ideas of the nonperturbative renormalization group approach using a watered-down form of the Kondo problem as an example. For reviews at the same level as the present review, there are Wilson and Kogut (1974), Ma (1973), Fisher (1974), the book by Toulouse and Pfeuty (1975), and a forthcoming book in the Domb and Green Series [Domb and Green (1975).] See further Wegner (1975) and Brout (1974). These reviews are all concerned with critical phenomena. For the block spin methods there is a recent review by Niemeier and Van Leeuwen (1975). There is no other review of the renormalization group solution of the Kondo problem reported in Sec. VII-XI except an earlier cryptic report by Wilson (1974a).

Renormalization group theory is technically more demanding than the theory of derivatives or Feynman diagrams. However, most of the unsolved problems in physics and theoretical chemistry are of the kind the renormalization group is intended to solve (other kinds of problems usually do not remain unsolved for long). It is likely that there will be a vast extension of the renormalization group over the next decade as the methods become more clever and powerful; there are very few areas in either elementary particle physics, solid state physics, or theoretical chemistry that are permanently immune to this infection.

I. DEFINITION OF A RENORMALIZATION GROUP TRANSFORMATION

A diagrammatic formulation of the renormalization group was introduced by [Wilson and Kogut (1974)]. In this section an alternative and more general motivation will be given for this formulation of the renormalization group than was provided in [Wilson and Kogut (1974)].

If one considers water, say, far from the critical point, there are microscopic fluctuations on an atomic scale (wavelengths of order 1 Å). If one increases the temperature and pressure towards the critical point, fluctuations (density fluctuations) become important at larger wavelengths. Sufficiently close to T_c and P_c (critical temperature and pressure) there are fluctuations on the scale of 100–10 000 Å which scatter ordinary light, and the water looks milky (this is critical opalescence). However the microscopic fluctuations (~ 1 Å) have not decreased in size: close to the critical point one has fluctuations at all wavelengths from 1 Å up to the correlation length ξ ; at the critical point $\xi = \infty$.

Most theoretical methods for handling fluctuations fail near the critical point. The reason is that most techniques require that only one order of magnitude range of wavelengths be important. This range could be the range ~ 1 Å (microscopic) or perhaps the range of wavelengths $\sim \xi$. There is trouble when all wavelengths between 1 Å and ξ

are important, too. (In field-theoretic diagrammatic methods the trouble takes the form of divergences when $\xi \rightarrow \infty$.)

The renormalization group approach is designed to handle fluctuations over many wavelengths. The renormalization group strategy is to divide the full range of wavelengths into subranges of manageable proportions and consider each subrange in sequence. For example, one can consider separately the ranges of wavelengths 1–2 Å, 2–4 Å, 4–8 Å, etc.

In statistical mechanics the properties of the system are determined by a partition function Z which is a sum over all possible configurations of the system. Taking into account a given range of wavelengths of fluctuations, say 10–20 Å, means in some sense performing the statistical averaging over fluctuations of these wavelengths. In this section a precise definition of “averaging over fluctuations” will be given; this will lead to a specific formulation of the renormalization group. For a more qualitative discussion see Wilson (1974a).

First it is useful to define a statistical averaging over all fluctuations with wave number greater than a cutoff Λ , i.e., wavelengths $< 2\pi/\Lambda$. It will be assumed that the wavelength cutoff $2\pi/\Lambda$ is much larger than 1 Å so that all microscopic fluctuations are included in this initial averaging. What this means intuitively is this. Consider the example of water. Consider a region of size $\sim 1/\Lambda$ in the water surrounding a point \mathbf{x} . Let $\rho(\mathbf{x})$ be the average density of this region. Then one would like to perform the statistical averaging over density fluctuations inside the region keeping the average density $\rho(\mathbf{x})$ fixed. Such fluctuations necessarily have wavelengths $\lesssim 1/\Lambda$. In a magnetic system one considers an analogous region and defines $M(\mathbf{x})$ to be the average magnetization of the region. One would like to perform the statistical average over all spin fluctuations in this region holding $M(\mathbf{x})$ fixed. The entire magnet can be divided into subregions of size $\sim 1/\Lambda$; and in each subregion one wants to perform the average over fluctuations which leave unchanged the average magnetization of the region.

This idea can be realized formally in the following fashion. Suppose, to be specific, that one starts with a magnetic system consisting of a lattice of spins s_n (spaced at about 1 Å). Let M_k be the Fourier transform of these spins

$$M_k = \sum_n \exp(i\mathbf{k} \cdot \mathbf{n}) s_n. \quad (I.1)$$

We can define an average magnetization density for regions of size $\sim 1/\Lambda$ by defining

$$M(\mathbf{x}) = \int_0^\Lambda \exp(-i\mathbf{k} \cdot \mathbf{x}) M_k d^3k / (2\pi)^3. \quad (I.2)$$

By only integrating to Λ instead of the maximum possible k one produces an $M(\mathbf{x})$ which involves only long wavelength fluctuations. This means $M(\mathbf{x})$ itself cannot vary much for changes in \mathbf{x} much smaller than the wavelength cutoff $1/\Lambda$: it acts much like an average magnetization. One would now like to perform a statistical mechanical averaging

holding $M(\mathbf{x})$ fixed for all \mathbf{x} . We can do this by holding M_k fixed for $|\mathbf{k}| < \Lambda$. This means we write

$$\exp(\mathcal{H}_\Lambda[M]) = \sum_{\{s_n\}} \left\{ \prod_{k=0}^\Lambda \delta(M_k - \sum_n \exp(i\mathbf{k} \cdot \mathbf{n}) s_n) \right\} \times \exp(-H_0/kT), \quad (I.3)$$

where $\sum_{\{s_n\}}$ is a sum over all configurations of the spins s_n , for all lattice sites \mathbf{n} . H_0 is the Hamiltonian of the system, and \mathcal{H}_Λ is an effective interaction depending on the variables M_k (with $0 < |\mathbf{k}| < \Lambda$), i.e. on the average magnetization function $M(\mathbf{x})$.

[A technical complication occurs here, namely that

$$\sum_n \exp(i\mathbf{k} \cdot \mathbf{n}) s_n$$

is complex. So one must define the δ function

$$\delta[M_k - \sum_n \exp(i\mathbf{k} \cdot \mathbf{n}) s_n].$$

This is done by defining the product

$$\delta[M_k - \sum_n \exp(i\mathbf{k} \cdot \mathbf{n}) s_n] \delta[M_{-k} - \sum_n \exp(-i\mathbf{k} \cdot \mathbf{n}) s_n]$$

to be

$$\delta[\text{Re}M_k - \sum_n \cos(\mathbf{k} \cdot \mathbf{n}) s_n] \delta[\text{Im}M_k - \sum_n \sin(\mathbf{k} \cdot \mathbf{n}) s_n].$$

Another problem is that for an infinite lattice the product \prod_k is a continuous product; one can avoid this problem by considering a large but finite lattice. See below.]

The partition function Z is originally defined as

$$Z = \sum_{\{s_n\}} \exp(-H_0/kT). \quad (I.4)$$

One can compute Z from the effective interaction \mathcal{H}_Λ by averaging over the long wavelength fluctuations M_k

$$Z = \prod_{k=0}^\Lambda \int_{-\infty}^{\infty} dM_k \exp[\mathcal{H}_\Lambda(M)]. \quad (I.5)$$

[Here also a definition is needed, namely

$$\int_{-\infty}^{\infty} dM_k \int_{-\infty}^{\infty} dM_{-k}$$

is a shorthand for

$$\int_{-\infty}^{\infty} d(\text{Re}M_k) \int_{-\infty}^{\infty} d(\text{Im}M_k).]$$

For a system of finite size there are a discrete set of allowed values for \mathbf{k} and the product $\prod_{k=0}^\Lambda$ is an ordinary product over all allowed vectors \mathbf{k} with magnitude less than Λ . In the infinite volume limit, \mathbf{k} is a continuous variable and Eq. (I.5) involves a functional integral. See Wilson and Kogut (1974) for background on functional integrals.

The renormalization group transformation can now be defined as a formula relating $\mathcal{H}_{\Lambda/2}$ to \mathcal{H}_Λ . One has

$$\exp(\mathcal{H}_{\Lambda/2}[M]) = \prod_{|k| > \Lambda/2} \int_{-\infty}^{\infty} dM_k \exp(\mathcal{H}_\Lambda[M]), \quad (I.6)$$

where \mathcal{H}_Λ is a functional (for the infinite volume case) depending on functions M_k ($0 < k < \Lambda$) and $\mathcal{H}_{\Lambda/2}[M]$ depends on M_k for $0 < k < \Lambda/2$. The factor 2 is arbitrary: one could also calculate $\mathcal{H}_{\Lambda/b}$ for any $b > 1$. This is not yet the renormalization group transformation. This is obtained by converting to dimensionless variables. For a given cutoff Λ one introduces a dimensionless momentum variable $\mathbf{q} = \mathbf{k}/\Lambda$, so that the range of \mathbf{q} is always $0 < |\mathbf{q}| < 1$. Furthermore one replaces M_k by a rescaled variable σ_q

$$M_k = z_\Lambda \sigma_q. \quad (I.7)$$

The scale factor z_Λ is introduced so that one can adjust the average magnitude of the fluctuation σ_q for some value of \mathbf{q} , to be of order 1 regardless of the average magnitude of M_k . Since the conversion to dimensionless variables is different for different Λ , one has different variables, say σ_q and σ'_q , appearing in \mathcal{H}_Λ and $\mathcal{H}_{\Lambda/2}$. These variables are related by

$$\sigma_q = \zeta \sigma'_{2q}, \quad (I.8)$$

where

$$\zeta = z_{\Lambda/2}/z_\Lambda. \quad (I.9)$$

The renormalization group transformation is precisely

$$\exp(\mathcal{H}_{\Lambda/2}[\sigma']) = \prod_{|\mathbf{q}| < 1} \int_{-\infty}^{\infty} d\sigma_q \exp(\mathcal{H}_\Lambda[\sigma]) \quad (I.10)$$

with σ' defined as above (Eq. I.8). The constant ζ is a parameter that will be determined later.

[A constant factor has been omitted from Eq. (I.10), namely the constant generated by changing the variables of integration from M_k to σ_q . The constant has been omitted because in this paper we will only be concerned with correlation functions; these involve statistical expectation values for which such constants are irrelevant. If one tries to calculate the free energy itself then such constants must be taken into account.]

The purpose of the change of variables is to make possible the existence of fixed points. A fixed point is a specific functional \mathcal{H}^* such that

$$\exp(\mathcal{H}^*[\sigma']) = \prod_{|\mathbf{q}| < 1} \int_{-\infty}^{\infty} d\sigma_q \exp(\mathcal{H}^*[\sigma]). \quad (I.11)$$

To have the same functional on both sides means in particular that σ_q and σ'_q must have the same range of \mathbf{q} ; they also must have the same average magnitude for the same value of \mathbf{q} . The change of momentum scale and the factor ζ are both needed to make this possible. This will be seen explicitly below.

What can one say about the interaction $\mathcal{H}_\Lambda[\sigma]$? Since they are defined by a complicated configuration sum one can make only general statements. It will be assumed that when σ is small, $\mathcal{H}_\Lambda[\sigma]$ has a power series expansion in σ .

In the case of a magnetic system with no external field there is a symmetry for $\sigma \rightarrow -\sigma$ which means only even terms in σ appear. Then the expansion can be written

$$\begin{aligned} \mathcal{H}_\Lambda[\sigma] = & -\frac{1}{2} \int_0^1 \int_0^1 \int_0^1 \mathbf{q} u_2(\mathbf{q}) \sigma_{\mathbf{q}} \sigma_{-\mathbf{q}} - \int_{\mathbf{q}_1} \int_{\mathbf{q}_2} \int_{\mathbf{q}_3} \\ & \times u_4(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, -\mathbf{q}_1 - \mathbf{q}_2 - \mathbf{q}_3) \sigma_{\mathbf{q}_1} \sigma_{\mathbf{q}_2} \sigma_{\mathbf{q}_3} \sigma_{-\mathbf{q}_1 - \mathbf{q}_2 - \mathbf{q}_3} \\ & - \int_{\mathbf{q}_1} \cdots \int_{\mathbf{q}_5} u_6(\mathbf{q}_1, \dots, \mathbf{q}_5, -\mathbf{q}_1 - \cdots - \mathbf{q}_5) \\ & \times \sigma_{\mathbf{q}_1} \cdots \sigma_{\mathbf{q}_5} \sigma_{-\mathbf{q}_1 - \cdots - \mathbf{q}_5} - \cdots, \end{aligned} \quad (I.12)$$

where $u_2(\mathbf{q})$, $u_4(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4)$ etc. are unknown functions (which also depend on Λ); there are only three independent momenta in u_4 due to momentum conservation (which is a consequence of translational invariance in position space). All integrals are restricted so that $|\mathbf{q}_i| < 1$ (note that $|\mathbf{q}_1 + \mathbf{q}_2 + \mathbf{q}_3|$ must also be less than 1 in u_4 , etc.). The functions $u_4(\mathbf{q}_1, \dots, \mathbf{q}_4)$ etc. will be required to be symmetric under interchanges of the \mathbf{q}_i (as one is free to demand); for $u_2(\mathbf{q})$ this means $u_2(\mathbf{q}) = u_2(-\mathbf{q})$.

A second assumption is that $\mathcal{H}_\Lambda[\sigma]$ has no long range forces. If critical behavior is caused by long wavelength fluctuations then hopefully no long range correlations are introduced by integrating out only the short wavelength fluctuations. A "long range force" means here a term in x space falling like a power of x , say

$$\int_x \int_y M(\mathbf{x}) M(\mathbf{y}) |\mathbf{x} - \mathbf{y}|^{-m}$$

for any m . Long range tails in position space are associated with singularities in momentum space, in this case a term $|\mathbf{q}|^{m-d}$ in $u_2(\mathbf{q})$ (if $m-d$ is 0 or a positive even integer then the term is $|\mathbf{q}|^{m-d} \ln |\mathbf{q}|$ which is still singular).

The absence of long range forces is one of the most tenuous of the assumptions of the renormalization group approach. For the particular renormalization group transformation described here, the sharp cutoff in momentum space leads to some particular long range terms, most notably a $|\mathbf{q}|$ term in $u_2(\mathbf{q})$. These terms are, I believe, manageable and are usually rather small. They will be ignored in this paper. See also Sak (1973). Otherwise no long range force terms occur in the explicit perturbation theory calculations of Sec. II-III.

Another general assumption is that the effective interaction $\mathcal{H}_\Lambda[\sigma]$ depends analytically on temperature and the parameters in the microscopic interaction H_0 . That is, the functions $u_2(\mathbf{q})$, $u_4(\mathbf{q}_1, \dots, \mathbf{q}_4)$, etc. will all be functions of T and any parameters in H_0 ; it is assumed that this dependence is analytic even at $T = T_c$ (for fixed Λ) where the thermodynamic quantities are nonanalytic functions of T . This will be discussed further in Sec. II.

In conclusion, the effective interactions $\mathcal{H}_\Lambda[\sigma]$ are assumed to be triply analytic: firstly, analytic in σ itself, e.g., no fractional powers of σ appear in \mathcal{H}_Λ ; secondly, the coefficient functions u_2 , u_4 , etc. analytic functions of momenta (apart from nonanalyticity caused by the sharp

momentum cutoff Λ). Finally, the coefficient functions u_2 , u_4 , etc. are analytic in the temperature T and parameters in the original interaction H_0 . All these assumptions are made even at $T = T_c$ where the thermodynamics is non-analytic. All these assumptions are verified within the perturbation theoretic calculations of Secs. II and III, but remain unproven for nonperturbative formulations of the renormalization group (see Sec. VI). See Wilson and Kogut (1974) and Wilson (1974a) for more discussion.

The effective interactions $\mathcal{H}_\Lambda[\sigma]$ have been defined as intermediate stages in the calculation of the partition function Z . They can also be used in the calculation of correlation functions. Consider the spin-spin correlation function of the original magnetic system

$$\Gamma_n = \frac{1}{Z} \sum_{\{s_n\}} s_n s_0 \exp(-H_0/kT). \quad (I.13)$$

Let $\Gamma_k = \sum_n \exp(i\mathbf{k} \cdot \mathbf{n}) \Gamma_n$ be the Fourier transform of Γ_n . One can obtain Γ_k directly as

$$\Gamma_k = \frac{1}{Z} \frac{1}{\delta(\mathbf{k} + \mathbf{k}_1)} \sum_{\{s_n\}} M_k M_{k_1} \exp(-H_0/kT) \quad (I.14)$$

[where M_k is the Fourier transform of s_n , and $\delta(\mathbf{k})$ means $(2\pi)^d \delta^d(\mathbf{k})$]. For $|\mathbf{k}| < \Lambda$, the quantities M_k and M_{k_1} ($\mathbf{k}_1 = -\mathbf{k}$) are held fixed in the integrations defining \mathcal{H}_Λ . Therefore one can calculate the expectation value of $M_k M_{k_1}$ using \mathcal{H}_Λ instead of H_0

$$\Gamma_k = \frac{1}{Z} \frac{1}{\delta(\mathbf{k} + \mathbf{k}_1)} \prod_{k'=0}^{\Lambda} \int dM_{k'} \{M_k M_{k_1} \exp(\mathcal{H}_\Lambda)\}. \quad (I.15)$$

It is convenient to define correlation functions for the rescaled spin variables σ_q , namely

$$\Gamma_{\Lambda,q} = \frac{1}{\delta(\mathbf{q} + \mathbf{q}_1)} \frac{1}{Z_\Lambda} \prod_{q'=0}^1 \int d\sigma_{q'} \sigma_q \sigma_{q_1} \exp(\mathcal{H}_\Lambda[\sigma]), \quad (I.16)$$

with

$$Z_\Lambda = \prod_{q=0}^1 \int d\sigma_q \exp(\mathcal{H}_\Lambda[\sigma]). \quad (I.17)$$

Then

$$\Gamma_k = z_\Lambda^2 \Lambda^d \Gamma_{\Lambda,q} \quad (\mathbf{q} = \mathbf{k}/\Lambda) \quad (k < \Lambda) \quad (I.18)$$

[the factor Λ^d comes from $\delta(\mathbf{k} + \mathbf{k}_1) = \delta(\Lambda\mathbf{q} + \Lambda\mathbf{q}_1) = \Lambda^{-d} \delta(\mathbf{q} + \mathbf{q}_1)$]. A related result is that the correlation function $\Gamma_{\Lambda,q}$ for $q < \frac{1}{2}$ can be expressed in terms of $\Gamma_{\Lambda/2,2q}$

$$\Gamma_{\Lambda,q} = 2^{-d} \zeta^2 \Gamma_{\Lambda/2,2q}. \quad (I.19)$$

Note that any constant term in \mathcal{H}_Λ (a term independent of σ) cancels in the calculation of $\Gamma_{\Lambda,q}$; this is the reason constant terms in \mathcal{H}_Λ are ignored in this paper.

In the special case that $\mathcal{H}_\Lambda[\sigma]$ involves only a quadratic term u_2 (i.e., u_4 , etc. are all 0), one can calculate $\Gamma_{\Lambda,q}$ explicitly

$$\Gamma_{\Lambda,q} = \frac{1}{u_2(\mathbf{q})} \quad (I.20)$$

[see Wilson and Kogut (1974)]. If u_4 , etc. are nonzero then $\Gamma_{\Lambda,q}$ has to be calculated by approximate procedures, and

if \mathcal{H}_Λ is a critical or near critical interaction this is difficult to do for the reasons already explained.

A quantity of particular interest is the correlation length ξ . This is defined in terms of the behavior of Γ_n

$$\Gamma_n \sim \exp(-|\mathbf{n}|/\xi) \quad (I.21)$$

for large $|\mathbf{n}|$, apart from powers of $|\mathbf{n}|$ (if Γ_n depends on the direction of the vector \mathbf{n} then one chooses the direction that maximizes ξ). The behavior of Γ_n for large $|\mathbf{n}|$ is related to the singularities of its Fourier transform Γ_k : for Γ_n to behave as in (I.21), Γ_k must have a singularity at $|\mathbf{k}| = i\xi^{-1}$. Consider now the correlation length ξ_Λ for the interaction \mathcal{H}_Λ (the location of the leading singularity of $\Gamma_{\Lambda,q}$ is at $|\mathbf{q}| = i\xi_\Lambda^{-1}$). From Eq. (I.18) one sees that $\Gamma_{\Lambda,q}$ has a singularity at $\mathbf{q} = \mathbf{k}/\Lambda$ if Γ_k has a singularity at \mathbf{k} , and hence

$$\xi_\Lambda = \Lambda \xi. \quad (I.22)$$

Therefore as one lowers the momentum cutoff Λ of \mathcal{H}_Λ , its correlation length also decreases. Note that ξ_Λ is a dimensionless correlation length, since it is defined in terms of the dimensionless variable \mathbf{q} . What ξ_Λ determines in physical terms is the correlation length of \mathcal{H}_Λ in units of the minimum wavelength Λ^{-1} permitted in \mathcal{H}_Λ .

Note also that

$$\xi_{(\Lambda/2)} = \frac{1}{2} \xi_\Lambda. \quad (I.23)$$

II. FIXED POINTS IN PERTURBATION THEORY

The formalism of the renormalization group approach begins with the idea of a fixed point of a renormalization group transformation, an example being a fixed interaction \mathcal{H}^* satisfying the transformation of Sec. I [see Eq. (I.11)]. Then one discusses the behavior of solutions \mathcal{H}_Λ of the transformation near a fixed point. This formalism will be illustrated in Sec. II and III by studying fixed points which are either Gaussian [only $u_2(q)$ is nonzero] or near to Gaussian (u_4 , u_6 , etc. present but small). The restriction to small non-Gaussian terms is a practical restriction; otherwise, the transformation of Sec. I is too complicated to solve (except by the rough procedures of the approximate recursion formula [Wilson (1971a), Wilson and Kogut (1974)]. The most important of the near Gaussian fixed points occurs near four space dimensions; this fixed point and its applications are discussed extensively in Wilson and Kogut (1974) and elsewhere. In this paper the more general question of when there are near-Gaussian fixed points will be discussed, along with some of their properties. Detailed applications will not be explained.

Only under special circumstances are fixed points near-Gaussian. It is important to be able to calculate fixed points even if they are far from Gaussian. Some current efforts to make such calculations are discussed in Sec. VI.

In this section the transformation of Sec. I will be written out in detail for near-Gaussian interactions. The perturbation theory which yields these results is explained in Wilson and Kogut (1974) and will not be repeated here. Then the Gaussian and near-Gaussian fixed points of the transformation will be considered.

Write the transformation in the form

$$\mathcal{K}'[\sigma'] = T[\mathcal{K}[\sigma], \zeta], \tag{II.1}$$

where \mathcal{K}' involves primed functions $u_2', u_4',$ etc. Then $u_2'(\mathbf{q}),$

$u_4'(\mathbf{q}_1, \dots, \mathbf{q}_4),$ etc., are given by a sum of Feynman graphs; the graphs for u_2' have two external lines, the graphs for u_4' have four external lines, etc. The first few graphs give the following equations:

$$u_2'(\mathbf{q}) = \zeta^2 2^{-d} \left\{ u_2 \left(\frac{\mathbf{q}}{2} \right) + 12 \int_{\frac{1}{2}}^1 \mathbf{p} \frac{u_4(\mathbf{q}/2, -\mathbf{q}/2, \mathbf{p}, -\mathbf{p})}{u_2(\mathbf{p})} + O(u_4^2, u_6, \text{etc.}) \right\}, \tag{II.2}$$

$$u_4'(\mathbf{q}_1, \dots, \mathbf{q}_4) = \zeta^4 2^{-3d} \left\{ u_4 \left(\frac{1}{2}\mathbf{q}_1, \frac{1}{2}\mathbf{q}_2, \frac{1}{2}\mathbf{q}_3, \frac{1}{2}\mathbf{q}_4 \right) - 12 \int_{\frac{1}{2}}^1 \mathbf{p} \frac{u_4 \left(\frac{1}{2}\mathbf{q}_1, \frac{1}{2}\mathbf{q}_2, \mathbf{p}, -\mathbf{p} - \frac{1}{2}\mathbf{q}_1 - \frac{1}{2}\mathbf{q}_2 \right) u_4 \left(\frac{1}{2}\mathbf{q}_3, \frac{1}{2}\mathbf{q}_4, -\mathbf{p}, \mathbf{p} + \frac{1}{2}\mathbf{q}_1 + \frac{1}{2}\mathbf{q}_2 \right)}{u_2(\mathbf{p}) u_2(-\mathbf{p} - \frac{1}{2}\mathbf{q}_1 - \frac{1}{2}\mathbf{q}_2)} \right. \\ \left. + \text{permutations} + 15 \int_{\frac{1}{2}}^1 \mathbf{p} \frac{u_6 \left(\frac{1}{2}\mathbf{q}_1, \frac{1}{2}\mathbf{q}_2, \frac{1}{2}\mathbf{q}_3, \frac{1}{2}\mathbf{q}_4, \mathbf{p}, -\mathbf{p} \right)}{u_2(\mathbf{p})} + O(u_4^3, u_4, u_6, u_8, \text{etc.}) \right\}. \tag{II.3}$$

$$u_6'(\mathbf{q}_1, \dots, \mathbf{q}_6) = \zeta_6^2 2^{-5d} \left\{ u_6 \left(\frac{1}{2}\mathbf{q}_1, \dots, \frac{1}{2}\mathbf{q}_6 \right) + O(u_4^2, u_4, u_6, u_8, \text{etc.}) \right\} \tag{II.4}$$

etc.

The first term in each of these formulae is trivial; for example, the u_4 term in the equation for u_4' is obtained by substituting $\zeta \sigma_{2q'}$ for σ_q in the expression

$$\int_{\mathbf{q}_1} \int_{\mathbf{q}_2} \int_{\mathbf{q}_3} u_4(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, -\mathbf{q}_1 - \mathbf{q}_2 - \mathbf{q}_3) \sigma_{\mathbf{q}_1} \sigma_{\mathbf{q}_2} \sigma_{\mathbf{q}_3} \sigma_{-\mathbf{q}_1 - \mathbf{q}_2 - \mathbf{q}_3}$$

(with all q 's restricted to be less than $\frac{1}{2}$) and changing variables so that \mathbf{q}_i is replaced by $\mathbf{q}_i/2$. The remaining terms come from diagrams: see Ref. 2.

As a first step in considering these rather complex equations consider what happens when $u_4, u_6,$ etc. are all 0, leaving only u_2 . In this case only u_2' is nonzero. When the transformation T is iterated one generates a sequence of functions $u_{2l}(q), (l = 0, 1, 2, \text{etc.})$. Suppose the initial cutoff Λ_0 associated with $u_{20}(\mathbf{q})$ is $(10 \text{ \AA})^{-1}$; then the cutoff Λ_l associated with u_{2l} is $2^{-l}\Lambda_0$, or $(2^l \times 10 \text{ \AA})^{-1}$. Near the critical point one is interested primarily in large wavelengths and therefore large l .

Let ζ be a fixed parameter independent of l . In this case the recursion formula for u_{2l} (dropping $u_4, u_6,$ etc.) is

$$u_{2, l+1}(\mathbf{q}) = \zeta^2 2^{-d} u_{2l} \left(\frac{\mathbf{q}}{2} \right). \tag{II.5}$$

The solution of this recursion formula is

$$u_{2l}(\mathbf{q}) = \zeta^{2l} 2^{-ld} u_{20} \left(\frac{\mathbf{q}}{2^l} \right). \tag{II.6}$$

The precise form of $u_{20}(q)$ is not known due to the complicated definition of \mathcal{K}_Λ (see Sec. I). However, for l large, one needs to know $u_{20}(\mathbf{q})$ only for $|\mathbf{q}| < 2^{-l}$. Assuming no long range forces are present in $u_{20}(\mathbf{q})$ (see Sec. I) one can expand $u_{20}(\mathbf{q})$ in powers of q . For simplicity suppose $u_{20}(\mathbf{q})$ depends only on q^2 . Then

$$u_{20}(q) = r_0 + z_0 q^2 + w_0 q^4 + \dots \tag{II.7}$$

This gives

$$u_{2l}(q) = \zeta^{2l} 2^{-ld} (r_0 + z_0 2^{-2l} q^2 + w_0 2^{-4l} q^4 + \dots). \tag{II.8}$$

If r_0 is not zero then $u_{2l}(q)$ is approximately a constant independent of q , for large l . In this case the correlation function $\Gamma_{l,q}$ for the interaction \mathcal{K}_l is also a constant approximately, and in particular has no singularity at $q = 0$ [see Eq. (I.20)]. Therefore one is not at a critical point.

The next possibility is that r_0 vanishes but $z_0 \neq 0$. Then for large enough l ,

$$u_{2l}(q) \simeq \zeta^{2l} 2^{-ld} 2^{-2l} z_0 q^2 \tag{II.9}$$

and if one chooses

$$\zeta = 2^{(d+2)/2} \tag{II.10}$$

one has

$$u_{2l}(q) \simeq z_0 q^2 \tag{II.11}$$

for large l , namely $u_{2l}(q)$ becomes independent of l for large l . This is an example of a fixed point. In fact the formula $u_{2l}(q) = z_0 q^2$ for any l is a special solution of the transformation Eq. (II.5), provided ζ is given by (II.10).

The correlation function corresponding to $u_{2l}(q) = z_0 q^2$ is given by

$$\Gamma_{l,q} = 1/(z_0 q^2). \tag{II.12}$$

The singularity at $\mathbf{q} = 0$ in $\Gamma_{l,q}$ means there is a singularity at $\mathbf{k} = 0$ in the original correlation function $\Gamma_{\mathbf{k}}$ (defined in Sec. I). Hence one is at a critical point.

The constant r_0 in $u_{20}(q)$ will not be zero, in general. To make r_0 be zero one must restrict in some way the initial microscopic interaction H_0 or temperature T . Since $r_0 = 0$ corresponds to a critical point (assuming the absence of $u_4, u_6,$ etc.), the corresponding restriction on (H_0, T) is that it also be at a critical point, i.e., one must put the temperature T equal to T_c .

There are higher order fixed point solutions, namely $u_{2l}(q) = w_0 q^4$, or $w_1 q^6$, etc. (each requiring an appropriate choice of ζ). The fixed point $w_0 q^4$ corresponds to a $(\nabla^2 M)^2$ term in a Landau theory replacing the usual $(\nabla M)^2$ term (see, e.g., [Wilson (1974b)]). To my knowledge these higher order fixed points have not been studied in detail.

Now consider fixed points with $u_4, u_6,$ etc. nonzero. Consider in particular those points for which $u_4, u_6,$ etc., are small so that one needs to consider only a few low order diagrams in constructing the fixed point.

A fixed point is a value ζ^* for ζ plus a set of functions $u_2^*(\mathbf{q}), u_4^*(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4),$ etc. which reproduce themselves through Eqs. (II.2)–(II.4). If $u_4^*, u_6^*,$ etc. are all small

then to a first approximation one can neglect u_4^* , u_6^* , etc. in the equation for u_2^* . One then has

$$u_2^*(\mathbf{q}) \simeq \zeta^{*2} 2^{-d} \{u_2^*(\mathbf{q}/2)\}. \quad (\text{II.13})$$

The fixed points of this equation have already been determined and only the solution $u_2^*(\mathbf{q}) \simeq z_0 q^2$ [with ζ^* given by Eq. (II.10)], will be considered here. There will be corrections to both $u_2^*(\mathbf{q})$ and ζ^* due to the u_4^* terms, etc. in Eq. (II.2), but these corrections will be assumed to be small.

The equation for u_4^* , to a first approximation, reads

$$u_4^*(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4) \simeq 2^{4-d} u_4^*(\mathbf{q}_1/2, \mathbf{q}_2/2, \mathbf{q}_3/2, \mathbf{q}_4/2). \quad (\text{II.14})$$

[Eq. (II.10) has been used for ζ^* .] One seeks solutions which are analytic in the \mathbf{q}_i 's for $\mathbf{q}_i \rightarrow 0$. It is not difficult to see that the solutions to this equation are monomials in the q_i , and that for each possible monomial, d is determined. For example, consider the case $u_4^*(q_1, \dots, q_4) \simeq u^*$, a constant. Then one has

$$u^* \simeq 2^{4-d} u^* \quad (\text{II.15})$$

which is possible only if $d \simeq 4$. Alternatively one might have $u_4^*(\mathbf{q}_1, \dots, \mathbf{q}_4) \simeq w^*(q_1^2 + q_2^2 + q_3^2 + q_4^2)$, giving

$$w^* \simeq 2^{2-d} w^*. \quad (\text{II.16})$$

[$u_4^*(\mathbf{q}_1, \dots, \mathbf{q}_4)$ cannot have a term linear in the \mathbf{q}_i : the requirements of symmetry in the \mathbf{q}_i plus momentum conservation rules out a linear term.] This equation can be satisfied for $d \simeq 2$. However the case $d \simeq 2$ is very awkward to discuss using the diagrammatic expansion of the renormalization group transformation; it will not be considered here.

Suppose one cannot satisfy Eq. (II.14), i.e., d is not near 2 or 4. There still can be fixed point solutions with u_4^* nonzero; however, there must be diagrams in the equation for u_4^* which are as large as u_4^* itself. The diagrams that can be as large as u_4^* (assuming u_6^* , etc., are also small) are the diagrams linear in u_6^* , u_8^* , etc. not involving u_4^* . Suppose, for example, that u_6^* is important but u_8^* is negligible. Then Eq. (II.14) for u_4^* need not be satisfied. However, there is now a simple equation for u_6^*

$$u_6^*(\mathbf{q}_1, \dots, \mathbf{q}_6) \simeq 2^{6-2d} u_6^*(\mathbf{q}_1/2, \dots, \mathbf{q}_6/2). \quad (\text{II.17})$$

The only solution for $d > 2$ of this equation is

$$u_6^*(\mathbf{q}_1, \dots, \mathbf{q}_6) = v^*$$

where v^* is a constant, in which case d must be $d \simeq 3$. Similarly if u_8^* is important but u_{10}^* is negligible then there is a simple equation for u_8^* with solutions for $8 - 3d \simeq 0$ and so forth.

Consider more particularly the case $d \simeq 4$. One now has an approximate solution for $u_2^*(\mathbf{q})$, $u_4^*(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4)$, and ζ^* , namely

$$u_2^*(\mathbf{q}) \simeq z_0 q^2, \quad (\text{II.18})$$

$$u_4^*(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4) \simeq u^*, \quad (\text{II.19})$$

$$\zeta^* \simeq 2^{1+d/2}, \quad (\text{II.20})$$

with z_0 and u^* arbitrary so far (u^* being assumed to be small).

The size of u^* , i.e., the strength of u_4^* , can be determined by considering nonlinear terms in the equation for u_4^* . A full discussion of the equation is rather complicated (see Ref. 2), so here the equation will be simplified by considering only the term involving u_4^2 in Eq. (II.3), and only an order of magnitude discussion will be given. With $u_4^*(\mathbf{q}, \dots, \mathbf{q}_3)$ approximately a constant u^* , the equation for u_4^* reads approximately

$$u^* = 2^\epsilon u^* + O(u^{*2}), \quad (\text{II.21})$$

where $\epsilon = 4 - d$. When ϵ is small, $(1 - 2^\epsilon)u^*$ is approximately $\epsilon(\ln 2)u^*$, i.e., of order ϵu^* . Thus the equation requires ϵu^* to be of order u^{*2} , which means u^* must be of order ϵ . Hence, for small ϵ , u^* is small.

In an accurate calculation of the fixed point, the functions u_6^* , u_8^* , etc., are nonzero but of higher order in ϵ ; likewise $u_4^*(\mathbf{q}_1, \dots, \mathbf{q}_4)$ has momentum dependence in order ϵ^2 and $u_2^*(\mathbf{q})$ is more complicated than $z_0 q^2$. To illustrate how this happens, consider a simplified order of magnitude equation for u_6^* neglecting its momentum dependence. There is a term of order $(u_4^*)^2$ in the equation for u_6^* so one has

$$u_6^* = 2^{6-2d} u_6^* + O(u^{*2}). \quad (\text{II.22})$$

With $d \simeq 4$ this becomes

$$u_6^*(1 - \frac{1}{4}) = O(u^{*2}). \quad (\text{II.23})$$

With u^* of order ϵ , this makes u_6^* of order ϵ^2 . This difference between this equation and the equation for u^* is that the linearized equation for u_6^* is not almost satisfied, i.e., the coefficient of u_6^* is of order 1 instead of order ϵ . Hence u_6^* is of order ϵ^2 rather than of order ϵ .

The complete set of equations for $u_2^*(\mathbf{q})$, $u_4^*(\mathbf{q}_1, \dots, \mathbf{q}_4)$, etc. now determine these functions completely, to any order in ϵ , except for the one constant z_0 . The constant z_0 cannot be determined; instead one determines the parameter ζ^* . To see this we show first that z_0 cannot be determined. The reason is a symmetry of the full renormalization group equations. Let $u_2^*(\mathbf{q})$, $u_4^*(\mathbf{q}_1, \dots, \mathbf{q}_4)$, $u_6^*(\mathbf{q}_1, \dots, \mathbf{q}_6)$, etc. be an exact fixed point for some value ζ^* of ζ . Then the functions

$$u_2^*(\mathbf{q}, z) = z^2 u_2^*(\mathbf{q}), \quad (\text{II.24})$$

$$u_4^*(\mathbf{q}_1, \dots, \mathbf{q}_4, z) = z^4 u_4^*(\mathbf{q}_1, \dots, \mathbf{q}_4), \quad (\text{II.25})$$

$$u_6^*(\mathbf{q}_1, \dots, \mathbf{q}_6, z) = z^6 u_6^*(\mathbf{q}_1, \dots, \mathbf{q}_6), \quad (\text{II.26})$$

also define a fixed point for any value of z . The transformations (II.24) to (II.26) are equivalent to a change of scale of the spin variables ϕ_q . This result is obvious for the terms exhibited explicitly in Eqs. (II.2)–(II.4) and is easily seen to be true for more complicated diagrams because z 's associated with the internal lines of any vertex cancel against the z 's in the propagator (see [Wilson and Kogut (1974)] for the rules for diagrams).

The arbitrariness in the fixed point just demonstrated can be used to impose an arbitrary value for z_0 , say $z_0 = 1$. This is to be understood, more precisely, to mean that when $u_2^*(q)$ is expanded in powers of q , the term proportional to q^2 has coefficient 1.

Now suppose one expands the equation for $u_2^*(\mathbf{q})$ in powers of \mathbf{q} and collects all terms linear in q^2 . The general

structure of this equation is (with ζ^* left arbitrary)

$$q^2 \{1 - \zeta^{*2} 2^{-(d+2)}\} = O(u^{*2})q^2, \tag{II.27}$$

where one term of order u^{*2} comes from the diagram shown in Fig. 1 and there are other terms of order ϵ^2 or higher. This equation determines ζ^* ; due to the order u^{*2} terms, ζ^* differs from $2^{(d+2)/2}$ in order ϵ^2 .

For some calculations to follow one needs to know the explicit value of u^* to order ϵ . This is most easily obtained by defining u^* to be $u_4^*(0,0,0)$ and considering Eq. (II.3) for all external momenta equal to 0. It is shown in Wilson and Kogut (1974) that the u_6^* term for this case is of order ϵ^3 instead of ϵ^2 and therefore is unimportant. One is left with the u_4^* term to calculate; using the first order equations (II.18)–(II.20), one obtains

$$u^* = 2\epsilon u^* - 9u^{*2}C, \tag{II.28}$$

with

$$C = 4 \int_{\frac{1}{2}}^1 \frac{1}{p^4} \frac{d^d p}{(2\pi)^d}. \tag{II.29}$$

The nontrivial solution of this equation is

$$u^* = \epsilon \ln 2 / (9C) \text{ to order } \epsilon. \tag{II.30}$$

In four dimensions precisely ($\epsilon = 0$) u^* is zero which means there is only the trivial fixed point. It is only for $d \simeq 4$ but not equal to 4 that the above calculation gives a nontrivial fixed point with u_4^* , u_6^* , etc., small. As ϵ gets large, say for $d \rightarrow 3$, the fixed point presumably still exists but u_4^* , etc. are not small any more. This means one can no longer neglect complicated diagrams in the equations for u_4^* , u_6^* , etc., which makes precise calculations hopeless. However, calculations in powers of ϵ are feasible to order ϵ^3 or higher [Brezin *et al.* (1973)] and these calculations give rather good results for $\epsilon = 1$.

As discussed earlier there are also nontrivial fixed points with u_4^* , u_6^* , etc., small for $d \simeq 3$, $3d \simeq 8$, $4d \simeq 10$, etc. The implications of these fixed points will be discussed after considering linearized behavior about these fixed points: see Sec. III. These fixed points are discussed in Chang and Stanley (1973) and in Stephen and McCauley (1973).

III. LINEARIZED BEHAVIOR NEAR FIXED POINTS

The next topic to be considered is the solution of linearized renormalization group equations near a fixed point. In particular, the eigenvalues and eigenoperators of the

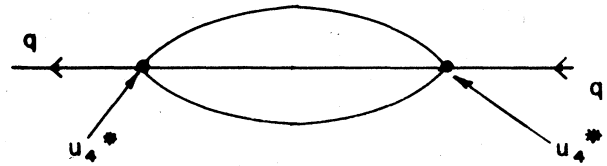


FIG.1. A diagram contributing to the q^2 term in $u_2^*(q)$.

linearized equations will be discussed for both the trivial and nontrivial fixed point. To start with, the general theory of linear perturbations about an arbitrary fixed point will be described. Then the example of the trivial fixed point will be discussed; this example will be used to show the relation between eigenvalues of the linearized transformation and the critical exponent ν . [The exponent ν gives the behavior of the correlation length ξ near the critical point, namely $\xi \propto (T - T_c)^{-\nu}$, where T_c is the critical temperature.] Finally some eigenvalues associated with the nontrivial fixed point near $d = 4$ will be calculated.

Suppose one has a fixed point $\mathcal{J}C^*$ of a renormalization group transformation T

$$\mathcal{J}C^* = T[\mathcal{J}C^*]. \tag{III.1}$$

Consider sequences of interactions $\mathcal{J}C_l$ generated by the transformation

$$\mathcal{J}C_{l+1} = T[\mathcal{J}C_l]. \tag{III.2}$$

By continuity, if $\mathcal{J}C_l$ is close to $\mathcal{J}C^*$ then $\mathcal{J}C_{l+1}$ will be close to $T[\mathcal{J}C^*]$, namely, close to $\mathcal{J}C^*$. Write

$$\mathcal{J}C_l - \mathcal{J}C^* = \delta\mathcal{J}C_l. \tag{III.3}$$

Consider

$$\delta\mathcal{J}C_{l+1} = T[\mathcal{J}C^* + \delta\mathcal{J}C_l] - \mathcal{J}C^*. \tag{III.4}$$

Expand this in powers of $\delta\mathcal{J}C_l$; one can then write

$$\delta\mathcal{J}C_{l+1} = L[\mathcal{J}C^*] \cdot \delta\mathcal{J}C_l + O[\delta\mathcal{J}C_l^2], \tag{III.5}$$

where $L[\mathcal{J}C^*]$ is a linear transformation, and the remainder is quadratic or higher in $\delta\mathcal{J}C_l$. To see what this equation means, it will be exhibited more explicitly in terms of u_2 , u_4 , etc. The fixed point $\mathcal{J}C_l$ is given by a set of functions $u_2^*(\mathbf{q})$, $u_4^*(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4)$, etc. The perturbation $\delta\mathcal{J}C_l$ is defined by another set of functions, say $\delta u_{2l}(\mathbf{q})$, $\delta u_{4l}(\mathbf{q}_1, \dots, \mathbf{q}_4)$, etc.; then $\mathcal{J}C_l$ itself is given by $u_2^* + \delta u_{2l}$, $u_4^* + \delta u_{4l}$, etc. The equations for $\delta u_{2, l+1}$, $\delta u_{4, l+1}$, etc. are obtained by substituting $u_2^* + \delta u_{2l}$ for u_2 , $u_4^* + \delta u_{4l}$ for u_4 , etc., in Eqs. (II.2)–(II.4). Consider in particular the equation for $u_{2, l+1}$. This now reads (assuming that ζ can be replaced by ζ^*)

$$u_2^*(\mathbf{q}) + \delta u_{2, l+1}(\mathbf{q}) = \zeta^{*2} 2^{-d} \left(u_2^*(\mathbf{q}/2) + \delta u_{2l}(\mathbf{q}/2) + 12 \int_{\frac{1}{2}}^1 \frac{u_4^*(\mathbf{q}/2, -\mathbf{q}/2, \mathbf{p}, -\mathbf{p}) + \delta u_{4l}(\mathbf{q}/2, -\mathbf{q}/2, \mathbf{p}, -\mathbf{p})}{[u_2^*(\mathbf{p}) + \delta u_{2l}^*(\mathbf{p})]} \right) + \text{contributions of other diagrams} \tag{III.6}$$

This equation can be expanded in powers of δu_{2l} , δu_{4l} , etc. The terms independent of the perturbation match by the

fixed point condition; the remaining terms give

$$\delta u_{2,l+1}(\mathbf{q}) = \zeta^{*2} 2^{-d} \left(\delta u_{2l}(\mathbf{q}/2) + 12 \int_{\frac{1}{2}}^1 \mathbf{p} \frac{\delta u_{4l}(\mathbf{q}/2, -\mathbf{q}/2, \mathbf{p}, -\mathbf{p})}{u_2^*(\mathbf{p})} - 12 \int_{\frac{1}{2}}^1 \mathbf{p} \frac{u_4^*(\mathbf{q}/2, -\mathbf{q}/2, \mathbf{p}, -\mathbf{p}) \delta u_{2l}(\mathbf{p})}{[u_2^*(\mathbf{p})]^2} + \text{other diagrams} + O(\delta^2) \right), \tag{III.7}$$

where $O(\delta^2)$ means terms of order $\delta u_{2l}^2, \delta u_{2l} \delta u_{4l}$, etc. This equation is explicitly linear in $\delta u_{2l}, \delta u_{4l}$, etc., except for the $O(\delta^2)$ terms. The coefficients of $\delta u_{2,l}, \delta u_{4,l}$, etc., depend on u_2^*, u_4^* , etc. It is this equation, plus the equations for $\delta u_{4,l+1}$, etc., that can be written abstractly in the form (III.5).

Now consider solutions of the equations for small $\delta \mathcal{C}_l$ so that the order $\delta \mathcal{C}_l^2$ terms can be neglected. One then has a linear equation; thus the general solution can be written as a linear combination of more special solutions. The simplest solutions are generated from eigenvalues and eigenoperators of the linear operator L . Suppose one has an eigenoperator O of L such that

$$\lambda O = L \cdot O \tag{III.8}$$

for some eigenvalue λ . By an "eigenoperator" O one means a set of functions $v_2(\mathbf{q}), v_4(\mathbf{q}_1, \dots, \mathbf{q}_4)$, etc.; the eigenvalue equation (III.8) written out is

$$\lambda v_2(\mathbf{q}) = \zeta^{*2} 2^{-d} \left(v_2(\mathbf{q}/2) + 12 \int_{\frac{1}{2}}^1 \mathbf{p} \frac{v_4(\mathbf{q}/2, -\mathbf{q}/2, \mathbf{p}, -\mathbf{p})}{u_2^*(\mathbf{p})} - 1, \int_{\frac{1}{2}}^1 \mathbf{p} \frac{u_4^*(\mathbf{q}/2, -\mathbf{q}/2, \mathbf{p}, -\mathbf{p})}{[u_2^*(\mathbf{p})]^2} v_2(\mathbf{p}) + \text{other diagrams} \right) \tag{III.9}$$

$$\begin{aligned} \lambda v_4(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4) &= \zeta^{*4} 2^{-3d} \left(v_4(\mathbf{q}_1/2, \dots, \mathbf{q}_4/2) \right. \\ &- 12 \int_{\frac{1}{2}}^1 \mathbf{p} \frac{u_4^*(\frac{1}{2}\mathbf{q}_1, \frac{1}{2}\mathbf{q}_2, \mathbf{p}, -\mathbf{p} - \frac{1}{2}\mathbf{q}_1 - \frac{1}{2}\mathbf{q}_2) v_4(\frac{1}{2}\mathbf{q}_3, \frac{1}{2}\mathbf{q}_4, -\mathbf{p}, \mathbf{p} + \frac{1}{2}\mathbf{q}_1 + \frac{1}{2}\mathbf{q}_2)}{u_2^*(\mathbf{p}) u_2^*(-\mathbf{p} - \frac{1}{2}\mathbf{q}_1 - \frac{1}{2}\mathbf{q}_2)} - 5(\text{similar terms}) \\ &+ 12 \int_{\frac{1}{2}}^1 \mathbf{p} \frac{u_4^*(\frac{1}{2}\mathbf{q}_1, \frac{1}{2}\mathbf{q}_2, \mathbf{p}, -\mathbf{p} - \frac{1}{2}\mathbf{q}_1 - \frac{1}{2}\mathbf{q}_2) u_4^*(\frac{1}{2}\mathbf{q}_3, \frac{1}{2}\mathbf{q}_4, -\mathbf{p}, \mathbf{p} + \frac{1}{2}\mathbf{q}_1 + \frac{1}{2}\mathbf{q}_2)}{[u_2^*(\mathbf{p})]^2 u_2^*(-\mathbf{p} - \frac{1}{2}\mathbf{q}_1 - \frac{1}{2}\mathbf{q}_2)} v_2(\mathbf{p}) \\ &\left. + (5 \text{ similar terms}) + \text{other diagrams} \right) \text{ etc.} \tag{III.10} \end{aligned}$$

Given a solution of the eigenvalue equation (III.8) one can construct a special solution of the linearized recursion formula, namely,

$$\delta \mathcal{C}_l = c \lambda^l O, \tag{III.11}$$

where c is an arbitrary constant independent of l . This formula satisfies

$$\delta \mathcal{C}_{l+1} = L \cdot \delta \mathcal{C}_l \tag{III.12}$$

as a consequence of Eq. (III.8).

If L were a Hermitian operator, the eigenoperators such as O would be complete, namely, the most general solution $\delta \mathcal{C}_l$ of the linearized equation (III.12) could be expanded as a sum over the eigenoperators

$$\delta \mathcal{C}_l = \sum_m c_m \lambda_m^l O_m, \tag{III.13}$$

where λ_m is the eigenvalue for the eigenoperator O_m . However L is not Hermitian, which raises two problems. The first is that to get a complete set of solutions of the linearized equations one may have to look for solutions behaving as $\lambda_m^l, l^2 \lambda_m^l$, etc., as well as the eigenoperator type of solution already discussed. No such solutions have occurred in any example studied to date but they might arise in the future.

The second problem is that L is an infinite dimensional operator, so that completeness of the eigenoperators is uncertain. In practical examples studied to date one has only *asymptotic completeness*, which means that an arbitrary solution $\delta \mathcal{C}_l$ of the linearized equations has an expansion (III.13) only for sufficiently large l ; for a given finite value of l there may not exist a convergent expansion in terms of eigenoperators. Examples of this will be given below.

As an example of eigenoperators and their usefulness the eigenoperators for the trivial fixed point will be discussed. In this case $u_2^*(q) = q^2$ (we set $z_0 = 1$ for convenience) and u_4^*, u_6^* , etc. are zero. There are several different classes of eigenoperators in this case. First there are eigenoperators with v_4, v_6 , etc. all zero, leaving only v_2 . Secondly, there are eigenoperators with v_2 and v_4 nonzero while v_6, v_8 , etc. are all zero. More generally there are eigenoperators with v_2, v_4, \dots, v_{2n} nonzero and v_{2n+2}, v_{2n+4} , etc., zero. Consider first the case that only v_2 is nonzero. The eigenvalue equation is (with $\zeta^* = 2^{(d+2)/2}$)

$$\lambda v_2(\mathbf{q}) = 4 v_2(\mathbf{q}/2). \tag{III.14}$$

The solutions of these equations are the homogeneous polynomials in q . Some examples with their eigenvalues are shown in Table I (q_i is a component of q).

Suppose now that one is interested only in eigenoperators which could contribute to an interaction \mathcal{H}_l . Suppose also that these interactions show cubic symmetry [as well as the previous requirement $v_2(\mathbf{q}) = v_2(-\mathbf{q})$]. Then all terms odd in \mathbf{q} (like q_i) are ruled out. Furthermore, the first term which is allowed and is not fully rotationally invariant is the $\sum_i q_i^4$ term.

The results for eigenoperators just obtained can be compared with the solution for \mathcal{H}_l obtained earlier (Eq. II.8) when only u_2 is nonzero. [One must choose $\zeta = 2^{(2+d)/2}$ in Eq. (II.8).] For an arbitrary $u_{20}(\mathbf{q})$ analytic about $\mathbf{q} = 0$ obeying cubic symmetry one has, for large l ,

$$u_{2l}(\mathbf{q}) = 4^l r_0 + z_0 q^2 + w_0 4^{-l} q^4 + w_0' 4^{-l} \sum_i q_i^4 + \dots \tag{III.15}$$

This formula can be understood as a special case of Eq. (III.13): the coefficients are r_0, z_0 , etc., and the l dependence is determined by the eigenvalues $\lambda = 4, 1, \frac{1}{4}$ of the eigenoperators.

For large l , the eigenoperators with eigenvalue $\lambda < 1$ make a negligible contribution to $u_{2l}(\mathbf{q})$. Such operators are called irrelevant. The operators with $\lambda > 1$ are called relevant; the coefficients of the relevant operators grow with l . In the expansion (III.15) there is one relevant operator, namely, the operator $v_2(\mathbf{q}) = 1$ with $\lambda = 4$. Finally, operators with $\lambda = 1$ precisely are called marginal; the coefficient of a marginal operator is constant with l . Note that at large l there is full rotational symmetry; the terms showing only cubic symmetry are irrelevant.

The eigenvalue of particular physical interest is the eigenvalue $\lambda = 4$ of the relevant operator; one can obtain the critical exponent ν from λ . The argument relating ν to λ is true independently of the nature of the fixed point (whether Gaussian, near-Gaussian or far-from-Gaussian).

The argument is as follows. Consider the sequence of interactions \mathcal{H}_l generated by the transformation T starting from some initial interaction \mathcal{H}_0 which is close to but not at the critical point, i.e., its correlation length ξ_0 is large but not infinite. Then the correlation lengths ξ_l for the interactions \mathcal{H}_l decrease with l : from Eq. (I.23), ξ_l is $2^{-l}\xi_0$, and when l is about $(\ln \xi_0 / \ln 2)$, ξ_l will be about 1, i.e., \mathcal{H}_l is far from any critical point.

If one is dealing with nontrivial interactions \mathcal{H}_l , then, as noted in Sec. I, it is very difficult to determine ξ_0 directly as a function of the parameters in \mathcal{H}_0 , due to the difficulties of solving a near critical interaction. This is no longer true of \mathcal{H}_l , for sufficiently large l . If l is large enough so that ξ_l is of order 1 then it is not so difficult to compute ξ_l , by solving the effective interaction \mathcal{H}_l . And once ξ_l is known for some value of l , one can determine ξ_0 from the relation $\xi_0 = 2^l \xi_l$.

To illustrate this, consider the Gaussian case under discussion. In this case \mathcal{H}_l is defined by $u_{2l}(\mathbf{q})$; for large l and a near critical $u_{20}(\mathbf{q})$, one has

$$u_{2l}(\mathbf{q}) \simeq 4^l r_0 + z_0 q^2. \tag{III.16}$$

The correlation function for \mathcal{H}_l is $[4^l r_0 + z_0 q^2]^{-1}$ [from Eq.

TABLE I. Some eigenoperators $v_2(\mathbf{q})$ and eigenvalues λ .

$v_2(\mathbf{q})$	λ
1	4
q_i	2
q^2	1
q^4	$\frac{1}{4}$
$\sum_i q_i^4$	$\frac{1}{4}$

(I.20)] which gives a correlation length

$$\xi_l = [z_0 / 4^l r_0]^{\frac{1}{2}} \tag{III.17}$$

and multiplication by 2^l gives $\xi_0 = [z_0 / r_0]^{\frac{1}{2}}$.

In the Gaussian case one can calculate ξ_0 directly, also using Eq. (I.20). However, even for the near-Gaussian case, calculating ξ_0 directly is not trivial. In the near-Gaussian case, \mathcal{H}_0 contains small u_4, u_6 , etc., terms, and one can think of calculating ξ_0 as an expansion in these terms. Unfortunately these expansions have divergent coefficients when $\xi_0 \rightarrow \infty$. In contrast, when ξ_l is of order 1 and \mathcal{H}_l is near-Gaussian there is no problem with using perturbation theory to compute ξ_l . (There do exist sophisticated methods for using perturbation theory to determine ξ_0 directly [see Wilson and Kogut (1974)]; even these methods fail when one is in a far-from-Gaussian situation).

How can one recognize the values of l for which \mathcal{H}_l is far from critical and ξ_l is small? The answer is simple: First, \mathcal{H}_l is far from critical when $\delta\mathcal{H}_l$ is large (and increasing), i.e., when one or more of the relevant operators contributing to $\delta\mathcal{H}_l$ in Eq. (III.13) have coefficients of order 1 rather than much less than 1. In the Gaussian example this means that ξ_l is of order 1 when $4^l r_0$ is of order 1; more generally $c_m \lambda_m^l$ must be of order 1 for one of the relevant operators O_m .

What "of order 1" means, more precisely, is this. Suppose there is only one relevant operator, to simplify the discussion (this is the case for ordinary critical phenomena in magnets in the absence of an external field: see later). If the coefficient $c\lambda^l$ of the relevant operator has a pre-assigned value \bar{c} , say, then the correlation length ξ_l will also have a specific value $\bar{\xi}$. This is true independently of the value l provided $c\lambda^l$ is \bar{c} . The reason is that if $c\lambda^l = \bar{c}$ is fixed, then $\delta\mathcal{H}_l$ is also fixed; namely it is $\bar{c}O$, where O is the relevant operator, apart from negligible irrelevant terms. Since \mathcal{H}^* is also a fixed operator, this means \mathcal{H}_l is itself fixed, and therefore $\xi_l = \bar{\xi}$ is fixed. Hence, in particular, if $c\lambda^l$ is 1 or near 1, ξ_l will be near to the fixed correlation length associated with $\bar{c} = 1$; ξ_l cannot be very large (or very small either).

For any given initial interaction \mathcal{H}_0 there will be a corresponding constant c multiplying $\lambda^l O$ in $\delta\mathcal{H}_l$. The constant c can be determined by expanding $\delta\mathcal{H}_l$, for some fixed value of l , in terms of eigenoperators. Then the analyticity principle of Sec. I requires that c be analytic in the temperature T . In the Gaussian case, c is r_0 , which we have argued is analytic in T . Once c is known, ν is determined as follows. $c\lambda^l$ is near 1 when $l \simeq -\ln c / \ln \lambda$.

For this value of l , ξ_l is a fixed number $\bar{\xi}$; therefore

$$\xi_0 = \exp\left\{-\frac{\ln c}{\ln \lambda} \ln 2\right\} \bar{\xi} = c^{-\nu} \bar{\xi}, \quad (\text{III.18})$$

where

$$\nu = \ln 2 / \ln \lambda. \quad (\text{III.19})$$

The critical point corresponds to $c = 0$; in this case $\delta\mathcal{F}_l$ goes to zero for $l \rightarrow \infty$, i.e., \mathcal{F}_l goes to the fixed point \mathcal{F}^* . For small c , analyticity in T suggests that c is linear in $T - T_c$, so that

$$\xi_0 \propto (T - T_c)^{-\nu}. \quad (\text{III.20})$$

Hence by a quite general argument, the correlation length exponent ν is given by (III.19). For the Gaussian example, λ is 4 and [from (III.19)] ν is $\frac{1}{2}$.

Return now to the classification of eigenoperators. So far only those perturbation with v_4, v_6 , etc., equal to zero have been discussed. It is also important to look at other eigenoperators of the linearized equations. Consider, for example, the eigenoperators with v_4 nonzero but $v_6 = v_8 = \dots = 0$. The fixed point is still the trivial fixed point so $u_4^* = u_6^* = \dots = 0$. The equation for v_4 is [from (III.10)]

$$\lambda v_4(\mathbf{q}_1, \dots, \mathbf{q}_4) = 2^{4-d} v_4(\mathbf{q}_1/2, \dots, \mathbf{q}_4/2). \quad (\text{III.21})$$

The solutions are homogeneous polynomials in the \mathbf{q} 's; considering only those which obey cubic symmetry and reflection invariance, one gets the following eigenvalues and eigenoperators:

$$\begin{aligned} \lambda &= 2^{4-d}: v_4 = 1 \\ \lambda &= 2^{2-d}: v_4 = \text{any quadratic form in} \\ &(\mathbf{q}_1 \cdot \dots \cdot \mathbf{q}_4) \text{ etc.} \end{aligned} \quad (\text{III.22})$$

For these eigenoperators v_2 is nonzero; the explicit form of v_2 will not be given here.

The eigenoperator with $v_4 = 1$ is relevant if $d < 4$, irrelevant if $d > 4$, and marginal for $d = 4$. The other eigenoperators are irrelevant for $d > 2$.

The fact that $v_4 = 1$ is a relevant eigenoperator for $d < 4$ means that for $d < 4$ one cannot reach the trivial fixed point for large l unless the $v_4 = 1$ perturbation is absent in \mathcal{F}_0 . Since one must also have the v_0 perturbation absent, this means two parameters in the initial microscopic interaction must be fixed at critical values to reach the trivial fixed point. This is in the absence of an external field. This means the trivial fixed point does not describe an ordinary critical point for $d < 4$, since an ordinary (magnetic) critical point is reached by fixing only one parameter. There are special critical points called tricritical points (there is a tricritical point in liquid ^3He - ^4He mixtures, for example) which do require fixing two parameters (for ^3He - ^4He mixtures there is a critical temperature and a critical concentration). Tricritical points have been discussed in detail using the renormalization group approach by Riedel and Wegner (1972, 1973, 1974), and are actively being studied experimentally as well.

There are also eigenoperators with v_6 nonzero, or v_6 and v_8 nonzero, etc. The interesting eigenoperators that result

are

$$\begin{aligned} \lambda &= 2^{6-2d}: v_6 = 1, v_4, \text{ and } v_2 \text{ nonzero} \\ \lambda &= 2^{8-3d}: v_8 = 1, v_6, v_4, \text{ and } v_2 \text{ nonzero, etc.} \end{aligned}$$

The $v_6 = 1$ eigenoperator is relevant for $d < 3$; the $v_8 = 1$ eigenoperator is relevant for $d < 2.67$, etc. This means that for $d < 3$ the trivial fixed point does not even describe a tricritical point since at least three parameters must be fixed in the initial interaction in order to reach the trivial fixed point. In the limit $d \rightarrow 2$ an infinite number of parameters must be fixed in order to reach the trivial fixed point.

At each dimension for which the number of relevant operators changes, i.e., the dimensions 4, 3, 2.67, etc., a nontrivial fixed point separates from the trivial fixed point. As will be seen shortly, for $d < 4$ it is the nontrivial fixed point that describes an ordinary critical point while the trivial fixed point (for $3 < d < 4$) describes a tricritical point. Similarly the nontrivial fixed point that separates from the trivial fixed point at $d = 3$ describes a tricritical point, etc.

To complete the discussion of eigenoperators one should consider eigenoperators odd in the spin variable, i.e., eigenoperators with v_1, v_3, v_5 , etc. nonzero instead of v_2, v_4, v_6 , etc. These eigenoperators are important for magnetic systems with external fields and for liquid-gas transitions which have no symmetry analogous to the symmetry spin \rightarrow - spin of magnetic systems. To save space the discussion of these operators is omitted.

Having discussed the role of relevant operators it is worth emphasizing the role of irrelevant operators. Suppose a particular microscopic interaction H_0 is critical, which means in the renormalization group framework that the resulting effective interactions \mathcal{F}_l approach a fixed point \mathcal{F}^* . Consider now an infinitesimal change in the original microscopic interaction: $H_0 \rightarrow H_0 + \delta H_0$. There will be a corresponding change in \mathcal{F}_l : $\mathcal{F}_l \rightarrow \mathcal{F}_l + \delta\mathcal{F}_l$. When l is large, \mathcal{F}_l will be essentially \mathcal{F}^* and therefore the infinitesimal $\delta\mathcal{F}_l$ will satisfy the linearized equations about the fixed point. This means (if asymptotic completeness is correct) that $\delta\mathcal{F}_l$ has an expansion for large l

$$\delta\mathcal{F}_l = \sum_m c_m \lambda_m^l O_m \quad (\text{III.23})$$

where the c_m 's are constants (depending on the choice of parameters in the critical perturbation $\delta\mathcal{F}_0$); the λ_m 's are the eigenvalues, and the O_m 's are the eigenvectors of the linearized equations.

If the perturbed interaction $H_0 + \delta H_0$ is still to be critical then the coefficients of all relevant operators in (III.23) must vanish. Suppose there is only one relevant operator, for example. Then one must fix one parameter in δH_0 to ensure the vanishing of the coefficient of this operator. For example, if H_0 is an Ising model interaction, δH_0 might consist of a nearest-neighbor coupling and several 2nd- and 3rd-neighbor couplings. Given the 2nd- and 3rd-neighbor couplings in δH_0 there will be a special value of the nearest-neighbor coupling in order to be at the critical point. Apart from fixing this one parameter,

the choice of δH_0 is irrelevant; for any choice, one reaches the same fixed point. If one then allows departures from criticality, the same critical exponents emerge, independently of the choice of second and third neighbor interactions in δH_0 . The reason for this is that critical exponents (such as ν) are determined by the eigenvalues λ_m and these are uniquely determined once \mathcal{H}^* is known.

Hence the role of the irrelevant variables is to establish *universality*, namely the independence of critical behavior on details of the original microscopic interaction. Specifically one establishes *local universality*, namely independence of critical behavior under infinitesimal changes in the interaction. More interesting in practice are questions of *global universality*, i.e., independence of critical behavior to large changes in the interaction (such as the change from a ferromagnet to a binary alloy or a liquid-gas transition). To establish global universality one must establish that very different initial interactions H_0 lead to the same fixed point \mathcal{H}^* . This has not been established to date, for cases of interest (e.g., $d = 3$). However, there is so far only one fixed point that seems suitable to describe ordinary critical phenomena with a scalar order parameter (liquid-gas transitions, uniaxial ferromagnets, etc.), namely the fixed point which is near-Gaussian for $d \simeq 4$.

To complete this section of the lectures, some eigenvalues will be calculated for the nontrivial fixed point for $d \simeq 4$. The eigenvalues will be computed to order $\epsilon = 4 - d$. The purpose of this calculation is to show that the eigenvalues for the nontrivial fixed point are different than those for the trivial fixed point, implying that critical exponents will also be different for the two fixed points.

For small ϵ the nontrivial fixed point is close to the trivial fixed point, so one expects the eigenvalues and eigenoperators of the nontrivial fixed point to be close to those for the trivial fixed point. Only two eigenoperators will be considered here, namely the operator with $v_2(\mathbf{q}) \simeq 1$ and the operator with $v_4(\mathbf{q}_1, \dots, \mathbf{q}_4) \simeq 1$.

Consider first the eigenoperator with $\lambda \simeq 4$, $v_2(\mathbf{q}) \simeq 1$. One cannot claim that v_4 is 0 since there is a term proportional to v_2 in Eq. (III.10) for v_4 . However, the v_2 term is of order ϵ^2 (since u_4^* is of order ϵ). The solution v_4 of the v_4 equation is therefore also of order ϵ^2 [the solution v_4 could be of order ϵ only if the homogeneous equation

$$\lambda v_4(\mathbf{q}_1, \dots, \mathbf{q}_4) = 2^{4-d} v_4(\mathbf{q}_1/2, \dots, \mathbf{q}_4/2) \tag{III.24}$$

had a solution for $\lambda \simeq 4$; but we know that such solutions exist only for $\lambda \lesssim 1$]. Since v_4 is of order ϵ^2 it will not affect the v_2 equation to order ϵ and we neglect it. Using the values of u_4^* , ζ^* , and u_2^* obtained earlier for the nontrivial fixed point, the v_2 equation (III.9) reads (to order ϵ)

$$\lambda v_2(\mathbf{q}) = 4 \left\{ v_2(\mathbf{q}/2) - \frac{12}{9C} \epsilon \ln 2 \int_{\frac{1}{2}}^1 \frac{1}{p^4} v_2(\mathbf{p}) \right\} \tag{III.25}$$

with C given by Eq. (II.29). This equation has a solution with $v_2(\mathbf{q}) = 1$; then

$$\lambda = 4[1 - (\epsilon \ln 2)/3]. \tag{III.26}$$

Note that the constant C has disappeared. The value of ν ,

to order ϵ , is

$$\begin{aligned} \nu &= (\ln 2)/\ln \lambda = (\ln 2)[2 \ln 2 - \epsilon(\ln 2)/3]^{-1} \\ &= 0.5 + \epsilon/12. \end{aligned} \tag{III.27}$$

In conclusion, the eigenvalue λ is not 4, and is ϵ dependent. This is the origin of the nonclassical values of critical indices in three dimensions for ordinary critical points.

Next consider the eigenoperator $\lambda \simeq 1$ and $v_4 \simeq 1$. To discuss this problem properly one should include in the v_4 equations the simplest diagram involving v_6 , but [as shown in Wilson & Kogut (1974)] this term turns out not to affect the calculation. To be precise, one finds that v_6 is of order ϵ , but the contribution of v_6 to the v_4 equation (III.10) is of order ϵ^2 for the special case $\mathbf{q}_1 = \mathbf{q}_2 = \mathbf{q}_3 = \mathbf{q}_4 = 0$. One expects $v_2(\mathbf{q})$ to be of order 1 since $v_2(\mathbf{q})$ is nonzero even for the trivial fixed point. However the term containing $v_2(\mathbf{q})$ in the v_4 equation is of order ϵ^2 and can be neglected in an order ϵ calculation. The eigenvalue λ can now be determined by setting $\mathbf{q}_1 = \mathbf{q}_2 = \mathbf{q}_3 = \mathbf{q}_4 = 0$ in the v_4 equation. We must assume that $v_4(\mathbf{q}_1, \dots, \mathbf{q}_4) = 1 + \text{order } \epsilon$ for any \mathbf{q}_i (i.e., v_4 has momentum dependence only in order ϵ); this is necessary to calculate the $u_4^* v_4$ terms. Finally, one has the equation (valid to order ϵ)

$$\lambda v_4(0 \dots 0) = 2^{4-d} v_4(0 \dots 0) - \frac{72 \epsilon \ln 2}{9C} \int_{\frac{1}{2}}^1 \frac{1}{p^4} \tag{III.28}$$

with $v_4(0 \dots 0) = 1 + \text{order } \epsilon$. Dividing out $v_4(0 \dots 0)$ one gets

$$\lambda = 2^4 - 2\epsilon \ln 2 = 1 - \epsilon \ln 2 \tag{III.29}$$

to order ϵ . It is straightforward to verify that momentum dependent terms in v_4 are of order ϵ (or higher) as assumed.

For $d < 4$ the eigenoperator $v_4 = 1$ of the Gaussian fixed point was found to be relevant (eigenvalue 2^4). Now one sees that the corresponding eigenoperator for the nontrivial fixed point is irrelevant. There are no further eigenoperators of the trivial fixed point which are relevant or even near to being relevant for $d \simeq 4$, so there is only one relevant operator for the nontrivial fixed point (in the absence of an external field). This is true for $d \simeq 4$; one is less certain how the eigenvalues and eigenoperators vary as d moves away from 4, down to 2 for example.

A note about the term "irrelevant." This is a technical term—its intuitive sense is not always correct. Operators are relevant or irrelevant only with respect to a particular fixed point; an irrelevant operator is irrelevant in the literal sense only if one is performing first order calculations about that fixed point. In nonlinear calculations about a fixed point (for example the calculation of the nontrivial fixed point near the trivial fixed point) "irrelevant" operators like v_6 are important, especially in high orders in ϵ .

IV. DIVERGENCES IN FIELD THEORY AND STATISTICAL MECHANICS; THE TRIANGLE OF RENORMALIZATION

The next topic is divergences in perturbation theory and the way they are treated in the renormalization group framework. In this section the divergences will be exhibited

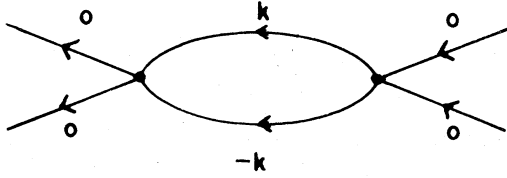


FIG. 2. Example of a diagram with divergences.

and the differences between field theoretic divergences and statistical mechanical divergences will be explained.

Consider the diagram of Fig. 2. When this diagram arises in a statistical mechanical context, it is a contribution to a four-spin correlation function (we are not considering its role in the renormalization group here). Suppose one is interested in the case that all external momenta are zero; let the propagator be $1/(r_0 + k^2)$. Then the diagram involves the integral

$$\int_0^\Lambda (r_0 + k^2)^{-2} d^d k$$

with the upper limit Λ on k being the inverse lattice spacing (before conversion to dimensionless momentum units). In the limit $r_0 \rightarrow 0$ (the critical limit) this integral is

$$\begin{aligned} &\text{convergent} && (d > 4) \\ &\text{logarithmically divergent (at } k = 0) && (d = 4) \\ &\text{divergent by a power of } k \text{ (at } k = 0) && (d < 4) \end{aligned}$$

In dimension $d < 4$ there are similar infrared divergences in more complicated graphs.

Consider now the same diagram but in a quantum field theoretic context, still with all external momenta equal to zero. Now r_0 is replaced by m^2 , where m is the particle mass, and the upper limit Λ on the k integration is infinite. In field theory d is the space-time dimension. Assuming m is not zero, the integral is:

$$\begin{aligned} &\text{divergent by a power of } k \text{ (} d > 4) \text{ (at } k = \infty) \\ &\text{logarithmically divergent (} d = 4) \text{ (at } k = \infty) \\ &\text{convergent} && (d < 4) \end{aligned}$$

The only point of agreement between field theorists and statistical mechanisms is that this graph is logarithmically divergent in four dimensions. It is these logarithmic divergences that will be emphasized in the following section. However, it is useful to understand why field theorists disagree with statistical mechanisms on what dimensions lead to power law divergences. It is a simple matter of a change of scale. In field theory the limit which produces divergences is a limit in which the upper limit Λ of the momentum variable k is taken to ∞ , holding all coupling constants (vertices) fixed. In statistical mechanics the upper limit on the momentum is fixed and one takes a limit in which the effective infrared cutoff r_0 goes to zero holding all vertices fixed. One can now make a scale change to make the field theorist's ultraviolet limit ($\Lambda \rightarrow \infty$) look like an infrared limit ($r_0 \rightarrow 0$). Namely one can replace the field theorist's variable k by a dimensionless variable $q = k/\Lambda$.

Then the field theorist's integral becomes

$$\Lambda^{d-4} \int_0^1 (m^2/\Lambda^2 + q^2)^{-2} d^d q.$$

The new integration momentum q has a fixed upper bound 1; the lower cutoff is m/Λ which goes to zero when $\Lambda \rightarrow \infty$. Thus one is back to the statistical mechanical type of divergences, except that now one has an extra factor Λ^{d-4} . For $d > 4$, this factor produces a divergence even though the integral is convergent. For $d < 4$, the factor Λ^{d-4} goes to zero as $\Lambda \rightarrow \infty$, cancelling the divergence of the integral. For $d = 4$, the factor is 1 and cannot affect convergence or divergence.

The factor Λ^{d-4} can be absorbed in a redefinition of the coupling constant (vertex); in a general graph one finds there is one integration and two propagators for each 4 point vertex. Therefore if one starts with field theoretic diagrams with 4 point coupling constant λ and cutoff Λ , the equivalent statistical mechanical graphs have coupling constant $\Lambda^{d-4}\lambda$ and cutoff 1.

The divergences discussed above provide one of the two basic problems that arise in solving either statistical mechanical or field theoretic problems by Feynman graphs. The other difficulty is that unless the coupling constants are small one has to calculate too many graphs. The renormalization group approach is designed to deal with both difficulties. The graphical formulation of the renormalization group discussed so far can deal only with the divergence difficulties (see below) but not the problem of too many graphs. To deal with strong coupling one must reformulate the renormalization group so as not to refer to diagrams; some ideas for doing this are discussed in Sec. VI; see also Secs. VII-X.

The renormalization group transformation in graphical form is completely free of divergences: since every momentum integral in the renormalization group equations ranges from momentum $\frac{1}{2}$ to momentum 1 there is no possibility for either infrared or ultraviolet divergences. Hence these divergences must occur only as a result of iterating the renormalization group transformation. This will be seen explicitly in the next lecture.

A problem for a field theorist with the renormalization group is that he wants to increase the maximum momentum Λ in his Lagrangian. Unfortunately, the renormalization group transformation of these lectures involves decreasing the cutoff rather than increasing it. The process of decreasing the cutoff cannot be avoided: it is necessary to solve the cutoff field theory. It is just as difficult to solve a Lagrangian with a large cutoff Λ as it is to solve a statistical mechanical problem near its critical temperature. In fact, we have seen that a simple change of momentum scale transforms one problem into the other. So in order to solve a Lagrangian with a large cutoff Λ , one uses the renormalization group transformation to generate effective Lagrangians with cutoff $\Lambda/2$, $\Lambda/4$, etc. This reduction procedure will be explained in a bit more detail before considering how to raise the cutoff.

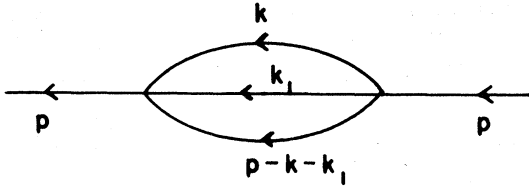


FIG. 3. Input diagram for a renormalization group calculation.

Consider the diagrams of a ϕ^4 field theory with cutoff Λ . Specifically, the cutoff is introduced by setting up the diagrams in a Euclidean metric ($k^2 = k_0^2 + \mathbf{k}^2$ instead of $k^2 = k_0^2 - \mathbf{k}^2$) and then restricting all virtual momenta k so that $k^2 < \Lambda^2$. The use of a Euclidean metric means one can calculate explicitly only vacuum expectation values for spacelike momenta, thereby excluding physical scattering amplitudes. But formal questions such as the problems of renormalization are still present in the Euclidean domain. A renormalization group approach similar to that of these lectures has yet to be developed for the Lorentz metric. See, however, the recent work of Kogut and Susskind (1974).

The idea behind the renormalization group transformation is to carry out all integrations over virtual momenta k only over the range $\Lambda/2 < k < \Lambda$, not doing any integrations for $k < \Lambda/2$. The results of these integrations can be expressed in terms of a set of effective vertices, which are then used to construct Feynman graphs with cutoff $\Lambda/2$. Consider, for example, the diagram of Fig. 3. In particular consider the part of this diagram with $|k| > \Lambda/2$, $|p - k - k_1| > \Lambda/2$ but $|k_1| < \Lambda/2$. Then one integrates over k and the line with momentum $p - k - k_1$ but not k_1 . The result of the k integration ($\Lambda/2 < k < \Lambda$ with the further restriction $\Lambda/2 < |p - k - k_1| < \Lambda$) can be expressed as an effective four-point vertex, after which one has the diagram of Fig. 4, with k_1 still to be integrated over but only over the range $0 < k_1 < \Lambda/2$.

One can start with any Lagrangian \mathcal{L}_0 and define Feynman graphs with cutoff Λ . The procedure illustrated above (when applied to all graphs and all possible ways of assigning momentum ranges within a graph) results in a set of effective vertices which define an effective Lagrangian \mathcal{L}_1 . Then when diagrams are constructed using \mathcal{L}_1 and cutoff $\Lambda/2$, one gets exactly the same vacuum expectation values as one obtains from diagrams using \mathcal{L}_0 and cutoff Λ .

It is convenient to make a change of scale and a renormalization in both \mathcal{L}_0 and \mathcal{L}_1 : if ϕ_k is the Fourier transform of the field $\phi(x)$, then one writes for \mathcal{L}_0

$$q = k/\Lambda, \tag{IV.1}$$

$$\phi_k = Z_0 \sigma_q, \tag{IV.2}$$

where Z_0 is an adjustable scale factor. One now writes $\mathcal{ZC}_0[\sigma] = \mathcal{L}_0[\phi]$. For \mathcal{L}_1 one writes

$$q = k/(\Lambda/2), \tag{IV.3}$$

$$\phi_k = Z_1 \sigma'_q, \tag{IV.4}$$

$$\mathcal{ZC}_1[\sigma'] = \mathcal{L}_1[\phi], \tag{IV.5}$$

where Z_1 is another arbitrary scale factor. One will normally use Z_0 and Z_1 to achieve a convenient normalization in \mathcal{ZC}_0

and \mathcal{ZC}_1 , for example that the term $\int q^2 \sigma_q \sigma_{-q}$ appears with coefficient 1. If $\zeta = Z_1/Z_0$, one can show that the relation of \mathcal{ZC}_1 to \mathcal{ZC}_0 is precisely that given in Sec. II [Eqs. (II.2)–(II.4)].

Let $G(k_1, \dots, k_m)$ be the Fourier transform of the vacuum expectation value $\langle \Omega | T \phi(x_1) \dots \phi(x_m) | \Omega \rangle$, where T is the time ordering symbol. Then there is a scale change in G corresponding to the scale change from \mathcal{L}_0 to \mathcal{ZC}_0 or \mathcal{L}_1 to \mathcal{ZC}_1 , namely

$$G(k_1, \dots, k_m) = \Lambda^d Z_0^m \Gamma_0(q_1, \dots, q_m), \tag{IV.6}$$

with $q_i = k_i/\Lambda$, where Γ_0 is the corresponding vacuum expectation value of the σ_q 's calculated for the interaction \mathcal{ZC}_0 [Λ^d is present because one divides out a factor $\delta^d(k_1 + \dots + k_m)$ from the Fourier transform in defining G just as was done in Sec. I to define $\Gamma_{\Lambda, q}$]. Similarly

$$G(k_1, \dots, k_m) = Z_1^m (\Lambda/2)^d \Gamma_1(q_1, \dots, q_m) \tag{IV.7}$$

with $q_i = 2k_i/\Lambda$ (provided that $k_i < \Lambda/2$).

Suppose one is interested in a problem with a renormalized mass of about 1 GeV; one has a given Lagrangian \mathcal{L}_0 and one wants to solve it in the limit of infinite cutoff Λ . This is done as follows. One considers a discrete set of cutoffs, say $\Lambda = 2^N$ for $N = 1, 2, 3, \dots, \infty$, in units of GeV. For each cutoff, i.e., each N , one applies the renormalization group transformation to produce a sequence of effective interactions $\mathcal{ZC}_0^N, \mathcal{ZC}_1^N, \mathcal{ZC}_2^N, \dots$ until one reaches \mathcal{ZC}_N^N . The cutoff for \mathcal{ZC}_N^N is $2^{N-N} = 1$, and in particular the cutoff for \mathcal{ZC}_N^N is 1 GeV for any N .

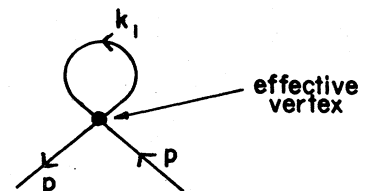
Any vacuum expectation value for momenta less than 1 GeV can be computed by computing the vacuum expectation values $\Gamma_N^N(q_1, \dots, q_m)$ generated by \mathcal{ZC}_N^N . Because the cutoff is 1 there is no change of scale from the q 's to the original momentum variables k ; however there are Z factors

$$G(k_1, \dots, k_m) = (Z_N^N)^m \Gamma_N^N(k_1, \dots, k_m). \tag{IV.8}$$

While the cutoff in \mathcal{ZC}_N^N is 1 GeV, the vacuum expectation values $G(k_1, \dots, k_m)$ calculated from \mathcal{ZC}_N^N through Eq. (IV.8) are the same as for the interaction \mathcal{ZC}_0^N with cutoff 2^N GeV. This means one is in a very fortunate position if one knows \mathcal{ZC}_N^N and Z_N^N for large N : the actual calculation of $G(k_1, \dots, k_m)$ can proceed using the 1 GeV cutoff appropriate to \mathcal{ZC}_N^N which means the calculations will be completely free of divergences (the mass was assumed to be of order 1 GeV also so there is no way for integrals to diverge). Nevertheless $G(k_1, \dots, k_m)$ will not have the unwanted cutoff dependence normally associated with a 1 GeV cutoff, since the cutoff in G is 2^N GeV not 1 GeV.

A field theorist is interested in the limit of infinite cutoff. This limit may not exist due to renormalization problems.

FIG. 4. Result of partial integration of diagram of Fig. 3: integration over k ($|k| > \Lambda/2$), with $|k_1| < \Lambda/2$, leads to new diagram shown with k_1 line remaining.



One way to ensure that this limit exists is for the sequence of effective interactions \mathcal{H}_N^N to have a limit for $N \rightarrow \infty$, which will be denoted \mathcal{H}_0^R . The interactions \mathcal{H}_N^N all have the same cutoff ($\Lambda = 1$ GeV); the vacuum expectation values $\Gamma_N^N(q_1, \dots, q_m)$ are computed by the same procedure for any N . Therefore, if \mathcal{H}_N^N has a limit for $N \rightarrow \infty$, so will the $\Gamma_N^N(q_1, \dots, q_m)$, and therefore so will the vacuum expectation values $G(k_1, \dots, k_m)$, except for the factors Z_N^N . In practice, Z_N^N is a product of an initial renormalization factor Z_0 , which is usually $\Lambda_0^{-(d+2)/2}$, times a sequence of ζ factors $\zeta_1^N \zeta_2^N, \dots, \zeta_N^N$. For example, for a ϕ^4 theory the integral $\int d^d x \mathcal{L}_0$, namely the initial action from which Feynman graphs are generated, is

$$\int_k (k^2 + m_0^2) \phi_k \phi_{-k} + \lambda_0 \int_{k_1} \int_{k_2} \int_{k_3} \phi_{k_1} \phi_{k_2} \phi_{k_3} \phi_{-k_1-k_2-k_3}.$$

To convert this to \mathcal{H}_0 starting from cutoff $2^N = \Lambda_0$, one writes $k = 2^N q$,

$$\phi_k = 2^{-N(d+2)/2} \sigma_q \tag{IV.9}$$

(corresponding to $Z_0 = \Lambda_0^{-(d+2)/2}$) and then one has

$$\begin{aligned} \mathcal{H}_0 &= \int_q (q^2 + m_0^2 \times 2^{-2N}) \sigma_q \sigma_{-q} + \lambda_0 2^{N(d-4)} \\ &\times \int_{q_1} \int_{q_2} \int_{q_3} \sigma_{q_1} \sigma_{q_2} \sigma_{q_3} \sigma_{-q_1-q_2-q_3}. \end{aligned} \tag{IV.10}$$

Hence the $q^2 \sigma_q \sigma_{-q}$ term is normalized to 1, which conforms to the conventions of Sec. II.

The factor Z_N^N is therefore

$$Z_N^N = \prod_{n=1}^N \{ \zeta_n^N / 2^{(d+2)/2} \}. \tag{IV.11}$$

In the limit $N \rightarrow \infty$ this product will diverge unless most of the ζ_n^N are approximately equal to $2^{(d+2)/2}$, i.e., unless $\zeta_n^N \simeq \zeta^*$ for the trivial fixed point [see Eq. (II.10)]. This is unlikely unless the \mathcal{H}_n^N are themselves mostly close to the trivial fixed point.

Fortunately the divergence of Z_N^N for $N \rightarrow \infty$ is of no importance: one simply defines a renormalized field ϕ_R by

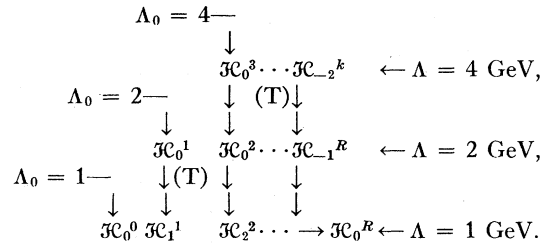
$$\phi_R = \lim_{N \rightarrow \infty} \phi_k / Z_N^N \tag{IV.12}$$

so that the vacuum expectation values of ϕ_R are precisely the Γ_N^N in the limit $N \rightarrow \infty$. One is free to make this renormalization because no physics depends on the normalization of the field ϕ .

Thus as long as one permits the wave function renormalization (IV.12), the existence of the limit $\mathcal{H}_N^N \rightarrow \mathcal{H}_0^R$ ensures the existence of the vacuum expectation values.

Strictly speaking, since \mathcal{H}_0^R has a cutoff of 1 GeV, one can only determine vacuum expectation values for $k < 1$ from \mathcal{H}_0^R . To compute for higher momenta one must look at the effective interactions with a general cutoff, say 2^n GeV. These are given by \mathcal{H}_{N-n}^N ; in the limit $N \rightarrow \infty$ these interactions must also have a limit, which can be denoted \mathcal{H}_{-n}^R .

The renormalization calculation thus produces a “triangle of renormalization”



One starts with a “bare” Lagrangian \mathcal{L}_0 , with a choice of cutoffs Λ_0 . For each cutoff $\Lambda_0 = 2^N$, the renormalization group transformation T is used to generate a column of effective interactions \mathcal{H}_M^N . The infinite cutoff limit is obtained as a limit of each row, moving to the right. If the limit defining \mathcal{H}_{-n}^R is well behaved, then the column of \mathcal{H}_0^R 's can also be generated by T , except that the column of renormalized interaction \mathcal{H}_{-n}^R is of infinite extent (n can be arbitrarily large) so it is difficult to start applying T at the top of this column.

One knows from ordinary renormalization theory that the vacuum expectation values of the renormalized theory do not exist (at least in perturbation theory) unless the bare parameters m_0 and λ_0 in \mathcal{L}_0 are permitted to depend on Λ_0 . This can also be seen by analogy with critical phenomena. Namely, the initial interactions \mathcal{H}_0^N must approach a critical interaction for $N \rightarrow \infty$. The reason for this is that the correlation functions of \mathcal{H}_N^N cannot approach a limit for $N \rightarrow \infty$ unless the correlation lengths ξ_N^N also approach a limit. This limit cannot be zero; the limiting two point function has to have a pole at $q^2 \sim -1$ to give a particle of mass of order 1 GeV, therefore ξ_N^N is of order 1. But this means that the correlation lengths ξ_0^N of \mathcal{H}_0^N must be of order 2^N [see Eq. (I.22)].

For a free field theory ($\lambda_0 = 0$) the critical interaction corresponds to $m_0 = 0$ ($r_0 = 0$ in the earlier notation). Once $\lambda_0 \neq 0$, however, there is a nonzero critical value r_{0c} for r_0 . This means that if \mathcal{H}_0^N is to go to a critical interaction then $m_0^2 \Lambda_0^{-2}$ (the parameter analogous to r_0) must approach r_{0c} as $\Lambda_0 \rightarrow \infty$, which is possible only if m_0 varies with Λ_0 , say $m_0^2 = r_{0c} \Lambda_0^2$.

Thus one sees that m_0^2 will be quadratically divergent, as one already knows from perturbation theory calculations. But the analogy to critical phenomena is valid even if perturbation theory is not valid. Thus one expects m_0^2 to be quadratically divergent in strongly coupled scalar field theories also.

It should be noted here, that in less than four dimensions, most field theorists hold λ_0 fixed when $\Lambda_0 \rightarrow \infty$ since coupling constant renormalization (see Sec. V) is not necessary in less than four dimensions. As argued earlier, the corresponding four point bare vertex in the statistical mechanical theory is $\Lambda_0^{d-4} \lambda_0$ which goes to zero as $\Lambda_0 \rightarrow \infty$ for $d < 4$. This means the statistical mechanical theory approaches the Gaussian case for $\Lambda_0 \rightarrow \infty$. This means r_{0c} is 0 and m_0 is no longer quadratically divergent in Λ_0 .

V. MARGINAL OPERATORS AND SECOND ORDER PERTURBATIONS ABOUT A FIXED POINT

The purpose of this lecture is to discuss higher order terms in perturbation theory about a fixed point. That is, the departure of an interaction \mathcal{H}_l from a fixed point \mathcal{H}^* will be calculated including terms of order $(\mathcal{H}_l - \mathcal{H}^*)^2$ and higher. This problem was first discussed by Wegner [1972].

Second order terms are important primarily when the linearized equations give rise to a marginal operator. In the linearized theory the strength of a marginal operator neither increases nor decreases as l changes. Second order terms can change this result, making the coefficient of the marginal operator either increase or decrease with l . This will be explained below.

In the case that there is one marginal operator, the result of the perturbation treatment described below is that the full renormalization group equations for all the interactions in \mathcal{H}_l are reduced to a single recursion formula [$g_{l+1} = V_g(1, g_l)$: see Eq. (V.37)] for the coefficient g_l of the marginal operator in \mathcal{H}_l . In the present analysis the recursion formula for g_l is a special simplification of the full equations; the reduction to a single recursion formula is possible only near a fixed point which has a marginal operator.

Historically the first formulations of the renormalization group equations by Stueckelberg and Petermann (1953) Gell-Mann and Low (1954), and Bogoliubov and Shirkov, (1959) were single variable recursion formulae (or more precisely differential equations for dg_l/dl ; see later) of this special type. These differential equations were derived from properties of renormalized Feynman diagram expansions, always for obscure reasons. The variable g_l in these equations was a coupling constant of some kind (originally, the electron charge), generalized to be momentum dependent (corresponding to l dependence in this paper). The Gell-Mann-Low theory will be discussed later in this section.

It turns out that there are a number of problems where there are fixed points with marginal variables, for which a Gell-Mann-Low theory can be developed. These include quantum field theory (the original case), the Kondo problem (see Secs. VII-IX), the Riedel-Wegner theory of tricritical points (Riedel and Wegner, 1972; 1973; 1974), one dimensional "superconductor" theory [see, e.g., Dyaloshinski and Larkin (1971) or Menyhard and Solym (1973)], etc. In all cases the fixed point is of a trivial nature so that Feynman diagram expansion are useful near the fixed point. But in almost all cases the coefficient of the marginal operator becomes large for some range of l , and when this happens the Gell-Mann-Low theory becomes useless for detailed calculations.

The main result of either Gell-Mann-Low theory or the perturbation theory of this section is a scaling law which shows that \mathcal{H}_l , considered as a function of both l and the initial marginal operator strength g_0 , actually depends on only a single variable built from l and g_0 . This scaling law will be rederived in this lecture.

Consider a renormalization group transformation T , a fixed point \mathcal{H}^* , and the linearized transformation L about the fixed point. Let $\mathcal{H}_l = \mathcal{H}^* + \delta\mathcal{H}_l$ be a solution of $\mathcal{H}_{l+1} = T[\mathcal{H}_l]$. One can rewrite the recursion formula

$$\delta\mathcal{H}_{l+1} = L \cdot \delta\mathcal{H}_l + N[\delta\mathcal{H}_l], \quad (\text{V.1})$$

where $N[\delta\mathcal{H}_l]$ contains all the second order terms and higher in $\delta\mathcal{H}_l$. Precisely

$$N[\delta\mathcal{H}_l] = T[\mathcal{H}^* + \delta\mathcal{H}_l] - \mathcal{H}^* - L \cdot \delta\mathcal{H}_l. \quad (\text{V.2})$$

In the linearized theory one could write $\delta\mathcal{H}_l$ as a sum over all the eigenoperators O_m of L . There are three distinct types of eigenoperators: relevant, marginal, and irrelevant. To set up the nonlinear theory it is convenient to make a partial expansion of $\delta\mathcal{H}_l$ in terms of the eigenoperators, namely to write

$$\delta\mathcal{H}_l = \sum_{m \in R} \mu_m O_m + \sum_{m \in M} g_m O_m + \delta W_l, \quad (\text{V.3})$$

where R is the set of all relevant operators, M is the set of all marginal operators, and δW_l is a remainder term. If the set of all O_m were complete, δW_l would be a sum over all irrelevant operators. With asymptotic completeness, δW_l is defined as an operator which contains only irrelevant terms after many iterations.

It is largely irrelevant for the discussion given below how many relevant, marginal, and irrelevant operators there are. So to illustrate the principles involved it will be imagined that there are only one of each, so there is one relevant parameter μ_l , one marginal parameter g_l and one irrelevant parameter w_l (the latter replacing the functional δW_l).

The nonlinear equations for μ , g , and w corresponding to Eq. (V.1) will be assumed to be the following

$$\mu_{l+1} = 4\mu_l + N_\mu[\mu_l, g_l, w_l], \quad (\text{V.4})$$

$$g_{l+1} = g_l + N_g[\mu_l, g_l, w_l], \quad (\text{V.5})$$

$$w_{l+1} = \frac{1}{4}w_l + N_w[\mu_l, g_l, w_l], \quad (\text{V.6})$$

where the functions N_μ , N_g , and N_w are both analytic and quadratic or higher in their arguments for small μ_l , g_l , and w_l , but are otherwise unrestricted. For convenience in writing formulae the eigenvalues of the relevant and irrelevant variables have been set equal to 4 and $\frac{1}{4}$, respectively: these values have no special significance.

The following discussion can be generalized to include an almost marginal variable instead of a marginal variable; this is done by allowing N_g to contain a term ϵg_l , where ϵ is a small number, in addition to quadratic terms and higher. This generalization is left to the reader.

The solution of the Eqs. (V.4)-(V.6) will be discussed assuming μ_l , g_l and w_l to be small enough so that the N terms can be treated as a perturbation.

The first step in a perturbative calculation is to solve the equations pretending that the N 's are known functions of l . Consider first the w_l equation. Let

$$x_l = 4^l w_l. \quad (\text{V.7})$$

Then (multiplying the w_l equation by 4^{l+1})

$$x_{l+1} = x_l + 4^{l+1}N_w[\mu_l, g_l, w_l]. \quad (\text{V.8})$$

The solution of this equation is

$$x_l = x_0 + \sum_{n=0}^{l-1} 4^{n+1}N_w[\mu_n, g_n, w_n] \quad (\text{V.9})$$

which gives

$$w_l = 4^{-l}w_0 + \sum_{n=0}^{l-1} 4^{n+1-l}N_w[\mu_n, g_n, w_n]. \quad (\text{V.10})$$

This is a well behaved equation, in the sense that if the inputs w_0 and N_w for any n are small then the output w_l is also small. In particular the N_w 's have a weighting factor 4^{n+1-l} which is small except for $l \simeq n$, which prevents the sum from being much larger than the individual N_w 's being summed.

The corresponding solution for μ_l is

$$\mu_l = 4^l\mu_0 + \sum_{n=0}^{l-1} 4^{l-(n+1)}N_\mu[\mu_n, g_n, w_n]. \quad (\text{V.11})$$

This equation is a disaster because the factor 4^l multiplying μ_0 and the weighting factors $4^{l-(n+1)}$ both make μ_l much larger than the input, and if μ_l becomes too large the N 's may also become large and cannot be treated perturbatively.

The difficulties with Eq. (V.11) can easily be avoided; instead of using μ_0 as an input one can use μ_L as an input where L is the maximum value of l for which the variables μ_l , g_l , and w_l are small (or perhaps the largest value of l of practical interest). The resulting equation (easily derived) is

$$\mu_l = 4^{l-L}\mu_L - \sum_{n=l}^{L-1} 4^{l-(n+1)}N_\mu[\mu_n, g_n, w_n]. \quad (\text{V.12})$$

This is a good equation, namely if μ_L and N_μ are small then μ_l will be small for any l .

It may happen that one knows μ_0 in advance of solving the equations while μ_L has to be calculated from the solution. Even in this case, the advantage of having a good equation are so important that one solves them as if μ_L were known, obtaining in particular the quantity μ_0 as a function of μ_L . The formula for μ_0 is then inverted to give μ_L as a function of μ_0 . This will be discussed further below.

Consider finally the equation for g_l . Solved in terms of g_0 one gets

$$g_l = g_0 + \sum_{n=0}^{l-1} N_g[\mu_n, g_n, w_n]. \quad (\text{V.13})$$

This equation is unsatisfactory: if l is large the sum involves a large number of terms with no damping factor. The sum is of order l times the mean value of N_g , and this can be arbitrarily large if l is large enough. For instance, if one tries to expand the solution g_l in powers of g_0 one finds the coefficient of g_0^2 is proportional to l , the coefficient of g_0^3 is proportional to l^2 , etc., provided that $N_g[\mu, g, w]$ includes a g^2 term when expanded in powers of μ , g , and w . These factors of l correspond to logarithmic divergences in Feyn-

man diagrams (the momentum scale corresponding to l is 2^{-l} so l itself is like the logarithm of a momentum). The factors of l are not, at first sight, as disastrous as the factors 4^l that appear in the bad form of the μ_l equation. However, the factors of l are less easily eradicated.

It does not help to use g_L as the input parameter instead of g_0 . In this case one has

$$g_l = g_L - \sum_{n=l}^{L-1} N_g[\mu_n, g_n, w_n], \quad (\text{V.14})$$

and the sum can be large if $L - l$ is large.

What can one do with the equation for g_l ? One can do the following. Suppose g_{l_0} is known for some value of l_0 . Then one can calculate g_l for l near l_0 . The equation is

$$g_l = g_{l_0} + \sum_{n=l_0}^{l-1} N_g[\mu_n, g_n, w_n]. \quad (\text{V.15})$$

If one considers only values of $l - l_0$ of order 1 (say $1 \leq l - l_0 \leq 10$) then the sum cannot become arbitrarily large and this is a good equation.

To make use of Eq. (V.15) one has to allow g_{l_0} to be input for a range of values of l_0 , and the simplest procedure is to permit l_0 to be arbitrary, i.e., let the whole set of constants g_l , $0 \leq l \leq L$, be input. This is for a first stage calculation in which the relevant and irrelevant variables μ_l and w_l are determined. In this first stage calculation one can also calculate g_{l+1} in terms of $(g_l, \mu_L, \text{ and } w_0)$, for any l . This means one gets a recursion formula for g_l alone, rather than coupled equations for g_l , μ_l , and w_l . The solution of this new recursion formula will constitute a second stage in the over-all calculation and will be discussed later.

The three equations to be solved are now

$$\mu_l = 4^{l-L}\mu_L - \sum_{n=l}^{L-1} 4^{l-(n+1)}N_\mu[\mu_n, g_n, w_n], \quad (\text{V.16})$$

$$w_l = 4^{-l}w_0 + \sum_{n=0}^{l-1} 4^{n+1-l}N_w[\mu_n, g_n, w_n], \quad (\text{V.17})$$

$$g_l = g_{l_0} + \sum_{n=l_0}^{l-1} N_g[\mu_n, g_n, w_n]. \quad (\text{V.18})$$

It is assumed that μ_L , w_0 , and g_{l_0} are small; we also assume L is large. The value of l_0 can be anywhere in the range, $0 \leq l_0 \leq L$.

One would like to consider only l with $l - l_0 \sim 1$ in order to use the g_l equation. Unfortunately, the equations for μ_l and w_l together involve sums over all n from 0 to L , which means one must know g_n for all n . Fortunately, this need is more apparent than real, for in the equations for μ_l and w_l the sums are dominated by values of n near l : values of n with $n - l$ large are suppressed by an explicit factor $4^{-|n-l|}$. What this means is one can use the equations (V.16)–(V.18) to compute μ_l , w_l , and g_l for l near l_0 and no difficulties will arise in practice, as will be seen shortly.

Equations (V.16)–(V.18) are now to be solved by iteration. The first approximation is obtained by neglecting the N terms completely, giving

$$\mu_l = 4^{l-L}\mu_L, \tag{V.19}$$

$$w_l = 4^{-l}w_0, \tag{V.20}$$

$$g_l = g_{l_0}. \tag{V.21}$$

The second approximation is obtained by substituting the first approximation inside N and recomputing w_l , μ_l , and g_l . The second approximation is

$$\mu_l = 4^{l-L}\mu_L - \sum_{n=l}^{L-1} 4^{l-(n+1)}N_\mu[4^{n-L}\mu_L, g_{l_0}, 4^{-n}w_0], \tag{V.22}$$

$$w_l = 4^{-l}w_0 + \sum_{n=0}^{l-1} 4^{n+1-l}N_w[4^{n-L}\mu_L, g_{l_0}, 4^{-n}w_0], \tag{V.23}$$

$$g_l = g_{l_0} + \sum_{n=l_0}^{l-1} N_g[4^{n-L}\mu_L, g_{l_0}, 4^{-n}w_0]. \tag{V.24}$$

The third approximation is obtained by substituting the second approximation inside N in Eqs. (V.16)–(V.18); similarly, one constructs approximations to any order.

Are large terms generated by these calculations? From the previous analysis one has to fear only the effects of the sum in Eq. (V.18). To be specific, let $N_\theta[\mu, g, w]$ include cg^2 where c is a constant, in the limit of small μ , g , and w . Then the second approximation to g_l includes a term

$$g_l = g_{l_0} + (l - l_0)cg_{l_0}^2 + \dots \tag{V.25}$$

and one has a large term when $l - l_0$ is large. This means g_n inside N_μ or N_w has a term $(n - l_0)cg_{l_0}^2$. Consider the effect of this term on μ_l . For this purpose consider two possible terms in $N_\mu[\mu, g, w]$, namely $c_1g^2 + c_2\mu g$, with c_1 and c_2 being constants. Substitute in these terms $(n - l_0)cg_{l_0}^2$ for g_n and $4^{n-L}\mu_L$ for μ_n (just to see the effects of these terms: this is not a complete calculation), and they give a contribution to μ_l of the form

$$- \sum_{n=l}^{L-1} 4^{l-(n+1)}[c_1c^2(n - l_0)^2g_{l_0}^4 + c_2cg_{l_0}^2(n - l_0)4^{n-L}\mu_L]. \tag{V.26}$$

The first term involves the sum

$$\sum_{n=l}^{L-1} 4^{l-n}(n - l_0)^2.$$

Since 4^{l-n} decreases with n much more rapidly than $(n - l_0)^2$ increases, this sum is of order $(l - l_0)^2$, i.e., not very large if $l - l_0 \sim 1$, even though the upper limit on n may be much larger than l_0 . In fact the upper limit on n could be changed to ∞ and the sum would still be finite. In summary, the $n - l_0$ terms in g_n do not produce large terms in μ_l (or w_l either) provided $l - l_0$ is not large. Now consider the $\mu_L g_{l_0}^2$ term: in this case the sum involved is

$$\sum_{n=l}^{L-1} 4^{l-L}(n - l_0)$$

which is of order $(L - l)2^{4^{l-L}}$ if $L \gg l$ and $L \gg l_0$. This is never numerically large either since the power of 4 more

than compensates the factor $(L - l)^2$. However, this term is important in computing the dependence of μ_0 on μ_L ; to a first approximation one has

$$\mu_0 = 4^{-L}\mu_L \tag{V.27}$$

but now higher approximations include terms like

$$\mu_0 = 4^{-L}\mu_L\{1 + c_2cg_{l_0}^2L^2\}. \tag{V.28}$$

The spirit of the calculation was that terms in the higher approximations should be negligible compared to the first approximation and this is no longer so.

Thus to determine μ_0 from μ_L a more sophisticated calculation is required: this will be discussed later. If it is not required to determine μ_0 the calculation described above is acceptable.

The method of successive approximations to Eqs. (V.16)–(V.18) determines μ_l , w_l , and g_l as functions of g_{l_0} , w_0 , and μ_L , for $l - l_0$ of order 1. There are further simplifications depending on the value of l_0 . For intermediate values of l_0 , namely $l_0 \gg 1$ and $L - l_0 \gg 1$ (and $l - l_0 \sim 1$), both w_0 and μ_L are negligible. This is because all terms involving w_0 also have a factor 4^{-l} , apart from powers of l ; all terms involving μ_L have a factor 4^{l-L} apart from powers of $L - l$. In contrast, terms involving only powers of g_{l_0} also involve only powers of $l - l_0$. Using this result the equations for μ_l , g_l , and w_l can themselves be simplified when l and l_0 are in the intermediate region: one can set w_0 and μ_L equal to zero and one can also extend the sums over n to $+\infty$ or $-\infty$:

$$\mu_l \simeq - \sum_{n=l}^{\infty} 4^{l-(n+1)}N_\mu[\mu_n, g_n, w_n], \tag{V.29}$$

$$w_l \simeq \sum_{n=-\infty}^{l-1} 4^{n+1-l}N_w[\mu_n, g_n, w_n], \tag{V.30}$$

$$g_l = g_{l_0} + \sum_{n=l_0}^{l-1} N_g[\mu_n, g_n, w_n]. \tag{V.31}$$

When these equations are solved by the iteration procedure defined above, it is easily verified that the sums over n converge and the error introduced by extending them to $\pm\infty$ is of order 4^{l-L} or 4^{-l} both of which are negligible.

The simplified equations (V.29)–(V.31) have a simple translational invariance in l and l_0 which means the solution for g_l has the general form

$$g_l = V_g[l - l_0, g_{l_0}], \tag{V.32}$$

where V_g is a function of only two arguments, not l , l_0 and g_{l_0} separately. Likewise $\mu_l = V_\mu[l - l_0, g_{l_0}]$, etc.

When l and l_0 are near L one can no longer neglect μ_L but w_0 can still be ignored. In this case one can simplify the equations to

$$\mu_l = 4^{l-L}\mu_L - \sum_{n=l}^{L-1} 4^{l-(n+1)}N_\mu[\mu_n, g_n, w_n], \tag{V.33}$$

$$w_l = \sum_{n=-\infty}^{l-1} 4^{n+1-l}N_w[\mu_n, g_n, w_n], \tag{V.34}$$

$$g_l = g_{l_0} + \sum_{n=l_0}^{l-1} N_g[\mu_n, g_n, w_n]. \tag{V.35}$$

The solution is a function of the form

$$g_l = V_\theta[l - l_0, L - l_0, g_{l_0}, \mu_L] \quad (\text{V.36})$$

etc.

Similarly, when l and l_0 are near 1, the solution depends on w_0 , but is independent of μ_L and L .

The next problem is to compute g_l , say in terms of g_0 . This is done as follows. Using the previous calculation one can determine g_l for small l , say $1 \leq l \leq 10$. The next step is to determine g_l for $l > 10$ and $L - l > 10$ in terms of g_{10} ; at this stage one can use Eq. (V.32). It is convenient to write this equation as a recursion formula

$$g_{l+1} = V_\theta[1, g_l] \quad (\text{V.37})$$

What is the form of the function $V_\theta[1, g]$? In the first approximation, one has $g_{l+1} = g_l$, i.e., $V_\theta[1, g] = g$. In the second approximation, one can write

$$V_\theta[1, g] = g + N_\theta[0, g, 0], \quad (\text{V.38})$$

and since the leading term in N_θ for small g must be quadratic or higher, the leading term is the cg^2 term introduced earlier. Hence

$$V_\theta[1, g] = g + cg^2 + O(g^3). \quad (\text{V.39})$$

In the third approximation, μ_n and w_n will be nonzero but only in order g_l^2 or higher, and will contribute to $V_\theta[1, g]$ in order g^3 or higher, through terms of order $\mu_n g_n$ or $w_n g_n$ in $N_\theta[\mu_n, g_n, w_n]$.

The end result is an equation

$$g_{l+1} = g_l + cg_l^2 + c_3 g_l^3 + c_4 g_l^4 + \dots \quad (\text{V.40})$$

valid for small g_l . When g_l is small, g_{l+1} will not be much different from g_l , i.e., g_l is a slowly varying function of l . This allows one to convert the recursion formula (V.40) into a differential equation. This is done as follows. Assuming g_l is a differentiable function of l , one writes

$$g_{l+1} = g_l + \frac{dg_l}{dl} + \frac{1}{2} \frac{d^2 g_l}{dl^2} + \dots \quad (\text{V.41})$$

One now writes Eq. (V.40) as

$$\frac{dg_l}{dl} = cg_l^2 + c_3 g_l^3 + c_4 g_l^4 + \dots - \frac{1}{2} \frac{d^2 g_l}{dl^2} - \frac{1}{6} \frac{d^3 g_l}{dl^3} - \dots \quad (\text{V.42})$$

The higher order derivatives can be removed by successive eliminations. The first step is to differentiate this equation, giving

$$\frac{d^2 g_l}{dl^2} = \{2cg_l + 3c_3 g_l^2 + \dots\} \frac{dg_l}{dl} - \frac{1}{2} \frac{d^3 g_l}{dl^3} - \dots \quad (\text{V.43})$$

Since dg_l/dl is of order g_l^2 , this shows that $d^2 g_l/dl^2$ is of order g_l^3 (further analysis shows that $d^3 g_l/dl^3$ is of order g_l^4 , etc.). To order g_l^3 one obtains

$$d^2 g_l/dl^2 = 2c^2 g_l^3. \quad (\text{V.44})$$

This allows us to determine dg_l/dl to order g_l^3

$$\frac{dg_l}{dl} = cg_l^2 + (c_3 - c^2)g_l^3 + \dots \quad (\text{V.45})$$

More complicated calculations of the same type give dg_l/dl to any order in g_l ; the result can be written

$$dg_l/dl = \psi(g_l). \quad (\text{V.46})$$

This equation is a simple differential equation which is easily integrated: let

$$F(g) = \int [1/\psi(g)] dg \quad (\text{V.47})$$

then (irrelevant variables will be ignored even for small l)

$$F(g_l) - F(g_0) = l. \quad (\text{V.48})$$

For small g , $F(g)$ is

$$\begin{aligned} F(g) &= \int \frac{1}{cg^2 + (c_3 - c^2)g^3 + \dots} dg \\ &= \int \left(\frac{1}{c} \frac{1}{g^2} - \frac{(c_3 - c^2)}{c^2} \frac{1}{g} + \text{order } 1 \right) dg \\ &= -\frac{1}{c} \frac{1}{g} - \left(\frac{c_3}{c^2} - 1 \right) \ln g + \text{order } g \end{aligned} \quad (\text{V.49})$$

So the equation for g_l is

$$\begin{aligned} -\frac{1}{cg_l} - \left(\frac{c_3}{c^2} - 1 \right) \ln g_l + \dots + \frac{1}{cg_0} + \left(\frac{c_3}{c^2} - 1 \right) \\ \times \ln g_0 - \dots = l. \end{aligned} \quad (\text{V.50})$$

Assuming g_l and g_0 are both small, this equation can be solved by successive approximations. The first approximation consists in keeping only $1/g_l$ and $1/g_0$

$$1/g_l = (1/g_0) - cl \quad (\text{V.51})$$

or

$$g_l = g_0 / (1 - clg_0). \quad (\text{V.52})$$

As a second approximation one uses this formula to compute $\ln g_l$, and then recalculate g_l , giving

$$g_l = g_0 / \left\{ 1 - clg_0 + \left(\frac{c_3}{c^2} - 1 \right) g_0 \ln(1 - clg_0) \right\}. \quad (\text{V.53})$$

There are now two cases to discuss. If cg_0 is negative, then one sees in the first approximation that g_l decreases with l , and for very large l ($l \gg 1/g_0c$) one has

$$g_l \sim -1/cl. \quad (\text{V.54})$$

In this case the second approximation is not much different from the first approximation.

The second case is the case $cg_0 > 0$. In this case, g_l increases with l . For $l \ll 1/cg_0$ this increase is not very noticeable: g_l is small if g_0 is small. But for $clg_0 \simeq 1$, g_l becomes much larger than g_0 , in fact g_l becomes too large for the perturbative approach of this lecture to be applicable.

There is a third case, namely $c = 0$. This will not be discussed here.

What one sees is that qualitatively the marginal coupling can behave either like a relevant variable or an irrelevant variable: if $cg_0 < 0$ then g_l decreases with l , eventually going to zero although not as fast as an ordinary irrelevant variable. If $cg_0 > 0$ the variable g_l increases with l until it becomes large enough so that perturbation theory about the fixed point is no longer valid.

It must be emphasized that the calculations described here are perturbative and therefore valid *only* when μ_l , g_l , and w_l are small over the *entire range* $0 \leq l \leq L$. In the case that g_l is increasing with l , one must choose L small enough so that g_L is still small. The input parameters w_0 and μ_L must also be small. This does not mean that anything fundamental goes wrong when g_l (or μ_l or w_l) is of order 1, just that a nonperturbative method of solving the renormalization group equations is required in this case.

One can now see the role of divergences in the renormalization group equations. When these are solved by a straightforward expansion in μ_0 , g_0 , and w_0 , the divergences appear as the factor 2^l multiplying μ_0 and l multiplying g_0^2 . These diverge for $l \rightarrow \infty$; in higher orders in g_0 and μ_0 there are more divergent factors.

By proper choice of the independent variables, namely μ_L , w_0 , and either g_0 or g_L (depending on the sign of cg_0) and by properly solving the g_l equation, all divergences disappear. There remains the problem of what to do if any of μ_L , w_0 , or (g_0 or g_L) is large, which is a problem of technique not of divergences.

It is now clear that the higher order calculation of this lecture is crucial for the discussion of marginal variables. In fact, in practical situations it is sometimes necessary to carry the perturbation calculation to quite high orders. An example is in connection with the numerical solution of the Kondo problem. See Secs. VII-IX.

In contrast, for fixed points with no marginal or near-marginal variables, the perturbations w_l and μ_l change so rapidly with l that there is less value to going beyond the linear theory about the fixed point. Once one gets to values of l for which the linear theory fails one gets almost immediately to a range of l for which nonperturbative methods are required.

The next topic to discuss is the physical significance of the constants g_l and the derivation of the Gell-Mann-Low renormalization group equations. The first thing to note is that the expansion of g_l in powers of g_0 is divergent for $clg_0 > 1$ [at least in the approximation (V.52)]. This means the expansion is useless for large enough l even if g_0 is small. This is because the coefficient of g_0^m is of order l^{m-1} . In contrast, g_{l+1} has a good expansion in terms of g_l ; this expansion is free of factors of l .

In the following a "good" expansion always means an expansion free of factors of l .

Now consider the multipin correlation functions $\Gamma_l(q_1, \dots, q_n)$ (vacuum expectation values for field theorists). If all the q_i are of order 1 (in particular, no q_i is much less than 1; one must also require that no partial sum of the q_i is small either) then one can argue that $\Gamma_l(q_1, \dots, q_n)$ has an expansion in terms of g_l which is also free of divergences (powers of l). In the specific case that the fixed point is the trivial fixed point in four dimensions and g_l is the four point vertex u_l , the argument is a diagrammatic one: the q_i provide an infrared cutoff on all diagrams so no divergences appear even if the constant r_l in the propagator $q^2 + r_l$ is zero (Gell-Mann and Low 1954). For example, a simple diagram that might occur is

$$\int_0^1 \frac{1}{(q+q')^2 + r_l} \frac{1}{q^2 + r_l} d^4q'$$

which is logarithmically divergent if *both* q and $r_l = 0$; but if q is of order 1 the whole integral is also of order 1 even for $r_l = 0$. An argument applicable to arbitrary fixed points is given in (Wilson and Kogut, 1974).

The correlation functions $\Gamma_l(q_1, \dots, q_m)$ are related to the initial correlation functions or vacuum expectation values $G(k_1, \dots, k_m)$ by a change of scale and a renormalization factor

$$G(k_1, \dots, k_m) = Z_l^m \left(\frac{\Lambda_0}{2^l} \right)^d \Gamma_l \left(\frac{2^l k_1}{\Lambda_0}, \dots, \frac{2^l k_m}{\Lambda_0} \right), \quad (\text{V.55})$$

where Λ_0 is the initial cutoff or maximum momentum. The factors Z_l are a nuisance; but these can be eliminated by considering a ratio of vacuum expectation values, say $G(k_1, \dots, k_m)/G(k)^{m/2}$, where $G(k)$ is the two point function. This is related by known factors of Λ_0 and 2^l , to the ratio $\Gamma_l(q_1 \dots q_m)/\Gamma_l(q)^{m/2}$ with $q_i = 2^l k_i/\Lambda_0$. Since the Γ_l have a good expansion in g_l for $q_i \sim 1$, this means so does the ratio of G 's, provided the momenta k_i are of order $2^{-l}\Lambda_0$.

Since g_l for sufficiently large l has a poor expansion in terms of g_0 (poor meaning powers of l are present), the fact that a ratio of G 's has a good expansion in terms of g_l means it has a poor expansion in terms of g_0 . In the special case that g_l is the coupling constant u_l , this means that for momenta of order $2^{-l}\Lambda_0$, u_l is the appropriate measure of the coupling strength and not u_0 . More generally, although only one of g_0 or g_l is an independent parameter, one should use g_l as the independent parameter for momenta of order $2^{-l}\Lambda_0$. See Wilson (1971b) for more discussion on this point.

Now consider the old Gell-Mann-Low theory. In the Gell-Mann-Low theory, say for a ϕ^4 field theory, there are a set of coupling constants g_l' , not the same as the constants g_l defined here, which however satisfy a differential equation similar to Eq. (V.46). The constant g_l' is defined in terms of vacuum expectation values at a momentum of order $2^{-l}\Lambda_0$ (the conventional variable is $\ln \Lambda$, where Λ is $2^{-l}\Lambda_0$, rather than l , but this is a trivial difference). What we now see is that if one takes the ratio $G(k_1, \dots, k_m)/G(k)^{m/2}$ and subtracts the value of the ratio at the fixed point [$G^*(k_1, \dots, k_m)/G^*(k)^{m/2}$] the difference will be proportional to g_l with corrections of order g_l^2 , etc., provided the k_i are fixed at values proportional to $2^{-l}\Lambda_0$. So if one

defines, e.g.,

$$g_l' = G(2^{-l}\Lambda_0, 2^{-l}\Lambda_0, 2^{-l}\Lambda_0, -3 \times 2^{-l}\Lambda_0)/G(2^{-l}\Lambda_0)^2 - (\text{fixed point value}), \quad (\text{V.56})$$

then g_l' will have a good expansion in g_l starting with a term proportional to g_l (we assume this term does not vanish).

Now g_{l+1}' also has a good expansion in terms of g_l since changing momenta by only a factor of 2 does not hurt these expansions. Furthermore the expansion of g_l' in terms of g_l can be inverted to give g_l in terms of g_l' and this will also be a good expansion since the inversion process cannot introduce powers of l . One can then proceed to express g_{l+1}' as an expansion in g_l' : this will also be a good expansion.

Finally one can see that g_{l+1}' is equal to g_l' up to terms of order $g_l'^2$. The reason for this is that the correlation functions $\Gamma_l(q_1, \dots, q_m)$ are uniquely and completely determined by \mathcal{H}_l . Now \mathcal{H}_l consists of \mathcal{H}^* (which is independent of l) plus perturbations of order g_l, μ_l , and w_l . Also $g_{l+1} = g_l$ up to order g_l^2 and (for intermediate values of l which is implicitly assumed in the above discussion) μ_l and w_l are of order g_l^2 or higher. Hence \mathcal{H}_{l+1} differs from \mathcal{H}_l only in order g_l^2 ; hence $\Gamma_{l+1}(q_1, \dots, q_m)$ differs from $\Gamma_l(q_1, \dots, q_m)$ (the same arguments q_i) only in order g_l^2 . Γ_l and Γ_{l+1} for the same arguments correspond to G 's with arguments differing by a factor 2. Hence one concludes that $g_{l+1}' - g_l'$ is of order $g_l'^2$, which means of order $g_l'^2$ since g_l' is of order g_l .

Hence g_{l+1}' can be written in terms of g_l' in a form exactly analogous to Eq. (V.40), except the constants c, c_3 , etc. may be different. One can again convert this equation to a differential equation and integrate as before.

Thus one can set up the Gell-Mann-Low equation using parameters g_l' defined in terms of vacuum expectation values (correlation functions) directly without reference to the renormalization group of this paper. In this case the function $\psi'(g_l')$ that replaces $\psi(g_l)$ may be calculated by computing both g_l' and dg_l'/dl as a power series in g_l' , then inverting the series for g_l' to give g_l as a series in g_l' , and finally expressing dg_l'/dl as a power series in g_l' . The only trouble with this procedure is that one starts with two bad expansions (the expansions of g_l' and dg_l'/dl in powers of g_l') and it seems miraculous that one winds up with a function $\psi'(g_l')$ which has a good expansion. Furthermore it is somewhat difficult to prove that $\psi'(g_l')$ has a good expansion beyond the terms one calculates. Gell-Mann and Low presented an argument to show that ψ' has a good expansion; unfortunately, the standard review of Bogoliubov and Shirkov ignores altogether the importance of distinguishing good expansions from bad.

The Gell-Mann-Low renormalization group [or its cousin, the Callan-Symanzik equations (Callan, 1970; Symanzik, 1970)] is thus a simplification of the complete renormalization group, in which one has a differential equation for only one variable g_l (if there is only one marginal variable) instead of recursion formulae for a functional \mathcal{H}_l . One pays a price for this: the Gell-Mann-Low equation is useful *only* for discussing small perturbations about a fixed point with a marginal variable. In the case that the couplings g_l increase with l one inevitably leaves the region of validity of the perturbation treatment for

sufficiently large l ; one is then forced to use an appropriate formulation of the complete theory. This is a phenomenological statement: there has never been a successful calculation based on the Gell-Mann-Low or Callan-Symanzik theory outside the region of small g_l . Meanwhile, the Kondo calculation discussed in later sections proves the usefulness of the complete renormalization group in nonperturbative calculations.

The remainder of this section is not important for any later section.

Finally, the problem of determining μ_0 , given μ_L , will be discussed further. To simplify the calculation, we assume μ_L is small and compute to first order in μ_L . Also, the dependence on the initial value w_0 of the irrelevant variable will be neglected. Neither of these simplifications remove the essential problem: the problem is the $L^2 g_{l_0}^2$ term in Eq. (V.28) and higher order terms of the same type: higher order in g_{l_0} but linear in μ_L and independent of w_0 .

The problem goes away if one calculates only μ_{L-1} instead of μ_0 , for then $L^2 g_{l_0}^2$ is replaced by $g_{l_0}^2$ and similarly in higher orders. This suggests that one revise the first stage calculation so that input variables are μ_{l_0+1}, g_{l_0} , and w_0 , while the output variables are μ_{l_0} and g_{l_0+1} . Then the second stage involves calculating both μ_l and g_l as a function of l instead of just g_l . This is still easier than computing μ_l, g_l , and w_l simultaneously, especially if there are a large number of irrelevant variables w_l .

The solution of the first stage calculation has already been defined: one sets $L = l_0 + 1$ in Eq. (V.36). The result is

$$\mu_{l_0} = V_\mu[0, 1, g_{l_0}, \mu_{l_0+1}], \quad (\text{V.57})$$

$$g_{l_0+1} = V_g[1, 1, g_{l_0}, \mu_{l_0+1}]. \quad (\text{V.58})$$

Both V_μ and V_g have power series expansions in g_{l_0} and μ_{l_0+1} .

To solve Eqs. (V.57)–(V.58) as they stand is painful, so a further simplification will be introduced. One solution of these equations is already known, namely, the solution (V.32). That is, the formulae $\mu_l = V_\mu[l - l_0, g_{l_0}]$ and $g_l = V_g[l - l_0, g_{l_0}]$ (and $w_l = V_w[l - l_0, g_{l_0}]$) satisfy the basic recursion formulae (V.4)–(V.6); in addition they satisfy the same boundary condition for g_{l_0} as (V.36); the only change is that μ_{l_0+1} is not arbitrary, since

$$\mu_{l_0+1} = V_\mu[1, g_{l_0}]. \quad (\text{V.59})$$

The simplification is to linearize about this solution. In particular, one writes

$$\mu_{l_0} = V_\mu[0, g_{l_0}] + \delta\mu_{l_0}, \quad (\text{V.60})$$

$$\mu_{l_0+1} = V_\mu[0, g_{l_0+1}] + \delta\mu_{l_0+1}. \quad (\text{V.61})$$

The quantities $\delta\mu_{l_0}$ and $\delta\mu_{l_0+1}$ will be assumed to be small and treated only to first order. Note that only μ_{l_0} is linearized with respect to the solution $V_\mu[l - l_0, g_{l_0}]$; for consistency μ_{l_0+1} is linearized about the solution $V_\mu[l - l_0 - 1, g_{l_0+1}]$ which satisfies a different boundary condition [g_{l_0+1} is specified instead of g_{l_0} ; g_{l_0+1} is related to g_{l_0} using Eq. (V.32)].

Next one calculates $\delta\mu_{l_0}$ as a function of $\delta\mu_{l_0+1}$. If $\delta\mu_{l_0+1}$ vanishes, then $\delta\mu_{l_0}$ also vanishes. The reason for this is

the following. One must compare three solutions of the basic recursion formulae (V.4)–(V.6). The first solution is given by $\mu_l = V_\mu[l - l_0, 1, g_{l_0}, \mu_{l_0+1}]$ (with similar formulae for g_l and w_l), where g_{l_0} and μ_{l_0+1} are specified. The second solution is $\mu_l = V_\mu[l - l_0, g_{l_0}]$, etc., where g_{l_0} is specified, and the boundary condition on μ_l is that it is bounded as $l \rightarrow \infty$ [this is evident from Eq. (V.29)]. The third solution is $\mu_l = V_\mu[l - l_0 - 1, g_{l_0+1}]$, where g_{l_0+1} is specified, and again μ_l is bounded as $l \rightarrow \infty$. (The boundary conditions on w_l are that w_l be bounded as $l \rightarrow -\infty$, for all three solutions.)

These three solutions are identical, provided that $\delta\mu_{l_0+1} = 0$ and that g_{l_0+1} agrees with the first solution [namely, g_{l_0+1} satisfies Eq. (V.58)]. These two provisos guarantee that the first and third solutions give identical values for μ_{l_0+1} and g_{l_0+1} . This in turn (combined with the boundary condition on w_l) determines a unique solution to Eqs. (V.4)–(V.6); hence the first and third solutions are identical. Now one knows that the first and third solutions give the specified value of g_{l_0} and have a bounded μ_l for $l \rightarrow \infty$, hence they are the same as the second solution. Hence μ_{l_0} is $V_\mu[o, g_{l_0}]$ and $\delta\mu_{l_0}$ vanishes.

One can now expand $\delta\mu_{l_0}$ in powers of $\delta\mu_{l_0+1}$. To first order the result can be written

$$\delta\mu_{l_0} = V_{1\mu}[g_{l_0}]\delta\mu_{l_0+1}. \quad (\text{V.62})$$

The function $V_{1\mu}[g_{l_0}]$ has a power series expansion in g_{l_0} ; this series can be determined once the functions $V_\mu[o, 1, g, \mu]$, $V_\mu[o, 1, g, \mu]$, and $V_\mu[o, g]$ are known.

To solve Eq. (V.62) is straightforward. One needs to know g_{l_0} only to o th order in $\delta\mu$, i.e. one can use the previous calculations [see, e.g. Eqs. (V.52) or (V.53)] to determine g_{l_0} . Equation (V.62) can now be turned into a linear differential equation for $\delta\mu_l$.

One can also generalize the Gell-Mann-Low methods to define quantities related to $\delta\mu_l$ which also satisfy linear differential equations analogous to (V.62) depending only on g^l . This has already been done in field theory by 't Hooft (1973), Weinberg (1973), and Symanzik (1971).

VI. BLOCK SPIN METHODS: THE TWO DIMENSIONAL ISING MODEL

There is a practical aspect to the renormalization group approach. This will be illustrated in this lecture using the two-dimensional Ising model as an example. The ϵ expansion does not work very well for two-dimensional systems; there is no other expansion that works for the two dimensional Ising or Heisenberg models. There is Onsager's exact solution for the nearest neighbor Ising model [see, e.g., Schultz, Mattis, and Lieb (1964)] but there are many systems in two dimensions that cannot be solved exactly.

In this section an approximate method for solving the two dimensional Ising model will be described. A renormalization group transformation T will be defined for the Ising model. The fixed point of the transformation will be obtained approximately. Likewise eigenvalues of the linearized transformation will be calculated; these give (approximately) the exponents of the two dimensional Ising model.

The emphasis will be on the general principles used in constructing the transformation and in defining practical approximations to the transformation.

There are many similar formulations of the transformation. Niemeier and Van Leeuwen (1973; 1974, 1975) were the first to define a transformation and carry out precise numerical calculations for the two dimensional Ising model. They studied a triangular lattice. Another transformation for the square lattice, due to L. Kadanoff (unpublished) and the author, will be discussed here. Other transformations have been considered by Nauenberg and Nienhuis (1974a, 1974b, 1975), Kadanoff and Houghton (1975), Hsu, Niemeier, and Gunton (1975), Subbarao (1975), and Kadanoff (1975). There is also an extensive literature on transformations for one dimensional Ising systems [Krinsky and Furman (1974), Balian and Toulouse (1974), Nauenberg (1975), Nelson and Fisher (1975), Weyland and Niemeier (1974)].

Consider the Ising model on a square lattice. The partition function is

$$Z = \sum_{\{s\}} \exp\{K \sum_{\mathbf{n}} \sum_i s_{\mathbf{n}} s_{\mathbf{n}+\hat{i}}\}. \quad (\text{VI.1})$$

Here $\mathbf{n} = (n_1, n_2)$ is a two dimensional vector with integral components labelling lattice sites on the square lattice; \hat{i} is a unit vector in the direction i ($i = 1$ or 2); $s_{\mathbf{n}}$ is the spin variable site \mathbf{n} , and $K = -J/kT$ where J is the spin-spin coupling strength. The spin $s_{\mathbf{n}}$ can have only the values ± 1 , and $\sum_{\{s\}}$ is a sum over all possible configurations of all the spins.

The most straightforward way to calculate Z would be to calculate the exponential numerically for each configuration and add up the results for all configurations. For a square lattice of size $N \times N$, there are 2^{N^2} configurations. The maximum reasonable size for a computer calculation at the present time is about 10^{10} multiplications (this requires a few hours on a CDC 7600). This means N^2 must be less than or near 33, i.e., $N \leq 6$, for a direct computer calculation. There are, of course, numerical tricks and procedures which allow one to increase this bound on N , but not by a large factor.

If K is close to the critical point value K_c there is a large correlation length ξ , say $\xi = 1000$ in units of the lattice spacing. In this case to obtain a sufficiently accurate approximation to the thermodynamic limit ($N \rightarrow \infty$) one must consider values of N of order ξ , and this is hopeless: there are of order $2^{1000000}$ configurations.

In the renormalization group approach the calculation of Z is broken down into a number of steps such that each step is a feasible calculation at least to a good approximation. For the Kadanoff method, this is done as follows. Instead of summing over all the spins on the lattice, the first step consists of a summation over half the spins, the other half being held fixed. The process of summing over half the spins is called "spin decimation." The spins that are held fixed lie on alternate diagonals; this is illustrated in Fig. 5. It is convenient to relabel the spins held fixed. Thus s_{11} becomes t_{10} , s_{02} becomes t_{11} , etc., as illustrated in Fig. 5. The advantage of this is that the set of "new spins"

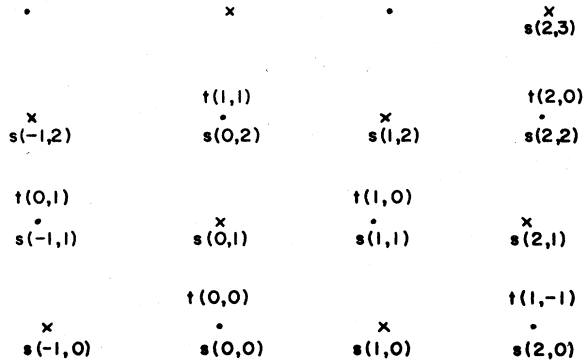


FIG. 5. Lattice of old spins (s) and new spins (t) for the Kadanoff transformation. New spins occur only at sites marked by a dot. Spins are written $s(\mathbf{n})$ rather than s_n .

t_m by themselves define a square lattice of Ising spins—the lattice label \mathbf{m} for the new spins runs through the same set of values as the original index \mathbf{n} (for an infinite size lattice).

The Kadanoff transformation can be written

$$\exp\{\mathcal{H}_1[t]\} = \sum_{\{s\}} \left\{ \prod_{\mathbf{m}} \delta_{t_m s_{\mathbf{n}(\mathbf{m})}} \right\} \exp\{\mathcal{H}_0[s]\}, \tag{VI.2}$$

where the sum $\sum_{\{s\}}$ is formally over all the spins s , but the set of Kronecker δ 's ensure that half the spins are held fixed. The function $\mathbf{n}(\mathbf{m})$ gives the label \mathbf{n} of the spin s_n in terms of the label \mathbf{m} of the corresponding new spin t_m [for $\mathbf{m} = (1,0)$, $\mathbf{n} = (1,1)$, etc.]. Also

$$\mathcal{H}_0[s] = K \sum_{\mathbf{n}} \sum_i s_n s_{\mathbf{n}+\mathbf{e}_i}. \tag{VI.3}$$

The result of the calculation is a function of the new spins t_m ; it is convenient to write the result in the form $\exp\{\mathcal{H}_1[t]\}$.

The partition function can be calculated from $\mathcal{H}_1[t]$ by summing over all the configurations of the spins t_m

$$Z = \sum_{\{t\}} \exp\{\mathcal{H}_1[t]\}. \tag{VI.4}$$

This calculation has exactly the same form as the original partition function (VI.1), since the t 's also lie on a square lattice and also have only the values ± 1 . However, if the original spins s_n are restricted to a finite, $N \times N$ lattice, the new spins t form a lattice with only $N^2/2$ sites, i.e., a lattice with fewer configurations.

The second and succeeding steps in the Kadanoff procedure are the same as the first. The second step consists of summing over half the spins of the t lattice holding alternate diagonals of spins fixed. These steps can be written as a recursion formula

$$\exp\{\mathcal{H}_i[t]\} = \sum_{\{s\}} \left\{ \prod_{\mathbf{m}} \delta_{t_m s_{\mathbf{n}(\mathbf{m})}} \right\} \exp\{\mathcal{H}_{i-1}[s]\}, \tag{VI.5}$$

where the old spins for the i th step are labeled by s , the new spins by t , and the old spins for the i th step are the new spins from the previous ($i - 1$ st) step.

If the original lattice was of finite size ($N \times N$) then after a finite number of steps (of order $\ln_2 N$) all the spins will be summed over and one will have the partition function. For an infinite lattice the recursion formula (VI.5) can be iterated indefinitely, but the renormalization group formalism makes this unnecessary.

An extensive formal discussion of transformations like (VI.5) has been given by Wegner (1972). The present lecture will be concerned with more practical aspects of the transformation.

The easiest way to demonstrate the advantages of the Kadanoff transformation is to compute explicitly the interaction $\mathcal{H}_1[t]$. This can be done exactly. It is not possible to compute \mathcal{H}_2 , \mathcal{H}_3 , etc., exactly: the approximations needed to compute \mathcal{H}_2 , \mathcal{H}_3 , etc., will be discussed later.

Consider a specific spin to be summed over in the first step (Eq. VI.2), namely s_{01} . This spin is coupled (in \mathcal{H}_0) to its four nearest-neighbor spins. As one can see from Fig. 5, these four nearest-neighbor spins are all held fixed. Thus s_{01} does not couple to any other of the spins being summed over. This means one can perform the sum over s_{01} independently of the sums over the other spins. The sum is

$$S = \sum_{s_{01}=\pm 1} \exp\{K s_{01}(t_{00} + t_{01} + t_{10} + t_{11})\} \\ = 2 \cosh K(t_{00} + t_{01} + t_{10} + t_{11}). \tag{VI.6}$$

The sums over all the other spins give the same type of result; thus $\exp\{\mathcal{H}_1[t]\}$ is simply an infinite product of cosh functions, each cosh depending on a different set of four t 's.

It is more convenient to have $\mathcal{H}_1[t]$ expressed as a polynomial in the t 's. This is easily accomplished. Using the fact that each t_m can only have values ± 1 , one can establish the following identity

$$2 \cosh K(t_{00} + t_{01} + t_{10} + t_{11}) \\ = \exp\{A(K) + B(K) \\ \times (t_{00}t_{01} + t_{00}t_{10} + t_{00}t_{11} + t_{01}t_{10} + t_{01}t_{11} + t_{10}t_{11}) \\ + C(K)t_{00}t_{10}t_{01}t_{11}\}. \tag{VI.7}$$

To prove this identity, one must show that it is true for all configurations of the spins t_{00}, \dots, t_{11} . Both sides of the identity are symmetric to interchanges of the t 's and to changing the signs of the t 's ($t_m \rightarrow -t_m$ for all \mathbf{m}).

This leaves three independent configurations to consider:

- (a) all $t_m = +1$,
- (b) one $t_m = -1$, the three others $+1$
- (c) two $t_m = -1$, two $t_m = +1$.

To satisfy the identity for these three configurations one must have

$$2 \cosh 4K = \exp\{A + 6B + C\}, \tag{VI.8}$$

$$2 \cosh 2K = \exp\{A - C\}, \tag{VI.9}$$

$$2 \cosh 0 = \exp\{A - 2B + C\}. \tag{VI.10}$$

These equations can be solved to give A , B , and C

$$A(K) = \ln 2 + \frac{1}{8} \{ \ln \cosh 4K + 4 \ln \cosh 2K \} \quad (\text{VI.11})$$

$$B(K) = \frac{1}{8} \ln \cosh 4K \quad (\text{VI.12})$$

$$C(K) = \frac{1}{8} \{ \ln \cosh 4K - 4 \ln \cosh 2K \}. \quad (\text{VI.13})$$

This proves the identity.

The interaction \mathcal{H}_1 now has the form

$$\begin{aligned} \mathcal{H}_1[\mathcal{I}] = & \frac{N^2}{2} A(K) + \sum_n \sum_i 2B(K) t_n t_{n+i} + \sum_n \sum_{\pm} B(K) \\ & \times t_n t_{n+\hat{1}\pm\hat{2}} + \sum_n C(K) t_n t_{n+\hat{1}} t_{n+\hat{2}} t_{n+\hat{1}+\hat{2}}. \end{aligned} \quad (\text{VI.14})$$

There is a constant term proportional to $A(K)$ with a coefficient $N^2/2$ which is proportional to the total area of the lattice. There is a new nearest-neighbor interaction with coefficient $2B(K)$. The factor 2 appears because each nearest-neighbor coupling, for example $t_{10}t_{11}$, is generated from two different sums ($t_{10}t_{11}$ comes both from the sum over s_{01} and the sum over s_{12}). There is a diagonal nearest-neighbor term ($t_{00}t_{11}$ is an example of a diagonal nearest-neighbor product) with coefficient $B(K)$, and a four spin coupling term with coefficient $C(K)$.

The critical value K_c for K is known to be 0.440687 from Onsager's solution. Eqs. (VI.11)–(VI.13) give

$$A(K_c) = 1.00376, \quad (\text{VI.15})$$

$$B(K_c) = 0.137327, \quad (\text{VI.16})$$

$$C(K_c) = -0.035960. \quad (\text{VI.17})$$

The calculation of $\mathcal{H}_1[\mathcal{I}]$ is evidently much simpler than the calculation of Z . Instead of computing 2^{N^2} configurations of N^2 spins, the problem reduced to summing over 1 spin coupled to four other fixed spins.

A similar but less trivial and only approximate simplification applies to subsequent steps. Consider the second step, to be specific. The diagonal nearest-neighbor and four spin interactions in $\mathcal{H}_1[s]$ couple different spins being summed over. For example, s_{01} couples to s_{12} through the diagonal nearest-neighbor coupling. Hence the sum over s_{01} cannot be separated from sums over the other spins exactly. However, the strengths of the diagonal nearest-neighbor and four spin couplings are small. Even at the critical point these couplings have strengths 0.14 and -0.04 as compared to the nearest-neighbor coupling 0.27 and the original nearest-neighbor coupling 0.44. This suggests that the diagonal nearest neighbor and four spin couplings can be treated by perturbation theory.

If the diagonal and four spin couplings are treated to first order, for example, then s_{01} couples to s_{12} but one can ignore s_{23} in calculating the sum over s_{01} (for reference see Fig. 5). The reason for this is that it takes the product of two diagonal couplings ($s_{01}s_{12}$ and $s_{12}s_{23}$) to obtain a term coupling s_{01} to s_{23} . In general, to any finite order in perturbation theory the sum over s_{01} involves only a finite number of other spins, even if the lattice and correlation length have a large size. Hence the calculation of $\mathcal{H}_2[\mathcal{I}]$

is also practical, provided one does not have to calculate too many terms in the perturbation expansion. For the nearest-neighbor Ising model, one knows that expansions in K have a radius of convergence given by $K_c \simeq 0.44$. Thus it seems likely that an expansion in the diagonal coupling strength 0.14 and four spin coupling -0.04 would be rapidly convergent.

In the second and subsequent steps, interactions of arbitrary complexity are generated, not just the nearest-neighbor, diagonal nearest-neighbor, and four spin couplings. One expects the couplings of complex interactions to be small because they arise only from the perturbation expansion. As long as the nearest-neighbor term dominates in \mathcal{H}_{i-1} and the remaining terms are small, one can use a perturbation expansion in the remaining terms to compute \mathcal{H}_i .

There are two general principles illustrated by the above discussion which are crucial to practical use of renormalization group methods. The first principle is that the interaction $\mathcal{H}_i[s]$ should be dominated by short-range couplings such as the nearest-neighbor coupling. For the Kadanoff transformation it is specifically the nearest-neighbor term that must dominate. For other transformations such as the one discussed by Niemeyer and Van Leeuwen (1973) one could tolerate large diagonal nearest neighbor coupling. Neither method is practical if couplings are important for fourth or fifth nearest neighbors, say.

The second principle is that in the calculation of a particular term in $\mathcal{H}_i[\mathcal{I}]$, for example the coefficient of $t_{10}t_{11}$, only a finite and small number of old spins s_n must be involved. In calculating \mathcal{H}_1 only two spins, s_{01} and s_{12} , contribute to the $t_{10}t_{11}$ term. In calculating \mathcal{H}_2 and subsequent interactions, more spins are involved, through the perturbation calculations. If too many more spins are involved there are too many configurations to sum over and the calculation is impractical.

The author has carried out calculations of the Kadanoff transformation involving no more than 15 spins simultaneously (15 spins have 65 000 configurations) which give critical exponents to an accuracy of about 0.2%. These calculations will be described briefly at the end of this lecture.

To illustrate the use of the renormalization group transformation to obtain critical exponents the Kadanoff transformation in the simplest approximation will be discussed. For this purpose $\mathcal{H}_{i-1}[s]$ will be truncated to two types of coupling: a nearest-neighbor coupling with strength K_{i-1} and a diagonal nearest-neighbor coupling with strength L_{i-1} . The interaction $\mathcal{H}_i[s]$ will be calculated as a perturbation expansion in both K_{i-1} and L_{i-1} , to second order in K_{i-1} and first order in L_{i-1} . The result is

$$K_i = 2K_{i-1}^2 + L_{i-1}, \quad (\text{VI.18})$$

$$L_i = K_{i-1}^2 \quad (\text{VI.19})$$

{The constant term in $\mathcal{H}_i[s]$ will not be studied; for a calculation including these constants see Nauenberg and Nienhuis (1974a,b)}. The terms $2K_{i-1}^2$ and K_{i-1}^2 follow from an expansion of $B(K)$ to order K^2 ; the L_{i-1} term

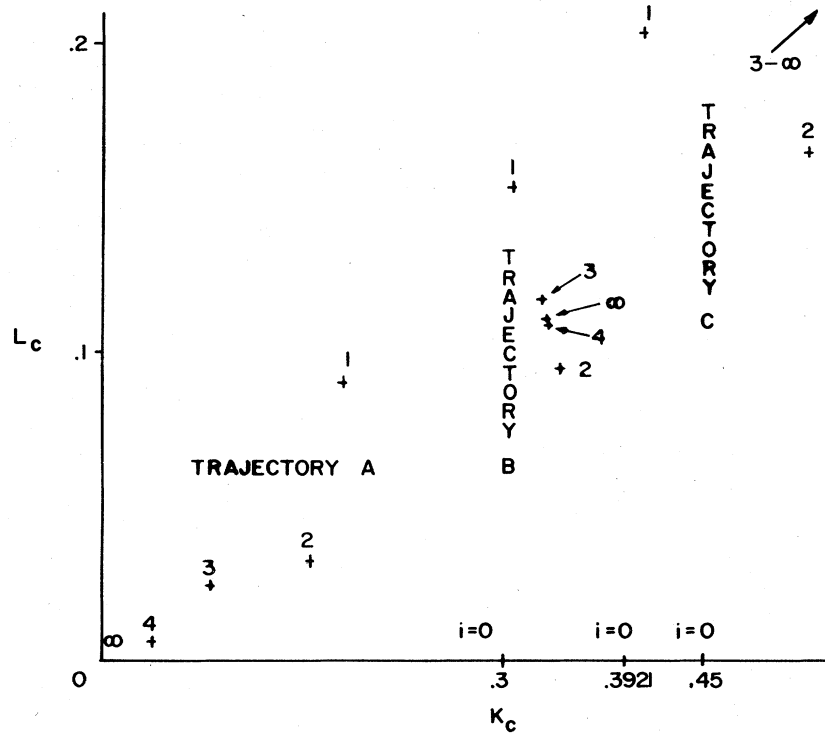


FIG. 6. Plot of the sequences (K_i, L_i) for three initial conditions. For trajectory (A), $K_0 = 0.3, L_0 = 0$. For (B), $K_0 = K_{0c} \approx 0.3921, L_0 = 0$. For (C), $K_0 = 0.45, L_0 = 0$. For $i \rightarrow \infty$, trajectory A goes to 0, trajectory B goes to the fixed point $K^* = \frac{1}{3}, L^* = 1/9$, and trajectory C goes to ∞ .

arises from the diagonal nearest neighbor s - s couplings which translate directly into nearest-neighbor t - t couplings (e.g., $s_{00}s_{11}$ becomes $t_{00}t_{10}$).

Suppose one specifies an initial nearest-neighbor coupling K_0 (with $L_0 = 0$) and iterates the equations (VI.18) and (VI.19). In the limit of large i , one finds (from actual calculation) three types of behavior. For illustration, see Fig. 6.

(a) If $0 < K_0 < 0.3921$ then $K_i \rightarrow 0$ and $L_i \rightarrow 0$ for large i . It is evident without calculation that this happens for very small K_0 .

(b) If $K_0 > 0.3921$ then $K_i \rightarrow \infty, L_i \rightarrow \infty$. This is evident if K_0 is very large.

(c) There is a unique value K_{0c} for K_0 , namely $K_{0c} \approx 0.3921$ for which K_i and L_i approach a nontrivial fixed point K^*, L^* for $i \rightarrow \infty$. The fixed point satisfies

$$K^* = 2K^{*2} + L, \tag{VI.20}$$

$$L^* = K^{*2}. \tag{VI.21}$$

These equations have three solutions: $K^* = L^* = 0, K^* = L^* = \infty$, and the nontrivial solution

$$K^* = \frac{1}{3}, \tag{VI.22}$$

$$L^* = \frac{1}{9}. \tag{VI.23}$$

To understand the significance of these three alternatives one must understand the meaning of K_i and L_i for

large i . A nearest-neighbor pair, say t_{00} and t_{01} , after i steps, corresponds to spins on the original lattice with separation $\sqrt{2}^i$ in units of the original lattice spacing. For example, after one step, t_{00} and t_{01} correspond to s_{00} and s_{11} which have separation $\sqrt{2}$; after two steps the new spin t_{00} and t_{10} correspond to s_{00} and s_{02} on the original lattice with separation 2, etc. The coupling between t_{00} and t_{01} becomes small (small compared to the fixed point value 0.333) when $\sqrt{2}^i$ becomes larger than the correlation length ξ , for in this situation there should be very little coupling between t_{00} and t_{01} . When $K_0 = K_{0c}$ the correlation length must be infinite since K_i never becomes small. Hence K_{0c} represents an approximate value for K_c , the critical value. There is about a 10% difference between K_c and K_{0c} . $K_0 > K_{0c}$ corresponds to temperatures below the critical temperature. The behavior $K_i \rightarrow \infty$ for $i \rightarrow \infty$ ensures the existence of spontaneous magnetization: for large K_i and L_i the dominant configuration is for t_n all to be +1 or t_n all to be -1.

As explained in Sec. III, the susceptibility exponent ν can be calculated from the leading eigenvalue λ of the linearized equations about the fixed point. Let $K = K^* + k, L = L^* + l$ with k and l small. Then the linearized equations giving λ are

$$\lambda k = 4K^*k + l, \tag{VI.24}$$

$$\lambda l = 2K^*k. \tag{VI.25}$$

This leads to the equation

$$\lambda^2 = \frac{4}{3}\lambda + \frac{2}{3} \tag{VI.26}$$

which has two solutions

$$\lambda_1 = 1.7208, \tag{VI.27}$$

$$\lambda_2 = -0.3874. \tag{VI.28}$$

The dominant eigenvalue is λ_1 . The calculation of ν from λ_1 is different from that of Sec. III because the length scale changes by $\sqrt{2}$ per iteration instead of 2; hence

$$\nu = (\ln\sqrt{2})/\ln\lambda_1 \tag{VI.29}$$

which gives $\nu = 0.638$.

The Onsager value for ν is 1; hence the result (VI.29) is not very good.

The transformation (VI.5) has a flaw. The difficulty arises in a study of the spin-spin correlation function Γ_n :

$$\Gamma_n = Z^{-1} \sum_{\{s\}} s_n s_0 \exp\{\mathcal{H}_0[s]\}. \tag{VI.30}$$

Let Γ_{in} be the correlation function for \mathcal{H}_i

$$\Gamma_{in} = Z^{-1} \sum_{\{s\}} s_n s_0 \exp\{\mathcal{H}_i[s]\}. \tag{VI.31}$$

There is a relation between Γ_i and Γ_{i-1} . This follows from the fact that new spins t_m and t_0 for stage i are the spins $s_{n(m)}$ and s_0 from stage $i - 1$. Hence

$$\begin{aligned} \Gamma_{im} &= Z^{-1} \sum_{\{t\}} t_m t_0 \exp\{\mathcal{H}_i[t]\} = Z^{-1} \sum_{\{t\} \{s\}} t_m t_0 \\ &\times \prod_{m'} \delta_{t_m, s_{n(m')}} \exp\{\mathcal{H}_{i-1}[s]\} = Z^{-1} \sum_{\{s\}} s_{n(m)} s_0 \\ &\times \exp\{\mathcal{H}_{i-1}[s]\} = \Gamma_{i-1, n(m)}. \end{aligned} \tag{VI.32}$$

At the critical point one knows from the Onsager-Yang calculation that

$$\Gamma_n = C/|\mathbf{n}|^{\frac{1}{2}} \tag{VI.33}$$

for $|\mathbf{n}|$ large, where C is a constant. From the recursion formula (VI.32) and the fact that

$$|\mathbf{n}(\mathbf{m})| = \sqrt{2} |\mathbf{m}|, \tag{VI.34}$$

one obtains (at the critical point)

$$\Gamma_{in} = C2^{-i/8}/|\mathbf{n}|^{\frac{1}{2}} \tag{VI.35}$$

for large $|\mathbf{n}|$. This formula contradicts the existence of a fixed point for large i : If $\mathcal{H}_i[s] \rightarrow \mathcal{H}^*[s]$ for large i then Γ_{in} should approach the correlation function Γ_n^* for $\mathcal{H}^*[s]$. According to (VI.35), $\Gamma_{in} \rightarrow 0$ for $i \rightarrow \infty$. One does not expect $\mathcal{H}^*[s]$ to have a zero correlation function, at least not if $\mathcal{H}^*[s]$ is a simple short-range type interaction.

Kadanoff (unpublished) has generalized the transformation (VI.5) to avoid this flaw. Kadanoff's generalization reads

$$\exp\{\mathcal{H}_i[t]\} = \sum_{\{s\}} \prod_m \frac{(1 + \rho t_m s_{n(m)})}{2} \exp\{\mathcal{H}_{i-1}[s]\}, \tag{VI.36}$$

where ρ is an arbitrary parameter in the transformation. The sum over $\{s\}$ is now a sum over all spins s , not just half of them, since there are no Kronecker deltas in Eq.

(VI.36). However, it is still true that

$$Z = \sum_{\{t\}} \exp\{\mathcal{H}_i[t]\} = \sum_{\{s\}} \exp\{\mathcal{H}_{i-1}[s]\} \tag{VI.37}$$

since the terms $\rho t_m s_{n(m)}$ sum to zero when the right-hand side of (VI.36) is summed over all t_m . (Note that each new spin t_m appears only in one of the factors of the product.) For the spin-spin correlation function one has

$$\begin{aligned} \Gamma_{im} &= Z^{-1} \sum_{\{t\}} t_m t_0 \exp\{\mathcal{H}_i[t]\} = Z^{-1} \sum_{\{t\} \{s\}} t_m t_0 \\ &\times \prod_{m'} \frac{(1 + \rho t_m s_{n(m')})}{2} \exp\{\mathcal{H}_{i-1}[s]\}. \end{aligned} \tag{VI.38}$$

The sums over the specific new spins t_m and t_0 result in factors $\rho s_{n(m)}$ and ρs_0 , respectively; thus

$$\Gamma_{im} = \rho^2 \Gamma_{i-1, n(m)}. \tag{VI.39}$$

Therefore, at the fixed point and for large $|\mathbf{m}|$

$$\Gamma_{im} = \rho^{2i} 2^{-i/8} (C/|\mathbf{m}|^{\frac{1}{2}}). \tag{VI.40}$$

This expression has a nontrivial limit if

$$\rho = \rho^* = 2^{1/16}. \tag{VI.41}$$

Thus if $\rho = 2^{1/16}$ there is no contradiction between the Onsager-Yang result (VI.33) and the existence of a fixed point. Alternatively, if one did not know the Onsager-Yang result, knowing only that

$$\Gamma_n = C/|\mathbf{n}|^\eta \tag{VI.42}$$

for large \mathbf{n} , then one would have

$$\Gamma_{im} = \rho^{2i} 2^{-i\eta/2} C/|\mathbf{m}|^\eta \tag{VI.43}$$

which has a nontrivial limit when

$$\rho = \rho^* = 2^{\eta/4}. \tag{VI.44}$$

The simple approximate recursion equations (VI.18) and (VI.19) are easily generalized to include ρ . Keeping to second order in K_{i-1} and first order L_{i-1} , one obtains

$$K_i = \rho^2 (2K_{i-1}^2 + L_{i-1}), \tag{VI.45}$$

$$L_i = \rho^2 K_{i-1}^2. \tag{VI.46}$$

The nontrivial fixed point is given by

$$K^* = (2\rho^2 + \rho^4)^{-1}, \tag{VI.47}$$

$$L^* = \rho^2 (2\rho^2 + \rho^4)^{-2}. \tag{VI.48}$$

In the exact theory one expects to find a fixed point only for $\rho = \rho^* = 2^{1/16}$, but in the approximate theory there is a fixed point for any ρ . The resolution of this contradiction is simple. In the approximate theory one can treat K^* as the independent variable instead of ρ : for any K^* there is a value for ρ , say $\rho = \rho^*(K^*)$ for which there is a fixed point. This is also true in the exact theory. However, in the exact theory $\rho^*(K^*) = 2^{1/16}$ is independent of K^* . In the simple approximation (VI.47), $\rho^*(K^*)$ is far from constant.

The statement that $\rho^*(K^*)$ is constant in the exact theory is equivalent to the assertion that for $\rho = \rho^*$ there is a fixed *line*, namely a one parameter set of fixed points. The parameter can be chosen to be K^* . The proof of the

TABLE II. Results of block spin computations (see text for definitions of K^* , etc.)

K^*	ρ^*	η	ν	ρ_s^*	h_1	h_6	h_7
0.2384	1.030	0.17	0.95	1.131	0.0287	-0.00470	-0.0029
0.2618	1.041	0.23	0.984	1.093	0.0085	-0.00110	-0.0005
0.2897	1.044	0.2486	1.002	1.054	0.0013	-0.00006	-0.0002
0.2992	1.04421	0.2497	0.998	1.041	0.0012	-0.00010	-0.0006
0.3104	1.04422	0.2497	0.988	1.027	0.0021	-0.00040	-0.0014
0.3306	1.04417	0.2494	0.97	1.003	0.0059	-0.00140	-0.0040

existence of the fixed line is as follows. Let $\mathcal{J}C^*$ be a fixed point solution:

$$\exp\{\mathcal{J}C^*[t]\} = \sum_{\{s\}} \prod_m \frac{(1 + \rho^* t_m s_{n(m)})}{2} \exp\{\mathcal{J}C^*[s]\}. \quad (VI.49)$$

Then there is a one parameter family of fixed points $\mathcal{J}C^*[s; q]$ given by

$$\exp\{\mathcal{J}C^*[t, q]\} = \sum_{\{s\}} \prod_n \frac{(1 + q t_n s_n)}{2} \exp\{\mathcal{J}C^*[s]\}. \quad (VI.50)$$

In this formula there is a new spin t_n for every s_n . The proof that $\mathcal{J}C^*[t; q]$ is a fixed point is straightforward. One must compute

$$\sum_{\{t\}} \prod_m \frac{[1 + \rho^* u_m t_{n(m)}]}{2} \exp\{\mathcal{J}C^*[t; q]\}. \quad (VI.51)$$

This must be $\exp[\mathcal{J}C^*(u; q)]$. From Eq. (VI.50), this is

$$\sum_{\{t\}} \sum_{\{s\}} \prod_m \frac{[1 + \rho^* u_m t_{n(m)}]}{2} \prod_n \frac{(1 + q t_n s_n)}{2} \exp\{\mathcal{J}C^*[s]\}. \quad (VI.52)$$

The sum over the t spins can be performed explicitly, giving

$$\sum_{\{s\}} \prod_m \frac{[1 + \rho^* q u_m s_{n(m)}]}{2} \exp\{\mathcal{J}C^*[s]\}. \quad (VI.53)$$

The formula (VI.53) is also obtained by summing the t spins in the following expression

$$\sum_{\{s\}} \sum_{\{t\}} \prod_m \frac{[1 + q u_m t_m]}{2} \prod_n \frac{[1 + \rho^* t_n s_{n(m)}]}{2} \exp\{\mathcal{J}C^*[s]\}. \quad (VI.54)$$

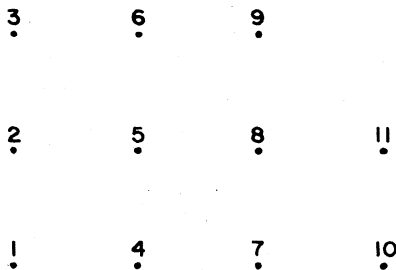


FIG. 7. Subset of lattice containing all interactions included in the computer calculations. The numbering of the lattice sites is for convenience.

This expression (VI.54) is

$$\sum_{\{t\}} \prod_m \frac{[1 + q u_m t_m]}{2} \exp\{\mathcal{J}C^*[t]\} = \exp\{\mathcal{J}C^*[u; q]\}.$$

Q.E.D. (VI.55)

Thus there is a fixed line $\mathcal{J}C^*[s; q]$ given by Eq. (VI.50) when $\rho = \rho^*$. Hence one expects that if one solves for ρ as a function of K^* in the exact theory, one will get $\rho = \rho^*$ independently of K^* : there will be no solutions for $\rho \neq \rho^*$.

The transformation of Niemeier and Van Leeuwen does not suffer the same flaw as the original Kadanoff transformation. The Niemeier-Van Leeuwen transformation involves a new spin which is defined to be the sign of the sum of three old spins. For this case one cannot relate the spin-spin correlation functions for different stages i , and no parameter ρ is required. The question of when the parameter ρ is required has been studied in detail (in a different context) by Bell and Wilson (1974).

The simple approximation to the Kadanoff transformation discussed above gives no hint that $\rho^*(K^*)$ should be independent of K^* . To test this (and other hypotheses) the author set up a much more accurate approximation to the Kadanoff approximation (see below). Calculations were performed on a CDC 7600 computer: one iteration of the transformation required 3.3 sec. Fixed points were obtained for values of K^* from 0.24 to 0.33. The results are as in Table II.

The value of η was calculated using Eq. (VI.44); the value of ν was obtained from the largest eigenvalue of the linearized transformation about the fixed point using Eq. (VI.29). The column ρ_s^* gives the value of $\rho^*(K^*)$ calculated from the simple approximation (VI.47). The columns labeled h_1 , h_6 , and h_7 give measures of the error of the computer calculation (see below).

The values of ρ^* for the computer calculation are close to constant: they vary by only 0.015 while the simple approximation gives a range 0.13. Where the computer error (measured by h_1 , h_6 , and h_7) is smallest (for $K^* = 0.2897$ and $K^* = 0.2992$) both ρ^* and ν vary hardly at all. The results for $K^* = 0.2897$ and 0.2992 are also in extraordinarily good agreement with the exact results (the exact numbers are $\eta = 0.25$ and $\nu = 1$).

The computer calculation will now be outlined. A number of details are omitted. In the computer calculation, $\mathcal{J}C_i[s]$ was represented by 217 different interaction constants including two-spin, four-spin, six-spin, eight-spin, and ten-

spin couplings. The couplings considered were all possible couplings that fit into a 3×3 square, plus a subset of couplings that fit into the set of 11 lattice sites shown in Fig. 7. (A typical coupling constant is the coefficient of the product $s_{15}s_6s_{10}$ using the numbering of Fig. 7.)

The perturbation approach described earlier in this lecture was not used in the computer calculation. Instead, the infinite lattice was approximated by a finite square lattice of size roughly 15×15 in units of the lattice spacing. The square boundaries were at 45° to the lattice axes (Fig. 8). Over most of this square, the new spin variables t_m were held fixed at $+1$. Within a smaller area in the center of the square (marked B in Fig. 8) containing 11 new spins (in the same geometry as Fig. 7) both values ± 1 were considered for the new spins.

The sum over all configurations of the old spins was carried out sequentially, starting with the spin numbered 1 in Fig. 8, and continuing along the row of spins numbered 2, 3, etc. Then the spins in the second row (12, 13, etc.) were summed over. This sequential pattern was used to sum over all the spins outside the inner area B. This calculation was not performed exactly. Consider a typical stage in this calculation, namely the sum over spin number 15. Spins 1 through 14 have already been summed over, leaving a function of the remaining 203 spins. This

function was written in exponential form, e.g.,

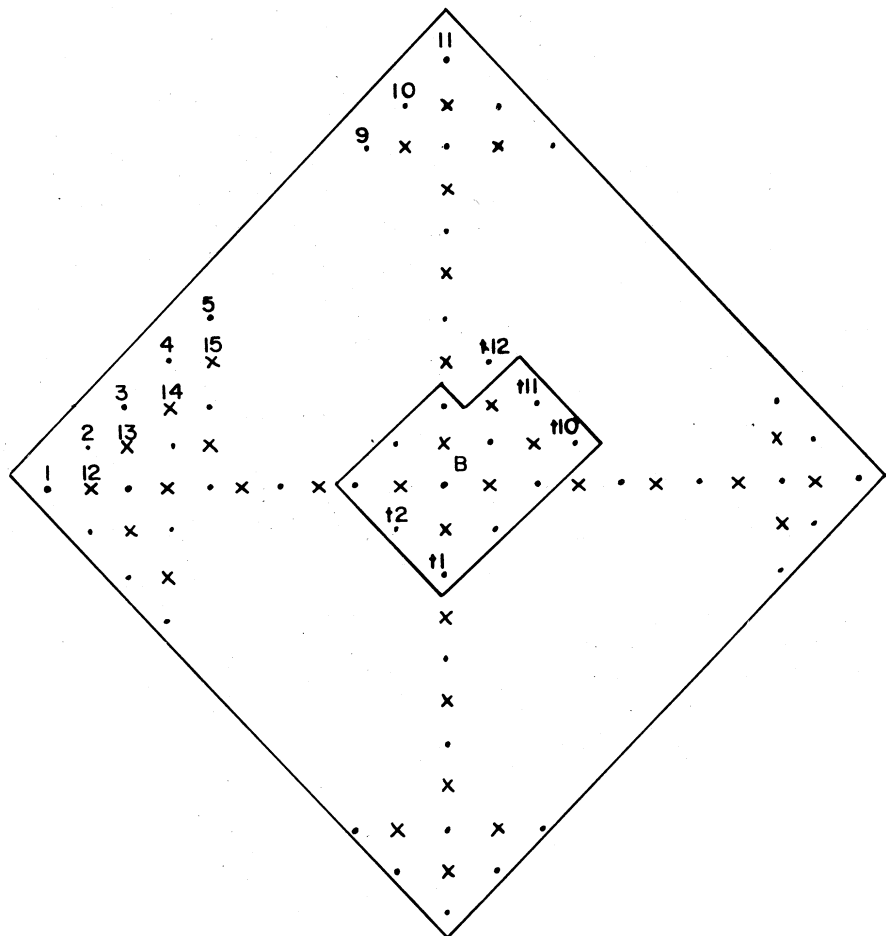
$$\exp\{\mathcal{H}_{\text{eff}}[s_{15}, \dots, s_{217}]\}.$$

The interaction \mathcal{H}_{eff} was written in polynomial form, and then only the coupling of s_{15} to nearby spins was included in the calculation. The "nearby spins" are those shown in Fig. 9. The sum over s_{15} was then performed for all configurations of the 25 nearby spins of Fig. 9; the remaining 178 spins were fixed at $+1$. (In practice it was possible to reduce the nearby spins to the subset C of 13 spins shown in Fig. 9. This was possible with the understanding that induced couplings over too long a distance were neglected, e.g., a coupling $s_{22}s_{26}$ was neglected. This is not a new approximation, since such a coupling would have been dropped anyways when the sum over s_{22} was performed. Further details of simplifications like this will not be reported here.)

Once all spins outside the area B in Fig. 8 were summed over, the second step was a sum over all the spins inside the area B. This sum was performed exactly for all configurations of the new spins t_1 to t_{11} .

The result of this calculation was $\exp\{\mathcal{H}_i[t]\}$ evaluated for all configurations in the t 's inside the area B, with all t 's outside B set equal to $+1$. In principle this would provide more than sufficient information to determine the

FIG. 8. Subset of lattice used in computer calculations. In region outside B, all new spins t_m were set equal to 1; inside B both values ± 1 were used.



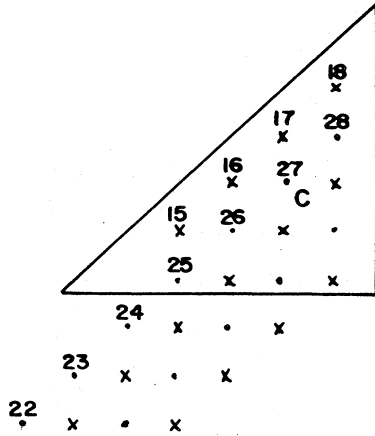


FIG. 9. Subset of spins included in the calculation of the sum over s_{15} . In the actual computer program it was possible to replace a single calculation over this subset of spins by several calculations over regions no larger than C.

217 coupling constants of \mathcal{H}_C , since there are 2048 configurations of the t 's in the region B. In practice this information is difficult to extract. The difficulty is due to terms in the interaction which contain products across the boundary of B, e.g., $t_{10}t_{12}$. These terms contribute to the dependence of $\mathcal{H}_C[t]$ on t_1, \dots, t_{11} when the spins t_{12} , etc. are equal to 1; this makes it difficult to separate out specific coupling constants. To simplify this and other parts of the calculations a different representation of the interaction was used.

First, to avoid exponentials in the computer program, the interaction was represented in product form. That is, terms like $\exp\{M s_1 s_6 s_7 s_{10}\}$ were replaced by the equivalent expression $(\cosh M) (1 + M' s_1 s_6 s_7 s_{10})$ where $M' = \tanh M$. This formula is equivalent to the exponential since $s_1 s_6 s_7 s_{10}$ can only be ± 1 . $\exp\{\mathcal{H}_{C_{i-1}}[s]\}$ was then a constant times a product of terms like $(1 + M' s_1 s_6 s_7 s_{10})$. The coupling constant M' was used in the program in place of M . This representation was used for input and output in the program. But in the actual calculation of \mathcal{H}_C from $\mathcal{H}_{C_{i-1}}$ a second representation was used. Let

$$s_m = 1 - 2\sigma_m, \tag{VI.56}$$

$$t_m = 1 - 2\tau_m. \tag{VI.57}$$

The values of σ_m and τ_m are 0 and 1. The second representation of $\mathcal{H}_{C_{i-1}}$ was as products of terms like $(1 + M'' \sigma_1 \sigma_6 \sigma_7 \sigma_{10})$. (The constant M'' is not a function only of M' ; instead M'' depends on all the couplings in $\mathcal{H}_{C_{i-1}}$ which contain the spins s_1, s_6, s_7 , and s_{10} . A special conversion program was needed to convert $\mathcal{H}_{C_{i-1}}$ from the s representation to the σ representation. After \mathcal{H}_C was calculated in the τ representation, another conversion program generated \mathcal{H}_C in the t representation.) The advantage of the σ and τ representation is that when s_m or t_m is equal to ± 1 then σ_m or τ_m is equal to 0. This means all products containing σ_m or τ_m are also 0.

The approximate form of the transformation described above had one unexpected benefit: it violates the symmetry to $s_n \rightarrow -s_n$ for all n . The reason for this violation is the specification $t_m = +1$ outside the area B of Fig. 8.

As a consequence, even if $\mathcal{H}_{C_{i-1}}[s]$ contained only symmetric interactions (i.e., only products of an even number of spins), the new interaction $\mathcal{H}_C[t]$ contains odd interactions. The strength of the odd interactions in $\mathcal{H}_C[t]$ provided a good over-all measure of the accuracy of the approximations. To prevent accumulation of error over many iterations, these odd interactions were dropped (in the t representation) before starting the next iteration. The size of these odd interactions dropped considerably as the computer program was improved to its final form (the author found many ways to increase the efficiency of the program, which in turn allowed increases in the number of "nearby" spins.) The columns labeled h_1, h_6 , and h_7 in the table above show the coefficients of the odd interactions $t_1, t_1 t_6 t_7$, and $t_1 t_2 t_3$, respectively (using the spin numbering of Fig. 7). The table shows the values of these odd interactions obtained from one iteration of the fixed point $\mathcal{H}^*(K^*, s)$.

To locate the fixed point $\mathcal{H}^*[s]$ for a given value of K^* , a combination of straight iteration and Newton's method was used. First a rough estimate of the fixed point was guessed at. Secondly, 18 coupling constants were picked out as being the most important (including the nearest- and next nearest-neighbor couplings). Then for fixed values of these 18 coupling constants and a fixed value of ρ , the transformation (VI.36) was iterated until the remaining couplings stabilized (3 iterations, in practice). (In this calculation the new values of the 18 picked couplings were thrown out and replaced by their previously assigned values after each iteration.) At the conclusion of this iteration the new values of the 18 picked couplings were computed and compared to the fixed values. Newton's method (using derivatives with respect to each of the 18 picked couplings, plus the derivative with respect to ρ) was then used to approach a fixed point for all couplings. Each derivative was computed numerically by recalculating the iteration with a small change in the appropriate coupling constant.

The eigenvalue λ was obtained by starting very close to the fixed point but not at the fixed point, and then iterating the transformation (VI.36) until all nonleading eigenoperators had negligible coefficients.

The largest couplings in the fixed point for $K^* = 0.2817$ are shown in Table III (using the spin numbering of Fig. 7). To give an indication of the sizes of the remaining couplings the table below gives the distribution of couplings by size, using factor of 2 bins, e.g., bin 5 contains all couplings whose absolute value lies between 2^{-5} and 2^{-6} . Table IV gives the maximum coupling for each bin (col-

TABLE III. Dominant spin couplings in the fixed point Hamiltonian for $K^* = 0.2817$. The spin numbering shown in Fig. 7.

Spin product	Coefficient	Spin product	Coefficient
$s_1 s_2$	0.281758	$s_1 s_2 s_5 s_6$	0.001762
$s_1 s_5$	0.095562	$s_1 s_2 s_5 s_6$	-0.001615
$s_1 s_2 s_4 s_5$	-0.017242	$s_2 s_4 s_6 s_7$	-0.001045
$s_1 s_3$	0.008422	$s_1 s_3 s_4 s_5$	-0.001023
$s_1 s_6$	0.004704	$s_1 s_2 s_3 s_4$	0.000736
$s_2 s_4 s_6 s_8$	-0.004008	$s_1 s_3 s_4 s_6$	-0.000612
$s_2 s_5 s_6 s_7$	0.001803	$s_1 s_5 s_6 s_7$	0.000575

TABLE IV. Distribution of coupling strengths in factor of 2 bins for the Hamiltonian with $K^* = 0.2817$. The column labeled Max gives the upper end of each bin.

Bin	Max	No. of couplings	Bin	Max	No. of couplings	Bin	Max	No. of couplings
1	0.5	1	9	0.00195	5	16	0.000015	31
2	0.25	0	10	0.00098	3	17	0.0000076	31
3	0.125	1	11	0.00049	9	18	0.0000038	32
4	0.0625	0	12	0.00024	9	19	0.0000019	23
5	0.0312	1	13	0.00012	9	20	0.0000009	9
6	0.0156	1	14	0.00006	11	21	0.0000005	11
7	0.0078	2	15	0.00003	20	22	0.00000024	3
8	0.0039	0						

um MAX) and the number of couplings per bin. The couplings in bins 11, 12, and 13 can be important because there are so many of them; the couplings in higher bins are probably negligible. This has not been investigated however.

There are many further questions about this calculation that will not be discussed here. The principal conclusion is however clear: one can do precise calculations using pure renormalization group methods with the only approximations being based on locality. The approximations were (i) to restrict the interactions in $\mathcal{H}_i[s]$ to sufficiently local interactions, and (ii) to restrict the range of effective interactions generated by the sequential summation of spins outside the region B in Fig. 8, as discussed earlier. No perturbation expansions were used.

The dominant interaction at the fixed point is the nearest-neighbor coupling. For accurate calculations one must include many more; but for a qualitative picture the nearest-neighbor constant K plus perhaps the next nearest-neighbor coupling L should be enough. The remaining couplings are at least a factor 5 smaller than L . This is also true in the Niemeier-Van Leeuwen calculations. Thus the old idea of Kadanoff that there would be effective nearest-neighbor Ising models for block spins is very close to the truth.

It will be much more difficult to do precise calculations for the two-dimensional Heisenberg model or the three-dimensional Ising model. The reason is a practical one: the number of configurations needed becomes astronomical. The three dimensional analogue of Fig. 7, for example, would contain over 30 spins corresponding to $2^{30} \sim 10^9$ configurations. Thus a study of various methods will be needed to find the most economical one. For example, Kadanoff and Houghton (1975) and Kadanoff (1975) have recently obtained more accurate results than mine using a method that may be generalizable to three dimensions.

VII. THE KONDO PROBLEM: INTRODUCTION AND DEFINITION OF BASIS

The remaining sections will be concerned with the Kondo problem. A short summary of this work appears in Wilson (1974a). The classical Kondo problem (defined below) will be solved by approximate numerical calculations within a renormalization group framework. This means one will define a renormalization group transformation T and a sequence of effective Hamiltonians H_N generated by the transformation. Approximate representa-

tions of the Hamiltonians H_N involving only a finite number of parameters will be defined; these parameters have been calculated numerically. The calculations were performed on a CDC 7600 in runs requiring about ten minutes each.

The Kondo problem is an important problem in its own right. In addition, the solution of the Kondo problem is the first example where the full renormalization program (as the author conceives it) has been realized: the formal aspects of the fixed points, eigenoperators, and scaling laws will be blended with the practical aspect of numerical approximate calculations of effective interactions to give a quantitative solution (the present accuracy is a few percent) to a problem that previously had seemed hopeless. The errors of the numerical calculation have been determined (although not rigorously) as part of the calculation and can be reduced by using more computing time.

There have also been numerical calculations within the renormalization group framework for critical phenomena. The first calculations, using the "approximate recursion formula" (Wilson, 1971) were only qualitative, and no systematic procedure is known for improving the accuracy of these calculations. Some calculations have already been performed on the two-dimensional Ising case as discussed in the previous lecture; these calculations only confirm the known solution of the two-dimensional Ising model. So at present the Kondo calculation sets the standards for what a renormalization group calculation can accomplish.

It will also become evident that numerous tricks have been used to obtain the most practical formulation of the renormalization group transformation for the Kondo problem and to squeeze the maximum amount of information from the calculations. This is also true of the work on critical phenomena. It is likely to be true of other problems solved by renormalization group methods. The renormalization group formalism can be set up in a fairly logical manner to reduce a problem to one involving a finite number of degrees of freedom at each iteration. Unfortunately for practical calculations success or failure can depend on whether the finite number is 1 or 10 or 100, and inevitably one must resort to tricks to achieve 1 or 10 in place of 10 or 100. If one finds this prospect discouraging, one should remember that the successful tricks of one generation become the more formal and more easily learned mathematical methods of the next generation.

The Kondo problem is concerned with magnetic impurities in a nonmagnetic metal, and more particularly,

the question of whether the magnetic moment of the impurity persists down to zero temperature. It is a part of the more general problem of how ferromagnetism develops, which is not very well understood.

As a preliminary, a brief review of the thermodynamics of an electron spin will be given. Consider a system with two spin states $+\frac{1}{2}$ and $-\frac{1}{2}$. The Hamiltonian is $-\mu HS_z$ where S_z is the spin, H is the external field, and μ is the magnetic moment. The magnetization at a temperature T is

$$M = \mu \frac{\text{Tr}[S_z \exp(\mu HS_z/kT)]}{\text{Tr}[\exp(\mu HS_z/kT)]} = (\mu/2) \tanh(\mu H/2kT). \quad (\text{VII.1})$$

The susceptibility in zero field is $\partial M/\partial H$ for $H = 0$; this is easily seen to be

$$\chi(H = 0) = \mu^2(kT)^{-1} \text{Tr}S_z^2/\text{Tr}1 = \frac{1}{4} \mu^2(kT)^{-1}. \quad (\text{VII.2})$$

In particular, the susceptibility behaves as T^{-1} as a function of temperature.

For comparison consider a system of two electrons with (in the absence of a field) two energy levels, one having spin one, the other having spin zero. Assuming the two levels have an energy separation δE , consider the thermodynamics for $kT \ll \delta E$. In this case the excited state is irrelevant. The susceptibility is therefore (S_z is the total spin of the two electrons)

$$\chi(H = 0) = \mu^2(kT)^{-1} \text{Tr}S_z^2/\text{Tr}1, \quad (\text{VII.3})$$

where the trace is over the ground state levels and μ is the ground state moment. If the ground state has spin 0 then $\chi(H = 0)$ is 0; the effect of the excited state is to give an exponential dependence $\exp(-\delta E/kT)$ to χ . If the ground state has spin 1, then χ again behaves as T^{-1} as $T \rightarrow 0$.

The Kondo problem to be discussed here involves a single spin $\frac{1}{2}$ impurity coupled to the conduction band of a nonmagnetic metal. The impurity is said to have a moment if the susceptibility due to the impurity shows a $1/T$ dependence. (To obtain the susceptibility due to the impurity one has to subtract the susceptibility of the pure metal from the total susceptibility of metal plus impurity.)

The problem as described here is a considerable idealization of the experimental situation. Experimental data come from dilute alloys such as Cu-Fe, Cu-Mn, Au-Vn, etc. (e.g., copper is the metal, iron or manganese is the impurity). The impurities are present in concentrations around 0.01%. One tries to use concentrations small enough so that the susceptibility due to the impurities is linear in the impurity concentration, in which case one can extract the susceptibility due to a single impurity. In practice there are ferromagnetic couplings between impurities which make it difficult to obtain linearity at very low temperatures. A very serious complication of the experimental situation is that the conduction band involves d -state electrons, which are coupled to d -band electrons of the impurity. The Anderson Hamiltonian (Anderson, 1961) which describes this d -band coupling is much more complicated than the Kondo Hamiltonian defined below,

in which the conduction band contains only s -wave electrons. A final simplification of the model is that ordinary potential scattering of conduction band electrons by the impurity is ignored. However, it is relatively straightforward to generalize the model to include s -wave scattering.

The question one is interested in both experimentally and theoretically is the zero temperature behavior of an impurity with weak antiferromagnetic coupling to the conduction band. The weak coupling means that at high temperatures the coupling is negligible (especially if kT is much larger than the coupling energy) and the susceptibility shows the T^{-1} temperature dependence. The question is whether the T^{-1} dependence continues to zero temperature. Experimentally the best evidence indicates that χ is a constant at zero temperature [see, e.g. G. T. Rado and H. Suhl (1973) and Boyce and Slichter (1974)]. The calculation reported here shows that χ is a constant at 0 temperature and determines this constant as a function of the initial coupling strength: see Sec. IX.

One is also interested in the specific heat near zero temperature, the zero temperature resistivity, etc. In this paper the ratio of the specific heat to the susceptibility at zero temperature will be discussed in detail: the specific heat is found to be linear in T for $T \rightarrow 0$. The resistivity will not be discussed; it is dependent on the strength of potential scattering which has been ignored in the calculations to date. Many other questions of practical interest are ignored also; the purpose of these lectures is to describe the method of solution rather than detailed results (the calculations so far have given only the susceptibility and specific heat; other quantities have not been determined).

There exist many good reviews of the Kondo problem in the literature. Therefore, no further background on the Kondo problem will be given here. A good review of older theoretical work is given by Kondo (1969). Volume V of the Rado-Suhl books on Magnetism [Rado and Suhl (1973)] contains many articles on the Kondo problem. Rizzuto (1974) gives a recent experimental review. There are also reviews by Grüner (1974) and Grüner and Zawadowski (1974).

The relation of the calculations described here to previous work on the Kondo effect is, briefly, as follows. Yuval and Anderson (1970) and Anderson, Yuval, and Hamann (1970a, 1970b) had argued using an analogy to a one dimensional Coulomb gas that the zero temperature susceptibility is a constant instead of behaving as T^{-1} . Prior to their approach, theoretical predictions at 0 temperature generally involved $\ln T$ terms coming from higher orders of perturbation theory [as in Eq. (IX.57)]. The approach of Anderson and Yuval did not lead to precise quantitative results at 0 temperature; to the author's knowledge, the best calculations based on the Anderson-Yuval approach are the Monte Carlo calculations of Schotte and Schotte (1971) which will be discussed further in Lecture IX.

Anderson (1970) has given a formulation of the renormalization group approach which is very similar in spirit to the method complicated than the Kondo Hamiltonian defined below, in which the conduction band contains only s -wave electrons. A final simplification of the model is that

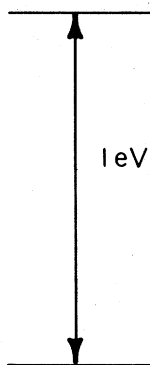


FIG. 10. Two level system illustrating the maximum energy scale of the conduction band.

ordinary potential scattering of conduction band electrons by the impurity is ignored. However, it is relatively straightforward to generalize the model to include s -wave scattering.

The question one is interested in both experimentally and theoretically is the zero temperature behavior of an impurity with weak antiferromagnetic coupling to the conduction band. The weak coupling means that at high temperatures the coupling is negligible (especially if kT is much larger than the coupling energy) and the susceptibility shows the T^{-1} temperature dependence. The question is whether the T^{-1} dependence continues to zero temperature. Experimentally the best evidence indicates that χ is a constant at zero temperature.³²

The calculation reported here shows that χ is a constant at 0 temperature and determines this constant as a function of the initial coupling strength: see Lecture IX.

One is also interested in the specific heat near zero temperature, the zero temperature resistivity, etc. In these lectures the ratio of the specific heat to the susceptibility at zero temperature will be discussed in detail: the specific heat is found to be linear in T for $T \rightarrow 0$. The resistivity will not be discussed; it is dependent on the strength of potential scattering which has been ignored in the calculations to date. The method developed here is more complex and more powerful than Anderson's but the basic ideas are the same. In addition, the methods developed here are a part of a continuing development of renormalization-group methods for Hamiltonian systems; earlier work is reported in Wilson (1965, 1970). Abrikosov and Migdal (1970) and Fowler and Zawadowski (1971) have applied the classical Gell-Mann-Low renormalization group methods to the Kondo problem, but they can only obtain high temperature results where perturbation theory can be used (see Sec. IX).

A survey of the Kondo calculation will now be given, before going into details. Another overview is presented in a talk by the author, published elsewhere (Wilson 1975). It is recommended (but not necessary) that the reader read Wilson (1975) before struggling with the remainder of this paper.

First a rough description of the energy scales in a conduction band will be given. Energy scales in quantum mechanics are determined by energy level spacings. The absolute value of an energy is unimportant (usually). The largest scale is a few electron volts; this is the excitation energy when an electron deep inside the Fermi surface is

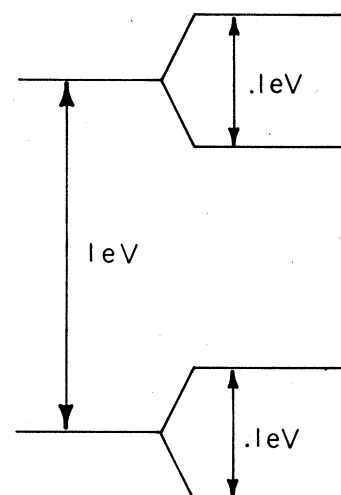


FIG. 11. Four level system for two different electrons in the conduction band with two different energy scales.

excited to well outside Fermi surface. This can be illustrated by a two level system with an energy level spacing of order 1 eV (Fig. 10). All energy scales below 1 eV exist also. For example, an electron reasonably close to the Fermi surface can be excited to a state somewhat above the Fermi surface with an excitation energy of only 0.1 eV. This electron can be excited independently of the first electron; the energy level structure now looks like Fig. 11. Similarly, by going closer and closer to the Fermi surface one finds electrons which can be excited with only 0.01 eV energy, or 0.001 eV energy, etc. The resulting energy level structure is shown in Fig. 12.

This structure is oversimplified: actually there are a continuum of excitation energies below 1 eV. However, order of magnitudes of energies are more important than their precise values. For qualitative purposes one can lump together all energy level spacings within a factor 2 or so of each other. The result of combining scales like this is that there are a large number of energy levels for each

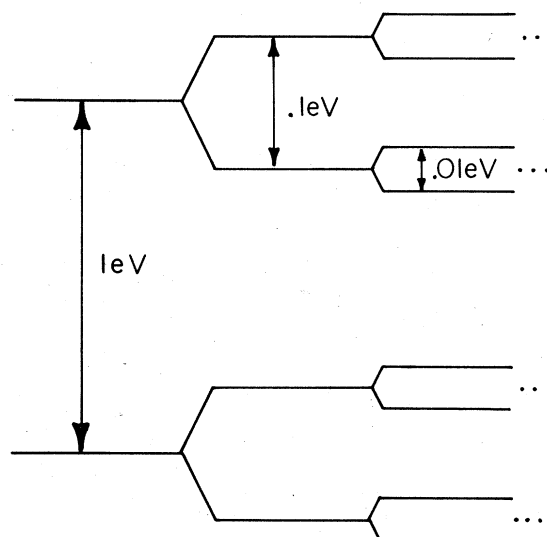


FIG. 12. Multiple energy scale structure of many-electron states of the conduction band.

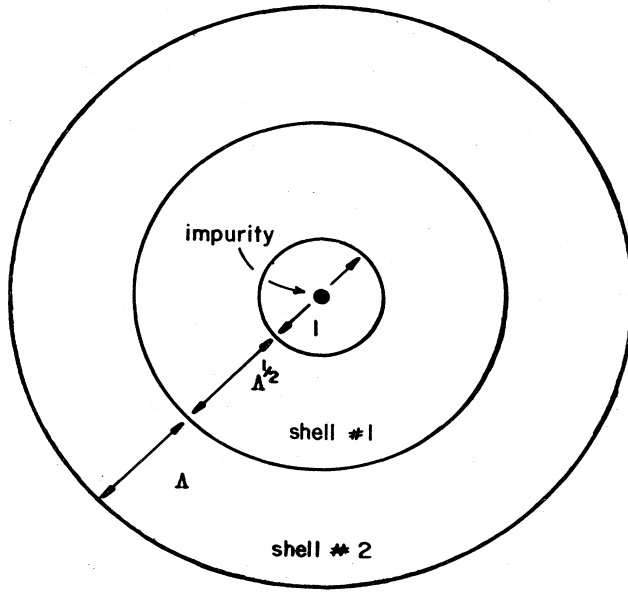


FIG. 13. Onion-like spherical shells giving the location of successive wave functions in the Kondo basis. The size of the smallest (inner) shell is a few Ångström units.

energy scale. Figure 12 with only two energy levels per scale is thus an oversimplification.

Here and throughout the Kondo calculations the emphasis is on properties of the conduction band rather than the impurity itself. One might think that the peculiar nature of the Kondo problem is due to the magnetic impurity, not the conduction band. The importance of the impurity is simple: it forces one to study the conduction band as a *many-electron* system. The cause of this is spin-flip scattering off the impurity, which is possible only when the impurity is magnetic (i.e., has a spin). Suppose two electrons, both with spin up, try to spin-flip scatter from a spin-down impurity. The first electron can spin-flip scatter, but the result is to leave the impurity with spin up. The second electron now *cannot* spin-flip scatter because this would violate spin conservation. Thus one cannot treat the electrons of the conduction band independently: one must treat the conduction band as a many electron system. Inevitably, as a many-electron system, the conduction band has the energy level structure indicated in Fig. 12, with each energy scale corresponding to a different set of electrons.

The normal description of individual electrons in the conduction band is in terms of plane wave or Bloch wave states. This description is poorly suited to the Kondo calculations. The trouble is that there are too many plane wave states with almost the same energy (due to the plane wave states being close to a continuum for a sample of macroscopic size). For the Kondo calculation it was necessary to define a new basis of states in the conduction band which emphasizes those states with the largest direct or indirect interaction with the impurity. The states chosen are closer to the localized Wannier states than the Bloch waves. The first state is (at least roughly) a Wannier state localized about the impurity, as localized as possible while still being in the conduction band. The remaining

states correspond (roughly) to spherical layers surrounding the impurity, as shown in Fig. 13. That is, the wave function for the second state of the Kondo basis is nonzero (except for small tails) only in the first spherical shell marked #1 in Fig. 13, with width $\Delta^{1/2}$. Here Δ is a parameter ≥ 1 ; it can be chosen arbitrarily. See below. For accidental reasons the width of the first layer is denoted $\Delta^{1/2}$ rather than Δ . The third state is predominantly contained in shell #2, etc. The shells increase in width; and correspondingly the momentum spread of succeeding states decreases. All wave functions in the Kondo basis are defined so that their average momentum is the Fermi momentum. As n increases the n th state is concentrated closer and closer to the Fermi surface; thus the energy scale for these states decreases being $\sim \Delta^{-n/2}$ for the n th state.

In this basis, electron states are neglected where the electron is far away from the impurity in position space *and* far away from the Fermi surface in momentum space. The motivation for this is as follows. The only reason for considering states far away from the impurity at all is that at very low temperatures only electrons very close to the Fermi surface are thermally excited. Being close to the Fermi surface in momentum space means the electrons have very broad wave functions in position space. At low temperatures the impurity interacts with these states near the Fermi surface; hence they must be included in the basis.

It is possible to add more states to the Kondo basis to make it complete. This is explained later in this Section. It turns out to be a good approximation (for static thermodynamic calculations) to neglect these extra states. See later in this Section for more discussion.

The conduction band is now approximated by an infinite set of discrete electron levels (called the "Kondo basis") arranged much like layers of an onion surrounding the impurity. The energy scale on the n th level is of order $\Delta^{-n/2}$.

The heart of the Kondo calculation is a solution of the Kondo Hamiltonian in the Kondo basis using numerical methods. This proceeds in steps. First one solves the impurity coupled to the first Kondo state (in the numbering of Sec. VIII this is the 0th step). The next step is to add the second layer of the onion, namely the terms involving the second Kondo state, and solve the combined coupling of the first and second conduction band states to the impurity. Then one adds the third state, then the fourth state, and so forth. This corresponds to solving for the eigenvalues at successively smaller and smaller energy scales in Fig. 12.

Beyond the first few steps this calculation cannot be done exactly: there are too many states. To be precise, there are 2^{2n+3} states in the n th step; when n is about 5 or higher this number is too large for an exact calculation. Therefore, an approximate method is used which generates only the lowest energy levels at each step; in practice this means calculating about the first 1000 energy levels. Note that for large n , say $n = 100$, the first 1000 energy levels are a negligible fraction of the total number of levels (2^{203} for $n = 100$).

The basic results of the Kondo calculation can be summarized in a geographical allegory. The sequence of Hamiltonians corresponding to adding successive layers of the onion to the impurity will be represented by a railroad track. The length of track from the beginning to the n th tie represents the Hamiltonian containing n conduction band single electron states (that is, the n th Hamiltonian contains n particle creation and destruction operators). There is a separate railroad track for each different strength of coupling to the impurity. The approximate numerical solution of this sequence of Hamiltonians is represented by a railroad car which travels down the track. Solving the n th Hamiltonian corresponds to having the railroad car at the n th tie on the track. The set of energy levels actually computed corresponds to the length of track covered by the railroad car; as the car moves down the track (i.e., as n increases) it covers a smaller and smaller fraction of the total track up to the n th tie.

The results of the calculation correspond to the arrangement of railroad tracks shown in Fig. 14. The tracks for two special cases: impurity coupling $\tilde{J} = 0$ and $\tilde{J} = -\infty$, are simple and featureless. In these cases the energy levels after n steps satisfy a simple scaling relation to the energy levels at $n + 2$ steps (except for very small n). This is explained in Sec. VIII. In renormalization group language, the iterations for $\tilde{J} = 0$ and $\tilde{J} = -\infty$ each lead to fixed

points. The interesting case is small negative \tilde{J} . For $\tilde{J} = -0.055$ (and $\Lambda = 2.25$) for example, the track stays close to the $\tilde{J} = 0$ track past the 30th tie; but near the 40th tie (40th iteration) the track moves away from the $\tilde{J} = 0$ track and moves over to the $\tilde{J} = -\infty$ track, which it approaches asymptotically for large tie number n . The crossover from the $\tilde{J} = 0$ track to the $\tilde{J} = -\infty$ track takes place for all negative values of \tilde{J} .

The full set of energy levels for the case $\tilde{J} = -0.055$ is very different from either the $\tilde{J} = 0$ or $\tilde{J} = -\infty$ energy levels. But the levels actually computed by computer stay close to the analogous $\tilde{J} = 0$ levels until near $n = 40$. Much beyond $n = 40$ the levels actually computed are similar to the analogous strong coupling levels. Large n ($n \gg 40$) corresponds to the railroad car being past the junction with the strong coupling track. Then the energy levels actually calculated for $\tilde{J} = -0.055$ (corresponding to the region of track covered by the railroad car) are indistinguishable from the corresponding levels of the strong coupling track. However, if one knew the entire set of energy levels for large n there would be significant differences between small \tilde{J} and large \tilde{J} at higher energies, corresponding to the large separation between the tracks at small n (large energy scales). For n near 40 (the crossover region) the computed levels for $\tilde{J} = -0.055$ are very different from either the $\tilde{J} = 0$ or $\tilde{J} = -\infty$ levels.

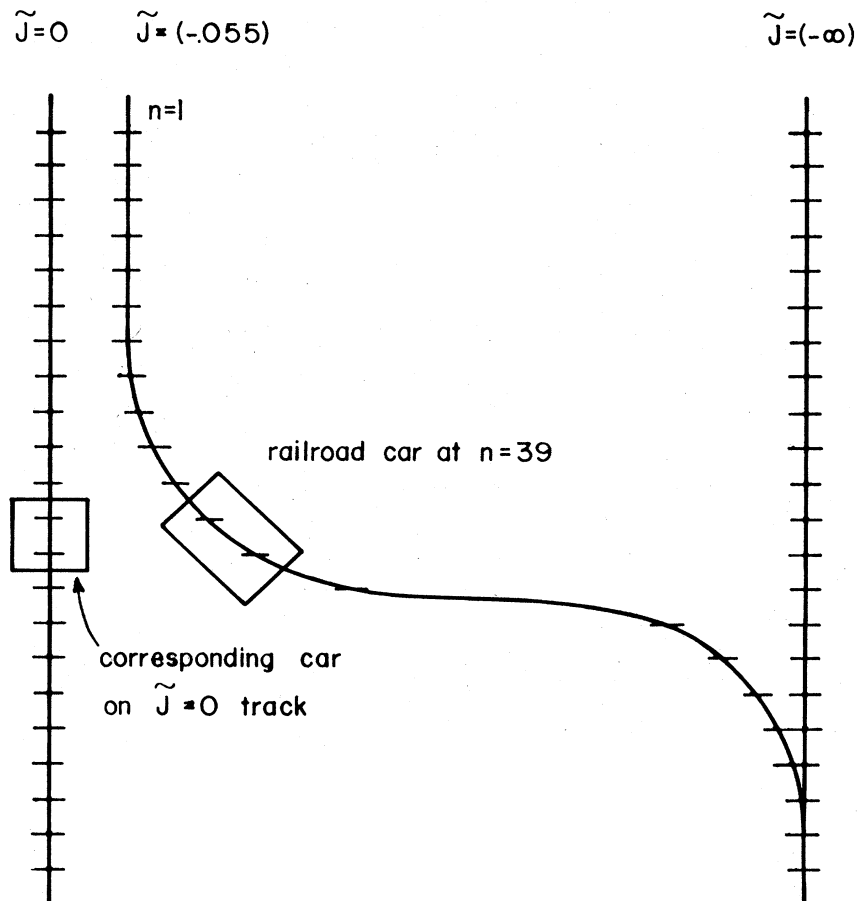


FIG. 14. Railroad track analogy for the Kondo calculation. Different tracks correspond to different initial values of \tilde{J} . A track from the top of the figure to the n th tie corresponds to the Kondo Hamiltonian with n electron states kept. The railroad cars illustrate the subset of energy levels actually kept in the numerical calculations.

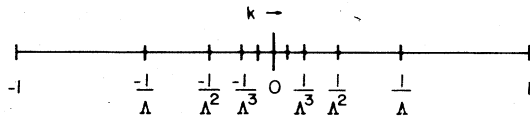


FIG. 15. Discretization on a logarithmic scale in k space.

The set of 1000 or so many-electron energy levels computed for given n is incomplete, both because the conduction band levels for higher n are omitted and because the levels above the first 1000 are omitted. Nevertheless, one can do some rough thermodynamic calculations using these levels. Thermodynamic calculations are based on the trace $\text{Tr} \exp -\beta H$, where $\beta = 1/kT$, and H is the Hamiltonian. If H were completely diagonalized, one would simply add up $\exp -\beta E_m$ for all eigenvalues E_m to compute the trace. Approximations of two types can be made in calculating the trace. First, energies E_m which are much larger than kT can be ignored since for these levels $\exp -\beta E_m$ is small. Secondly, terms in $H \ll kT$ can be neglected because they do not change E_m by enough to matter. In consequence one can use the levels after n iterations for approximate thermodynamic calculations, provided the value of n is chosen appropriately. Since the energy scales less than $\Lambda^{-n/2}$ are neglected in the n th iteration, $\beta \Lambda^{-n/2}$ must be $\ll 1$. In the actual calculations, the highest energies kept are of order $7 \times \Lambda^{-n/2}$; for higher energies, to be negligible, one must have $\beta \cdot 7 \cdot \Lambda^{-n/2} \gg 1$. This means the best value for β is around $(1/2.5)\Lambda^{n/2}$. Thus the numerical results for the n th iteration will determine the thermodynamics for temperatures kT or order $\Lambda^{-n/2}$.

From the map in Fig. 15 it is seen that for very low temperatures or zero temperature ($kT \sim \Lambda^{-n/2}$ with $n > 40$ if $\tilde{J} = -0.055$) the thermodynamics for weak coupling is very similar to the thermodynamics for strong coupling ($\tilde{J} = -\infty$). This is the principal qualitative result of the Kondo calculation. Precise numbers are obtained in Sec. IX.

The range of energies kept in the numerical calculations is meager (from $\Lambda^{-n/2}$ to about $7 \times \Lambda^{-n/2}$). This is despite the fact that over 1000 states are computed. The problem is that only a few electrons are needed to generate 1000 different levels so 1000 energy levels can only cover a small range of energies. This makes it hard to do accurate thermodynamic calculations using only the states computed numerically. Therefore, the numerical calculations have been supplemented by more complete analytic calculations whenever this is possible. Analytic calculations have been done where the track is near the $\tilde{J} = 0$ track, using perturbation expansions in \tilde{J} . The thermodynamics can be calculated directly from an expansion in \tilde{J} . Also individual energy levels can be computed and compared with numerical calculations. These analytic calculations are reported in Sec. VIII and IX and an Appendix. Analytic calculations were also performed for very large n where the small \tilde{J} track is close to the strong coupling track. These are reported also in Sec. VIII and IX. Once again one can calculate thermodynamic quantities or individual energy levels in these analytic calculations. In the "crossover region" where the track crosses over from the weak to the strong coupling tracks, no analytic cal-

culations are possible and one relies entirely on the numerical computations.

In the strong coupling theory the moment of the impurity disappears. The impurity moment combines with an electron spin from the conduction band to form a singlet state and there is no $1/T$ term in the susceptibility. For weak coupling the consequence of the crossover to strong coupling is that the impurity susceptibility increases as $1/T$ until T reaches roughly a temperature T_K corresponding to the crossover (i.e., kT_K is of order $\Lambda^{-n/2}$ where n is in the crossover region). Below T_K the impurity susceptibility shows strong coupling behavior, that is, it is a constant (of order $1/T_K$) instead of increasing further.

Why the crossover from weak to strong coupling takes place will not be explained. The author has no simple physical argument for it. It is the result of a complicated numerical calculation. There are enough checks on this calculation (reported in Sec. VIII and IX) so that it is very difficult to challenge the result. Nevertheless, it is not explained.

The renormalization group tools developed in Sec. I-V will be important for analyzing the behavior of the railroad tracks near the two extremes of weak and strong coupling. On the weak coupling side, the beginning of the crossover is governed by a single marginal eigenoperator of the kind discussed in Sec. V. It will be necessary to define a phenomenological coupling constant (called z_n) to parameterize the marginal variable, which varies with n . The variable z_n is proportional to \tilde{J} in lowest order. The dependence of z_n on n will be studied both analytically and numerically. On the strong coupling side the end of the crossover is governed by two separate "irrelevant" operators. The coupling constants of these operators will be obtained by matching to the numerical calculations. The two coupling constants will then be used in analytic calculations of the zero temperature behavior of the susceptibility and specific heats.

Finally, the susceptibility as a function of temperature near T_K was calculated directly from the numerical eigenvalues: see Sec. IX.

The Kondo Hamiltonian, in the form to be discussed in this paper, is

$$H_K = \int_{-1}^1 dk a_k^\dagger a_k - JA^\dagger \sigma A \cdot \tau, \quad (\text{VII.4})$$

where

$$A = \int_{-1}^1 a_k dk,$$

and

$$\{a_k, a_{k'}^\dagger\} = \delta(k - k').$$

The operator a_k is a conduction band electron destruction operator for an electron in an s -wave state about the origin of momentum k . The angular momentum indices are not indicated because there is no coupling between

s -wave electrons and higher partial waves (the kinetic energy is diagonal and by assumption higher partial waves will not couple to the impurity). However a_k does have a spin index μ : one should write $a_{k\mu}$ but in practice the μ will be omitted. The $a_k^\dagger a_k$ term is the kinetic energy of the conduction band. The energy k should have been written $\epsilon_k - \mu$, where μ is the chemical potential, and ϵ_k is the electron energy. However, this has been simplified, first by measuring the momentum relative to the Fermi momentum k_F for which $\epsilon_{k_F} = \mu$. That is, the true momentum is $k + k_F$ not k . Secondly, $\epsilon_k - \mu$ has been linearized about $k = 0$: $\epsilon_k - \mu \propto k$ for k near 0. Only the linear term has been kept. Thirdly, k is measured in units of the band edge momentum so the maximum of $|k|$ is 1; it is also assumed that the band is symmetric about the Fermi momentum. Finally the energy scale has been chosen to avoid a constant factor in front of k .

The operator A destroys an electron in the vicinity of the impurity (the impurity is assumed to lie at the origin). A also has a spin index μ ; for example, $A^{+\sigma_x} A$ means $\sum_{\mu\nu} A_{\mu}^{+\sigma_x \mu\nu} A_{\nu}$, where σ_x is a Pauli spin matrix. The matrices $\tau = (\tau_x, \tau_y, \tau_z)$ are the Pauli matrices for the impurity. J is the coupling strength of the impurity to the conduction band.

(There may be confusion between the s wave of “ s -wave conduction band” and the s wave of “electron in an s -wave state about the impurity.” An s -wave conduction band is built from s -wave states of each atom in the metal. The resulting conduction band states can have any orbital angular momentum about the impurity. We consider here only s states about the impurity. A d -wave conduction band is built of d -wave atomic orbitals: since there are 10 of these (counting spin) for each atom, a d -wave conduction band is described by operators $a_{k\mu}$, where μ has 10 values instead of two.)

When J is small, the susceptibility of the model (VII.4) can be calculated by perturbation theory in J ; the result to order J^2 is (assuming unit magnetic moment and unit g factors for electron and impurity)

$$\chi = \frac{1}{4}(1/kT)\{1 + 4J + 16J^2 \ln kT + cJ^2 + \dots\}. \quad (\text{VII.5})$$

(c is a constant). If one looks at only the J term it would appear that χ is proportional to T^{-1} for $T \rightarrow 0$. However, for $T \rightarrow 0$ the $\ln kT$ term spoils the argument; instead for $\ln kT \sim 1/J$ the perturbation expansion ceases to be valid because the second order term (and higher orders as well) are as large as the first order term. Thus the true zero temperature behavior cannot be determined from Eq. (VII.5).

In the renormalization group analysis one finds that the logarithm in order J^2 is symptomatic of the existence of a marginal operator. Correspondingly one can define a temperature dependent coupling which to a first approximation is

$$J(T) = 1/(1 - 4J \ln kT) \quad (\text{VII.6})$$

and the susceptibility $\chi(T)$ has a good expansion in terms of $J(T)$ (see Sec. V for the term “good”):

$$\chi(T) = \frac{1}{4}(1/kT)\{1 + 4J(T) + cJ^2(T) + \dots\}, \quad (\text{VII.7})$$

where this expansion is free of logarithms. This much has already been learned using the Gell-Mann–Low renormalization group and related diagram summation techniques.

For ferromagnetic coupling ($J > 0$) the function $J(T)$ goes to zero as $T \rightarrow 0$; in this case $\chi(T)$ behaves as T^{-1} for $T \rightarrow 0$ and nothing further needs to be said, at least for small J .

For antiferromagnetic coupling $J(T)$ becomes large for $\ln kT \sim 1/J$ and a nonperturbative calculation is needed to obtain the zero temperature limit. But if J is small, it is only at extremely small temperatures that perturbation theory fails. This is seen experimentally. The energy scale set by the band edge corresponds to 10^4 K, while specific heat measurements in Cu Cr show a maximum at about 1° K [Triplett and Phillips (1971)]. (The specific heat increases as $J(T)$ increases as long as perturbation theory is valid; a maximum is a nonperturbative effect.) More generally the temperature T_K at which $J(T)$ becomes of order 1 is called the Kondo temperature. This is a qualitative definition of T_K : various quantitative definitions exist (Kondo, 1969). Experimental values of T_K vary from much less than 1° K to more than 300° K.

The next topic is to set up an approximation to H_K which will be used in the numerical calculation. The Hamiltonian which will be investigated numerically has the form

$$H = \sum_{n=0}^{\infty} \Lambda^{-n/2} (f_n^\dagger f_{n+1} + f_{n+1}^\dagger f_n) - \tilde{J} f_0^\dagger \sigma f_0 \cdot \tau, \quad (\text{VII.8})$$

where f_n are a set of discrete electron destruction operators to be defined later; \tilde{J} is proportional to J , and Λ will be explained below. This Hamiltonian is called a “hopping” Hamiltonian (because the coupling of f_n to f_{n+1} is reminiscent of nearest-neighbor coupling models on a lattice which are used as models of conduction). There are no diagonal terms (i.e., $f_n^\dagger f_n$) in Eq. (VII.8) because the average energy of the state created by f_n will be the Fermi energy, and the Fermi energy has already been subtracted (by a chemical potential term) from the original Hamiltonian of Eq. (VII.4).

Three steps are needed to obtain the hopping Hamiltonian from H_K . The first step is a discretization of the operators a_k : The continuum of points k ($-1 < k < 1$) will be replaced by a discrete set. The discretization will be done on a logarithmic scale (Fig. 5): the discrete values of k are $1, 1/\Lambda, 1/\Lambda^2$, etc. and $-1, -1/\Lambda, -1/\Lambda^2$, etc., where Λ is an arbitrary parameter > 1 . Calculations have been done for $\Lambda = 2, 2.25, 2.5$, and 3 ; the continuum limit is the limit $\Lambda \rightarrow 1$. For $\Lambda \neq 1$ the Hamiltonian obtained from the discretization is only an approximation to H_K .

One can use the discretization to define a sequence of intervals: the m th interval is $\Lambda^{-m-1} < k < \Lambda^{-m}$ (there is an equivalent interval for negative k also). Inside each interval one can construct a complete set of wave functions ψ_{m_i} . The first wave function $\psi_{m_0}(k)$ we choose to be a constant; after normalization one obtains

$$\psi_{m_0}(k) = \Lambda^{m/2} (1 - \Lambda^{-1})^{-\frac{1}{2}} \quad (\Lambda^{-m-1} < k < \Lambda^{-m}); \quad (\text{VII.9})$$

outside the interval $\psi_{m0}(k)$ is defined to be zero. The wave functions $\psi_{ml}(k)$ can be chosen to be

$$\psi_{ml}(k) = \Lambda^{m/2}(1 - \Lambda^{-1})^{-1/2} \times \exp(i\omega_m k l) (\Lambda^{-m-1} < k < \Lambda^{-m}), \quad (\text{VII.10})$$

with

$$\omega_m = 2\pi\Lambda^m/(1 - \Lambda^{-1}); \quad (\text{VII.11})$$

again these wave functions vanish outside the interval.

The wave functions $\psi_{ml}(k)$ together with $\psi_{ml}(-k)$ form a complete, orthogonal, discrete set of wave functions for the interval $-1 < k < 1$, for any value of $\Lambda > 1$. One can therefore expand a_k in these wave functions:

$$a_k = \sum_m \sum_l \{a_{ml}\psi_{ml}(k) + b_{ml}\psi_{ml}(-k)\}. \quad (\text{VII.12})$$

The operators a_{ml} and b_{ml} define a complete set of independent, discrete, electron destruction operators satisfying the standard anticommutation rules

$$\{a_{ml}, a_{m'l'}\} = \delta_{mm'}\delta_{ll'} \text{ etc.} \quad (\text{VII.13})$$

We now approximate H_k by neglecting a_{ml} and b_{ml} for $l > 0$. The operators a_{m0} and b_{m0} will be denoted a_m and b_m for short; likewise ψ_n denotes ψ_{n0} . The resulting approximation to H_K is

$$H_K \simeq \frac{1}{2}(1 + \Lambda^{-1}) \sum_{m=0}^{\infty} \Lambda^{-m}(a_m^+ a_m - b_m^+ b_m) - JA^+ \sigma A \cdot \tau \quad (\text{VII.14})$$

with

$$A = (1 - \Lambda^{-1})^{1/2} \sum_{m=0}^{\infty} \Lambda^{-m/2}(a_m + b_m). \quad (\text{VII.15})$$

(The factor $\frac{1}{2}(1 + \Lambda^{-1})\Lambda^{-m}$ in H_K is the integral

$$\int_{-1}^1 k \psi_m^2(k) dk;$$

the factor $(1 - \Lambda^{-1})^{1/2}\Lambda^{-m/2}$ in A is the integral

$$\int_{-1}^1 \psi_m(k) dk.)$$

What has been neglected in this approximation? The formula for A is exact because the integrals

$$\int_{-1}^1 \psi_{ml}(k) dk$$

are 0 for $l \neq 0$. However, the exact conduction band energy includes terms involving a_{ml} and b_{ml} for $l \neq 0$. The crucial terms are those coupling a_{ml}^+ or b_{ml}^+ for $l \neq 0$ to a_m and b_m . The strengths of these terms are given by the integrals

$$\int_{-1}^1 k \psi_{ml}^*(k) \psi_m(k) dk.$$

If the factor k were replaced by a constant, these integrals would vanish, by orthogonality. Let k_m be the mean value

of k in the m th interval; these integrals are equal to

$$\int_{-1}^1 (k - k_m) \psi_{ml}^*(k) \psi_m(k) dk.$$

If Λ is close to 1 so that the intervals are small, then $k - k_m$ is small and the coupling of a_{ml} and b_{ml} to a_m and b_m is small. If this coupling can be neglected, then only the operators a_m and b_m couple to the impurity, and when the impurity susceptibility is calculated the remaining operators a_{ml} and b_{ml} do not contribute. For $\Lambda = 2$ to 3 it is not so obvious that the a_{ml} and b_{ml} can be neglected, but actual calculations (given later) have shown that the approximation is good to a few percent accuracy even for $\Lambda = 3$ and arguments will be given later to explain this result.

Now the motivation for the logarithmic discretization will be discussed. As background, it is useful to consider a much simpler quantum mechanical problem. Suppose one has a Hamiltonian $H = H_0 + H_1$, where H_0 is a Hamiltonian with energy level spacing of order 1, while H_1 is of order 0.01. Suppose that H_0 has a degenerate ground state, consisting for example of two states $|0\rangle$ and $|1\rangle$. Suppose also that H_0 is a nontrivial Hamiltonian which can be solved only approximately and that in practice one can calculate its energy levels only to about 5% accuracy. Finally, suppose that H_1 splits the degenerate ground states, and it is the resulting energy splitting that one wants to calculate. For example one might be computing the ground state fine structure splitting for a complex atom. This is a standard problem: one first diagonalizes H_0 as best one can, obtaining approximate wave functions for the states $|0\rangle$ and $|1\rangle$. The ground state splitting is now obtained by degenerate perturbation theory, for which one must calculate the matrix elements $\langle 0|H_1|0\rangle$, $\langle 1|H_1|0\rangle$, etc. The accuracy of the calculation depends on the accuracy with which these matrix elements can be calculated; if the wave functions are known to 5% and H_1 is of order 0.01, then these matrix elements will be known with an absolute error of about 0.0005, and this will be the error in the ground state splitting.

Suppose that one were idiotic enough to diagonalize $H = H_0 + H_1$ directly without first diagonalizing H_0 . This is no easier than diagonalizing H_0 , so one cannot expect to calculate the eigenvalues of H to better than 5% accuracy. But in this case one might miss altogether the ground state splitting, or else get a result but with 500% error!

What is the lesson of this example? It is that when one is interested in small energy level splittings for a complicated Hamiltonian H , it is essential to identify and treat separately, by the perturbation technique, the small term H_1 in H , which causes these splittings. This does not mean one can ignore the larger term H_0 for one has to diagonalize H_0 in order to set up a perturbation calculation.

A more complicated situation that can arise is that H consists of three parts $H = H_0 + H_1 + H_2$, where H_1 is 1% of H_0 in size, and H_2 is 1% of H_1 in size (H_2 is only

10^{-2} percent of H_0). This happens in hyperfine structure calculations. In this case the sensible procedure is first to diagonalize H_0 , second to treat H_1 as a perturbation to H_0 , and thirdly to treat H_2 as a perturbation to H_1 . Any departure from this procedure (e.g., treating $H_1 + H_2$ as a single perturbation) will lead to disaster unless one can do calculations to 0.05% accuracy instead of 5% accuracy.

A corollary to these results should also be noted. As long as one *does* use the perturbation treatment, the energy levels of H_0 need only be determined to rough accuracy even though this means an error much larger than the splittings one is finally interested in.

In the Kondo problem one has a situation which is similar to the examples cited above, only worse. The operators a_k^+ for $k \sim 1$ create electrons with energies of order 1, i.e., they generate states with energy level spacing of order 1. The operators a_k^+ with k of order ± 0.01 generate states with energy level spacing 0.01; but since the a_k^+ with $k \sim 0.01$ can act on a state with any number of electrons of energy 1, the effect of the a_k^+ with $k \sim 0.01$ is to split the energy levels produced by the a_k^+ with $k \sim 1$. Hence the $ka_k^+a_k$ terms with $k \sim 1$ are analogous to H_0 ; the $ka_k^+a_k$ terms with $k \sim 0.01$ are analogous to H_1 . Also the $ka_k^+a_k$ terms with $k \sim 10^{-4}$ are analogous to H_2 . Since k ranges from 1 to 0 there are even smaller terms with $k \sim 10^{-6}$, $k \sim 10^{-8}$, etc. According to the lesson of the example each of these terms should be isolated and treated as a perturbation relative to the previous term.

The difficulty with this approach is the presence of all energy scales in between $k \sim 1$ and $k \sim 0.01$, due to k being a continuous variable, or more generally the presence of all energy scales from 1 to 0. Because of this there is no obvious separation of H_K into $H_0 + H_1 + H_2 + \dots$ analogous to the separation of fine structure and hyperfine structure in atomic physics. Nevertheless, the terms with energy scales 1, 0.01, 10^{-4} , etc. are present in H_K , and the presence of the other energy scales as well *in no way invalidates* the lesson of the example. The logarithmic discretization is simply a way of setting up a separation of H_K into separate terms associated with separate energy scales. The ratio of energies of successive terms is Λ instead of 100, and there is no *a priori* best choice for Λ . In fact any choice of Λ in the actual range used (2 to 3) ensures that the crucial separations are made; the terms in H_K which differ by a factor of 100 or more in energy are now separated. This would not be true if one had used a linear discretization instead of a logarithmic one, for in a linear discretization one finite size interval contains the point $k = 0$, say $0 < k < \epsilon$. This one interval contains an infinite number of different energy scales.

Obviously one cannot treat successive terms in the discrete form of H_K by perturbation theory; this difficulty will be overcome by the approximate renormalization group calculation described later.

In the example one was interested in the energy level splittings due to H_1 and H_2 . Likewise in the Kondo problem one is interested in the splittings due to the small k electrons (i.e., electrons near the Fermi surface). These splittings are crucial for the low temperature thermodynamics.

The reason is that the low temperature thermodynamics involves expressions like

$$\text{Tr} S_z^2 \exp(-\beta H_K) / \text{Tr} \exp(-\beta H_K),$$

where β is $1/kT$, and S_z is the total spin. These ratios are determined primarily by the energy level splittings of order kT , i.e., the energy differences between the ground state and those excited states with energies of order kT above the ground state. Only energy level differences come in because the ground state energy itself cancels out in the ratio. Energies much greater than kT above the ground state are exponentially suppressed. So if kT is small, then small energy level splittings are important.

The analogy between H_K and the example may be imperfect, but it is used only to motivate the use of a logarithmic division of momentum space. The only true justification for using the logarithmic division is that a successful calculation results. For a more clear exposition of the above ideas see (Wilson, 1975).

One final comment. The author's original plan was to fix Λ , at $\Lambda = 2$, for example, rather than varying Λ to study the limit $\Lambda \rightarrow 1$. To achieve reasonable accuracy the idea was to take the operators a_{nl} and b_{nl} for $l \neq 0$ into account by treating their coupling to the a_n and b_n as a perturbation. This has proven to be unnecessary; at least for susceptibility and specific heat calculations. However, future calculations, for example the resistivity at finite temperatures, may require the use of such a perturbation approach. As will be discussed later, the alternative of considering values of Λ less than 2 is impractical.

The second step in obtaining the hopping form (VII.8) is a transformation from the operators a_m and b_m to a new orthogonal basis f_n . The operators are defined as follows. First one defines f_0 to be A itself apart from a normalization factor chosen so that $\{f_0, f_0^+\} = 1$. The result is $f_0 = 1/(A\sqrt{2})$. Having defined f_0 to be $(1/A\sqrt{2})$ there is no hope of obtaining a free electron energy diagonal in the f 's. As the next best thing one insists that the free energy contain only "nearest-neighbor" couplings:

$$H_K = \sum_{n=0}^{\infty} \epsilon_n (f_n^+ f_{n+1} + f_{n+1}^+ f_n) - 2J f_0^+ \sigma f_0 \cdot \tau, \quad (\text{VII.16})$$

where the ϵ_n will be determined below. Consider now the general structure of an orthogonal transformation. One can write

$$f_m = \sum_n \{u_{nm} a_m + v_{nm} b_m\}. \quad (\text{VII.17})$$

If this is a real orthogonal transformation, then the inverse transformation is

$$a_m = \sum_n u_{nm} f_n, \quad (\text{VII.18})$$

$$b_m = \sum_n v_{nm} f_n. \quad (\text{VII.19})$$

This follows from the orthonormality requirements on an orthogonal transformation: The complete set of such re-

quirements in the present case are

$$\sum_n u_{nm}u_{nm'} = \delta_{mm'}, \tag{VII.20}$$

$$\sum_n u_{nm}v_{nm'} = 0, \tag{VII.21}$$

$$\sum_n v_{nm}v_{nm'} = \delta_{mm'}, \tag{VII.22}$$

$$\sum_m (u_{nm}u_{n'm} + v_{nm}v_{n'm}) = \delta_{nn'}. \tag{VII.23}$$

In practice the operators f_1, f_2 , etc., will be defined in turn as orthogonal operators, so that the last of the orthogonality requirements (VII.23) will be satisfied by construction. The first three orthogonality requirements then are equivalent to the requirement that the whole set $\{f_n\}$ is complete and this will be proven in the Appendix to this lecture. We shall proceed assuming that the first three requirements will be satisfied.

The formula $f_0 = (1/\sqrt{2})A$ determines u_{0m} and v_{0m}

$$u_{0m} = v_{0m} = (1/\sqrt{2})(1 - \Lambda^{-1})^{\frac{1}{2}}\Lambda^{-m/2}. \tag{VII.24}$$

This means that one knows the coefficient of f_0 in the formulae (VII.18)–(VII.19) for a_m and b_m . Now according to the requirement (VII.16), the only term in H_K involving f_0 is $\epsilon_0 f_1^+ f_0$. However, the coefficient of f_0 in H_K can be determined explicitly by considering Eq. (VII.14) for H_K and substituting $u_{0m}f_0$ for a_m and $v_{0m}f_0$ for b_m . This gives

$$H_K = \frac{(1 + \Lambda^{-1})}{2} \sum_{m=0}^{\infty} \frac{1}{\sqrt{2}} (1 - \Lambda^{-1})^{\frac{1}{2}} \Lambda^{-3m/2} (a_m^+ - b_m^+) f_0$$

+ non- f_0 terms. (VII.25)

Thus one must have

$$\epsilon_0 f_1^+ = \frac{(1 + \Lambda^{-1})}{2} \sum_{m=0}^{\infty} \frac{1}{\sqrt{2}} (1 - \Lambda^{-1})^{\frac{1}{2}} \Lambda^{-3m/2} (a_m^+ - b_m^+). \tag{VII.26}$$

This formula plus the normalization requirement $\{f_1, f_1^+\} = 1$ determines both f_1^+ and ϵ_0

$$f_1^+ = \frac{1}{\sqrt{2}} (1 - \Lambda^{-3})^{\frac{1}{2}} \sum_{m=0}^{\infty} \Lambda^{-3m/2} (a_m^+ - b_m^+), \tag{VII.27}$$

$$\epsilon_0 = \frac{(1 - \Lambda^{-1})^{\frac{1}{2}} (1 + \Lambda^{-1})}{(1 - \Lambda^{-3})^{\frac{1}{2}} 2}. \tag{VII.28}$$

Note that f_1 is orthogonal to f_0 because f_1 involves differences $a_m - b_m$ while f_0 involves sums $a_m + b_m$.

Now one knows that

$$u_{1m} = -v_{1m} = (1/\sqrt{2})(1 - \Lambda^{-3})^{\frac{1}{2}}. \tag{VII.29}$$

This means one knows the coefficient of f_1 in the formulae for a_m and b_m . This means one can determine explicitly the coefficient of f_1 in H_K ; one obtains:

$$H_K = \sum_{m=0}^{\infty} \frac{\Lambda^{-5m/2} (1 - \Lambda^{-3})^{\frac{1}{2}}}{\sqrt{2}} (a_m^+ + b_m^+) f_1 \frac{(1 + \Lambda^{-1})}{2}$$

+ non- f_1 terms. (VII.30)

Consider the coefficient of f_1 . One knows that this coefficient can always be written in the form $\alpha f_0^+ + f_2^+$ where f_2 is an operator orthogonal to f_0 , and α and β are parameters to be determined. Furthermore, the coefficient only involves sums $a_m^+ + b_m^+$ so it, and therefore f_2 , will both be orthogonal to f_1 .

Explicit calculation shows that

$$\alpha = (1 - \Lambda^{-1})^{\frac{1}{2}} (1 - \Lambda^{-3})^{-\frac{1}{2}} \frac{(1 + \Lambda^{-1})}{2}, \tag{VII.31}$$

$$\beta = \frac{\Lambda^{-\frac{1}{2}} (1 - \Lambda^{-2})}{(1 - \Lambda^{-3})^{\frac{1}{2}} (1 - \Lambda^{-5})^{\frac{1}{2}}} \frac{(1 + \Lambda^{-1})}{2}, \tag{VII.32}$$

$$f_2 = \Lambda^{\frac{1}{2}} \frac{(1 - \Lambda^{-5})^{\frac{1}{2}}}{\sqrt{2} (1 - \Lambda^{-1})} \sum_{m=0}^{\infty} \{ (1 - \Lambda^{-3}) \Lambda^{-5m/2} - (1 - \Lambda^{-1}) \Lambda^{-m/2} \} (a_m + b_m), \tag{VII.33}$$

where f_2 is also properly normalized. Now the coefficient of f_1 in H_K is $\alpha f_0^+ f_1 + \beta f_2^+ f_1$. Clearly one now defines $\epsilon_1 = \beta$. Since H_K is Hermitian, the $f_0^+ f_1$ term must have the same coefficient as the $f_1^+ f_0$ term calculated earlier, i.e., α must be ϵ_0 , which it is.

Continuing with the same type of calculation, one can construct all the f_n and all the constants ϵ_n . One technical observation: the Hermiticity of H_K ensures that the coefficient of f_n in H_K is not a totally arbitrary sum of $f_{n+1}^+, f_{n-1}^+, f_{n-3}^+$, etc.; the Hermiticity requires that the coefficient of f_n be simply $\epsilon_{n-1} f_{n-1}^+ + \epsilon_{n+1} f_{n+1}^+$, where ϵ_{n+1} is to be determined, but ϵ_{n-1} and f_{n-1} are known from prior calculation. So one takes the coefficient of f_n , subtracts $\epsilon_{n-1} f_{n-1}^+$, and one has $\epsilon_{n+1} f_{n+1}^+$ which is easily separated into f_{n+1}^+ and ϵ_{n+1} .

It turns out that one can obtain analytic expressions for ϵ_n, u_{nm} , and v_{nm} for all n and m . The u_{nm} and v_{nm} are obtained in the form of a generating functional

$$U_n(z) = \sum_{m=0}^{\infty} u_{nm} z^m. \tag{VII.34}$$

Then the analytic expressions are

$$\epsilon_n = \Lambda^{-n/2} [1 - \Lambda^{-(n+1)}] [1 - \Lambda^{-2n+1}]^{-\frac{1}{2}} \times [1 - \Lambda^{-(2n+3)}]^{-\frac{1}{2}} [1 + \Lambda^{-1}]/2, \tag{VII.35}$$

$$v_{nm} = (-1)^n u_{nm}, \tag{VII.36}$$

$$U_n(z) = \left(\frac{1 - \Lambda^{-(2n+1)}}{2\Lambda^{n(n-1)/2}} \right)^{\frac{1}{2}} \left(\frac{1}{1 - \Lambda^{-(n+1/2)}z} \right) \times \prod_{r=0}^{r_M} \frac{(1 - \Lambda^{a+2r}z)}{(1 - \Lambda^{-a+2r}z)}, \tag{VII.37}$$

where

$$a = \frac{1}{2} (n \text{ even}) \text{ or } \frac{3}{2} (n \text{ odd}), \tag{VII.38}$$

$$r_M = \frac{n - \frac{3}{2} - a}{2}, \tag{VII.39}$$

and the product from $r = 0$ to r_M is omitted for $n < 2$. For proof of these formulae see the Appendix to this section.

One now has a Hamiltonian H_K with a set of constants ϵ_n . For large n , one has

$$\epsilon_n = \frac{(1 + \Lambda^{-1})}{2} \Lambda^{-n/2}. \tag{VII.40}$$

The third step in deriving the hopping Hamiltonian H of Eq. (VII.8) is rather trivial: one first redefines the ϵ_n to be $[(1 + \Lambda^{-1})/2\Lambda^{-n/2}]$ for all n , and secondly one divides the total Hamiltonian by $(1 + \Lambda^{-1})/2\Lambda^{-n/2}$. One finally gets Eq. (VII.8) with

$$\tilde{J} = 4J(1 + \Lambda^{-1})^{-1}. \tag{VII.41}$$

The redefinition of ϵ_n only changes the ϵ_n for small n ; it will become clear later that the values of the ϵ_n for small n act like irrelevant variables in the renormalization group sense and do not affect the low temperature calculations. The rescaling of H_K is equivalent in thermodynamic calculations to a renormalization of temperature scale and this is easily corrected for in the final answers.

APPENDIX TO VII

In this Appendix the generating functional $U_n(z)$ of Eq. (VII.37), Eq. (VII.35) for ϵ_n and Eq. (VII.36) are derived.

Continuing the analysis of Sec. VII, one determines explicitly the coefficient of f_n in H_K by substituting $u_{nm}f_n$ for a_m and $v_{nm}f_n$ for b_m in Eq. (VII.14). The result is

$$H_K = \frac{(1 + \Lambda^{-1})}{2} \sum_{m=0}^{\infty} \Lambda^{-m}(u_{nm}a_m^+ - v_{nm}b_m^+)f_n + \text{non-}f_n \text{ terms} + \tilde{J} \text{ term}. \tag{VII.42}$$

The coefficient of f_n must be equal to $\epsilon_{n-1}f_{n-1}^+ + \epsilon_n f_{n+1}^+$; using Eq. (VII.17), this means that, for all m ,

$$\epsilon_{n-1}u_{n-1,m} + \epsilon_n u_{n+1,m} = \Lambda^{-m}[(1 + \Lambda^{-1})/2]u_{nm}, \tag{VII.43}$$

$$\epsilon_{n-1}v_{n-1,m} + \epsilon_n v_{n+1,m} = -\Lambda^{-m}[(1 + \Lambda^{-1})/2]v_{nm}. \tag{VII.44}$$

One can set $v_{nm} = (-1)^n u_{nm}$; then the second equation reduces to the first. The first equation can be multiplied by z^m and summed over m , giving

$$\epsilon_{n-1}U_{n-1}(z) + \epsilon_n U_{n+1}(z) = [(1 + \Lambda^{-1})/2]U_n(z/\Lambda). \tag{VII.45}$$

If $n = 0$, $\epsilon_{n-1}U_{n-1}(z)$ is replaced by 0. Some tedious algebra shows that the formulae (VII.35) and (VII.37) give a solution to Eq. (VII.45) for all z .

Now the orthonormality conditions (VII.20)–(VII.23) must be verified. Using Eq. (VII.36) one finds that these

conditions are equivalent to the following:

$$\sum_{n \text{ even}} u_{nm}u_{nm'} = \frac{1}{2}\delta_{mm'}, \tag{VII.46}$$

$$\sum_{n \text{ odd}} u_{nm}u_{nm'} = \frac{1}{2}\delta_{mm'}, \tag{VII.47}$$

$$\sum_m u_{nm}u_{n'm} = \frac{1}{2}\delta_{nn'} \text{ (if } n-n' \text{ is even)}. \tag{VII.48}$$

The last condition is equivalent to the requirement

$$\frac{1}{2\pi i} \oint \frac{dz}{z} U_n(z)U_{n'}(z^{-1}) = \frac{1}{2}\delta_{nn'} \tag{VII.49}$$

for even $n-n'$; the contour integral runs around the unit circle in the complex z plane. (It is easily verified that the power series in z for $U_n(z)$ converges for z on the unit circle.) It is sufficient to consider the case $n \geq n'$. For $n > n'$ one finds that the product $z^{-1}U_n(z)U_{n'}(z^{-1})$ is analytic inside the unit circle so the integral vanishes. For $n' = n$ the same product has one pole inside the unit circle, at $z = \Lambda^{-(n+1/2)}$. One obtains (VII.49) by computing the residue at this pole.

The relation (VII.46) is equivalent to

$$\sum_{n \text{ even}} U_n(z)U_n(z') = \frac{1}{2} \frac{1}{1 - zz'}. \tag{VII.50}$$

The sum over n can be written

$$\sum_{n \text{ even}} U_n(z)U_n(z') = R_0(z,z') + R_0(z,z')R_2(z,z') + R_0(z,z')R_2(z,z')R_4(z,z') + \dots \tag{VII.51}$$

where

$$R_0(z,z') = (1 - \Lambda^{-1})/[2(1 - \Lambda^{-1/2}z)(1 - \Lambda^{-1/2}z')], \tag{VII.52}$$

and for $n \geq 2$

$$R_n(z,z') = \frac{[1 - \Lambda^{(n-1/2)}z][1 - \Lambda^{(n-1/2)}z']}{[1 - \Lambda^{-(n+1/2)}z][1 - \Lambda^{-(n+1/2)}z']} \times \frac{[1 - \Lambda^{1-2n}]\Lambda^{3-2n}}{[1 - \Lambda^{3-2n}]} \tag{VII.53}$$

Write

$$\Sigma_n(z,z') = 1 + R_{n+2}(z,z') + R_{n+2}(z,z')R_{n+4}(z,z') + \dots \tag{VII.54}$$

Then

$$\Sigma_{n-2}(z,z') = 1 + R_n(z,z') \Sigma_n(z,z'). \tag{VII.55}$$

This recursion formula has an analytic solution, namely

$$\Sigma_n(z,z') = [1/(1 - zz')][1 - \Lambda^{-(n+1/2)}z] \times [1 - \Lambda^{(n+1/2)}z][1 - \Lambda^{-(2n+1)}]^{-1}. \tag{VII.56}$$

One verifies by substitution that this solves Eq. (VII.55). To show that the explicit formula (VII.56) is equal to the sum (VII.54) we must study the convergence of the sum.

For n large (remember that $\Lambda > 1$), one has

$$R_n(z, z') \simeq zz'. \quad (\text{VII.57})$$

The product $R_0 R_2 \cdots R_n$, therefore, behaves roughly as $(zz')^{n/2}$; thus the sum converges if $|zz'| < 1$. Using the recursion formula (VII.55), the analytic expression (VII.56) satisfies

$$\Sigma_0(z, z') = 1 + R_2 + R_2 R_4 + \cdots R_2 R_4 R_6 \cdots R_m \Sigma_m \quad (\text{VII.58})$$

for any even m . For $m \rightarrow \infty$, the last term behaves as $(zz')^{m/2}$ and therefore vanishes. Thus for $m \rightarrow \infty$ one obtains Eq. (VII.54) for $n = 0$. Eqs. (VII.51), (VII.52), (VII.54), and (VII.56) now give Eq. (VII.50).

There is a similar proof of the odd n equation (VII.47). This completes the proof of the orthonormality relations.

VIII. RENORMALIZATION GROUP TRANSFORMATION FOR THE KONDO HAMILTONIAN

In this section the solution of the hopping Hamiltonian of Eq. VII.8 by renormalization group methods will be discussed. The actual numbers computed in this section have rather little relation to the physics of the Kondo Hamiltonian; the physics will not be extracted until the following section where the susceptibility and specific heat will be computed. There is nothing surprising in this. In this section the aim will be to set up a renormalization group transformation and calculate the resulting effective Hamiltonians. These effective Hamiltonians depend on how the transformation is defined, just as in the case of critical phenomena, so one expects to have to do a further calculation in order to get some physics from the effective Hamiltonians. For an elementary discussion of the method of this lecture see Wilson (1975).

It is convenient to define a set of Hamiltonians H_N as follows:

$$H_N = \Lambda^{(N-1)/2} \left\{ \sum_{n=0}^{N-1} \Lambda^{-n/2} (f_n^+ f_{n+1} + f_{n+1}^+ f_n) - \tilde{J} f_0^+ \sigma f_0 \cdot \tau \right\}. \quad (\text{VIII.1})$$

Then the original Hamiltonian H of (VII.8) is

$$H = \lim_{N \rightarrow \infty} \Lambda^{-(N-1)/2} H_N. \quad (\text{VIII.2})$$

The purpose of the factor $\Lambda^{(N-1)/2}$ is so that the smallest term in H_N is of order 1, the smallest term being $f_{N-1}^+ f_N + f_N^+ f_{N-1}$. The H_N satisfy a recursion formula:

$$H_{N+1} = \Lambda^3 H_N + f_{N+1}^+ f_N + f_N^+ f_{N+1} \quad (\text{VIII.3})$$

which will be used to define the renormalization group transformation.

In the example of atomic hyperfine structure discussed in the previous lecture, one had a Hamilton $H = H_0 + H_1 + H_2$ (with no relation to the H_N defined just above). The approved strategy was first to diagonalize the largest term H_0 , second to set up the Hamiltonian $H_0 + H_1$ as a matrix in terms of the eigenstates of H_0

and then diagonalize $H_0 + H_1$, and finally to set up the full Hamiltonian H in terms of the eigenstates of $H_0 + H_1$ and then diagonalize H . The purpose of defining the Hamiltonians H_N is so that one can apply the same strategy to the hopping Hamiltonian H . H_0 contains the largest terms in H (for this purpose the J term will be lumped together with the $f_0^+ f_1$ and $f_1^+ f_0$ terms). The subsequent Hamiltonians H_1, H_2 , etc., are obtained by bringing in the successively smaller terms $\Lambda^{-1/2}(f_1^+ f_2 + f_2^+ f_1)$, $\Lambda^{-1}(f_2^+ f_3 + f_3^+ f_2)$, etc., from the original Hamiltonian H , until for $N \rightarrow \infty$ one has recovered the full H [except for the scale factor $\Lambda^{(N-1)/2}$]. Thus the strategy analogous to that used on the hyperfine structure consists of diagonalizing each H_N , in turn, and then calculating the matrix elements of H_{N+1} in the representation in which H_N is diagonal. One then diagonalizes H_{N+1} .

Consider now the practical aspects of diagonalizing H_N . The Hamilton H_N is a $2^{2N+3} \times 2^{2N+3}$ matrix. The number of states is 2^{2N+3} because H_N involves $2N + 2$ independent electron operators $f_{n\mu}$ ($0 \leq n \leq N, \mu = \pm \frac{1}{2}$) plus τ : each single electron state can be occupied or empty so the total number of many electron states is 2^{2N+2} . The impurity has two states so the total number of states is 2^{2N+3} . When N is large (in the calculations reported later N will be as large as 180) the total number of states is unmanageable: one can consider in practice only a small subset of the eigenstates of H_N .

There are no analytic methods known for diagonalizing H_N except for the special cases $\tilde{J} = 0$ and $\tilde{J} = \infty$ (which will be discussed later). So one has to diagonalize H_N numerically. This means one has to truncate the matrix for H_N from its original size $2^{2N+3} \times 2^{2N+3}$ down to a more manageable size (in practice manageable means roughly 1000×1000 : see later).

What energy levels of H_N is one interested in? The ultimate aim of the calculation is to determine the thermodynamics at very small temperatures T ; suppose to be specific that one is considering $kT \sim \Lambda^{-70}$. Then one is interested in the energy levels of H which are of order Λ^{-70} above the ground state. One has to consider at least the first 140 terms in H in order to include the terms which produce an energy level spacing Λ^{-70} . Thus when one is diagonalizing H_N for $N \ll 140$, it is only the ground state that is directly relevant physically: the excited states are too high in energy to be important thermodynamically. However, in order to calculate the ground state of H_{N+1} one will have to know the excited states of H_N as well (see below). For $N \simeq 140$ one needs also the first few excited states of H_N for the thermodynamics, which is manageable. However for $N \gg 140$ one is including terms in H with much smaller energies than kT , and now many excited states of H_N are important, too many to be calculated numerically.

In this lecture we consider only the problem of calculating the lowest 1000 or so eigenstates of H_N , as is appropriate for $\Lambda^{-N/2} \gtrsim kT$. The problem of calculating many excited states for $\Lambda^{-N/2} \ll kT$ will be handled analytically (in the limit of small T); this will be explained in the next lecture.

The strategy proposed earlier is an iteration procedure: the N th iteration consists in expressing H_{N+1} as a matrix in terms of eigenstates of H_N , and then diagonalizing this matrix to determine the eigenvalues and eigenvectors of H_{N+1} . The truncation procedure used to bring this matrix down to manageable size will be to consider only the first 1000 or so eigenstates of H_N in the calculation instead of all 2^{2N+3} eigenstates of H_N . (This is a truncation only for $N > 3$; for $N \leq 3$ H_N does not have 1000 eigenstates.)

Numerical tests of the accuracy of this truncation procedure show that the low excited state energy levels of H_N are determined to an accuracy of about 1% for $\Lambda = 2$, and a fraction of a percent for $\Lambda = 2.5$ and 3; this even includes accumulated errors from many iterations (~ 50 to 80). Detailed error estimates are given in the next lecture. To understand this accuracy one has to look more closely at the matrix defining H_{N+1} . It is known from the numerical calculations that the first 1000 energy levels of H_N include energies up to about 6 to 10 (these are actual numbers) above the ground state, while the first excited state energy is usually < 1 above the ground state. (The factor $\Lambda^{(N-1)/2}$ was included in H_N so that these numbers would be free of powers of Λ .) The energy levels of H_N appear as diagonal elements of the matrix for H_{N+1} ; the energy levels of H_N are also multiplied by a factor $\Lambda^{1/2}$ as in Eq. (VIII.3). Thus the diagonal elements of the truncated matrix H_{N+1} range from < 1 to about 8–17 (depending on the size of $\Lambda^{1/2}$; the range $2 < \Lambda < 3$ of the numerical calculations is being considered here). The off diagonal matrix elements of H_{N+1} come from the term $f_N^+ f_{N+1} + f_{N+1}^+ f_N$. The operators f_N already appear in H_N , and they have a rather complicated set of matrix elements between eigenstates of H_N . (These matrix elements are computed from a knowledge of the eigenstates of H_N .) However there is a sum rule which ensures that no matrix element of f_N is greater than 1. Let $|k\rangle$ be a complete set of states: then the anticommutation relations give

$$\sum_{k'} \langle k' | f_{N\mu} | k' \rangle \langle k' | f_{N\mu}^+ | k \rangle + \langle k | f_{N\mu}^+ | k' \rangle \langle k' | f_{N\mu} | k \rangle = 1. \quad (\text{VIII.4})$$

This is

$$\sum_{k'} |\langle k | f_{N\mu} | k' \rangle|^2 + |\langle k' | f_{N\mu} | k \rangle|^2 = 1 \quad (\text{VIII.5})$$

so no individual term can be greater than 1. In addition, the matrix elements of f_N tend to be largest for nearby eigenstates of H_N . This means in particular that the matrix elements of f_N between the low excited states of H_N and the more highly excited states being neglected are $\ll 1$. This has been seen in the numerical calculations; the reason for this will be explained shortly.

The operators f_{N+1} and f_{N+1}^+ do not appear in H_N . This allows one to work in a simple representation of those operators. The first effect of these operators is that each eigenstate $|k\rangle$ of H_N becomes four states when f_{N+1}^+ and f_{N+1} are taken into account; these four states can be denoted

$$|\Omega; k\rangle = |k\rangle, \quad (\text{VIII.6})$$

$$|\frac{1}{2}; k\rangle = f_{N+1, \frac{1}{2}}^+ |k\rangle, \quad (\text{VIII.7})$$

$$|-\frac{1}{2}; k\rangle = f_{N+1, -\frac{1}{2}}^+ |k\rangle, \quad (\text{VIII.8})$$

$$|\frac{1}{2}, -\frac{1}{2}; k\rangle = f_{N+1, \frac{1}{2}}^+ f_{N+1, -\frac{1}{2}}^+ |k\rangle. \quad (\text{VIII.9})$$

These states form an orthonormal basis for H_{N+1} , since the electron states created by $f_{N+1, \mu}^+$ are orthogonal to any of the electron states involved in the states $|k\rangle$. It is straightforward to compute matrix elements of $f_{N+1, \mu}$ and $f_{N+1, \mu}^+$ in this basis using the anticommutation relations of $f_{N+1, \mu}$ and $f_{N+1, \mu}^+$ plus the fact that $f_{N+1, \mu} |k\rangle = 0$ and the conjugate result $\langle k | f_{N+1, \mu}^+ = 0$. For example,

$$\begin{aligned} & \langle k' | \Omega | f_{N+1, \mu} | \frac{1}{2}; k \rangle \\ &= \langle k' | f_{N+1, \mu} f_{N+1, \frac{1}{2}}^+ | k \rangle \\ &= \delta_{\mu, \frac{1}{2}} \langle k' | k \rangle - \langle k' | f_{N+1, \frac{1}{2}}^+ f_{N+1, \mu} | k \rangle \\ &= \delta_{\mu, 1/2} \delta_{k', k} \end{aligned} \quad (\text{VIII.10})$$

It is clear from this example that f_{N+1} and f_{N+1}^+ only have matrix elements between states with the same k value, i.e., the same eigenvalue for H_N . Thus the products $f_{N+1}^+ f_N$ and $f_N^+ f_{N+1}$ primarily connect nearby energy levels of H_N (since f_N only connects nearby energy levels). So if one orders the states $|k\rangle$ in order of increasing energy the matrix for H_{N+1} looks like Fig. 16; the diagonal elements starting with $\Lambda^{1/2} E_0 = 0$ (the ground state energy of H_N is subtracted from H_{N+1} as a practical convenience) repeated four times along the diagonal, then $\Lambda^{1/2} E_1$ four times (E_1 is the first excited state energy of H_N) etc. The important matrix elements of $f_{N+1}^+ f_N$ and $f_N^+ f_{N+1}$ lie near the principal diagonal: far away from the principal diagonal the matrix elements of H_{N+1} are very small.

There are now two effects that make the truncation down to a thousand states a good approximation. The first is that the states neglected have diagonal elements > 8 to 17 while their couplings to the states kept are less than 1. This means the importance of these states is reduced by a factor or 1/10 or more due to energy denominators in the perturbation theoretic sense. But even more important is the fact that the states neglected couple mainly to the highest excited states kept; the direct coupling of the states neglected to the lowest few states of H_N is extremely small. The result of this is that the first few states of H_{N+1} are built mainly from the first few states of H_N , and even the higher states kept contribute little; the states neglected are negligible.

In the numerical calculations one diagonalizes numerically the truncated matrix for H_{N+1} . The diagonalization procedure produces both eigenvalues and eigenvectors, the eigenvectors being linear combinations of the states $|\Omega; k\rangle$, $|\frac{1}{2}; k\rangle$, etc. Given the eigenstates in this form, it is straightforward to calculate the matrix elements of f_{N+1} and f_{N+1}^+ between these eigenstates. This means one has all the necessary information to set up the matrix for H_{N+2} . Note that if one keeps 1000 states of H_N , one is actually truncating the matrix for H_{N+1} down to 4000×4000 , so one can calculate 4000 eigenstates of H_{N+1} . But one will keep only the lowest 1000 of these, if one is keeping 1000 states for each N .

One consequence of the form (Fig. 16) for H_{N+1} is that the matrix elements of f_{N+1} and f_{N+1}^+ are small for well-separated eigenstates of H_{N+1} . The reason is that well-separated eigenstates of H_{N+1} will involve mainly non-overlapping states of H_N . Since f_{N+1} has no matrix elements between different eigenstates of H_N , this means the matrix elements of f_{N+1} and f_{N+1}^+ are small, as claimed.

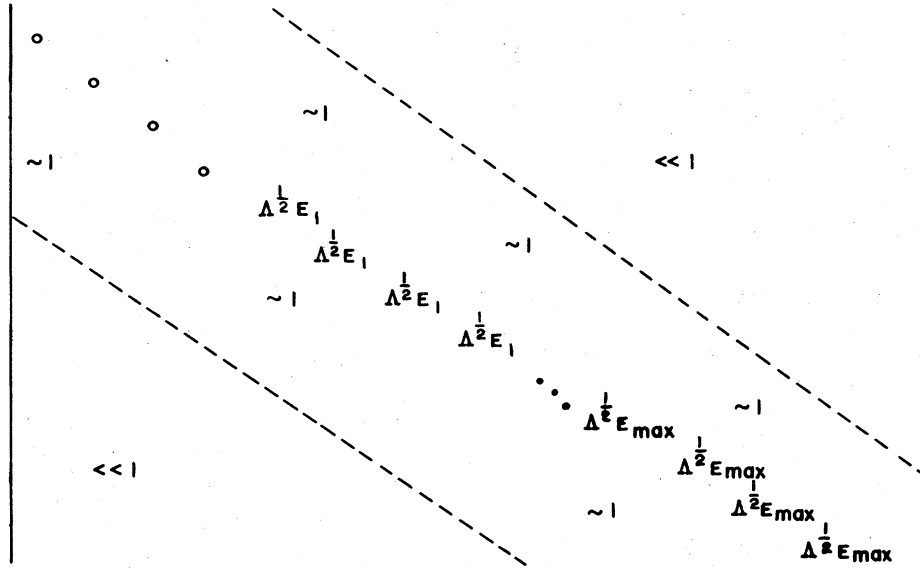


FIG. 16. Form of matrix for H_{N+1} .

This means that in the next iteration the input matrix elements are small except near the diagonal, as was stated.

There is one further comment about the actual calculations. The total spin \mathbf{S} of the electrons plus impurity commutes with H_N , as well as the total charge Q of the electrons. Hence in diagonalizing H_{N+1} one can consider independently each subspace of given S^2 , S_z , and Q . This reduces the size of the matrices being diagonalized from 4000×4000 to a maximum of about 100×100 , enormously reducing the computer time required to complete the diagonalizations. Computations were performed keeping either 526 or 1620 states and compared to test the accuracy of the truncation; a single iteration with 1620 states kept required less than 7 seconds on the CDC 7600.

The recursion formula VIII.3 defines the transformation T , except for a ground state energy subtraction. To be precise the transformation is

$$H_{N+1} = T[H_N] = \Lambda^{1/2} H_N + f_{N+1}^+ f_N + f_N^+ f_{N+1} - E_{G,N+1}, \tag{VIII.11}$$

where $E_{G,N+1}$ is chosen so that the ground state energy of H_{N+1} is 0. In the calculations, the input to the transformation T consists of the eigenvalues of H_N and the matrix elements of f_N between these eigenstates; the output of the transformation consists of the eigenvalues of H_{N+1} and the matrix elements of f_{N+1} .

The renormalization group transformation has two fixed points, both of which are important for the interpretation of the numerical results. The fixed points are both rather trivial and can be obtained by the study of two limiting cases: $\tilde{J} = 0$ and $\tilde{J} = -\infty$. This study will be described before presenting numerical results. For $\tilde{J} = 0$, H_N is a free electron Hamiltonian. That is, H_N is a quadratic form in f_n and f_n^+ , which can be diagonalized by a suitably chosen set of single electron operators. Unfortunately it is not sufficient to reverse the transformation used initially to define the f_n 's, because that transformation was defined

for an infinite set of f_n 's while H_N contains only a finite subset of f_n 's.

The quadratic form can be written in matrix form. Let f be the vector (f_0, f_1, \dots, f_N) . Then $H_N (J = 0) = f^+ \mathcal{H}_N f$ where \mathcal{H}_N is an $(N + 1) \times (N + 1)$ matrix with elements

$$\mathcal{H}_{N,n,n+1} = \Lambda^{(N-1-n)/2} = \mathcal{H}_{N,n+1,n} \tag{VIII.12}$$

$(0 \leq n \leq N - 1)$; all other elements are 0. The matrix \mathcal{H}_N is a real symmetric matrix. Therefore there exists a real orthogonal matrix M which diagonalizes \mathcal{H}_N

$$\tilde{M} \mathcal{H}_N M = \eta_N, \tag{VIII.13}$$

with

$$\tilde{M} M = M \tilde{M} = 1, \tag{VIII.14}$$

so that

$$\mathcal{H}_N M = M \eta_N, \tag{VIII.15}$$

where η_N has only diagonal elements η_{Nu} . The elements η_{Nu} are the eigenvalues of \mathcal{H}_N , while the matrix M is built from the eigenstates of \mathcal{H}_N . (Since

$$\sum_m \mathcal{H}_{N,nm} M_{ml} = M_{nl} \eta_{Nu} \tag{VIII.16}$$

the vectors $M_l = (M_{0l}, M_{1l}, \dots, M_{Nl})$ are the eigenvectors.) Once M has been obtained one defines new operators g_l as

$$g_l = \sum_n M_{nl} f_n \tag{VIII.17}$$

or

$$f_n = \sum_l M_{nl} g_l \tag{VIII.18}$$

and H_N is

$$H_N = \sum_l \eta_{Nu} g_l^+ g_l. \tag{VIII.19}$$

The matrix M and the eigenvalues η_{Nl} have to be computed numerically which is trivial for $N \lesssim 40$. From the results of the calculation one can easily extrapolate to any larger N . What happens is that for large N the eigenvalues of \mathcal{H}_N approach a limit, except that there are two separate limits, one for even N and another for odd N . For example, for $\Lambda = 2$ one obtains the eigenvalues

$$N \text{ even and large: } \eta = 0, \pm 1.297, \pm 2.827, \pm 4\sqrt{2}, \pm 8\sqrt{2}, \dots, \pm 2^{l-1}\sqrt{2}, \dots \quad (\text{VIII.20})$$

$$N \text{ odd and large: } \eta = \pm 0.6555, \pm 1.976, \pm 4, \pm 8, \pm 16, \dots, 2^l, \dots \quad (\text{VIII.21})$$

There is a symmetry (particle-hole symmetry) which ensures that for every eigenvalue η of \mathcal{H}_N there is also an eigenvalue $-\eta$, except when η is 0. The symmetry for the complete Kondo Hamiltonian is the following:

$$f_n \leftrightarrow (-1)^n f_n^+ \quad (\text{VIII.22})$$

$$\tau \leftrightarrow -\tau^* \quad (\text{VIII.23})$$

(where $\tau_x^* = \tau_x, \tau_y^* = -\tau_y, \tau_z^* = \tau_z$). Under this symmetry $H_N \rightarrow -H_N$. Hence for every positive eigenvalue of H_N there must be an equal and opposite eigenvalue. Since \mathcal{H}_N has $N + 1$ eigenvalues, it has an odd number of eigenvalues for even N ; the odd eigenvalue must be 0 as it is. For odd N no eigenvalue 0 is required, and there is none. The large eigenvalues for even N are approximately $\Lambda^{l-1}\sqrt{\Lambda}$ and for odd N are $\simeq \Lambda^l$, except for the very largest eigenvalues (l near N). This is seen in the numerical calculations and can be justified by setting up and analyzing the formal equations defining the fixed point.

Due to the even-odd alternation one does not have a fixed point of T but rather of T^2 ; T^2 is the transformation which takes H_N to H_{N+2} . The even and odd N limits define two distinct fixed points of T^2 . For T itself the even-odd alternation is called a limit cycle.

One practical comment. H_N , in diagonal form, includes terms of type $\eta g^+ g$ where η is negative. In this case the electron state created by g^+ is filled in the ground state of H_N . It is then useful to think in terms of holes instead of electrons. The operator which creates a hole (removes an electron) is g . So one defines $h^+ = g$ and $h = g^+$; using the anticommutation rules, and remembering that g has two spin components, one has

$$\eta g^+ g = -\eta h^+ h + 2\eta. \quad (\text{VIII.24})$$

The coefficient of $h^+ h$ is positive. This particle-hole transformation can be used to rewrite H_N so that it involves only positive single-electron energies.

The fixed points of T^2 involve an infinite set of eigenvalues, behaving as either $\Lambda^{l-1/2}$ or Λ^l as $l \rightarrow \infty$. However the \mathcal{H}_N for finite N have only a finite number of eigenvalues. If one looks at the largest eigenvalues of \mathcal{H}_N , they do not approach the fixed point eigenvalues even for $N \rightarrow \infty$. The precise statement of the fixed point limit is that the l th eigenvalue of \mathcal{H}_N approaches the l th eigenvalue of the fixed point for $N \rightarrow \infty$, but this approach is not uniform in l : the limit is reached only for $N \gg l$.

Now one can see the true reason for introducing the factor $\Lambda^{(N-1)/2}$ in the definition of H_N , namely this factor is necessary to have a fixed point in the large N limit for the low eigenvalues of \mathcal{H}_N . If no factor $\Lambda^{(N-1)/2}$ were introduced one would still get a fixed point, but the fixed point would be a limit of the largest eigenvalues of \mathcal{H}_N instead of the smallest eigenvalues.

Consider now the case $\tilde{J} = -\infty$, or more precisely $\tilde{J} \ll -1$. In this case the \tilde{J} term in H_N is much larger than the rest of H_N . The \tilde{J} term is $-\Lambda^{(N-1)/2} \tilde{J} f_0^+ \sigma f_0 \cdot \tau$; one wants to diagonalize this term and then treat the rest of H_N as a perturbation. The eigenstates of $f_0^+ \sigma f_0 \cdot \tau$ contain either 0, 1, or 2 electrons since there are two electron spin states available. The states with 0 or 2 electrons have total spin 0, i.e., $f_0^+ \sigma f_0$ is 0 for these states. The states with one electron have spin $\frac{1}{2}$; the operator $f_0^+ \sigma f_0$ acts on this spin $\frac{1}{2}$ multiplet like the matrices σ themselves. The product $\sigma \cdot \tau$ has two eigenvalues: +1 for total spin 1 and -3 for total spin 0. Since $-\Lambda^{(N-1)/2} \tilde{J}$ is positive, the spin 0 state has the lowest energy.

In the limit of large negative \tilde{J} , the excited states of the \tilde{J} term are negligible. Since the singlet state expectation value of f_0 and f_0^+ is zero, the effect of neglecting the excited states is to rub out the $f_0^+ f_1$ and $f_1^+ f_0$ terms in H_N , as well as replacing the \tilde{J} term by its eigenvalue for the singlet state. One is left with

$$H_N(\tilde{J} \rightarrow -\infty) = \text{constant} + \Lambda^{(N-1)/2} \times \sum_{n=1}^{N-1} \Lambda^{-n/2} (f_n^+ f_{n+1} + f_{n+1}^+ f_n). \quad (\text{VIII.25})$$

Again one has a quadratic form to diagonalize, with the only difference from the $\tilde{J} = 0$ case being that there are only N operators f_n instead of $N + 1$. In fact the quadratic form for H_N with $\tilde{J} = -\infty$ is precisely the same (apart from the labeling of the f 's) as the quadratic form for H_{N-1} at $\tilde{J} = 0$. As a result the same fixed points are obtained for large N , but the even-odd alternation is reversed: the single electron eigenvalues are (for $\Lambda = 2$)

$$\tilde{J} = -\infty, N \text{ even and large: } \pm 0.6555, \pm 1.976, \text{ etc.},$$

$$\tilde{J} = -\infty, N \text{ odd and large: } 0, \pm 1.297, \text{ etc.}$$

Some results of the computations will now be reported. Calculations were performed for $\tilde{J} = -0.024$ and $\Lambda = 2, 2.25, 2.5,$ and 3 ; the results are similar for the four cases and to be specific the results for $\Lambda = 2$ will be described. In this calculation the lowest 1620 energy levels of H_N were computed through $N = 180$. Some examples of the first and second excited state energies are as given in Table V.

TABLE V. Two excited state energies from the computer iterations (N : iteration number).

N	20	22	108	110	130	180
E_1	0.0314	0.0321	0.313	0.363	0.6541	0.6555
E_2	0.0419	0.0428	0.446	0.529	1.3055	1.3110

For small N , the solution is near the $\tilde{J} = 0$ values; for even N this means there is a single electron state with 0 energy. The numerical solution gives many electron state energies; in particular there are 0-, 1- and 2-electron states of zero energy, when $\tilde{J} = 0$. Combined with the impurity spin, the zero energy states are states with total spin $\frac{1}{2}$, 1, 0, and $\frac{1}{2}$: the energies E_1 and E_2 correspond to a spin $\frac{1}{2}$ state and spin 1 state, respectively (the ground state has spin $\frac{1}{2}$). For $N = 20$ or 22 one can see from Table V that these energies are still small.

As N increases the eigenvalues move away from their $\tilde{J} = 0$ values. For large N the solution approaches the $\tilde{J} = -\infty$ fixed point. This means the lowest single electron energy is 0.6555 (for N even) and the impurity spin is no longer involved explicitly (the impurity spin is tied up with f_0 in making a singlet state) so the states of energy 0.6555 above the ground state have only spin $\frac{1}{2}$. To make a state of spin 1 (energy level E_2) requires two electrons and therefore has an energy 1.311. By $N = 180$ the eigenvalues are indistinguishable from the eigenvalues of the $\tilde{J} = -\infty$ fixed point.

For all values of Λ , for both choices of the number of states kept, and for all values of \tilde{J} investigated (all such values were negative) the same result is seen: a crossover from near the $\tilde{J} = 0$ eigenvalues for small N (if \tilde{J} is small) to the $\tilde{J} = -\infty$ eigenvalues for large N .

The remainder of this section is devoted to a detailed and technical discussion of the behavior of the eigenvalues of H_N for very large N where H_N is near the $\tilde{J} = -\infty$ fixed point, followed by a similar discussion of moderately small N where H_N is near the $\tilde{J} = 0$ fixed point. This analysis has two purposes. One is that it provides a very thorough check on the validity and accuracy of the numerical calculations. The other purpose is to provide crucial information for the susceptibility and specific heat calculations described in Sec. IX. For very large N where the energy level spacing is much less than kT even for small T , one needs to know many more energy levels than are computed numerically; the analytic calculations for large N will be used in Sec. IX to give this information. The perturbation theory near $\tilde{J} = 0$ will be used to compute H_N for arbitrarily small \tilde{J} , given H_N for $\tilde{J} = -0.024$. All that is involved here is a translation: H_N for $\tilde{J} = -0.024$ is H_N , for very small \tilde{J} where $N' = N + (\text{function of } \tilde{J})$. The function of \tilde{J} will be determined in Sec. IX, based on information obtained in this lecture.

For further analysis of the numerical results it is necessary to identify the leading eigenoperators and eigenvalues of the linearized renormalization group transformation about each of the fixed points. A shortcut will be used to obtain these eigenoperators and eigenvalues. Consider a Hamiltonian $H' = H + \delta H$ where δH depends only on τ and the first few f_n . Also put $\tilde{J} = 0$ in H . Define

$$H_{N'} = H_N + \Lambda^{(N-1)/2} \delta H. \quad (\text{VIII.26})$$

Starting from the fact that $H_{N'}$ and H_N both satisfy the recursion formula VIII.3, and assuming that δH is infinitesimal, the difference $H_{N'} - H_N$ satisfies a linearized renormalization group equation with a linear operator L

depending on H_N . For large N , H_N is essentially H^* , where H^* is the $\tilde{J} = 0$ fixed point, and therefore the operator $L[H_N]$ is essentially $L[H^*]$, the linear operator whose eigenoperators and eigenvalues are sought. Note that even for N large, H_N is not *exactly* H^* , and if δH is small enough $H_{N'} - H_N$ may be much smaller than $H_N - H^*$. This does not change the fact that the linearized equation satisfied by $H_{N'} - H_N$ is close to the linearized equation about the fixed point, for N large. Therefore, for large N one can expand $H_{N'} - H_N$ in terms of eigenoperators and eigenvalues, giving

$$H_{N'} - H_N = \sum_m c_m O_m \lambda_m^N, \quad (\text{VIII.27})$$

where the O_m are the eigenoperators, λ_m the eigenvalues, and c_m are constants depending on the choice of δH .

For simple choices of δH , the Hamiltonians $H_{N'}$ can be diagonalized explicitly to first order in δH and the eigenvalue differences of $H_{N'} - H_N$ determined. From the N dependence of these eigenvalues, one can determine the operators O_m and eigenvalues λ_m , except for eigenoperators for which c_m is 0. By studying a sufficiently wide range of interactions δH , one should be able to identify all eigenoperators O_m of interest.

For the purposes of the renormalization group transformation, an operator H_N is specified by a list of energies (its eigenvalues) and a set of matrix elements (matrix elements of f_N). Correspondingly an eigenoperator O_m is also specified by a list of energies and a list of the matrix elements. In the following only the first few energies will be considered explicitly; these will be sufficient to identify the eigenvalues λ_m , and only the first few energies of the eigenoperators will be needed for later discussion.

The diagonalization of H_N for $\tilde{J} = 0$ has already been discussed. The eigenvalues of $H_{N'}$ are needed to first order in δH . By standard perturbation theory the eigenvalue differences of $H_{N'} - H_N$ are given by the matrix elements $\langle k | \delta H | k \rangle$ where the states $|k\rangle$ are the eigenstates of H_N .

Given the single particle operators g_i which diagonalize H_N , and their eigenvalues η_i (in the limit of large N , η_{Nu} reduces to N -independent values η_i), one can construct the multielectron eigenstates of H_N . It is convenient to replace those g_i with negative eigenvalues η_i by the corresponding hole operators $h_i = g_i^+$. In the following, the g_i will be only operators with positive eigenvalues η_i . For simplicity only the odd N case will be discussed; in this case there is both a particle operator g_i and a hole operator h_i for each eigenvalue η_i : there are $(N + 1)/2$ eigenvalues η_i . These eigenvalues will be arranged in increasing order: $\eta_1 < \eta_2$, etc.

The ground state of H_N is the state $|0\rangle$ destroyed by all the operators g_i and h_i . This is actually two states because of the two possible states for the impurity. The first excited states are the single particle (or hole) states $g_{1\mu}^+ |0\rangle$ and $h_{1\mu}^+ |0\rangle$. Then there are further single particle states, two particle states, etc. such as $g_{2\mu}^+ |0\rangle$, $g_{1\mu}^+ h_{1\nu}^+ |0\rangle$, etc.

In order to calculate matrix elements of δH between these eigenstates, it is convenient to express δH in terms of the operators g_i and h_i . Since δH depends on f_0, f_1 , etc. we need the expansion of f_0, f_1 , etc. in terms of g_i and h_i (actually h_i^+). This expansion has been determined numerically for N less than about 40. For $\Lambda = 2$ one finds for large N

$$f_0 = 2^{-(N-1)/4}\{0.588(g_1 + h_1^+) + 0.629(g_2 + h_2^+) + 0.861(g_3 + h_3^+) + \dots\}, \tag{VIII.28}$$

$$f_1 = 2^{-3(N-1)/4}\{0.386(g_1 - h_1^+) + 1.243(g_2 - h_2^+) + 3.446(g_3 - h_3^+) + \dots\}. \tag{VIII.29}$$

For $\Lambda \neq 2$, the N dependence is $\Lambda^{-(N-1)/4}$ and $\Lambda^{-3(N-1)/4}$ instead of $2^{-(N-1)/4}$ and $2^{-3(N-1)/4}$, and the numerical coefficients are different. This N dependence can only be approximately verified from the calculations since only $N < 40$ was considered; but the coefficients are independent of N to the accuracy cited. The power of Λ is not surprising: the operators g_i and h_i are roughly analogous to the operators a_m and b_{m_i} of the original discretization, with $m = (N - 1)/2$, in the sense that the energy Λ^{-m} associated with a_m and b_m is analogous to the energy $\eta_1 \sim 1$ of g_1 and h_1 when one takes into account the scale factor $\Lambda^{(N-1)/2}$ in the definition of H_N . The coefficients of a_m and b_m with $m = (N - 1)/2$ in the original formulae for f_0 and f_1 behave precisely as $\Lambda^{-(N-1)/4}$ and $\Lambda^{-3(N-1)/4}$, respectively, which is the same as the N dependence obtained from the numerical calculations.

For f_2, f_3 , etc. one finds that: f_2, f_4 , etc. are proportional to f_0 except for terms behaving as $\Lambda^{-5(N-1)/4}$, while f_3, f_5 , etc. are proportional to f_1 apart from terms behaving as $\Lambda^{-7(N-1)/4}$.

Consider now a particular example of an operator δH , say $\delta H = \epsilon(f_0^+ f_1 + f_1^+ f_0)$ where ϵ is a small number. The perturbation to H_N is $\Lambda^{(N-1)/2} \delta H$. Consider only matrix elements $\langle k | \delta H | k \rangle$ where $|k\rangle$ involves only the operators g_i^+ and h_i^+ through g_3^+ and h_3^+ acting on $|0\rangle$. Any term in δH involving g_4, h_4, g_5, h_5 , etc. on the right of an expression gives zero acting on any of these states $|k\rangle$; likewise any operator g_i^+, h_i^+ , on the left of an expression gives zero acting on any of the conjugate states $\langle k|$. (For example the matrix element $\langle k | h_4^+ h_4 | k \rangle$ vanishes). Also since one uses the same state $|k\rangle$ on both sides, the only operators with nonzero matrix elements are those that do not change the number of particles in any state. For example, $g_1 g_1^+$ has a matrix element but not $g_1 h_2^+$. However, if two different states $|k\rangle$ and $|k'\rangle$ have the same energy one must also consider the off-diagonal matrix elements $\langle k | \delta H | k' \rangle$ as part of a degenerate perturbation calculation. This means one is interested in any term in δH which connects states of equal energy.

The result of these considerations is that

$$\Lambda^{(N-1)/2} \delta H = \epsilon(\Lambda^{-(N-1)/2})\{c_1[2g_1^+ g_1 - 2h_1 h_1^+] + c_2[2g_2^+ g_2 - 2h_2 h_2^+] + \dots\}, \tag{VIII.30}$$

where c_1, c_2 , etc. are Λ dependent constants. (For $\Lambda = 2$, $c_1 = 0.227$, $c_2 = 0.782$, etc.) For $l > 3$ one can replace $2g_l^+ g_l - 2h_l h_l^+$ by $2g_l^+ g_l + 2h_l^+ h_l - 4$ and then neglect $2g_l^+ g_l + 2h_l^+ h_l$. Thus for our restricted set of states $|k\rangle$,

one has

$$\Lambda^{(N-1)/2} \delta H = \epsilon(\Lambda^{-(N-1)/2})\{c_1[2g_1^+ g_1 + 2h_1^+ h_1] + c_2[2g_2^+ g_2 + 2h_2^+ h_2] + c_3[2g_3^+ g_3 + 2h_3^+ h_3]\} + \text{constant}. \tag{VIII.31}$$

The constant is neglected because it is eliminated when one subtracts the ground state energies of both H_N and $H_{N'}$ (this is part of the renormalization group transformation). The c_1, c_2 , and c_3 terms have no ground state matrix elements. For excited states $|k\rangle$, the matrix elements of $\Lambda^{(N-1)/2} \delta H$ behave as $\Lambda^{-(N/2)}$ as a function of N . This means [cf., Eq. (VIII.27)] one can identify $\Lambda^{-1/2}$ as an eigenvalue of the linearized form of T , or more properly, since only odd N is being discussed, Λ^{-1} is an eigenvalue of the linearized form of T^2 about the odd N fixed point. The eigenoperator is the coefficient of $\epsilon \Lambda^{-N/2}$ in Eq. (VIII.31) apart from an arbitrary normalization constant which can be chosen to be 1. The eigenoperator has for its list of energies the diagonal matrix elements of $\sqrt{\Lambda} c_1(2g_1^+ g_1 + 2h_1^+ h_1) + \dots$.

Using $f_0^+ f_1 + f_1^+ f_0$ as the initial δH , one generates an irrelevant eigenoperator. Are there any relevant or marginal operators? Evidently one wants to use the minimum possible number of operators f_0 and f_1 in δH , since each f_0 or f_0^+ contributes a factor $\Lambda^{-N/2}$ and each f_1 or f_1^+ contributes a factor $\Lambda^{-3N/2}$ to the N dependence (and using f_2, f_3 , etc., is no better). Also, we will be interested only in eigenoperators with the same symmetries as H itself, namely δH must conserve charge and total spin and particle-hole symmetry. This means a minimum of two f 's; to achieve a maximum N dependence this means f_0^+ and f_0 . The only such operator satisfying all the symmetry requirements is $f_0^+ \sigma f_0 \cdot \tau$, the same operator that multiplies \tilde{J} in H . The alternative $f_0^+ f_0$ violates particle-hole symmetry. With $\delta H = f_0^+ \sigma f_0 \cdot \tau$, the operator $\Lambda^{(N-1)/2} \delta H$ is independent of N when expressed in terms of g_i, h_i , etc.; thus $f_0^+ \sigma f_0 \cdot \tau$ generates a marginal eigenoperator of the odd N fixed point.

Note that the operator $f_0^+ \sigma f_0 \cdot \tau$ exists only for the $\tilde{J} = 0$ fixed point. For the $\tilde{J} = -\infty$ fixed point, there is no τ operator (no doubling of the electron states due to the impurity), and hence there are no relevant or marginal operators for the $\tilde{J} = -\infty$ fixed point. For future use it will be necessary to know the dominant irrelevant operators for the $\tilde{J} = -\infty$ fixed point. Since products $f_0^+ f_0, f_0^+ f_2$, etc. are all ruled out by particle hole symmetry, it is not difficult to see that there are two dominant irrelevant operators (both having the same eigenvalue) namely the eigenoperator generated from $f_0^+ f_1 + f_1^+ f_0$, already discussed, and the eigenoperator generated from $(f_0^+ f_0 - 1)^2$. It is easily seen that $(f_0^+ f_0 - 1)$ is odd under particle-hole symmetry so $(f_0^+ f_0 - 1)^2$ is even. At first sight this would appear to be a marginal operator due to the $f_0^+ f_0$ term when $(f_0^+ f_0 - 1)^2$ is multiplied out. However, the $f_0^+ f_0$ term is canceled when the operators in $f_0^+ f_0 f_0^+ f_0$ are rearranged in normal ordered form. Normal ordering means all g_i^+ and h_i^+ operators are the left of all g_i and h_i operators. The reason for using normal ordering is to ensure that all terms with any g_i, h_i, g_i^+ , or h_i^+ for $l > 3$ have zero matrix element for the states $|k\rangle$ being considered. (The algebra demonstrating the cancella-

TABLE VI. Eight eigenstates and eigenvalues near the strong coupling fixed point (see text for explanations).

State	Q	S	E(H*)	Change in energy due to:	
				O ₂	O ₂
g ₁ ⁺ 0⟩	1	½	η ₁	λ*	0
g ₂ ⁺ 0⟩	1	½	η ₂	3.444λ*	0
g ₁ ⁺ g ₁ ⁺ h ₁ ⁺ 0⟩	1	½	3η ₁	3λ*	0
g ₁ ⁺ g ₁ ⁺ 0⟩	2	0	2η ₁	2λ*	w*
g ₁ ⁺ g ₂ ⁺ 0⟩	2	0	η ₁ + η ₂	4.444λ*	2.285w*
g ₁ ⁺ g ₂ ⁺ 0⟩	2	1	η ₁ + η ₂	4.444λ*	0
g ₁ ⁺ h ₁ ⁺ 0⟩	0	0	2η ₁	2λ*	w*
g ₁ ⁺ h ₁ ⁺ 0⟩	0	1	2η ₁	2λ*	-w*

tion of $f_0^+f_0$ is as follows: write $f_0 = (1/\sqrt{2})(g + h^+)$, where g is the sum of all g_i terms in f_0 , and h^+ is the sum of all the h_i^+ terms. It is easily seen from Eq. (VIII.28) that $\{g, g^+\} = \{h, h^+\}$; since $\{f_0, f_0^+\} = 1$ it follows that $\{g, g^+\} = \{h, h^+\} = 1$ also. Now

$$f_0^+f_0 - 1 = \frac{1}{2}(g^+ + h)(h^+ + g) - 1 = \frac{1}{2}(g^+h^+ + g^+g + hh^+ + hg) - 1. \quad (\text{VIII.32})$$

Using

$$hh^+ = \sum_{\mu=\pm\frac{1}{2}} h_{\mu}h_{\mu}^+ = 2 - h^+h, \quad (\text{VIII.33})$$

one has

$$f_0^+f_0 - 1 = \frac{1}{2}(-h^+h + g^+g + g^+h^+ + hg). \quad (\text{VIII.34})$$

Therefore

$$(f_0^+f_0 - 1)^2 = \frac{1}{4}(-h^+h + g^+g + g^+h^+ + hg)^2. \quad (\text{VIII.35})$$

When this square is multiplied out and put into normal ordered form (using relations like $h^+hh^+h = \sum_{\mu}\sum_{\nu}h_{\mu}^+h_{\nu}^+h_{\mu}h_{\nu} + 2h^+h$) one finds only quartic terms (like h^+h^+hh) and constants remain: there are no quadratic terms anymore.) With the $f_0^+f_0$ term canceled it is clear that for $\delta H = \epsilon(f_0^+f_0 - 1)^2$, $\Lambda^{(N-1)/2}\delta H$ behaves as $\Lambda^{-(N-1)/2}$ in terms of g_i, h_i , etc., just as in the case of $f_0^+f_1 + f_1^+f_0$. For $\Lambda = 2$ some of the terms in $\Lambda^{(N-1)/2}\delta H$ are

$$\Lambda^{(N-1)/2} H = 2^{-(N-1)/2}\epsilon\{ (0.588)^4 \sum_{\mu} \sum_{\nu} (h_{1\mu}^+h_{1\nu}^+h_{1\nu}h_{1\mu}) + g_{1\mu}^+g_{1\nu}^+g_{1\nu}g_{1\mu} - 2g_{1\mu}^+h_{1\mu}^+g_{1\nu}h_{1\nu} - 2h_{1\mu}^+h_{1\mu}g_{1\nu}^+g_{1\nu} + \dots \}. \quad (\text{VIII.36})$$

In particular the eigenoperator associated with $(f_0^+f_0 - 1)^2$ affects only two particle states and higher, not the one particle states. Once again the eigenoperator will be defined to be the coefficient of $2^{-N/2}\epsilon$ in Eq. (VIII.36).

Warning: in the above expression $h_{1\frac{1}{2}}^+$ creates a hole with spin $-\frac{1}{2}$ ($h_{1\frac{1}{2}}^+$ destroys an electron with spin $+\frac{1}{2}$).

Now a more detailed analysis of the numerical results for nonzero \tilde{J} will be given, concentrating on the small and large N limits where the Hamiltonians H_N are close to a fixed point. For very large N , H_N is close to the $\tilde{J} = -\infty$ fixed point: only even N will be discussed here so the fixed point H^* is the even N , $\tilde{J} = -\infty$ fixed point. For large but finite N , $H_N - H^*$ is small and should be approximately a solution of the linearized equation about

the fixed point. This means one should have (for even N)

$$H_N - H^* = \sum_m w_m(\tilde{J})\lambda_m^N O_m, \quad (\text{VIII.37})$$

where $w_m(\tilde{J})$ are coefficients depending on the initial choice of \tilde{J} (and also on Λ). For large N , the dominant terms will be the eigenoperators with the largest eigenvalue λ_m , and as shown above there are two dominant operators (call them O_1 and O_2) having the same eigenvalue: $\lambda_1 = \lambda_2 = \Lambda^{-\frac{1}{2}}$. Thus:

$$H_N - H^* = \{w_1(\tilde{J})O_1 + w_2(\tilde{J})O_2\}\Lambda^{-N/2}. \quad (\text{VIII.38})$$

There are two unknown constants $w_1(\tilde{J})$ and $w_2(\tilde{J})$ in this expression. These constants can be determined by fitting the formula (VIII.38) to the numerical results for the first few eigenvalues of H_N . In principle one needs to consider only two eigenvalues of H_N since only two constants are to be determined; in practice, eight eigenvalues were used in order to be sure that Eq. (VIII.38) is correct (and also as one of the checks on the computer program). The eight eigenstates were chosen as shown in Table VI. For each eigenstate, the charge Q , total spin S , and energy E (with respect to H^*) are given; charge and energy are both relative to the ground state.

In Table VI, products $g_1^+g_1^+$ always mean $g_{1\frac{1}{2}}^+g_{1-\frac{1}{2}}^+$ which has spin 0. Products like $g_1^+g_2^+$ and $g_1^+h_1^+$ are ambiguous; when $S = 1$ is specified then one is referring to spin 1 combinations like $g_{1\frac{1}{2}}^+g_{2\frac{1}{2}}^+$; when $S = 0$ is specified one is referring to the spin zero state

$$(\sqrt{2})^{-1}[g_{1\frac{1}{2}}^+g_{2-\frac{1}{2}}^+ - g_{1-\frac{1}{2}}^+g_{2\frac{1}{2}}^+]|0\rangle.$$

For $\Lambda = 2$ the energies η_1 and η_2 are 0.6555129 and 1.9760024, respectively. The constants λ^* and w^* are proportional to w_1 and w_2 , respectively

$$\lambda^* = 2\sqrt{2}(0.227)w_1 = 0.642w_1, \quad (\text{VIII.39})$$

$$w^* = 0.338w_2. \quad (\text{VIII.40})$$

The results for O_1 and O_2 were computed by computing matrix elements using the coefficients of $\epsilon\Lambda^{-N/2}$ in Eqs. (VIII.31) (O_1) and (VIII.36) (O_2). The numbers quoted above are for $\Lambda = 2$.

A problem that arises in trying to fit Eq. (VIII.38) to the numerical calculations is that the eigenvalues of H^* given in Table VI do not quite match the limit for $N \rightarrow \infty$ of the numerical calculations. This is because the numerical calculations involve truncation while the eigenvalues of Table VI are exact (although the values obtained for η_1 and η_2 are also only an approximation). The numerical energy for all even iterations with $N > 164$ ($\Lambda = 2$, $\tilde{J} = -0.024$) was 0.6555132 compared with exact value 0.6555129 (exact to seven decimal places). For the state $g_1^+g_2^+|0\rangle$ ($S = 0$) the computer gave 2.631526 while $\eta_1 + \eta_2$ is 2.631515; the difference 0.00001 was the largest found among the states of the table.

The linearized behavior (VIII.38) is true for the numerical results only if H^* is the numerically determined H^* , in the limit of large N . Furthermore, the truncations of the numerical calculations result in small modifications of the eigenvalues and eigenoperators of the linearized theory.

TABLE VII. Comparison of numerical calculations with analytic predictions from the strong coupling analysis.

State	Q	S	$E_\lambda(H^*)$	$E(H_N) - E(H^*)$ $N = 132 \times 10^{-4}$:		$E(H_N) - E(H^*)$ $N = 144 \times 10^{-5}$:	
				Computed	Theory	Computed	Theory
$g_1^+ 0\rangle$	1	$\frac{1}{2}$	0.6555132	-7.015	input	-1.101	input
$g_2^+ 0\rangle$	1	$\frac{1}{2}$	1.9760066	-21.833	-21.826*	-3.791	-3.787*
$g_1^+g_1^+h_1^+ 0\rangle$	1	$\frac{1}{2}$	1.9665404	-23.384	-23.380*	-3.307	-3.309*
$g_1^+g_1^+ 0\rangle$	2	0	1.31102801	-0.185	-0.181	-0.031	-0.034
$g_1^+g_2^+ 0\rangle$	2	0	2.6315262	0.459	0.468	0.065	0.061
$g_1^+g_2^+ 0\rangle$	2	1	2.6315174	-31.160	-31.175	-4.892	-4.894
$g_1^+h_1^+ 0\rangle$	0	0	1.3110274	-0.181	input	-0.034	input
$g_1^+h_1^+ 0\rangle$	0	1	1.3110256	-27.884	-27.880	-4.370	-4.371
λ^*				$-7.015 \times 10^{-4} \times 2^{66}$		$-1.101 \times 10^{-5} \times 2^{72}$	
w^*				$13.850 \times 10^{-4} \times 2^{66}$		$2.1687 \times 10^{-5} \times 2^{72}$	

To test Eq. (VIII.38) for the difference $H_N - H^*$, the difference $H_N - H^*$ (H^* given by iteration $N = 180$) was calculated from the numerical results for two values of N : $N = 132$ and $N = 144$. The results were compared with the theoretical formulae of the table, with λ^* and w^* determined independently for each N . Two of the eight eigenvalues in Table VI were used as input to determine λ^* and w^* ; the remaining 6 eigenvalues could then be predicted from the formulae of the Table. The results are shown in Table VII:

The numbers marked with an asterisk in Table VII had to be computed specially by degenerate perturbation theory. The two states involved have the same quantum numbers and for $N = 132$ the perturbation due to O_1 and O_2 is about the same order as the difference of the unperturbed energies. So it was necessary to take into account the off diagonal matrix element of the perturbation and diagonalize the two by two matrix from ($H^* +$ perturbation) to obtain the correct theoretical predictions. The off diagonal matrix element is $1.069 w^*$. The agreement of the computer results to theoretical predictions is excellent. The values of λ^* and w^* for $N = 132$ and $N = 144$ agree to within 0.5%. This means the eigenvalues of the linearized numerical transformation have to be within 0.1% of 2. One concludes that the theoretical analysis leading to Eq. (VIII.38) is correct and complete and that the coefficients $w_1(\tilde{J})$ and $w_2(\tilde{J})$ can be determined to about 0.5% accuracy. [Eqs. (VIII.39) and (VIII.40) relate w_1 and w_2 to λ^* and w^* .]

One final point for large N . The constants w_1 and w_2 have been determined for an example of small negative \tilde{J} , namely $\tilde{J} = -0.024$. They can also be determined for $\tilde{J} = -\infty$. Even for $\tilde{J} = -\infty$, the interactions H_N for finite N are not equal to H^* and for large N the form (VIII.38) should still be valid. For $\tilde{J} = -\infty$, H_N is a quadratic form in the g 's and h 's, so the nonquadratic term $(f_0^+f_0 - 1)^2$ cannot occur in the linearized approximation; hence $w_2(J)$ (i.e., w^*) must vanish. The value of w_1

TABLE VIII. Single particle energies for $\tilde{J} = -\infty$, vs N .

	$N = 30$	$N = 32$	$N = \infty$	$(N = 30) - (N = \infty)$
η_1	0.65552674	0.65551984	0.65551294	1.38×10^{-5}
η_2	1.97605015	1.97602630	1.97600245	2.385×10^{-5}
η_3	4.00006262	3.99997201	3.99988140	9.061×10^{-5}

(i.e., λ^*) can be determined from the N dependence of the single particle energies of H_N . From numerical diagonalizations one finds the N dependence ($\Lambda = 2$) (N refers to H_N with $\tilde{J} = -\infty$) shown in Table VIII. This gives $\lambda_* \simeq 0.44$.

This concludes the analysis of the numerical results for N very large. The next topic is the analysis for reasonably small values of N , where the $\tilde{J} = 0$ fixed point is relevant (assuming the initial value of \tilde{J} is small). First a formal discussion will be given. Near the $\tilde{J} = 0$ fixed point there is one marginal eigenoperator; the rest are irrelevant, with the largest irrelevant eigenvalue being $\Lambda^{-1/2}$. From the general discussion of marginal operators given in Sec. V, one expects the following. For sufficiently large N , large enough so that the irrelevant operators can be neglected, H_N will be completely determined except for one parameter (the parameter g_l of Sec. V). This is true independently of the original value of \tilde{J} . This means, as remarked in Sec. V, that there will be discrete sequences of initial values \tilde{J} which lead to the same set of effective interactions H_N , except for a translation in N . This prediction has been confirmed numerically. For example, two different initial values of \tilde{J} , for $\Lambda = 2.25$, gave the results for the first and second excited state energies and even N shown in Table IX.

The second run had a smaller value of \tilde{J} than the first run. Starting with iteration 20 there is virtually perfect matching. For small N , irrelevant operators are important, and the two runs diverge from each other. The same results for $N = 20$ apply to every eigenvalue of the numerical calculations, not just the examples cited.

TABLE IX. Verification (from numerical results) that a change in \tilde{J} is equivalent to a change in N when N is large. The first three columns show two excited state energies vs N for $\tilde{J} = -0.055016$, the last three columns are for $\tilde{J} = -0.02424$.

N	$\tilde{J} = -0.055016$		N	$\tilde{J} = -0.02424$	
	E_1	E_2		E_1	E_2
6	0.075257	0.10067	62	0.07261	0.09714
8	0.07777	0.10407	64	0.07662	0.10255
10	0.08156	0.10919	66	0.08109	0.10858
20	0.11367	0.15275	76	0.11377	0.15290
30	0.18440	0.25076	86	0.18450	0.25092
40	0.38734	0.56332	96	0.38737	0.56338
50	0.68168	1.31203	106	0.68166	1.31199

TABLE X. Comparison of numerical calculation of z_N with perturbation formula $V(z_{N-2}; 2)$.

N	z_N	$V(z_{N-2}; 2)$
39	0.0386566	
41	0.0397257	0.0397258
43	0.0408546	0.0408547
45	0.0420484	0.0420485

The fact that the eigenvalues of H_N for the first run are essentially the same as the eigenvalues of H_{N+56} for the second run, means, for example, that $w_1(\tilde{J})$ and $w_2(\tilde{J})$ for the second run are larger by a factor 2.25^{56} than w_1 and w_2 for the first run, and in particular the ratio w_2/w_1 is the same for both values of \tilde{J} . More generally one can construct a decreasing (in absolute value) sequence of initial values for \tilde{J} , say $\tilde{J}_0, \tilde{J}_2, \tilde{J}_4, \dots$ such that H_{N+l} for \tilde{J}_l is the same as H_N for \tilde{J}_0 ; for all these values of \tilde{J} , the ratio w_2/w_1 is constant.

Note however that w_2/w_1 is different for $\tilde{J} = -\infty$ than its value for \tilde{J} small, and therefore w_2/w_1 cannot be strictly constant even for small \tilde{J} . The deviations from constancy for small \tilde{J} are caused by irrelevant variables near the $\tilde{J} = 0$ fixed point. The size of the irrelevant variable effects is of order $\Lambda^{-N/2}$ in H_N . These effects can be estimated to still be of order $\Lambda^{-N/2}$ even for very large N , which means they have the same N dependence for very large N as the irrelevant terms involving w_1 and w_2 , in fact they cause corrections to w_1 and w_2 of absolute order 1. But for weak \tilde{J} , w_1 and w_2 are both very large (in the example with $\tilde{J} = -0.024$ and $\Lambda = 2$, w_1 and w_2 are of order 10^{17}) so corrections of absolute order 1 to these numbers are negligible. But when \tilde{J} is large, w_1 and w_2 are not large and deviations from constancy of the ratio are large.

One can only show that w_2/w_1 is constant for discrete sequences of \tilde{J} which produce the same discrete sequences of H_N apart from a translation in N . There are other values of \tilde{J} whose eigenvalues of H_N only interpolate between the eigenvalues of H_N of the original value \tilde{J}_0 , i.e., correspond in a rough sense to a non-integral translation in N . There is no guarantee, from the scaling law alone, that w_2/w_1 will be the same constant in these cases. This is part of a general problem of "wobble" caused by the discrete nature of the renormalization group transformation T . If one were dealing with an infinitesimal transformation so that N were a continuous variable, this problem would not arise. Various measures of this wobble will be discussed later, but the result is that it is a minor problem for $\Lambda = 3$ and 2.5 and negligible for $\Lambda = 2.25$ and 2 .

The scaling law discussed above allows one to extend the numerical results obtained for $\tilde{J} = -0.024$, say, to much smaller values of \tilde{J} without having to redo the calculations. However the numerical results for $\tilde{J} = -0.024$ and $\Lambda = 2$ are only useful for $N > 20$ or so; otherwise irrelevant variables are important. $N = 20$ for $\tilde{J} = -0.024$ will correspond to very large N for values of \tilde{J} much smaller than -0.024 . It will be useful to know the eigenvalues of H_N for all N (at least all $N > 20$) for arbitrarily

small values of \tilde{J} . Pretend for example that for $\tilde{J} = -0.014$, H_{80} corresponds to H_{20} of $\tilde{J} = -0.024$. One would like to know H_{20} through H_{79} for $\tilde{J} = -0.014$.

As \tilde{J} decreases the number of iterations required before H_N reaches H_{20} for $\tilde{J} = -0.024$ becomes enormous. The reason is that as long as H_N is near the $\tilde{J} = 0$ fixed point it is also very slowly varying with N . This is due to the presence of the marginal operator (generated by $f_0^+ \sigma f_0^-$) which to a first approximation does not change with N . This makes it hard to compute numerically for small \tilde{J} , because numerical errors accumulate over many iterations and because too much computer time is required. Fortunately, the range of N for which H_N is near the $\tilde{J} = 0$ fixed point can be handled by perturbation theory. Initially one calculates perturbation expansions in \tilde{J} , but for large N there are factors of N in these expansions (the factors of N are precisely the l factors that occur in Sec. V). We know from the general discussion of Sec. V that these factors of N can be removed at the price of introducing a set of expansion parameters, one for each N , in place of \tilde{J} itself. These expansion parameters will be chosen here to be

$$z_N = \eta_1 - E_{1N}, \quad (\text{VIII.41})$$

where E_{1N} is the first excited state energy (odd N only is considered), and η_1 is the lowest single particle energy of the odd N fixed point. From the general discussion of Sec. V one expects that z_{N+2} will have an expansion in z_N which is free of factors of N , and this is confirmed by actual calculations. The result is a formula

$$z_{N+2} = V[z_N; \Lambda] \quad (\text{VIII.42})$$

valid for large N , where $V[z_N]$ has a power series in z_N . In order to make accurate comparisons with the numerical calculations it was necessary to compute $V[z_N; \Lambda]$ to order z_N^4 . This was done by expanding both z_N and z_{N+2} in powers of \tilde{J} to order \tilde{J}^4 by standard perturbation methods (see the Appendix), and then inverting the series for z_N to express \tilde{J} as a function of z_N . The formulae for z_N in powers of \tilde{J} involves sums over all the eigenvalues η_{Nl} of H_N (a single sum in order \tilde{J}^2 , a double sum in order \tilde{J}^3 , etc.); these sums were computed by computer. As an example of the results, for $N = 31$ and $\Lambda = 2$ the expansion of z_N in a parameter J' proportional to \tilde{J} was found to be

$$z_N = 1.5J' + 27.2808J'^2 + 469.3577J'^3 + 7661.9091J'^4. \quad (\text{VIII.43})$$

By comparison the expansion of z_{N+2} in powers of z_N , for $N > 29$ and $\Lambda = 2$ was found to be

$$z_{N+2} = z_N + 0.71436z_N^2 + 0.14220z_N^3 - 2.9214z_N^4 = V(z_N; 2). \quad (\text{VIII.44})$$

(Calculations were performed using 30 decimal places to avoid the obvious problems of roundoff error.) This simple recursion formula can be compared with numerical results: for $\Lambda = 2$, and $\tilde{J} = -0.024$, the numerically determined values of z_N and corresponding theoretical prediction (calculating z_N from z_{N-2}) are given in Table X. The difference between the theoretical and numerical results is negligible.

One can now use Eq. (VIII.44) to extrapolate backwards the numerical results from $N = 39$. Define an infinite sequence of numbers z_N , for odd N and $-\infty < N \leq 39$, such that $z_{N+2} = V(z_N; 2)$ as given by Eq. (VIII.44), and such that $z_{39} = 0.0386566$. This sequence can be calculated by successive approximations (using this method on a pocket computer it was found that $z_{-29} = 0.020101$, for example) or it can be calculated analytically by first converting the recursion formula (VIII.44) to a differential equation (using the methods of Sec. V), and then integrating the differential equation (assuming z_N is small, which it is). The result of this procedure is the equation

$$(N - N')/2 = \Psi(z_{N'}) - \Psi(z_N), \quad (\text{VIII.45})$$

with (for $\Lambda = 2$)

$$\Psi(z) = (0.71436z)^{-1} - 0.72135 \ln(0.71436z) - 7.7310(0.71436z) \quad (\text{VIII.46})$$

neglecting order z^2 . By setting $N' = 39$ and $z_{N'} = 0.0386566$, one obtains the equation

$$(39 - N)/2 = \Psi(z_N) - 38.5885 \quad (\text{VIII.47})$$

which can be solved implicitly to give z_N as a function of N .

The extrapolation defined here has no relevance for $\tilde{J} = -0.024$, for negative values of N have no meaning for this case. But for smaller values of \tilde{J} the whole sequence of Hamiltonians H_N are translated in N , and negative values of N for $\tilde{J} = -0.024$ correspond to large positive N for sufficiently small \tilde{J} ; in this case the sequence $\{z_N\}$ defined by Eq. (VIII.47) completes the construction of effective Hamiltonians for small \tilde{J} except for the region N small ($N < 20$ or so) where irrelevant variables are important.

This completes the analysis of the numerical results. The emphasis was on results for $\Lambda = 2$ where the errors of the numerical calculations are largest (for $\Lambda = 2.25, 2.5$, and 3 the truncation errors are much smaller). Even for $\Lambda = 2$ the errors are remarkably small and fits to perturbation theory about the $\tilde{J} = 0$ and $\tilde{J} = -\infty$ fixed points work very well.

IX. IMPURITY SUSCEPTIBILITY AND SPECIFIC HEAT FOR THE KONDO HAMILTONIAN

In this section various methods for calculating the susceptibility and specific heat of the Kondo Hamiltonian will be discussed; the final result will be the zero temperature limit of both. The first part of the section is a discussion for the hopping Hamiltonian H for fixed Λ ; in the second part the continuum limit $\Lambda \rightarrow 1$ is examined.

To begin with the susceptibility $\chi(T)$ must be defined. The susceptibility one wants is the impurity susceptibility, namely the susceptibility of metal plus impurity minus the susceptibility of the normal metal. For the full hopping Hamiltonian H , both these susceptibilities are infinite due to the infinite number of free electron states in H . So one has to define a limiting procedure in order to obtain a finite result for the difference. The obvious choice of limit is to take the limit from a finite number of terms in H ; this

leads to the definition

$$\chi(T) = \frac{1}{kT} \lim_{M \rightarrow \infty} \left\{ \frac{\text{Tr} S_{Mz}^2 \exp[-\Lambda^{-(M-1)/2} H_M(\tilde{J})/kT]}{\text{Tr} \exp[-\Lambda^{-(M-1)/2} H_M(\tilde{J})/kT]} - \frac{\text{Tr} S_{Mz}^2 \exp[-\Lambda^{-(M-1)/2} H_M(\tilde{J} = 0)/kT]}{\text{Tr} \exp[-\Lambda^{-(M-1)/2} H_M(\tilde{J} = 0)/kT]} + \frac{1}{4} \right\} \quad (\text{IX.1})$$

(M is used as a substitute for N here to avoid confusion later on), where S_{Mz} is the z component of the total spin of the impurity plus the electrons in H_M ; it is assumed that the electrons and impurity have the same magnetic moment and g factor and both these terms have been set equal to 1; the final term $\frac{1}{4}$ is necessary to remove the impurity susceptibility at $\tilde{J} = 0$ from the $\tilde{J} = 0$ term in order that the $\tilde{J} = 0$ term subtract only the pure metal susceptibility. The factor $\Lambda^{-(M-1)/2}$ multiplying H_M is necessary to remove the factor $\Lambda^{(M-1)/2}$ used in defining H_M .

For any finite temperature T , the limit $M \rightarrow \infty$ means in practice that one must consider values of M such that $\Lambda^{-(M-1)/2}$ is much less than kT . This means that very highly excited states of H_M will be important, and these have not been calculated in the numerical calculations. Suppose, however, that one splits H_M into two parts: one can write

$$H_M = \Lambda^{(M-N)/2} H_N + R_{M,N}, \quad (\text{IX.2})$$

where N is any integer less than M , and the remainder $R_{M,N}$ is

$$R_{M,N} = \Lambda^{(M-1)/2} \sum_{n=N}^{M-1} \Lambda^{-n/2} (f_n^+ f_{n+1} + f_{n+1}^+ f_n). \quad (\text{IX.3})$$

Let H_N consist of those terms which, in H , are of order kT or greater, while $R_{M,N}$ includes the terms in H smaller than kT . This means that N is chosen so that $\Lambda^{-(N-1)/2}$ is of order kT , and that $\Lambda^{-(M-1)/2} R_{M,N}$ consists of terms each of which is smaller than kT . The remainder term $R_{M,N}$ can be split further:

$$R_{M,N} = \Lambda^{(M-N)/2} H_I + R_{M,N+1}, \quad (\text{IX.4})$$

where

$$H_I = \Lambda^{-1/2} (f_N^+ f_{N+1} + f_{N+1}^+ f_N). \quad (\text{IX.5})$$

The point of this splitting is that the operators which appear in $R_{M,N+1}$, namely f_{N+1}, \dots, f_M , are completely independent of the operators which appear in H_N . If it were not for the coupling term H_I , the trace calculations for H_N and $R_{M,N+1}$ could be done independently, using formulae such as

$$\begin{aligned} & \text{Tr} \exp\{-\Lambda^{-(N-1)/2} H_N/kT - \Lambda^{-(M-1)/2} R_{M,N+1}/kT\} \\ &= \text{Tr} \exp\{-\Lambda^{-(N-1)/2} H_N/kT\} \\ & \times \text{Tr}\{-\Lambda^{-(M-1)/2} R_{M,N+1}/kT\}. \end{aligned} \quad (\text{IX.6})$$

(Expectation values with S_{Mz} will be discussed later.) The coupling term must be taken into account; however, if N is not too small H_I can be treated as a perturbation. H_I appears inside exponentials multiplied by $\Lambda^{-(N-1)/2}/kT$, and H_I is itself of order 1. Thus if one chooses N so that $\Lambda^{-(N-1)/2}/kT$ is somewhat smaller than 1, the operator $\Lambda^{-(N-1)/2} H_I/kT$ will also be somewhat smaller than 1 and

one can expand in powers of this operator using standard techniques. (One might be able to treat the full remainder $R_{M,N}$ as a perturbation, but the large number of states associated with the operators $f_N^+ \cdots f_M^+$ may cause problems.) The perturbation expansion in H_I has been set up in practice (to order H_I^2) and the results are reported later. The results are not sufficiently accurate to check the convergence of the limit $\Lambda \rightarrow 1$. It is useful as a preliminary analysis to study the susceptibility neglecting the H_I term altogether. One can split S_{Mz} also

$$S_{Mz} = S_{Nz} + \frac{1}{2} \sum_{n=N+1}^M f_n^+ \sigma f_n \quad (\text{IX.7})$$

$$= S_{Nz} + S_{M,N+1,z}. \quad (\text{IX.8})$$

Thus

$$S_{Mz}^2 = S_{Nz}^2 + 2S_{Nz}S_{M,N+1,z} + S_{M,N+1,z}^2. \quad (\text{IX.9})$$

By total spin conservation the traces

$$\text{Tr } S_{Nz} \exp\{-\Lambda^{-(N-1)/2} H_N/kT\}$$

and $\text{Tr } S_{M,N+1,z} \exp\{-\Lambda^{-(M-1)/2} R_{M,N+1}/kT\}$ both vanish. Therefore

$$\begin{aligned} & \text{Tr } S_{Mz}^2 \exp\{-\Lambda^{-(M-1)/2} H_M/kT\} \\ & \simeq \text{Tr } S_{Nz}^2 \exp\{-\Lambda^{-(N-1)/2} H_N/kT\} \\ & \times \text{Tr } \exp\{-\Lambda^{-(M-1)/2} R_{M,N+1}/kT\} + \text{Tr } S_{M,N+1,z}^2 \\ & \times \exp\{-\Lambda^{-(M-1)/2} R_{M,N+1}/kT\} \text{Tr} \\ & \times \exp\{-\Lambda^{-(N-1)/2} H_N/kT\} \end{aligned} \quad (\text{IX.10})$$

neglecting H_I . (The traces are over the states of the impurity plus electron states 0 to N for H_N and over electron states $N+1$ to M for $R_{M,N+1}$.) The same breakup is true for any \tilde{J} , and $R_{M,N+1}$ is independent of \tilde{J} . The result of this approximation for $\chi(T)$ is that the terms involving $R_{M,N+1}$ cancel and one is left with Eq. (IX.1) but with the fixed number N replacing M .

With $\Lambda^{-(N-1)/kT}$ of order 1, only those excited states of H_N with energies of order 1 are important in the trace; higher energies are exponentially damped. Hence for a rough calculation one can keep only the energies calculated numerically. However it is more important to note the two limiting cases where one can calculate $\chi(T)$ by perturbation theory. If kT is extremely small, then N will be large enough so that H_N is approximately given by the $\tilde{J} = -\infty$ fixed point; if more accuracy is needed, one can use the fit to $H_N - H^*$ in terms of the leading irrelevant operators. If N is not too large, i.e., kT is not too small, then H_N will be close to the $\tilde{J} = 0$ fixed point, and one can use a perturbation expansion in \tilde{J} to compute $\chi(T)$. A better procedure for small N is to expand $\chi(T)$ in powers of z_N instead of \tilde{J} (N chosen so that $\Lambda^{-(N-1)/2}/kT \sim 1$). The low lying energy levels of H_N have an expansion in terms of z_N which is free of powers of N , and therefore $\chi(T)$ should also have a good expansion in terms of z_N . This turns out to be the case.

From the numerical calculations it is found that there is only a small range of values of N for which neither perturbation theory works at least roughly. Correspondingly there is a single order of magnitude range of T for which neither expansion is good; this range of T qualitatively defines T_K , the Kondo temperature. A more precise definition of T_K will be given later.

Now the perturbation expansions for $\chi(T)$ will be discussed in detail, first considering the large N case ($T \ll T_K$). In these calculations the exact formula (IX.1) will be used, not the approximation $M = N$.

The first approximation to $\chi(T)$ is obtained by replacing $H_M(\tilde{J})$ by $H^*(\tilde{J} = -\infty)$. Similarly, for large M one can replace $H_M(\tilde{J} = 0)$ by the $\tilde{J} = 0$ fixed point Hamiltonian. Both fixed point Hamiltonians are free electron Hamiltonians, and can be written as sums of independent terms of the form $\eta g^+ g$ ($g \equiv g_\mu$). For a single term of this type (summed over the spin index μ)

$$\text{Tr } S_z^2 \exp(-\beta H) = \frac{1}{2} e^{-\beta \eta}, \quad (\text{IX.11})$$

$$\text{Tr } S_z \exp(-\beta H) = 0, \quad (\text{IX.12})$$

$$\text{Tr } \exp(-\beta H) = (1 + e^{-\beta \eta})^2. \quad (\text{IX.13})$$

From these results it is easy to construct $\chi(T)$. Let η_1, η_2 , etc., be the single particle eigenvalues of the N odd, $\tilde{J} = 0$, fixed point Hamiltonian (each eigenvalue occurring twice). Let $0, \eta'_1, \eta'_2$, etc., be the single particle eigenvalues of the N even $\tilde{J} = 0$ fixed point Hamiltonian. There are now two cases to consider: M odd or M even. The M odd $\tilde{J} = -\infty$ fixed point is the same as the N even $\tilde{J} = 0$ fixed point except to omit the impurity states. Let M be odd. Let $\bar{\beta}_M = \Lambda^{-(M-1)/2}/kT$. Using the three results (IX.11)–(IX.13), one obtains

$$\begin{aligned} \chi(T) &= \frac{1}{kT} \lim_{M \rightarrow \infty} \left\{ \frac{1}{8} + \sum_{l=1}^{\infty} \frac{(\exp - \bar{\beta}_M \eta'_l)}{[1 + \exp(-\bar{\beta}_M \eta'_l)]^2} \right. \\ &\quad \left. - \sum_{l=1}^{\infty} \frac{\exp(-\bar{\beta}_M \eta_l)}{[1 + \exp(-\bar{\beta}_M \eta_l)]^2} \right\}. \end{aligned} \quad (\text{IX.14})$$

[The explicit $\frac{1}{4}$ in Eq. (IX.1) is cancelled by the impurity susceptibility term in the $\tilde{J} = 0$ term of Eq. (IX.1).] The $\frac{1}{8}$ term comes from the 0 energy states of the N even $\tilde{J} = 0$ fixed point. Similarly, for M even one obtains

$$\begin{aligned} \chi(T) &= \frac{1}{kT} \lim_{M \rightarrow \infty} \left\{ \sum_{l=1}^{\infty} \frac{\exp(-\bar{\beta}_M \eta_l)}{[1 + \exp(-\bar{\beta}_M \eta_l)]^2} \right. \\ &\quad \left. - \frac{1}{8} - \sum_{l=1}^{\infty} \frac{\exp(-\bar{\beta}_M \eta'_l)}{[1 + \exp(-\bar{\beta}_M \eta'_l)]^2} \right\}. \end{aligned} \quad (\text{IX.15})$$

Except for small l , η_l is Λ^{l-1} and η'_l is $\Lambda^{l-1/2}$; but for large enough M , $\bar{\beta}_M \eta_l$ is $\ll 1$ for small l , and can be replaced by 0, or equally well by $\bar{\beta}_M \Lambda^{l-1}$. Hence the formulae $\eta_l = \Lambda^{l-1}$ and $\eta'_l = \Lambda^{l-1/2}$ may be used for all l . Furthermore if one combines the two sums over l , the first few terms cancel completely for large enough M , and one can replace the lower limit $l = 1$ by $l = -\infty$ without changing the result. Hence

$$\begin{aligned} kT\chi(T) &= \chi_0(\bar{\beta}, \Lambda) \\ &= \lim_{\substack{M \rightarrow \infty \\ M \text{ even}}} \left\{ \frac{1}{8} + \sum_{l=-\infty}^{\infty} \left[\frac{\exp(-\bar{\beta}_M \Lambda^{l-1})}{[1 + \exp(-\bar{\beta}_M \Lambda^{l-1})]^2} \right. \right. \\ &\quad \left. \left. - \frac{\exp(-\bar{\beta}_M \Lambda^{l-1/2})}{[1 + \exp(-\bar{\beta}_M \Lambda^{l-1/2})]^2} \right] \right\} \end{aligned} \quad (\text{IX.16})$$

with a similar expression for M odd.

Expression (IX.16) for $\chi(T)$ is invariant to the change $\bar{\beta}_M \rightarrow \Lambda \bar{\beta}_M$, since this can be compensated by the change of summation variable $l \rightarrow l - 1$. This means the sum over l only needs to be calculated for a factor of Λ range of $\bar{\beta}_M$. The sum has been calculated numerically; some typical results are given in Table XI.

The results are small and decrease as Λ decreases. One can show that for $\Lambda \rightarrow 1$, $kT\chi(T) \rightarrow 0$; this will be proven later. The same is true for the alternate formula for M odd.

The above result shows that substituting $H^*(\vec{J} = 0)$ for H_M gives 0 for $kT\chi(T)$, at least in the limit $\Lambda \rightarrow 1$. Hence, it is important to study the effect of corrections to H^* .

It was shown in Sec. VIII that the leading corrections to H^* have the form

$$H_M \simeq H^* + \Lambda^{-M/2} [w_1(\vec{J})O_1 + w_2(\vec{J})O_2], \quad (\text{IX.17})$$

where w_1 and w_2 are parameters and O_1 and O_2 are the leading irrelevant operators. In Sec. VIII the corrections for the low lying eigenstates of H_M were discussed explicitly.

More complete formulae for O_1 and O_2 are needed for the susceptibility calculation. When $M \rightarrow \infty$, $\bar{\beta}_M \rightarrow 0$ and therefore highly excited states of H_M contribute to the thermodynamics. To start with a more complete form of Eqs. VIII.28–VIII.29 is required. One can write

$$f_0 = \Lambda^{-(N-1)/4} \sum_l \alpha_l (g_l + h_l^+), \quad (\text{IX.18})$$

$$f_1 = \Lambda^{-3(N-1)/4} \sum_l \gamma_l (g_l - h_l^+). \quad (\text{IX.19})$$

The quantities α_l and γ_l are known from numerical calculation for $l \leq 20$. For intermediate values of l they are found to behave as

$$\alpha_l \simeq \alpha \Lambda^{(l-1)/2}, \quad (\text{IX.20})$$

$$\gamma_l \simeq \gamma \Lambda^{3(l-1)/2}, \quad (\text{IX.21})$$

where α and γ are constants independent of l (but varying with Λ). This form is found to fit the numerical calculations except very near the endpoints of the range of l (the endpoints are 1 and $(N+1)/2$). For $\Lambda = 2$, α and γ are both 0.4307.

The formulae (VIII.31) and (VIII.36) must also be generalized. The results are

$$w_1(J)O_1 = \lambda^*(J) \sum_l \frac{\alpha_l \gamma_l}{\alpha_1 \gamma_1} (g_l^+ g_l + h_l^+ h_l) \quad (\text{IX.22})$$

$$w_2(J)O_2 = \frac{1}{2} w^*(J) \sum_{i_1} \cdots \sum_{i_4} \sum_{\mu} \sum_{\nu} (\alpha_1)^{-4} \alpha_{i_1} \alpha_{i_2} \alpha_{i_3} \alpha_{i_4} \\ \times \{g_{i_1\mu}^+ g_{i_2\nu}^+ g_{i_3\nu} g_{i_4\mu} + h_{i_1\mu}^+ h_{i_2\nu}^+ h_{i_3\nu} h_{i_4\mu} \\ + 2g_{i_1\mu}^+ h_{i_2\mu}^+ h_{i_3\nu} g_{i_4\nu} - 2g_{i_1\mu}^+ g_{i_2\mu} h_{i_3\nu}^+ h_{i_4\nu}\}. \quad (\text{IX.23})$$

(The indices μ and ν are spin indices.) These formulae are obtained by expressing $f_0^+ f_1 + f_1^+ f_0$ and $(f_0^+ f_0 - 1)^2$ in terms of g_l and h_l . For O_1 , only the diagonal terms are listed since only these affect the energy levels of H_M . All terms are included for O_2 . The parameters λ^* and w^* are normalized so that the energy level shifts of states created by g_l^+ and h_l^+ are simple [see Table VI, following Eq.

TABLE XI. Values of $kT\chi(T)$ very near $T = 0$ [using Eq. (IX.16)].

Λ	$\bar{\beta}_M$	$kT\chi(T)$
2	0.5	0.00001016
2.5	0.46	-0.000285
3	0.433	-0.000883

(VIII.38)]. This normalization is accomplished by the denominators $\alpha_1 \gamma_1$ and α_1^4 in Eqs. (IX.22) and (IX.23).

The above formulae define O_1 and O_2 when M is even [in Eq. (IX.17)]. For odd M one must use the even N , $\vec{J} = 0$ fixed point for E^* . The operators O_1 and O_2 still are derived from the operators $f_0^+ f_1 + f_1^+ f_0$ and $(f_0^+ f_0 - 1)^2$, but the expansions of f_0 and f_1 in terms of g_l and h_l have new coefficients. [Also there is a single creation operator for 0 energy, say g_0 , in addition to the pairs (g_1, h_1) , (g_2, h_2) , etc.]. The author has not calculated the operators O_1 and O_2 explicitly for odd M .

The calculation of $\chi(T)$ is straightforward. The O_1 and O_2 terms are to be treated in lowest order perturbation theory. Hence one has to compute

$$Z_M = \text{Tr} \exp(-\bar{\beta}_M H_M) \simeq \text{Tr} \exp(-\bar{\beta}_M H^*) \\ - \bar{\beta}_M \Lambda^{-M/2} \text{Tr} \{ (w_1 O_1 + w_2 O_2) \exp(-\bar{\beta}_M H^*) \}, \quad (\text{IX.24})$$

and

$$\text{Tr} S_{Mz^2} \exp(-\bar{\beta}_M H_M) \simeq \text{Tr} S_{Mz^2} \exp(-\bar{\beta}_M H^*) \\ - \bar{\beta}_M \Lambda^{-M/2} \text{Tr} \{ S_{Mz^2} (w_1 O_1 + w_2 O_2) \exp(-\bar{\beta}_M H^*) \}. \quad (\text{IX.25})$$

The traces involving O_1 and O_2 are calculated by standard methods. The result for $\chi(T)$ is that

$$kT\chi(T) = kT\chi_0 - \Lambda^{-M/2} \bar{\beta}_M \lambda^* \\ \times \sum_l \frac{\alpha_l \gamma_l \exp(-\bar{\beta}_M \eta_l) [1 - \exp(-\bar{\beta}_M \eta_l)]}{\alpha_1 \gamma_1 [1 + \exp(-\bar{\beta}_M \eta_l)]^3} \\ + 2\bar{\beta}_M w^* \Lambda^{-M/2} \left\{ \sum_l \frac{\alpha_l^2 \exp(-\bar{\beta}_M \eta_l)}{\alpha_1^2 [1 + \exp(-\bar{\beta}_M \eta_l)]^2} \right\}^2 \quad (\text{IX.26})$$

[$\chi_0(\beta, \Lambda)$ is defined in Eq. (IX.16).]

For $M \rightarrow \infty$, $\bar{\beta}_M$ becomes small and the terms for l near 1 are negligible in the sums, compared to the terms for large l . The reason for this is that $\alpha_l \gamma_l$ and α_l^2 increase with l . The dominant terms in the sums are those with $\bar{\beta}_M \eta_l \sim 1$. For very large l , $\bar{\beta}_M \eta_l$ is large and the terms are exponentially small. For large l one can use the asymptotic forms (IX.20) and (IX.21) for α_l and γ_l . Thus for sufficiently large M (sufficiently small $\bar{\beta}_M$) one has the following. Let

$$\chi_1(\bar{\beta}_M, \Lambda) = (\bar{\beta}_M)^2 \ln \Lambda \sum_{l=-\infty}^{\infty} \Lambda^{l-1} \\ \times \frac{\exp(-\bar{\beta}_M \eta_l) [1 - \exp(-\bar{\beta}_M \eta_l)]}{[1 + \exp(-\bar{\beta}_M \eta_l)]^3}, \quad (\text{IX.27})$$

$$\chi_2(\bar{\beta}_M, \Lambda) = \bar{\beta}_M^2 (\ln \Lambda)^2 \left\{ \sum_{l=-\infty}^{\infty} \frac{\Lambda^{l-1} \exp(-\bar{\beta}_M \eta_l)}{[1 + \exp(-\bar{\beta}_M \eta_l)]^2} \right\}^2, \quad (\text{IX.28})$$

TABLE XII. χ_1 and χ_2 for various values of $\bar{\beta}_M$.

$\Lambda = 2: \bar{\beta}_M$	0.5	0.6	0.77	0.93	1
χ_1	0.49989	0.49999	0.50015	0.49998	0.49989
χ_2	0.249988	0.250011	0.250002	0.249984	0.249988

where η_l is replaced by Λ^{l-1} . Then

$$kT\chi(T) \simeq \chi_0(T) - \frac{\Lambda^{-M/2}}{\ln\Lambda} \lambda^* \left(\frac{\alpha\gamma}{\alpha_1\gamma_1} \right) \frac{1}{\bar{\beta}_M} \chi_1(\bar{\beta}_M) + 2 \frac{\Lambda^{-M/2}}{(\ln\Lambda)^2} w^* \left(\frac{\alpha}{\alpha_1} \right)^4 \frac{1}{\bar{\beta}_M} \chi_2(\bar{\beta}_M). \tag{IX.29}$$

The reason for including the factor $\bar{\beta}_M^2$ in χ_1 and χ_2 is that they make the functions χ_1 and χ_2 invariant when $\bar{\beta}_M \rightarrow \Lambda\bar{\beta}_M$. The invariance is obtained by letting $\bar{\beta}_M \rightarrow \Lambda\bar{\beta}_M$ and $l \rightarrow l - 1$ in Eqs. (IX.27) and (IX.28). Hence χ_1 and χ_2 need be calculated only for one factor of Λ range of $\bar{\beta}_M$. The factors of $\ln\Lambda$ are included in χ_1 and χ_2 so that χ_1 and χ_2 have a limit for $\Lambda \rightarrow 1$. It will be shown later that $\chi_1 \rightarrow 0.5$ and $\chi_2 \rightarrow 0.25$ for $\Lambda \rightarrow 1$.

The functions χ_1 and χ_2 are easily calculated. For example, for $\Lambda = 2$ one obtains the results shown in Table XII. Clearly the sums for $\Lambda = 2$ are very close to the limit for $\Lambda \rightarrow 1$. The reason for this will be explained later. There is a small variation in χ_1 and χ_2 as $\bar{\beta}_M$ varies. The results for $\Lambda = 2, 2.25, 2.5,$ and 3 are summarized in Table XIII. As Λ increases the average values of χ_1 and χ_2 remain 0.5 and 0.25 , but the variation of $\chi_1(\bar{\beta}_M, \Lambda)$ and $\chi_2(\bar{\beta}_M, \Lambda)$ with $\bar{\beta}_M$ is larger, the maximum variation being $\pm 4\%$ for χ_1 at $\Lambda = 3$.

The variation of χ_1 and χ_2 with $\bar{\beta}_M$ is an artifact of the discretization for $\Lambda \neq 1$. It is one of the sources of error when $\chi(T)$ for $\Lambda = 1$ is approximated by $\chi(T)$ for $\Lambda \neq 1$. Thus there is a minimum error of about 4% in $\chi(T)$ when calculated using $\Lambda = 3$; for $\Lambda = 2$ this source of error is negligible. There are similar but smaller variations in χ_0 about 0 .

Neglecting the variation of χ_1 and χ_2 with $\bar{\beta}_M$, replacing $\chi_0(T)$ by 0 , and expressing $\bar{\beta}_M$ in terms of kT , one has

$$kT\chi(T) \simeq -0.5 \frac{kT}{\Lambda^{\frac{1}{2}} \ln\Lambda} \left\{ \lambda^* \left(\frac{\alpha\gamma}{\alpha_1\gamma_1} \right) - \frac{w^*}{\ln\Lambda} \left(\frac{\alpha}{\alpha_1} \right)^4 \right\}. \tag{IX.30}$$

Thus $\chi(T)$ is a constant independent of T .

The above calculation is valid only for small T , namely $T \ll T_K$, where T_K is the Kondo temperature. The reason for this is that $\Lambda^{-M/2}(w_1O_1 + w_2O_2)$ is a small perturbation only for low lying excited states of H_M . To be precise, the size of $\Lambda^{-M/2}(w_1O_1 + w_2O_2)$ is given essentially by the factors

TABLE XIII. Variation in χ_1 and χ_2 , vs $\bar{\beta}_M$, for various values of Λ .

Λ	2	2.25	2.5	3
χ_1	$0.5 \pm 0.03\%$	$0.5 \pm 0.2\%$	$0.5 \pm 1\%$	$0.5 \pm 4\%$
χ_2	$0.25 \pm 0.006\%$	$0.25 \pm 0.06\%$	$0.25 \pm 0.3\%$	$0.25 \pm 1.3\%$

$\Lambda^{-M/2}\lambda^*\alpha_l\gamma_l$ and $\Lambda^{-M/2}w^*\alpha_l^4$, which increase rapidly with l . These factors are to be compared to η_l , the unperturbed energies of H^* . Now $\alpha_l\gamma_l$ and α_l^4 increase as Λ^{2l} , while η_l increases only as Λ^l . The sums over l determining $\chi(T)$ are cut off when $\bar{\beta}_M\eta_l$ becomes $\gg 1$. Hence $\Lambda^{-M/2}(w_1O_1 + w_2O_2)$ can be treated as a perturbation in the calculation of $\chi(T)$ only if $\Lambda^{-M/2}\lambda^*\alpha_l\gamma_l$ and $\Lambda^{-M/2}w^*\alpha_l^4$ are smaller than η_l when l is chosen to make $\bar{\beta}_M\eta_l$ of order 1 . Consider for example the example of Sec. VIII with λ^* and w^* of order 2^{56} ($\Lambda = 2$). Then we must have $2^{56-M/2}2^{2l} \ll 2^l$ when $2^{-M/2}2^l/kT \geq 1$. This is true if $2^{56}kT \ll 1$. This means that $\bar{\beta}_M$ itself can be small only for $M \gg 112$. Referring to the table of numerical results (following Eq. VIII.25), $M \sim 112$ is about the lowest value of M for which H_M is close to H^* ($\bar{J} = -\infty$). Many people are upset by the large absolute magnitude of λ^* and w^* (2^{56} in the example). This analysis shows that the O_1 and O_2 terms still can legitimately be treated as a perturbation if M is large enough and $\bar{\beta}_M$ not too small; the basic reason is that λ^* and w^* are multiplied by $\Lambda^{-M/2}$ and the products $\lambda^*\Lambda^{-M/2}$, $w^*\Lambda^{-M/2}$ are small for large enough M .

The specific heat of the impurity can also be calculated, for small T . The calculation will be described briefly. The specific heat $C(T)$ for a general system is defined as

$$C(T) = \frac{d}{dT} [\text{Tr } H \exp(-H/kT)] / [\text{Tr } \exp(-H/kT)] = \frac{d}{dT} kT^2 \frac{d}{dT} \ln \text{Tr } \exp(-H/kT). \tag{IX.31}$$

The specific heat of the impurity is defined as the difference of (specific heat of metal plus impurity) - (specific heat of pure metal). Using the same limiting procedure as for the susceptibility, one obtains

$$C(T) = k \frac{d}{dT} T^2 \frac{d}{dT} \lim_{M \rightarrow \infty} \{ \ln \text{Tr} \times \exp[-\Lambda^{-(M-1)/2} H_M(\bar{J})/kT] - \ln \text{Tr} \times \exp[-\Lambda^{-(M-1)/2} H_M(\bar{J} = 0)/kT] + \ln 2 \}. \tag{IX.32}$$

(The $\ln 2$ term is to remove the free impurity contribution from the trace for $\bar{J} = 0$.) Here $C(T)$ calculated to first order in $w_1O_1 + w_2O_2$ gives

$$C(T) = k \frac{d}{dT} T^2 \frac{d}{dT} \ln Z(T), \tag{IX.33}$$

with

$$Z(T) = \frac{1}{4} \left\{ \prod_{l=-\infty}^{\infty} [1 + \exp(-\bar{\beta}_M\eta_l)]^4 [1 + \exp(-\bar{\beta}_M\eta_l')]^4 \right\} \times \left\{ 1 - \frac{4kT}{\Lambda^{\frac{1}{2}} \ln\Lambda} \lambda^* \frac{\alpha\gamma}{\alpha_1\gamma_1} I(\bar{\beta}_M, \Lambda) \right\}, \tag{IX.34}$$

where

$$I(\bar{\beta}_M, \Lambda) = \ln \Lambda \sum_{l=-\infty}^{\infty} \bar{\beta}_M^2 \Lambda^{2l-2} \frac{\exp(-\bar{\beta}_M\eta_l)}{[1 + \exp(-\bar{\beta}_M\eta_l)]}. \tag{IX.35}$$

In these formulae, η_l is Λ^{l-1} , and η_l' is $\Lambda^{l-1/2}$. There is no contribution to Z from the $w_2 O_2$ term. The over-all factor $\frac{1}{4}$ is due to the 0 energy mode present in the N even, $\bar{J} = 0$ fixed point. The sum I has been defined to be invariant to the change $\beta_M \rightarrow \Lambda \beta_M$ and to have a finite limit for $\Lambda = 1$, namely $I(\beta_M, 1) = \pi^2/12$ (see later). For $\Lambda = 2, 2.25, 2.5$, and 3 , one obtains $I = \pi^2/12$ with variation as shown below:

Λ	2	2.25	2.5	3
error in I	$\pm 0.006\%$	$\pm 0.06\%$	$\pm 0.3\%$	$\pm 1.5\%$

The product over l gives 4, except for similar but smaller fluctuations as a function of β_M . So, neglecting these fluctuations,

$$Z(T) = 1 - \frac{\pi^2}{3} \left(\frac{\lambda^*}{\Lambda^{\frac{1}{2}} \ln \Lambda} \frac{\alpha \gamma}{\alpha_1 \gamma_1} \right) kT \tag{IX.36}$$

for small T ; this gives

$$C(T) = -\frac{2\pi^2}{3} k^2 T \frac{\lambda^*}{\Lambda^{\frac{1}{2}} \ln \Lambda} \frac{\alpha \gamma}{\alpha_1 \gamma_1} \tag{IX.37}$$

The formulae obtained above for $\chi(T)$ and $C(T)$ can be used for specific values of \bar{J} for which numerical calculations are performed and values obtained for λ^* and w^* . A better and more general result can be obtained using a further set of formulae for $\chi(T)$ valid for large T [formulae for $C(T)$ for large T will not be needed]. For $T \gg T_K$ one can use perturbation theory in \bar{J} to compute $\chi(T)$, provided \bar{J} is small. When \bar{J} is small, the Kondo temperature T_K is very small (see later) so T can be much larger than T_K and still have $kT \ll 1$. In this case the perturbation expansion of χ involves logarithms: $J \ln kT$, $J^2 \ln^2 kT$, etc. i.e., the series for $\chi(T)$ is a bad series in the language of Sec. V. In the formula defining $\chi(T)$, the dominant states are those with energies of order kT above the ground state, and in the rough calculation described at the beginning of this section, one can determine these states by studying H_N with N chosen so that $\Lambda^{-N/2} \sim kT$. The eigenvalues of H_N have a good expansion in terms of z_N [z_N was defined in Eq. (VIII.41)]. Thus one anticipates that $\chi(T)$ has a good expansion in terms of z_N with N chosen to give $\Lambda^{-N/2} \sim kT$. This is the case. For example, let $\Lambda = 2$ and let $\beta_N = 2^{-(N-1)/2}/kT$ be $\sqrt{2}/2$. In the limit of large N a perturbation calculation (see the Appendix) gives

$$kT\chi(T) = 0.25 - 0.25766z_N + 0.31472z_N^2 - 1.1300z_N^3 \dots \tag{IX.38}$$

This is true for any large N as long as $kT = 2 \cdot 2^{-N/2}$. For large N , kT is very small but there are no $\ln kT$ terms in this formula; it is a good expansion. [Equation (IX.38) was obtained by expanding $\chi(T)$ in powers of \bar{J} and then substituting the expansion of \bar{J} in terms of z_N . For detailed formulae, see the Appendix. Both these calculations were performed for large but finite N . A typical set of formulae are

$$kT\chi(T) = -0.3864841J' - 9.219866J'^2 - 213.31641J'^3, \tag{IX.39}$$

$$J' = 0.6666667z_N - 11.41686z_N^2 + 202.52955z_N^3, \tag{IX.40}$$

where J' is proportional to \bar{J} ; substituting z_N for J' gives

the expansion (IX.38). The large numbers in the expansions (IX.39) and (IX.40) are due to $\ln kT$ terms in the expansion of $\chi(T)$ and terms proportional to N and N^2 in the expansion of J' .)

To understand the result (IX.38) being independent of T or N as long as $\beta_N = 0.7071$ is fixed, consider the definition of $\chi(T)$, Eq. (IX.1). First consider the qualitative calculation where $M \rightarrow \infty$ is replaced by $M = N$. In this case one has two traces to calculate. The first trace is $\exp\{-\beta_N H_N(\bar{J})\}$. This trace depends only on z_N , and not otherwise on N , for large N . The reason for this is that all the low lying energy levels of $H_N(\bar{J})$ are determined once z_N is given (see Sec. VIII). The highly excited states of H_N may depend on N but are exponentially damped in the calculation of $\chi(T)$. The other trace involves $\beta_N H_N(\bar{J} = 0)$. For large N , $H_N(\bar{J} = 0)$ can be replaced by a fixed point Hamiltonian which does not depend on N .

To be more accurate, one must consider $M > N$. Let $M = N + L$; the first trace now involves $\beta_N \Lambda^{-L/2} H_{N+L}(\bar{J})$. Once z_N is specified, H_N is determined; from the recursion formula for H_{N+1} , the energy levels of H_{N+1} , H_{N+2} , etc. are all uniquely determined also: they are independent of N if z_N is held fixed. So the trace is again independent of N . Likewise for the trace with $\bar{J} = 0$. Thus $kT\chi(T)$ depends on T and N only through β_N and z_N .

There is a consistency check one can perform on the expansion for $\chi(T)$. Namely, holding T fixed one can increase N , say to $N + 2$ and expand in z_{N+2} . One obtains $\beta_{N+2} = \beta_N/\Lambda$. For the example one has $\beta_{N+2} = 0.3535$. The expansion of $\chi(T)$ calculated for $\beta_{N+2} = 0.3535$ is found to be

$$kT\chi(T) = 0.25 - 0.25766z_{N+2} + 0.49878z_{N+2}^2 - 1.80601z_{N+2}^3 + \dots \tag{IX.41}$$

The expansion of z_{N+2} in terms of z_N was given earlier [Eq. (VIII.44)]. Using this expansion, Eq. (IX.41) reduces to the previous form (IX.38), as required.

The perturbation expansion for χ , Eq. (IX.38) can be used to determine the temperature dependence of $\chi(T)$, for $T \gg T_K$ and $kT \ll 1$. Specifically one proceeds as follows. Let $kT_0 = 2 \cdot 2^{-N/2}$ and $T = T_0 \cdot 2^{-L}$. Then Eq. (IX.38) can be used for both $\chi(T_0)$ and $\chi(T)$ (if T and T_0 are both larger than T_K). One has

$$kT_0\chi(T_0) = 0.25 - 0.25766z_N + 0.31472z_N^2 - 1.1300z_N^3 \tag{IX.42}$$

$$kT\chi(T) = 0.25 - 0.25766z_{N+2L} + 0.31472z_{N+2L}^2 - 1.1300z_{N+2L}^3 \tag{IX.43}$$

From Eq. (VIII.45), one has

$$\Psi(z_{N+2L}) - \Psi(z_N) = -L \tag{IX.44}$$

One can invert Eqs. (IX.42) and (IX.43) to express z_N and z_{N+2L} in terms of $\chi(T_0)$ and $\chi(T)$. It will be convenient to use a variable

$$y_0 = 2kT_0\chi(T_0) - 0.5, \tag{IX.45}$$

in place of $\chi(T_0)$ itself. In terms of y_0 ,

$$z_N = -1.9406y_0 + 4.5998y_0^2 + 10.2447y_0^3 + \dots \tag{IX.46}$$

TABLE XIV. Values of E , F , and G in Eq. (IX.58).

Λ	2.0	2.25	2.5	3
E	$-0.5 \pm 0.004\%$	$-0.5 \pm 0.04\%$	$-0.5 \pm 0.2\%$	$-0.5 \pm 1\%$
F	-0.5	-0.5	-0.5	-0.5
G	$3.16485 \pm 0.01\%$	$3.1648 \pm 0.1\%$	$3.165 \pm 0.3\%$	$3.165 \pm 1.5\%$

The expansions for z_N and z_{N+2L} can be substituted in Eq. (IX.44). Writing $y_L = 2kT\chi(T) - 0.5$, the result is

$$\Phi(y_L) - \Phi(y_0) = -L \ln 2 = \ln(T/T_0), \quad (\text{IX.47})$$

where

$$\Phi(y_0) = (\ln 2)\Psi(z_N) + 1.0018 = (-1/2y_0) - 0.5 \ln|2y_0| + 3.1648y_0 + O(y_0^2). \quad (\text{IX.48})$$

The constant 1.0018 was added in Eq. (IX.48) so that $\Phi(y_0)$ would have no constant term. Now one has

$$\Phi[2kT\chi(T) - 0.5] - \Phi[2kT_0\chi(T_0) - 0.5] = \ln(T/T_0) \quad (\text{IX.49})$$

which means that if one defines B to be

$$B = \ln kT - \Phi[2kT\chi(T) - 0.5], \quad (\text{IX.50})$$

then B is independent of T so long as T is an inverse power of 2.

The quantity B depends on the initial coupling constant J . In order to make connections with previous work on the Kondo problem we shall return to the constant J of previous workers in place of \tilde{J} . In the conventional continuum Kondo theory discussed in (Kondo, 1969), $\chi(T)$ has an expansion of the form

$$kT\chi(T) = 0.25\{1 + 2J\rho + 4J^2\rho^2 \ln(kT/D) + \dots\}, \quad (\text{IX.51})$$

where ρ is the density of states at the Fermi surface, and D is the width (in energy) of the conduction band. (In the model of Sec. VII, ρ is 2, and D is 1.) Only the leading logarithm has been calculated in order $J^2\rho^2$.

Consider B as a function of $J\rho$. It is convenient to write B in terms of a new constant \tilde{D}

$$B(J\rho) = \ln \tilde{D}(J\rho) - \Phi(J\rho). \quad (\text{IX.52})$$

The point of this equation is that $\tilde{D}(J\rho)$ has a power series expansion in $J\rho$ while $B(J\rho)$ does not. This follows from the fact that $\chi(T_0)$ for fixed T_0 has a power series expansion in $J\rho$, and in particular $y_0 = 2kT_0\chi(T_0) - 0.5$ has the form

$$y_0 = J\rho + O(J^2\rho^2). \quad (\text{IX.53})$$

From Eq. (IX.50) one has

$$\ln[kT_0/\tilde{D}(J\rho)] = \Phi[y_0] - \Phi[J\rho]. \quad (\text{IX.54})$$

The difference $\Phi(y_0) - \Phi(J\rho)$ contains the differences $[y_0^{-1} - (J\rho)^{-1}]$ and $\ln y_0 - \ln(J\rho)$. Using Eq. (IX.53) it follows that each of these differences has a power series expansion in $J\rho$. Using the second order term in Eq. (IX.51), one obtains

$$\tilde{D}(J\rho) = D(c_0 + c_1J\rho + c_2J^2\rho^2 + \dots). \quad (\text{IX.55})$$

The constants c_0 , c_1 , etc. are determined by the constant

(nonlogarithmic) terms in the expansion (IX.51) in order $J^2\rho^2$, $J^3\rho^3$, etc. An example will be given later.

The temperature dependence of $\chi(T)$ for $T \gg T_K$ is now given by

$$\Phi[2kT\chi(T) - 0.5] = \Phi[J\rho] + \ln[kT/\tilde{D}(J\rho)]. \quad (\text{IX.56})$$

Solving this equation as an expansion in $J\rho$, using Eq. (IX.48), one obtains

$$\begin{aligned} kT\chi(T) = 0.25\{ & 1 + 2J\rho + 4J^2\rho^2 \ln(kT/\tilde{D}) \\ & + J^3\rho^3[8 \ln^2 kT/\tilde{D}] + 4 \ln(kT/\tilde{D}) \\ & + J^4\rho^4[16 \ln^3(kT/\tilde{D}) + 20 \ln^2(kT/\tilde{D}) \\ & - 21.33 \ln(kT/\tilde{D})] + \dots\}. \end{aligned} \quad (\text{IX.57})$$

To the author's knowledge the nonleading logarithms in this expansion have not been previously calculated to this order in $J\rho$. The leading logarithms agree with earlier calculations.

The function $\Phi(y)$ in principle should be calculated in the limit $\Lambda \rightarrow 1$. We need to know the errors that occur for $\Lambda \geq 2$. The general form of $\Phi(y)$ is

$$\Phi(y) = (E/y) + F \ln|2y| + Gy, \quad (\text{IX.58})$$

where E , F , and G depend on Λ ; except for F they also depend on the value of $\tilde{\beta}$ chosen in calculating the expansion of χ in terms of z_N . These constants have been calculated for a range of values of $\tilde{\beta}$, with the result shown in Table XIV.

Once again there is rapid convergence as Λ decreases; this will be explained later.

The equations for the temperature dependence of $\chi(T)$ obtained above were derived assuming $kT = \Lambda^{-(N-1)/2}/\tilde{\beta}$ for integer N and fixed $\tilde{\beta}$; they are not (as derived) valid for all T . To cover a continuous range of T one must relate the calculations of $\chi(T)$ for a continuous range of $\tilde{\beta}$. This can be done in practice as follows. For two values of $\tilde{\beta}$, say $\tilde{\beta}$ and $\tilde{\beta}'$, one can calculate $\chi(T)$ for fixed N . Thus one calculates $\chi(T)$ for $kT = \Lambda^{-(N-1)/2}/\tilde{\beta}$ and $\chi(T')$ with $kT' = \Lambda^{-(N-1)/2}/\tilde{\beta}'$. The result can be written

$$\Phi[2kT\chi(T) - 0.5] = (\ln \Lambda)\Psi(z_N) - H(\tilde{\beta}), \quad (\text{IX.59})$$

$$\Phi[2kT'\chi(T') - 0.5] = (\ln \Lambda)\Psi(z_N) - H(\tilde{\beta}'), \quad (\text{IX.60})$$

where $H(\tilde{\beta})$ is a constant for given $\tilde{\beta}$. In these formulae the small variations of Φ with $\tilde{\beta}$ have been ignored. Numerical calculations of $H(\tilde{\beta})$ show that to a good approximation

$$H(\tilde{\beta}) = + \ln \tilde{\beta} + H, \quad (\text{IX.61})$$

where H is independent of $\tilde{\beta}$. The results for H are:

Λ	2	2.25	2.5	3
H	-0.65528	-0.6424	-0.628	-0.596
	$\pm 0.03\%$	$\pm 0.2\%$	$\pm 0.8\%$	$\pm 2\%$

Thus, one has, for N fixed,

$$\begin{aligned} \Phi[2kT\chi(T) - 0.5] - \Phi[2kT'\chi(T') - 0.5] \\ = -\ln(\bar{\beta}/\beta') = \ln(kT/kT') \end{aligned} \quad (\text{IX.62})$$

which is the previous formula (IX.56) but now valid for a continuous range of values of T and T' .

This completes the study of $\chi(T)$ for large T ($T \gg T_K$, $1 \gg kT$). The stage is set for the calculation of the zero temperature susceptibility and specific heat as a function of $J\rho$. This means computing λ^* and w^* as a function of $J\rho$. The computer calculation gives λ^* and w^* for a single value of $J\rho$. More precisely the information taken from the computer calculation is as follows. For some initial value N_i of N , the value $z_i = z(N_i)$ is obtained from the computer output. In the example discussed in Sec. VIII, N_i can be chosen to be 39; then z_i is 0.038657. Then values of λ^* and w^* are calculated by matching to the energies calculated for a large value N_f for N ($N = 132$ for the example of Sec. VIII). One calculates λ^* from the formula

$$\lambda^* = \Lambda^{N_f/2} [E_{1f} - E_1(\infty)], \quad (\text{IX.63})$$

where E_{1f} is the energy of the state $g_1^+|0\rangle$ (see Sec. VIII) for iteration N_f , obtained from the computer calculation. Likewise

$$w^* = \Lambda^{N_f/2} [E_{2f} - E_2(\infty)] - 2\lambda^*, \quad (\text{IX.64})$$

where E_{2f} is the energy of the state $g_1^+h_1^+|0\rangle$ with total spin O for iteration N_f .

For an arbitrary value of $J\rho$, one cannot expect $z_N = z_i$ for $N = N_i$. However one can calculate the value of N for which $z_N = z_i$. Namely, from Eqs. (IX.56) and (IX.59), one has

$$(\ln\Lambda)\Psi(z_N) - H - \ln\bar{\beta} = \Phi(J\rho) + \ln(kT/\bar{D}) \quad (\text{IX.65})$$

as long as $kT = \Lambda^{-(N-1)/2}/\bar{\beta}$. This means

$$(N-1)/2 \ln\Lambda = -\ln\bar{\beta} - \ln kT. \quad (\text{IX.66})$$

If z_N is required to be z_i then N is

$$\begin{aligned} N = N_0(J\rho) = 1 + (2/\ln\Lambda)\{\Phi(J\rho) + H - \ln\bar{D}\} \\ - 2\Psi(z_i). \end{aligned} \quad (\text{IX.67})$$

Suppose for a moment that $N_0(J\rho)$ is an integer. The computer calculation tells us that $N_f - N_i$ further iterations are required in order that the first excited state energy $E_1(N)$ be equal to E_{1f} . In other words, if $z_{N_0} = z_i$ for initial coupling $J\rho$, then $E_1(N_0 + N_f - N_i)$ will be equal to E_{1f} . The reason for this is that $E_1(N_0 + N_f - N_i)$ is determined once z_{N_0} is given (as explained earlier) irregardless of the value of N_0 (provided N_0 is sufficiently large).

Now one obtains $\lambda^*(J\rho)$ from a generalization of Eq. (IX.63)

$$\lambda^*(J\rho) = [E_{1f} - E_1(\infty)]\Lambda^{(N_0+N_f-N_i)/2}. \quad (\text{IX.68})$$

A similar formula holds for $w^*(J\rho)$. Both λ^* and w^* depend on $J\rho$ only through the factor $\Lambda^{N_0/2}$; in the ratio w^*/λ^* this factor is removed. Thus w^*/λ^* is independent of $J\rho$. (This is true only for small $J\rho$, as was explained in Sec. VIII.)

One can now combine Eqs. (IX.67) and (IX.68) to give the zero temperature susceptibility

$$\chi(0) = \chi_z \exp\{\Phi(J\rho) - \ln\bar{D}\}, \quad (\text{IX.69})$$

where χ_z is a constant independent of $J\rho$

$$\begin{aligned} \chi_z = -\frac{1}{2} \frac{1}{\ln\Lambda} \{E_{1f} - E_1(\infty)\} \frac{\alpha\gamma}{\alpha_1\gamma_1} \left\{ 1 - \frac{w^*}{\lambda^*} \frac{\gamma_1}{\gamma} \left(\frac{\alpha}{\alpha_1}\right)^3 \frac{1}{\ln\Lambda} \right\} \\ \times \exp\left\{ H + \frac{(N_f - N_i)}{2} \ln\Lambda - \ln\Lambda\Psi(z_i) \right\}. \end{aligned} \quad (\text{IX.70})$$

In this formula, N_f , N_i , z_i , E_{1f} , $E_1(\infty)$, and w^*/λ^* are determined from computer output; α , γ , α_1 , γ_1 are determined from free electron ($\vec{J} = 0$) calculations; the constant H and functions $\Phi(J\rho)$ and $\Psi(z)$ are determined from perturbation theoretic calculations.

The ratio of specific heat to susceptibility at zero temperature has a simpler form. The ratio one studies is the zero temperature limit of $C(T)/T\chi(T)$ and this is

$$\lim_{T \rightarrow 0} \frac{C(T)}{T\chi(T)} = \frac{4\pi^2 k^2}{3R}, \quad (\text{IX.71})$$

with

$$R = \left\{ 1 - \frac{w^*}{\lambda^*} \frac{1}{\ln\Lambda} \left(\frac{\gamma_1}{\gamma}\right) \left(\frac{\alpha}{\alpha_1}\right)^3 \right\}. \quad (\text{IX.72})$$

The results for χ_z and R are shown in Table XV.

In these calculations, a maximum of 1620 states were kept (for $\Lambda = 2, 2.25$, and 2.5), broken down as shown in Table XVI.

(For $S \neq 0$ this table gives the number of states kept for a given value of S_z .) The calculations were performed on a CDC 7600 and required ~ 7 seconds for 1 iteration. The calculations for $\Lambda = 3$ were performed keeping only 526 states, as shown in Table XVII.

TABLE XV. Results of the Kondo calculation.

Λ	2.0	2.25	2.5	3
R	1.998	1.9994	2.00006	2.0001
χ_z	0.10345	0.10315	0.1029	0.1026
N_i	43	35	31	31
z_i	0.0408546	0.0409743	0.0417065	0.0303433
N_f	142	130	114	120
$E_{1f} - E_1(\infty)$	-2.2027×10^{-5}	-2.5359×10^{-6}	-1.3746×10^{-6}	-2.5902×10^{-4}

TABLE XVI. Number of states retained for different values of Q and S in the full calculation (1620 states total) counting different values of S_z .

S	0	$\frac{1}{2}$	1	$1\frac{1}{2}$	2	$2\frac{1}{2}$	3	$3\frac{1}{2}$
$Q = 0$	26	38	34	21	10	4	2	1
± 1	21	33	26	19	7	3	1	1
± 2	11	16	14	9	5	2	1	
± 3	6	8	6	3	1	1	1	
± 4	4	4	4	2	1	1		
± 5	2	2	2	1	1			
± 6	1	1	1	1	0	0	0	0

If $N_0(J\rho)$ calculated from Eq. (IX.67) is not an integer then one must change the value of z_i until $N_0(J\rho)$ is an integer, and then redo the computer calculations with the new value of z_i . This means there will be a new value of E_{1f} also. But for $\Lambda \rightarrow 1$ there should be a unique value for $\chi(0)$, which means the product $\{E_{1f} - E_1(\infty)\} \times \exp\{-\ln\Lambda\Psi(z_i)\}$ should not vary appreciably as z_i varies (if N_i and N_f are held fixed). Likewise the ratio w^*/λ^* should not vary. To test this, calculations for $\Lambda = 2.25$ were carried out for four values of z_i ; only a 0.1% variation was found for the product $\{E_{1f} - E_1(\infty)\} \exp\{-\ln\Lambda\Psi(z_i)\}$ and the ratio w^*/λ^* .

The values of the susceptibility and specific heat are subject to intrinsic errors due to fluctuations for $\Lambda \neq 1$ as noted in various tables, but for $\Lambda = 2$ and 2.25 these fluctuations are a small fraction of 1% and are unimportant. For $\Lambda = 2.5$ they are of order 1% to 2%, and <5% for $\Lambda = 3$. In addition there are numerical truncation errors which could approach 1% for χ_z for $\Lambda = 2$; for all the other results these truncation errors are much less than 1%. A fuller discussion of errors will be given later.

The results for R and χ_z are the central results of the Kondo calculation. R is remarkably close to 2; one can hardly avoid the conclusion that it is 2 exactly. Recently Nozieres (1974) has argued that the factor R is 2 precisely. Nozieres combines qualitative results from the present work with known properties (in perturbation theory) of the Kondo Hamiltonian under a shift of the Fermi energy. Such a shift produces a violation of particle-hole symmetry; the effect of such a violation on the numerical calculations has not been studied yet. An even simpler argument for R being 2 is given by Yamada (1975).

The first step in the error analysis is to explain the rapid convergence of the sums χ_1, χ_2 , and I as $\Lambda \rightarrow 1$ and likewise the rapid convergence of E, F , and G [the coefficients in $\Phi(y)$]. The coefficients E, F , and G also involve sums over the free particle energies. E is just $-\chi_2^{1/2}$ while F involves a

TABLE XVII. Number of states kept in the fast calculation (526 states total).

S	0	$\frac{1}{2}$	1	$1\frac{1}{2}$	2	$2\frac{1}{2}$
$Q = 0$	13	18	14	6	1	1
± 1	11	15	13	4	1	1
± 2	6	8	6	2	1	1
± 3	3	3	2	1	1	0
± 4	1	1	1	1	0	0

double sum and G involves a triple sum, both of which the author has not tried to write down explicitly.

The quantity $I_2 = (\chi_2^{1/2})$ can be used to illustrate the rapid convergence. In the limit $\Lambda \rightarrow 1$, I_2 becomes an integral. This is seen as follows. Let $x = (l - 1) \ln\Lambda$. For Λ near 1, $\ln\Lambda$ is small and δx is small for a unit change in l : $\delta x = \ln\Lambda$. Write

$$u_l = \bar{\beta}^{l-1} = \bar{\beta}e^x. \tag{IX.73}$$

Then from Eq. (IX.28)

$$I_2(\beta, \Lambda) = \ln\Lambda \sum_{l=-\infty}^{\infty} \frac{u_l \exp(-u_l)}{[1 + \exp(-u_l)]^2} \simeq \int_{-\infty}^{\infty} dx \bar{\beta}e^x \exp(-\bar{\beta}e^x) [1 + \exp(-\bar{\beta}e^x)]^{-2}. \tag{IX.74}$$

Let $z = \exp(-\bar{\beta}e^x)$ giving $dz = -\bar{\beta}e^x dx$. Then

$$I_2(\bar{\beta}, \Lambda) \simeq \int_0^1 dz \frac{1}{(1+z)^2} = 0.5. \tag{IX.75}$$

For $\Lambda \rightarrow 1$, $\ln\Lambda \rightarrow 0$, and this formula becomes exact.

The sum in Eq. (IX.74) is a very naive approximation to the integral. If the limits of integration were finite, the error in approximating the integral by a sum would be of order $\delta x = \ln\Lambda$. But because the limits are infinite the error is exponentially small, namely of order $\exp\{-\pi^2/\ln\Lambda\}$. This is the reason for the rapid convergence. To show this result we review a standard but poorly known (in the U.S.) theorem about approximation of infinite integrals by infinite sums [Fisher and Barker (1972)]. Let $f(x)$ be analytic in the complex x plane in a strip $-\tau - \text{Im}x < \text{Re}x < \tau$ about the real axis. Also assume $f(x)$ goes to zero rapidly as $x \rightarrow \pm\infty$ in the strip. Then let

$$I = \int_{-\infty}^{\infty} f(x) dx. \tag{IX.76}$$

An approximation to this is

$$I(a) = a \sum_{l=-\infty}^{\infty} f(al) \tag{IX.77}$$

for small a . One relates $I(a)$ to I as follows. One can write

$$I(a) = a \int_{-\infty}^{\infty} dx f(x) \sum_{l=-\infty}^{\infty} \delta(x - al). \tag{IX.78}$$

Now use the formula

$$a \sum_{l=-\infty}^{\infty} \delta(x - al) = \frac{1}{\exp(2\pi i x/a) - 1 + \epsilon} - \frac{1}{\exp(2\pi i x/a) - 1 - \epsilon}, \tag{IX.79}$$

where ϵ is as usual an infinitesimal positive quantity. {The difference in Eq. (IX.79) vanishes unless $x/a \simeq l$ with integer l ; for $x/a = l$ the difference is $[2\pi i(x/a - l) + \epsilon]^{-1} - [2\pi i(x/a - l) - \epsilon]^{-1}$ which is $\delta(x/a - l) = a\delta(x - al)$.}

Hence

$$\begin{aligned}
 I(a) &= \int_{-\infty}^{\infty} dx f(x) \left\{ \frac{1}{\exp(2\pi ix/a) - 1 + \epsilon} \right. \\
 &\quad \left. \frac{1}{\exp(2\pi ix/a) - 1 - \epsilon} \right\} \\
 &= \int_{-\infty}^{\infty} dx f(x) \left\{ \frac{1}{\exp(2\pi ix/a) - 1 + \epsilon} \right. \\
 &\quad \left. \frac{\exp(2\pi ix/a)}{\exp(2\pi ix/a) - 1 - \epsilon} + 1 \right\}. \tag{IX.80}
 \end{aligned}$$

Now one can move the contour of integration to $\text{Im}x = \pm r$ as long as one avoids poles of the integrand. One can obtain (x is real in the formula below)

$$\begin{aligned}
 I(a) &= I + \int_{-\infty}^{\infty} dx \frac{f(x - ir)}{\exp(2\pi r/a) \exp(2\pi ix/a) - 1} \\
 &\quad \frac{f(x + ir)}{1 - \exp(2\pi r/a) \exp(-2\pi ix/a)}. \tag{IX.81}
 \end{aligned}$$

For $a \rightarrow 0$ the integrals each have a factor $\exp(-2\pi r/a)$ (from the denominator); hence

$$|I(a) - I| \sim \exp(-2\pi r/a). \tag{IX.82}$$

This theorem of exponential convergence applies only to infinite integrations. If the integral has finite limits then the contours of integration must always begin and end on the real axis and one cannot obtain the exponential factor near the endpoints of these contours. The contribution of these endpoints is an error of order a , usually.

Applying this result to the sum for I_2 , we have $a = \ln\Lambda$. Then the differences between the sum for I_2 and the corresponding integral behaves as $\exp(-2\pi r/a)$ where r is the width of the strip of analyticity of $ue^{-u}/(1 + e^{-u})^2$, with $u = \bar{\beta}e^x$. The only singularities are due to vanishing of the denominator, which occurs when $u = 2\pi im + \pi i$ for integral m . This means

$$x = \ln(2m + 1)\pi + \frac{\pi i}{2} + 2\pi ip - \ln\bar{\beta}$$

for integral p , which is possible only if $|\text{Im}x| \geq (\pi/2)$. Hence $r = (\pi/2)$ and the error $|I_2(\bar{\beta}, \Lambda) - I_2(\bar{\beta}, 1)|$ is of order $\exp\{-\pi^2/\ln\Lambda\}$. For $\Lambda = 2$ this gives an error of order 6×10^{-7} . For $\Lambda = 2.25$ the error is of order 5×10^{-6} , almost a factor of 10 larger. There are other factors in the error so 6×10^{-7} is not precise, but the factor of 10 improvement between $\Lambda = 2.25$ and $\Lambda = 2$ is evident in the result found earlier for all the quantities $\chi_1, \chi_2, I_3, E, F, G$, and H .

The exponential convergence applies only for infinite sums. Hence it applies to the Kondo calculations only where the sums over l are infinite sums. This excludes any calculations for small N and calculations for kT near the band edge, for which there is a finite maximum value to l which cannot be ignored. But all the susceptibility and specific heat calculations for small T involve infinite sums, and the exponential convergence is evident in the results obtained for $\chi_1, \chi_2, I_3, E, F, G$, and H , as a function of Λ .

The limit for $\Lambda \rightarrow 1$ of all the strong coupling functions can be obtained analytically, just as for I_2 . One obtains

$$\lim(\Lambda \rightarrow 1)\chi_1(\bar{\beta}, \Lambda) = 0.5, \tag{IX.83}$$

$$\lim(\Lambda \rightarrow 1)I(\bar{\beta}, \Lambda) = \int_0^1 \frac{(-\ln z)}{1+z} dz = \pi^2/12. \tag{IX.84}$$

(The latter result comes from the CRC tables of definite integrals.) For the zeroth order susceptibility χ_0 one has

$$\begin{aligned}
 \chi_0(\bar{\beta}, \Lambda) &= -\frac{1}{8} + \sum_{l=-\infty}^{\infty} \left\{ \frac{\exp(-u_l)}{[1 + \exp(-u_l)]^2} \right. \\
 &\quad \left. \frac{\exp(-\Lambda^{1/2}u_l)}{[1 + \exp(-\Lambda^{1/2}u_l)]^2} \right\}, \tag{IX.85}
 \end{aligned}$$

with $u_l = \bar{\beta}_M \Lambda^{l-1}$. For Λ near 1 the sum is very well approximated by an integral

$$\begin{aligned}
 \chi_0(\bar{\beta}, \Lambda) &= -\frac{1}{8} + \frac{1}{\ln\Lambda} \int_{-\infty}^{\infty} dx \\
 &\quad \times \left\{ \frac{\exp(-u)}{[1 + \exp(-u)]^2} - \frac{\exp(-\Lambda^{1/2}u)}{[1 + \exp(-\Lambda^{1/2}u)]^2} \right\} \\
 &\quad + O\left[\frac{1}{\ln\Lambda} \exp(-2\pi^2/\ln\Lambda) \right], \tag{IX.86}
 \end{aligned}$$

where $u = \bar{\beta}e^x$. The error term goes to zero as $\Lambda \rightarrow 1$. For $\Lambda = 1 + \lambda$ with λ small the integral can be expanded in powers of λ . Disregarding terms which vanish as $\lambda \rightarrow 0$, one has

$$\begin{aligned}
 \chi_0(\bar{\beta}, \Lambda) &= -\frac{1}{8} + \frac{1}{\lambda} \int_{-\infty}^{\infty} dx \left(-\frac{\lambda}{2} \right) \frac{d}{du} \\
 &\quad \times \left\{ \frac{\exp(-u)}{[1 + \exp(-u)]^2} \right\} + O(\lambda). \tag{IX.87}
 \end{aligned}$$

But $du = u dx$ so $ud/du = d/dx$;

$$\chi_0(\bar{\beta}, \Lambda) = -\frac{1}{8} + \frac{1}{2} \left. \frac{-\exp(-u)}{[1 + \exp(-u)]^2} \right|_{x=-\infty}^{x=\infty} = 0. \tag{IX.88}$$

A similar analysis shows that the infinite product in the equation for $Z(T)$ is 4 in the limit $\Lambda \rightarrow 1$: one writes

$$\begin{aligned}
 Z_0(\bar{\beta}, \Lambda) &= \prod_{l=-\infty}^{\infty} [1 + \exp(-u_l)]^4 [1 + \exp(-\Lambda^{1/2}u_l)]^{-4} \\
 &= \exp \sum_{l=-\infty}^{\infty} \{ 4 \ln[1 + \exp(-u_l)] \\
 &\quad - 4 \ln[1 + \exp(-\Lambda^{1/2}u_l)] \} \\
 &\simeq \exp \frac{1}{\ln\Lambda} \int_{-\infty}^{\infty} dx \{ 4 \ln[1 + \exp(-u)] \\
 &\quad - 4 \ln[1 + \exp(-\Lambda^{1/2}u)] \} \\
 &= \exp \left\{ -\frac{1}{2} \left[4 \ln[1 + \exp(-u)] \right] \right. \\
 &\quad \left. \right\}_{x=-\infty}^{x=\infty} + O(\lambda) \\
 &= 4 + O(\lambda). \tag{IX.89}
 \end{aligned}$$

TABLE XVIII. Variations in χ_0 and Z_0 , is Λ .

Λ	2	2.25	2.5	3
χ_0	± 0.00002	± 0.00015	± 0.0005	± 0.002
Z_0	$\pm 10^{-7}\%$	$\pm 0.01\%$	$\pm 0.05\%$	$\pm 0.5\%$

The fluctuations in $\chi_0(\bar{\beta}, \Lambda)$ and $Z_0(\bar{\beta}, \Lambda)$ for $2 \leq \Lambda \leq 3$ are shown in Table XVIII (absolute values, not percent, are shown for χ_0).

Now comes the big question: Do the calculations of R and χ_z , which involve much more than sums over l , also show exponential convergence as $\Lambda \rightarrow 1$? The author's suspicion is that they do and this accounts for the remarkable agreement of the results for $\Lambda = 2, 2.25, 2.5$, and 3 . However there are various sources of error in the results for χ_z and R which must now be discussed. The errors to be discussed are numerical errors for a fixed value of Λ . These errors come from the truncation of the Hamiltonian H_N to a finite number of states. (Round-off error seems not to be a problem.) There is also an error in the calculation of $\Psi(z_i)$ [in Eq. (IX.70) for χ_z] because $\Psi(z_i)$ is known only to order z_i from the perturbation calculation. The order z_i^2 term in $\Psi(z_i)$ has been estimated from the computer calculations and was found to make a 0.5% change in χ_z . This correction is included in the numbers reported for χ_z .

There are various error tests one can make on the computer calculations, some of which have already been mentioned in Sec. VIII. They give the following results:

(1) Agreement of the numerical $\bar{J} = -\infty$ fixed point with the single particle eigenvalues: for η_1 , 0.00005% , for η_2 , 0.0002%

(2) Agreement of the leading irrelevant eigenvalues with Λ^{-1} (eigenvalue of T^2): this varies with Λ , being 0.1% , 0.025% , and 0.01% for $\Lambda = 2, 2.25$, and 2.5 , respectively.

(3) Comparison of energy differences with predictions in terms of λ^* and w^* : 0.1% .

(4) Departure of marginal eigenvalue from 1 for weak coupling: 0.003% ($\Lambda = 2$) and 0.0002% ($\Lambda = 2.25$).

(5) Agreement of z_{N+2} with perturbation expansion in z_N to order z_N^4 [cf. Eq. (VIII.44)] (for $z_N \sim 0.04$): 0.001% .

TABLE XIX. Comparison of 1620 state calculation and 526 state calculation for first excited state energy.

526 states		1620 states		Difference
N	Energy	N	Energy	
31	0.514605	87	0.514569	0.000036
33	0.487647	89	0.487619	0.000028
35	0.453728	91	0.453709	0.000019
37	0.410697	93	0.410689	0.000008
39	0.356266	95	0.356270	0.000004
41	0.289271	97	0.289287	-0.000016
43	0.212729	99	0.212754	-0.000025
45	0.137212	101	0.137238	-0.000022
47	0.077129	103	0.077147	-0.000018

(6) Comparison of calculations with 526 states kept with the calculations with 1620 states kept: This is the only way one can check the calculations in the middle of the crossover region where H_N is far from both fixed points. A precise comparison was made for $\Lambda = 2$ and for $\Lambda = 2.25$. For 15 iterations through the crossover region the accumulated error was 1.5% for $\Lambda = 2$ while for $\Lambda = 2.25$ the accumulated error over 15 iterations was 0.2% . It is worth exhibiting the comparison for $\Lambda = 2.25$. The 526 state calculation was for $\bar{J} = -0.055016$ while the 1620 state calculation was for $\bar{J} = -0.02424$. The precise values of \bar{J} were chosen to make iterations 35 and 91 agree.

The comparison of the first excited state over 15 iterations is shown in Table XIX.

Table XIX also illustrates universality, namely the same sequence of energies are obtained for two different initial values of \bar{J} . The errors at first sight seem of order 0.03% at most, but the error that one needs is an error in N , i.e., how much must N be changed to make the two series agree exactly. For example, compare $N = 31$ (526 states) with $N = 87$ (1620 states). The difference is 0.000036 . But the difference between $N = 31$ and $N = 33$ is only 0.027 , so an energy difference 0.000036 corresponds to a change in N of $2 \times (0.000036/0.027) \simeq 0.0027$. Now in the calculation of χ_z , the difference $N_f - N_i$ is multiplied by $\ln(2.25)/2 \simeq 0.4$ so an error of 0.0027 in N_f or N_i means an error of 0.1% in χ_z .

The principal sources of error seem to be (1) errors in the estimation of the z^2 term in Ψ , and (2) accumulated error due to departure of the leading irrelevant eigenvalue from Λ^{-1} in the strong coupling region. This error accumulates from the end of the crossover region to the final value N_f of N ; it is estimated to be ~ 10 times the error in the eigenvalue, namely 1% ($\Lambda = 2$), 0.25% ($\Lambda = 2.25$), and 0.1% ($\Lambda = 2.5$). The accumulated error in the weak coupling region is negligible due to the very small error in each iteration (error #4). These three errors affect only χ_z ; the ratio R is presumably much more accurately determined.

The above discussion is for fixed values of Λ . Now the errors involved in extrapolating to $\Lambda = 1$ must be considered. For the perturbation theory sums ($\chi_1, \chi_2, I_3, E, F, G$, and H) the errors for $\Lambda = 2.25$ and $\Lambda = 2$ are negligible as noted earlier due to exponential convergence. For $\Lambda = 2.5$ the errors are 1% in the worst case (the worst case is χ_1). Even for $\Lambda = 3$ there is at most a 4% error. If there is exponential convergence for χ_z and R , then one would also expect their values for $\Lambda = 2.25$ and 2 to be very close to the $\Lambda = 1$ limit. The results for R seem to confirm this; there is only a 0.1% change in R from $\Lambda = 2$ to $\Lambda = 2.5$. The 0.1% change from $\Lambda = 2.25$ to $\Lambda = 2$ is easily accounted for by numerical errors in the $\Lambda = 2$ calculation, so the intrinsic difference between $\Lambda = 2$ and $\Lambda = 2.25$ could be smaller than 0.1% . The most plausible conclusion is that R is 2 exactly for $\Lambda = 1$ and already very close to 2 for $\Lambda = 2.25$. It may in fact be 2 for all Λ ; this possibility has to be investigated. For χ_z the results for $\Lambda = 2, 2.25, 2.5$, and 3 vary by 0.8% ; this variation is easily accounted for by numerical errors and if the convergence to the $\Lambda = 1$ limit is exponential then the result 0.10315 for $\Lambda = 2.25$ is, one would guess, within 0.5% of the $\Lambda = 1$ limit. If the convergence is not

exponential then there could be extrapolation errors to $\Lambda = 1$ several times the variation from $\Lambda = 2$ to $\Lambda = 2.5$, or an extrapolation error of order 2% say. There is not much hope of repeating the calculation with $\Lambda < 2$ because the truncation errors of the numerical calculation increase rapidly as Λ decreases (in the crossover region they increased by a factor 7 from $\Lambda = 2.25$ to $\Lambda = 2$).

There is no experimental reason to seek better than 5% to 10% accuracy and for this purpose the simpler $\Lambda = 3$ calculation is good enough. Thus much of the future work on the Kondo problem could be done with $\Lambda = 3$. In this case one can get good results keeping only ~ 300 states. The purpose of the high accuracy calculation reported here was only to make sure there are no subtle difficulties in the renormalization group approach to the Kondo problem, difficulties that might be overlooked in a rough calculation. No such difficulties arose.

How universal are the results for $\chi(T)$ and $C(T)$? First, it is useful to collect the results:

(1) Large T ($D \gg kT \gg kT_K$ where D is the conduction band width and T_K the Kondo temperature): [the complete formula is Eq. (IX.57)]

$$kT\chi(T) = 0.25\{1 + 2J\rho + 4J^2\rho^2 \ln(kT/\bar{D}) + O(J^3\rho^3)\} \quad (\text{IX.90})$$

where J is the impurity coupling and ρ the density of states at the Fermi surface, in the notation of [Kondo (1969)]. The variable \bar{D} is proportional to D , depends on $J\rho$, and is defined so that there are no constant (non-logarithmic) terms in this expansion beyond the $J\rho$ term, i.e., no $cJ^2\rho^2$ term with constant c . Terms of order kT , $(kT)^2$, etc. for small T are neglected. The usual factors, namely the impurity moment μ and g factor, have been set equal to 1; the conduction band electrons have been given the same g factor as the impurity.

Note that some authors use $-J\rho$ for $2J\rho$ or other more obscure notations, but one can correct for this if one can compare the above formula for the high temperature susceptibility with the analogous formula in the obscure notation. Not all authors give the conduction band electrons the same g factor as the impurity, which is a complication not considered here [see Kondo (1969)]. The formulae given here are valid *only* for the case that the g factors are the same.

(2) Small T ($T \ll T_K$)

$$\chi(T) = \frac{(0.1032 \pm 0.0005)}{\bar{D}(J\rho)} \times \exp\left\{\frac{1}{|2J\rho|} - \frac{1}{2} \ln|2J\rho| - 1.5824|2J\rho| + O(J^2\rho^2)\right\} \quad (\text{IX.91})$$

$$\frac{C(T)}{T\chi(T)} = \frac{2\pi^2}{3} k^2. \quad (\text{IX.92})$$

(3) The Kondo temperature T_K . We agree with previous work that the Kondo temperature is, in order of magnitude,

$$T_K = D|2J\rho|^{+1/2} \exp\{-1/|2J\rho|\}. \quad (\text{IX.93})$$

The basic use of the Kondo temperature is in a scaling formula, e.g., $kT\chi(T)$ is a function only of T/T_K [Yuval and Anderson (1970); Anderson, Yuval, and Hamann (1970a, 1970b)]. This follows, e.g., from Eq. (IX.56), which also gives

$$T_K = \bar{D}(J\rho) \exp\{-\Phi(J\rho)\} = \bar{D}(J\rho) \times \exp\left\{\frac{-1}{|2J\rho|} + \frac{1}{2} \ln|2J\rho| + 1.5824|2J\rho| + O(J^2\rho^2)\right\}. \quad (\text{IX.94})$$

More generally one can define T_K to be $c\bar{D} \exp\{-\Phi(J\rho)\}$, where c is an arbitrary constant independent of $J\rho$, and $kT\chi(T)$ will still depend only on T/T_K . Since $\bar{D}(J\rho)$ has a power series expansion in $J\rho$, it cannot be confused with the first two terms in the exponential, but the third ($J\rho$) term in the exponential is not very useful unless $\bar{D}(J\rho)$ is known to order $J\rho$. To determine this one must compute the expansion of $\chi(T)$ in powers of $J\rho$ to order $(J\rho)^3$ for the particular model one is interested in. In any case the order of magnitude of (IX.94) agrees with (IX.93) apart from a numerical constant (namely the ratio \bar{D}/D for $J\rho = 0$ which can be determined if the expansion of $\chi(T)$ is known to order $J^2\rho^2$).

(4) The susceptibility near the Kondo temperature (see later).

The essential result giving universality and the scaling law (namely that $kT\chi(T)$ depends only on T/T_K) is that there is a unique trajectory of effective Hamiltonians H_N , parametrized by z_N . For any small initial value of \bar{J} , the H_N for large N lie on this trajectory. The temperature dependence of $\chi(T)$ and $C(T)$ have all been determined from properties of this trajectory and the recursion formula for H_{N+1} which connects successive members of the trajectory. Thus any initial Hamiltonian which leads to effective Hamiltonians on this trajectory will give the same behavior for the susceptibility and specific heat, for small T , including $T \sim T_K$.

There are some changes in the initial Hamiltonian which necessarily are universal, i.e., still lead to the same family of effective Hamiltonians. As was shown in Sec. VIII, the operators $f_0^+f_1 + f_1^+f_0$, $f_1^+f_2 + f_2^+f_1$, etc. are all irrelevant operators; adding these terms with small coefficients to the initial Hamiltonian will not change the H_N for large N . This means the replacement of ϵ_n by 1 in Sec. VII should have no effect on the results reported here. This has not been checked explicitly. This is true only for small \bar{J} ; for large \bar{J} the irrelevant variables generated by the ϵ_n are important. This does *not* mean one can add infinite sums such as $\sum \Lambda^{-n/2}(f_n^+f_{n+1} + f_{n+1}^+f_n)$ to the Hamiltonian; adding such infinite sums changes the recursion formula satisfied by H_N for large N , and any such changes spoil universality.

There are some other changes in the initial Hamiltonian which are obviously nonuniversal. These are of three types. First there are changes which change the recursion formula itself; in particular adding d -wave electrons from the conduction band coupled to the impurity is nonuniversal. Here d -wave means in a d -wave state about the impurity. Then there are 10 more operators f_N for each N due to the d -wave electrons and these appear in the recursion formula.

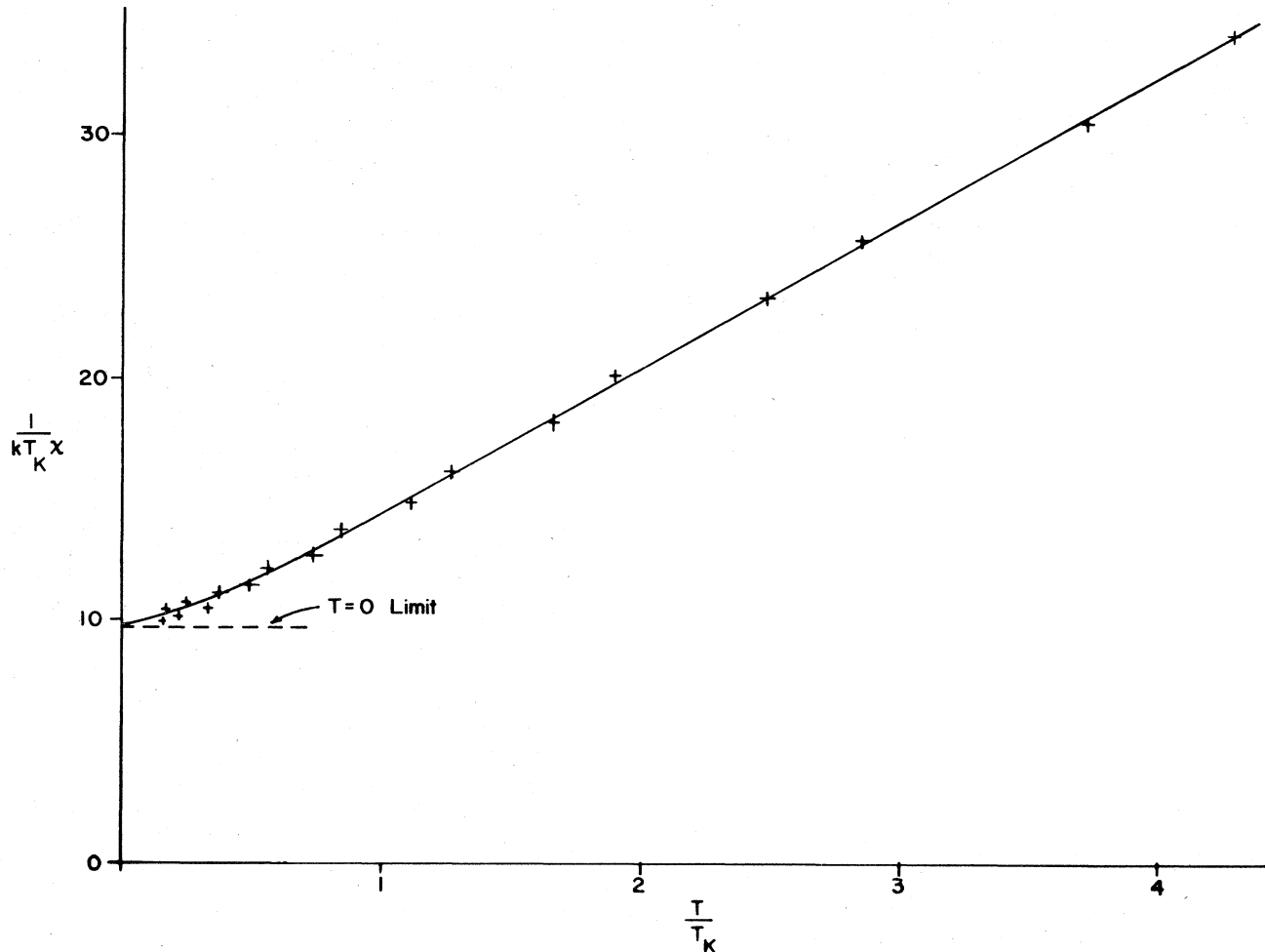


FIG. 17. Plot of inverse susceptibility $\chi^{-1}(T)$ vs T from the computer calculations. The magnetic moment and g factor of both the impurity and the conduction band electrons are set equal to 1. The plot actually shows $(kT_K\chi)^{-1}$ vs T/T_K , where T_K was defined in Eq. (IX.94). The crosses represent results for two different calculations (both with $\Lambda = 2.25$, but $\tilde{J} = 0.024$ for one and 0.02412 for the other); the scatter at small T is due to truncation errors. The zero temperature value is taken from Eq. (IX.69) and Table XVI.

Secondly, there are changes which change the $\tilde{J} = 0$ fixed point. In particular, changing the spin of the impurity, from spin $\frac{1}{2}$ to spin 1 or spin 2, for example, changes the $\tilde{J} = 0$ fixed point (by increasing the degeneracy of each energy level). Thirdly, a change of the Hamiltonian which adds the marginal eigenoperator $f_0^\dagger f_0$ violates universality. This term violates particle-hole symmetry. But particle-hole symmetry has no experimental basis. In particular, any change of the Fermi energy will generate this marginal term.

The formulae given here for $\chi(T)$ are both independent of a scale change in energy. A scale change in energy is equivalent to a scale change in temperature. In the high temperature formula, a scale change in temperature is compensated for by the same scale change in \tilde{D} . This scale change in \tilde{D} then gives the correct low temperature formula for $\chi(T)$. In the formula for $C/T\chi$, a change $H \rightarrow sH$ leads to a change, namely to $C/T\chi = s^2 t^2 (2\pi^2)/3$.

There are modifications to the Hamiltonian which might be universal but must be studied further. One example

would be macroscopic changes in the model near the band edge, for example modification of the single electron energy from the $k - k_F$ law assumed here. These changes must be studied further because the universality guaranteed by renormalization group is only for *infinitesimal* modifications of the initial Hamiltonian.

Anderson and Yuval (1970a,b), Hamann (1970), and Schotte (1971) have developed an analogy between the solution of the Kondo Hamiltonian and the partition function of a one dimensional Coulomb gas. The analogy is not exact; a "long-time" approximation is used. This approximation preserves the leading logarithmic behavior of the Kondo Hamiltonian. However it appears that the full universal properties of the Kondo Hamiltonian are *not* preserved by this analogy. In particular the susceptibility formula given by Schotte and Schotte (1971) disagrees with the universal formula (IX.57) in order $(J\rho)^3 \ln kT/\tilde{D}$. The calculation which shows this will be outlined below. The formula of Schotte and Schotte through order $(J\rho)^3$ is $(J\tau$ in the Schotte and Schotte paper is $-2J\rho$ in the Kondo notation and τ is D^{-1} in the Kondo notation; this transcrip-

tion was used in obtaining the formula below)

$$kT\chi(T) = (1 + J\rho)^2 \left\{ 0.25 + \frac{(J\rho)^2}{\beta^2} \int_0^\beta d\tau_2 \int_0^{\tau_2} d\tau_1 \right. \\ \left. \times [(\tau_2 - \tau_1 - \beta/2)^2 - \beta^2/4] \right. \\ \left. \times \left\{ \frac{\sin[\pi(\tau_2 - \tau_1)/\beta]}{\sin(\pi/\beta D)} \right\}^{-(2+2J\rho)} \right\} \quad (\text{IX.95})$$

To avoid a divergent integral, the ratio of sines is replaced by 1 if either $\tau_2 - \tau_1 < D^{-1}$ or $\tau_2 - \tau_1 > \beta - D^{-1}$. This formula has been evaluated in the limit of large $\beta = 1/kT$ (partly by numerical methods); the result is, to order $(J\rho)^3$,

$$kT\chi(T) = 0.25(1 + J\rho)^2 \{ 1 + 4J^2\rho^2 \ln(kT/D) \\ + 1.3515(J\rho)^2 + 8J^3\rho^3 \ln^2(kT/D) + J^3\rho^3 c_3 \\ = 0.25\{ 1 + 2J\rho + 4J^2\rho^2 \ln(kT/D) + 2.3515(J\rho)^2 \\ + 8J^3\rho^3 \ln^2(kT/D) + 8J^3\rho^3 \ln(kT/D) \\ + J^3\rho^3(c_3 + 2.703) \}, \quad (\text{IX.96})$$

where c_3 has not been evaluated. One must now define D to eliminate all constant terms. The result is

$$\tilde{D}(J\rho) = 0.5555D\{1 + c_4 J\rho\}, \quad (\text{IX.97})$$

where c_4 is a constant. In terms of \tilde{D}

$$kT\chi(T) = 0.25\{ 1 + 2J\rho + 4J^2\rho^2 \ln(kT/\tilde{D}) \\ + 8J^3\rho^3 \ln^2(kT/\tilde{D}) - 1.406(J\rho)^3 \ln(kT/\tilde{D}) \}. \quad (\text{IX.98})$$

The coefficient -1.406 multiplying $(J\rho)^3 \ln(kT/\tilde{D})$ disagrees with the universal result $+4$ of Eq. (IX.57).

Schotte and Schotte (1971) have calculated the susceptibility $\chi(T)$ by Monte Carlo methods for the Coulomb analog. The violation of universality means that the value of T_K they calculate is unreliable because the $J\rho$ dependence they obtain does not have the $(J\rho)^{+1/2}$ contained in the exact result. [It can be shown that a change in the $(J\rho)^3 \ln(kT/D)$ term amounts to a change in the coefficient of $\ln|2J\rho|$ in $\Phi(J\rho)$.] If one allows an adjustable scale factor in the temperature scale of Schotte and Schotte then their calculations can be brought into agreement with numerical calculations of the susceptibility reported below, to within $\sim 20\%$ except for $T \ll T_K$ where the Schotte and Schotte calculations are unreliable.

The susceptibility through the Kondo temperature was calculated by using second order perturbation theory in the perturbation H_I of Eq. (IX.4), as part of the computer calculations. The results have an absolute accuracy of about 0.001 in $kT\chi(T)$, according to comparison of numerical results with weak and strong coupling results. Results for $\Lambda = 2, 2.25, 2.5$, and 3 agree to about this accuracy. The results are shown in Fig. 17. The results fit well to a Curie-Weiss form near the Kondo temperature, in agreement with the earlier results of Schotte and Schotte (1971) and Gotze and Schlottmann (1973):

$$\chi(T) = \frac{1}{k} \left(\frac{0.17}{T + 2T_K} \right) \quad (0.5T_K < T < 16T_K). \quad (\text{IX.99})$$

For $T/T_K > 16$ the high temperature formula

$$\Phi[2kT\chi(T) - 0.5] = \ln(T/T_K) \quad (\text{IX.100})$$

(inverted numerically to determine $\chi(T)$ from T/T_K) gives $\chi(T)$ to better than 1%.

In conclusion, the renormalization group approach described in Sec. VII-IX appears clearly superior to all previous methods of solving the Kondo Hamiltonian. The principal approximation—the discretization—does not spoil perturbation theory even in third order so long as $\Lambda < 2.5$ or so; the numerical calculations agree excellently with perturbation theory when perturbation theory is valid. Nothing qualitative is thrown out with the discretization, and the comparison of results for $\Lambda = 2$ to 3 suggests that one is very close to the continuum limit over this whole range of Λ .

Future renormalization group studies are likely to include (1) the effects of potential scattering, (2) studies of the Anderson model (the impurity spin is replaced by an extra electron state), (3) studies of dynamic properties (resistivity, etc.), (4) attempts to include d -wave coupling to the impurity, and (5) study of impurity-impurity interactions. There are good prospects for solving all these problems, especially if one learns to work with large values of Λ so that only a few states need be kept in the numerical calculations.

APPENDIX: PERTURBATION EXPANSIONS FOR THE KONDO HAMILTONIAN

In this appendix the weak coupling expansions used in Sec. VIII and IX will be reported in more detail. No derivations will be given; only some further results are quoted here. The expansions discussed here are the expansion of z_{N+2} in powers of z_N [Eq. (VIII.44)] to order z_N^4 and the expansion of $\chi(T)$ in powers of z_N [Eq. (IX.38)] to order z_N^3 . The formulae which determine these expansions will be given below, for arbitrary Λ , followed by some tables of results.

The first step in constructing the expansions is to diagonalize the free electron Hamiltonian H_N of Eq. (VIII.1) (for $\tilde{J} = 0$) obtaining in particular the single electron energies η_l and the expansion coefficients α_l of Eq. (IX.18). These were computed only for N odd and sufficiently large (e.g., $N \approx 35$ for $\Lambda = 2$). In order to calculate z_{N+2} one must also diagonalize H_{N+2} , giving energies η'_l and expansion coefficients α'_l . The quantities η_l and α_l exist for $1 \leq l \leq (N+1)/2$; the quantities η'_l and α'_l exist for $1 \leq l \leq (N+3)/2$. In order to compute $\chi(T)$ one must also know the corresponding η''_l and α''_l for the Hamiltonian

$$H = \Lambda^{(N-1)/2} \sum_{n=0}^{\infty} \Lambda^{-n/2} (f_n + f_{n+1} + f_{n+1} + f_n). \quad (\text{A1})$$

Here the Hamiltonian contains an infinite sum but is still rescaled by the finite factor $\Lambda^{(N-1)/2}$. The only reason for this factor is to make thermodynamic calculations involve $\tilde{\beta} = \Lambda^{-(N-1)/2}/(kT)$ in a natural way. In practice one uses asymptotic properties of the solution of H_N for large N to determine the values of η''_l and α''_l . It is convenient to let the label l range from $-\infty$ to $(N+1)/2$, with the eigenvalues η''_l arranged to increase with l . Then there

are two asymptotic results of interest. The first is that

$$\eta_l'' = \eta_l \quad (l \text{ large, } N \text{ large}) \tag{A2}$$

$$\alpha_l'' = \alpha_l \quad (l \text{ large, } N \text{ large}) \tag{A3}$$

in the limit of large N and large l .

For example, this result can be used to determine η_l'' and α_l'' for large N and $l \geq (N + 1)/4$. The second result [see Eqs. (VIII.21) and (IX.20)] is

$$\eta_l'' = \eta_m'' \Lambda^{(l-m)} \quad (N \text{ large, } l \ll N, m \ll N) \tag{A4}$$

$$\alpha_l'' = \alpha_m'' \Lambda^{(l-m)/2} \quad (N \text{ large, } l \ll N, m \ll N) \tag{A5}$$

for large N , $l \ll N$, and $m \ll N$ (including the cases that l and/or m are negative). These formulae can be used to determine η_l'' and α_l'' for any $l < (N + 1)/4$, given results for η_l'' and α_l'' for $l = (N + 1)/4$ (or the next integer above $(N + 1)/4$).

The next step is to construct expansions in \tilde{J} for z_N , z_{N+2} , and $\chi(T)$. The formulae for z_N are as follows:

$$z_N = R_1 \tilde{J} + R_2 \tilde{J}^2 + R_3 \tilde{J}^3 + R_4 \tilde{J}^4 \tag{A6}$$

with

$$R_1 = -3\alpha_1^2/2; \tag{A7}$$

$$R_2 = 0.75\alpha_1^2 \sum_{i=1}^L \alpha_i^2 \left\{ \frac{1}{\eta_1 + \eta_i} + \frac{3(1 - \delta_{i1})}{\eta_i - \eta_1} \right\}, \tag{A8}$$

TABLE A.1. Coefficients in Eq. (A.10).

i	C_i	D_i
1	192	$(\eta_m + \eta_n)^2(\eta_m + \eta_i)$
2	30	$(\eta_1 + \eta_m)(\eta_m + \eta_n)(\eta_n + \eta_i)$
3	-18	$(\eta_1 - \eta_m)(\eta_m + \eta_n)(\eta_n + \eta_i)$
4	27	$(\eta_1 + \eta_m)(\eta_m + \eta_n)(\eta_m + \eta_i)$
5	99	$(\eta_1 - \eta_m)(\eta_m + \eta_n)(\eta_m + \eta_i)$
6	-42	$(\eta_1 + \eta_m)(\eta_1 + \eta_n)(\eta_m + \eta_i)$
7	-54	$(\eta_1 - \eta_m)(\eta_1 - \eta_n)(\eta_m + \eta_i)$
8	-39	$(\eta_1 + \eta_m)(\eta_1 + \eta_n)(\eta_1 + \eta_i)$
9	81	$(\eta_1 - \eta_m)(\eta_1 - \eta_n)(\eta_1 - \eta_i)$
10	3	$(\eta_1 + \eta_m + \eta_n + \eta_i)(\eta_m + \eta_n)(\eta_m + \eta_i)$
11	3	$(\eta_1 - \eta_m - \eta_n - \eta_i)(\eta_m + \eta_n)(\eta_m + \eta_i)$
12	30	$(\eta_1 + \eta_m + \eta_n + \eta_i)(\eta_m + \eta_n)^2$
13	6	$(\eta_1 - \eta_m - \eta_n - \eta_i)(\eta_m + \eta_n)^2$
14	12	$(\eta_1 + \eta_m + \eta_n + \eta_i)(\eta_1 + \eta_m)(\eta_n + \eta_i)$
15	36	$(\eta_1 - \eta_m - \eta_n - \eta_i)(\eta_1 - \eta_m)(\eta_n + \eta_i)$
16	30	$(\eta_1 + \eta_m + \eta_n + \eta_i)(\eta_1 + \eta_m)(\eta_m + \eta_n)$
17	18	$(\eta_1 - \eta_m - \eta_n - \eta_i)(\eta_1 - \eta_m)(\eta_m + \eta_n)$
18	-18	$(\eta_1 + \eta_m + \eta_n + \eta_i)(\eta_1 + \eta_m)^2$
19	54	$(\eta_1 - \eta_m - \eta_n - \eta_i)(\eta_1 - \eta_m)^2$
20	-21	$(\eta_1 + \eta_m + \eta_n + \eta_i)(\eta_1 + \eta_m)(\eta_1 + \eta_n)$
21	27	$(\eta_1 - \eta_m - \eta_n - \eta_i)(\eta_1 - \eta_m)(\eta_1 - \eta_n)$
22	18	$(\eta_1 + \eta_m)(\eta_n + \eta_i)^2$
23	-54	$(\eta_1 - \eta_m)(\eta_n + \eta_i)^2$
24	18	$(\eta_1 + \eta_m)^2(\eta_n + \eta_i)$
25	54	$(\eta_1 - \eta_m)^2(\eta_n + \eta_i)$
26	-90	$(\eta_m + \eta_n)^3$
27	-18	$(\eta_1 + \eta_m)(\eta_m + \eta_n)^2$
28	-54	$(\eta_1 - \eta_m)(\eta_m + \eta_n)^2$
29	-18	$(\eta_1 + \eta_m)^2(\eta_m + \eta_n)$
30	54	$(\eta_1 - \eta_m)^2(\eta_m + \eta_n)$
31	99	$(\eta_1 + \eta_m)^2(\eta_1 + \eta_n)$
32	-243	$(\eta_1 - \eta_m)^2(\eta_1 + \eta_n)$
33	-27	$(\eta_1 + \eta_m)^2(\eta_1 - \eta_n)$
34	27	$(\eta_1 - \eta_m)^2(\eta_1 - \eta_n)$
35	-27	$(\eta_1 + \eta_m)^3$
36	81	$(\eta_1 - \eta_m)^3$

where $L = (N + 1)/2$; and

$$R_3 = -0.375\alpha_1^2 \sum_{m=1}^L \sum_{l=1}^L \alpha_m^2 \alpha_l^2 \times \left\{ \frac{8}{(\eta_m + \eta_l)^2} - \frac{5}{(\eta_m + \eta_l)(\eta_l + \eta_1)} + \frac{2}{(\eta_1 + \eta_m)(\eta_l + \eta_m)} - \frac{6(1 - \delta_{l1})}{(\eta_l - \eta_1)(\eta_l + \eta_m)} - \frac{9(1 - \delta_{l1})(1 - \delta_{m1})}{(\eta_l - \eta_1)(\eta_m - \eta_1)} \right\} - 0.375\alpha_1^2 \sum_{l=1}^L \left\{ \frac{3}{(\eta_1 + \eta_l)^2} + \frac{9(1 - \delta_{l1})}{(\eta_1 - \eta_l)^2} \right\}. \tag{A9}$$

R_4 has the following form:

$$R_4 = \frac{1}{16}\alpha_1^2 \sum_{l=1}^L \sum_{m=1}^L \sum_{n=1}^L \alpha_l^2 \alpha_m^2 \alpha_n^2 \left\{ \sum_{i=1}^{25} \frac{C_i}{D_i} \right\} + \frac{1}{16}\alpha_1^4 \sum_{m=1}^L \sum_{n=1}^L \alpha_m^2 \alpha_n^2 \sum_{i=26}^{34} C_i/D_i + \frac{1}{16}\alpha_1^6 \sum_{m=1}^L \alpha_m^2 \sum_{i=35}^{36} C_i/D_i, \tag{A10}$$

where the numerators and denominators C_i and D_i are listed in Table A.I

Terms with $D_i = 0$ are to be omitted from the sum in Eq. (A.10) (e.g., the $m = 1$ term is omitted when D_i contains $\eta_1 - \eta_m$).

The expansion of z_{N+2} in powers of \tilde{J} is obtained by replacing α by α' , η by η' , and L by $L + 1$ throughout the above formulae.

The expansion of $\chi(T)$ has the form

$$\chi(T) = \frac{1}{4} + Y_1(\tilde{\beta})\tilde{J} + Y_2(\tilde{\beta})\tilde{J}^2 + Y_3(\tilde{\beta})\tilde{J}^3 \tag{A11}$$

with

$$Y_1(\tilde{\beta}) = +\tilde{\beta} \sum_{l=-\infty}^L (\alpha_l'')^2 \exp(-\tilde{\beta}\eta_l'') [1 + \exp(-\tilde{\beta}\eta_l'')]^{-2} \tag{A12}$$

$$Y_2(\tilde{\beta}) = \frac{1}{8}(\tilde{\beta})^2 \sum_{n=-\infty}^L \sum_{m=-\infty}^L [1 + (\eta_m'' \rightarrow -\eta_m'')] \times (\alpha_m'')^2 (\alpha_n'')^2 h(\tilde{\beta}\eta_m'') h(\tilde{\beta}\eta_n'') \{ (3[1 - \exp(-\tilde{\beta}\eta_m'')]^2 \times h^2(\tilde{\beta}\eta_m'') + [1 + \exp(-\tilde{\beta}\eta_m'')] \exp(-\tilde{\beta}\eta_n'')] \times h(\tilde{\beta}\eta_m'') h(\tilde{\beta}\eta_n'') - 4) G(\tilde{\beta}\eta_m'' + \tilde{\beta}\eta_n'') + 4 \exp(-\tilde{\beta}\eta_n'') \exp(-\tilde{\beta}\eta_m'') h(\tilde{\beta}\eta_m'') h(\tilde{\beta}\eta_n'') \}, \tag{A13}$$

where $[1 + (\eta_m'' \rightarrow -\eta_m'')]$ means that for each m there are two terms, the first calculated as shown, the second calculated by replacing η_m'' by $-\eta_m''$ wherever η_m'' appears. The auxiliary functions G and h are

$$G(x) = (1 - e^{-x})/x, \tag{A14}$$

$$h(x) = 1/(1 + e^{-x}). \tag{A15}$$

Finally

$$\begin{aligned}
 Y_3(\bar{\beta}) &= \frac{1}{96}(\bar{\beta})^3 \sum_{l=-\infty}^L [1 + (\eta_l'' \rightarrow -\eta_l'')] \\
 &\times \sum_{m=-\infty}^L [1 + (\eta_m'' \rightarrow -\eta_m'')] \\
 &\times \sum_{n=-\infty}^L [1 + (\eta_n'' \rightarrow -\eta_n'')] (\alpha_l'')^2 (\alpha_m'')^2 (\alpha_n'')^2 \\
 &\times h(\bar{\beta}\eta_l'') h(\bar{\beta}\eta_m'') h(\bar{\beta}\eta_n'') \{ [-32 \exp(-\bar{\beta}\eta_n'') \\
 &\times \exp(-\bar{\beta}\eta_m'') h(\bar{\beta}\eta_n'') h(\bar{\beta}\eta_m'') - 24 \exp(-2\bar{\beta}\eta_n'') \\
 &\times h^2(\bar{\beta}\eta_n'') + 16 \exp(-\bar{\beta}\eta_m'') h(\bar{\beta}\eta_n'') h(\bar{\beta}\eta_m'') \\
 &+ 72 \exp(-\bar{\beta}\eta_n'') h^2(\bar{\beta}\eta_n'') - 48 h^2(\bar{\beta}\eta_n'') + 56 \\
 &\times \exp(-\bar{\beta}\eta_n'') h(\bar{\beta}\eta_n'') + 16 h(\bar{\beta}\eta_n'')] K(\bar{\beta}\eta_n'' + \bar{\beta}\eta_m'', \\
 &\bar{\beta}\eta_n'' + \bar{\beta}\eta_l'') + \exp(-\bar{\beta}\eta_l'') h(\bar{\beta}\eta_l'') [-64 \exp(-\bar{\beta}\eta_n'') \\
 &\times h(\bar{\beta}\eta_n'') + 32 h(\bar{\beta}\eta_n'') + 40] K(\bar{\beta}\eta_n'' + \bar{\beta}\eta_m'', 0) \\
 &- 36 \exp(-\bar{\beta}\eta_l'') h(\bar{\beta}\eta_l'') G(\bar{\beta}\eta_l'' + \bar{\beta}\eta_m'') \}, \quad (A16)
 \end{aligned}$$

where

$$K(x,y) = \frac{1}{xy} + \frac{e^{-x}}{x(x-y)} + \frac{e^{-y}}{y(y-x)}. \quad (A17)$$

For the limit $y \rightarrow 0$, one gets

$$K(x,0) = (e^{-x} - 1 + x)/x^2. \quad (A18)$$

To obtain expansions in powers of z_N one must invert Eq. (A6) to give \bar{J} in powers of z_N . The result is

$$\begin{aligned}
 \bar{J} &= (R_1^{-1})z_N - (R_2/R_1^3)z_N^2 + [(2R_2^2 - R_1R_3)/R_1^5]z_N^3 \\
 &+ [(-5R_2^3 + 5R_1R_2R_3 - R_1^2R_4)/R_1^7]z_N^4. \quad (A19)
 \end{aligned}$$

It is now straightforward to obtain z_{N+2} and $\chi(T)$ as expansions in z_N , once the expansions in \bar{J} for z_{N+2} and $\chi(T)$ have been calculated.

The coefficients of the \bar{J} expansion were calculated on a computer. The computer program was not completely straightforward. The worst problem was calculating $K(x,y)$ for x small, or y small, or $x - y$ small; Eq. (A17) used directly in the computer gives large round-off errors and alternative formulae had to be obtained.

Some results of the calculations are shown in Table A.II.

TABLE A.II Numerical results for perturbation expansions.

$z_{N+2} = z_N + Az_N^2 + Bz_N^3 + Cz_N^4,$	(A20)				
$\chi(T) = \frac{1}{4} + F(\bar{\beta})z_N^2 + G(\bar{\beta})z_N^3 + H(\bar{\beta})z_N^4,$	(A21)				
$\Lambda =$	2	2.25	2.5	3	2.5
A	0.71436	0.77059	0.81931	0.89773	
B	0.14220	0.22768	0.30497	0.43913	
C	-2.9214	-2.6467	-2.4463	-2.1338	
$\bar{\beta}$	0.70711	0.5	0.46	0.57	0.58
F	-0.25766	-0.23763	-0.22375	-0.20384	-0.22333
G	0.31472	0.32527	0.29055	0.18306	0.24322
H	-1.1300	-1.0536	-0.88256	-0.51805	-0.74841

REFERENCES

- Abrikosov, A. A., and A. B. Migdal, 1970, *J. Low Temp. Phys.* **3**, 519.
 Anderson, P. W., 1961, *Phys. Rev.* **124**, 41.
 Anderson, P. W., 1970, *J. Phys. C* **3**, 2346.
 Anderson, P. W., G. Yuval, and D. R. Hamann, 1970a, *Phys. Rev. B* **1**, 4664.
 Anderson, P. W., G. Yuval, and D. R. Hamann, 1970b, *Solid State Commun.* **8**, 1033.
 Balian, R., and G. Toulouse, 1974, *Ann. of Phys. (N. Y.)* **83**, 28.
 Bell, T. L., and K. G. Wilson, 1974, *Phys. Rev. B* **10**, 3935.
 Bell, T. L., and K. G. Wilson, 1975, *Phys. Rev. B* **11**, 3431.
 Bogoliubov, N. N., and D. V. Shirkov, 1959, *Introduction to the Theory of Quantized Fields* (Interscience, New York), Ch. VIII.
 Boyce, J. B., and C. P. Slichter, 1974, *Phys. Rev. Lett.* **32**, 61.
 Brézin, E., J. C. Le Guillou, J. Zinn-Justin, and B. G. Nickel, 1973, *Phys. Lett.* **44A**, 227.
 Callan, C. G., 1970, *Phys. Rev. D* **2**, 1541.
 Chang, T. S., and H. E. Stanley, 1973, *Phys. Rev. B* **8**, 4435.
 —, 1962, *Chemical Rubber Company Handbook of Mathematics*, edited by C. D. Hodgman (Chemical Rubber Publishing Co., Cleveland), p. 314, Formula 439.
 Domb, C., and M. S. Green, *Phase Transitions and Critical Phenomena* (Academic, London) Vol. 6, in press.
 Dyaloshinski, I. E., and A. I. Larkin, 1971, *Zh. Eksp. Teor. Fiz.* **61**, 791 [English Translation: 1972, *Sov. Phys.-JETP* **34**, 422].
 Fisher, M. E., 1974, *Rev. Mod. Phys.* **46**, 597.
 Fisher, M. E., and M. N. Barber, 1972, *Arc. Ration. Mech. Anal.* **43**, 205.
 Fowler, M., and A. Zawadowski, 1971, *Solid State Commun.* **9**, 471.
 Gell-Mann, M., and F. E. Low, 1954, *Phys. Rev.* **95**, 1300.
 Gotze, W., and P. Schlottman, 1973, *Solid State Commun.* **13**, 861.
 Gross, D., and F. Wilczek, 1973a, *Phys. Rev. Lett.* **30**, 1343.
 Gross, D., and F. Wilczek, 1973b, *Phys. Rev. D* **8**, 3633.
 Gross, D., and F. Wilczek, 1974, *Phys. Rev. D* **9**, 980.
 Grüner, G., 1974, *Adv. Phys.* **23**, 941.
 Grüner, G., and A. Zawadowski, 1974, *Repts. Prog. Phys.* **37**, 1497.
 Hamann, D., 1970, *Phys. Rev. B* **2**, 1373.
 Hsu, S., Th. Niemeier, and J. D. Gunton, 1975, *Phys. Rev. B* **11**, 2699.
 Kadanoff, L., 1975, *Phys. Rev. Lett.* **34**, 1005.
 Kadanoff, L., and A. Houghton, 1975, *Phys. Rev. B* **11**, 377.
 Kogut, J., and L. Susskind, 1974, *Phys. Rev.* **9**, 697.
 Kondo, J., 1969, in *Solid State Physics*, edited by F. Seitz, D. Turnbull, and H. Ehrenreich (Academic, New York), Vol. 23, p. 183.
 Krinsky, S., and D. Furman, 1974, *Phys. Rev. Lett.* **32**, 731.
 Menyhard, N., and J. Solyom, 1973, *J. Low Temp. Phys.* **12**, 529.
 Nauenberg, M., and B. Nienhuis, 1974a, *Phys. Rev. Lett.* **33**, 944.
 Nauenberg, M., and B. Nienhuis, 1974b, *Phys. Rev. Lett.* **33**, 1598.
 Nauenberg, M., and B. Nienhuis, 1975, *Phys. Rev. B* **11**, 4152.
 Nelson, D., and M. E. Fisher, 1975, *Ann. Phys. (N.Y.)* **91**, 226.
 Niemeier, Th., and J. M. J. Van Leeuwen, 1973, *Phys. Rev. Lett.* **31**, 1411.
 Niemeier, Th., and J. M. J. Van Leeuwen, 1974, *Physica (Utr.)* **71**, 17.
 Niemeier, Th., and J. M. S. Van Leeuwen, in Domb and Green (to be published).
 Noziers, P., 1974, *J. Low Temp. Phys.* **17**, 31.
 Stueckelberg, E. C. G., and A. Peterman, 1953, *Helv. Phys. Acta* **26**, 499.
 Politzer, H., 1973, *Phys. Rev. Lett.* **30**, 1346.
 Rado, G. T., and H. Suhl, 1973, *Magnetism* (Academic P., New York), Vol. V, Part II.
 Riedel, E. K., and F. J. Wegner, 1972, *Phys. Rev. Lett.* **29**, 349.
 Riedel, E. K., and F. J. Wegner, 1973, *Phys. Rev. B* **7**, 248.
 Riedel, E. K., and F. J. Wegner, 1974, *Phys. Rev. B* **9**, 294.
 Rizzuto, C., 1974, *Rep. Prog. Phys.* **37**, 147.
 Sak, J., 1973, *Phys. Rev. B* **8**, 281.
 Schotte, K. D., 1970, *Z. Phys.* **230**, 99.
 Schotte, K. D., and U. Schotte, 1971, *Phys. Rev. B* **4**, 2228.
 Schultz, T. D., D. C. Mathis, and E. H. Lieb, 1964, *Rev. Mod. Phys.* **36**, 856.
 Stephen, M. J., and J. L. McCauley, Jr., 1973, *Phys. Lett.* **44A**, 89.
 Subbarao, K., 1975, *Phys. Rev. B* **11**,
 Symanzik, K., 1970, *Commun. Math. Phys.* **18**, 227.
 Symanzik, K., 1971, *Commun. Math. Phys.* **23**, 49.
 t'Hooft, G., 1973, *Nucl. Phys. B* **61**, 455.

- Triplet, B. B., and N. E. Phillips, 1971, *Phys. Rev. Lett.* **27**, 1001.
Wegner, F., 1972, *Phys. Rev. B* **5**, 4529.
Weinberg, S., 1973, *Phys. Rev. D* **8**, 3497.
Weyland, A., and Th. Niemeier, 1974, *Phys. Lett.* **46A**, 301.
Wilson, K. G., 1965, *Phys. Rev.* **140**, B445.
Wilson, K. G., 1970, *Phys. Rev. D* **2**, 1438.
Wilson, K. G., 1971a, *Phys. Rev. B* **4**, 3184.
Wilson, K. G., 1971b, *Phys. Rev. D* **3**, 1818.
Wilson, K. G., 1972, *Phys. Rev. D* **6**, 419.
Wilson, K. G., 1974a, in *Nobel Symposia—Medicine and Natural Sciences* (Academic P., New York), Vol. 24, p. 68.
Wilson, K. G., 1974b, *Physica (Utr.)* **73**, 119.
Wilson, K. G., and J. Kogut, 1974, *Phys. Rep.* **12C**, 75.
Wilson, K. G., 1975, *Adv. Math.* **16**, 444.
Yamada, K., 1975, *Progr. Theor. Phys. (Kyoto)* **53**, 970.
Yuval, G., and P. W. Anderson, 1970, *Phys. Rev. B* **1**, 1522.