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						

umn 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 Niemeyer–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 representations 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 onei nvolving 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 $-\mu H S_z$ where S_z is the spin, H is the external field, and μ is the magnetic moment. The magnetization at a temperature T is

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

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

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

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)

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

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 $T \to 0$.

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 constants 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 $T \to 0$. 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

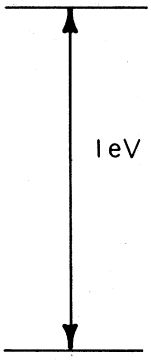

FIG. 10. Two level system illustrating the maximum energy scale of the conduction band.

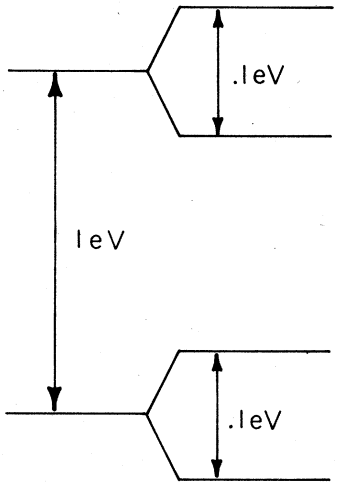

FIG. 11. Four level system for two different electrons in the conduction band with two different energy scales.

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 $T \to 0$. 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 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

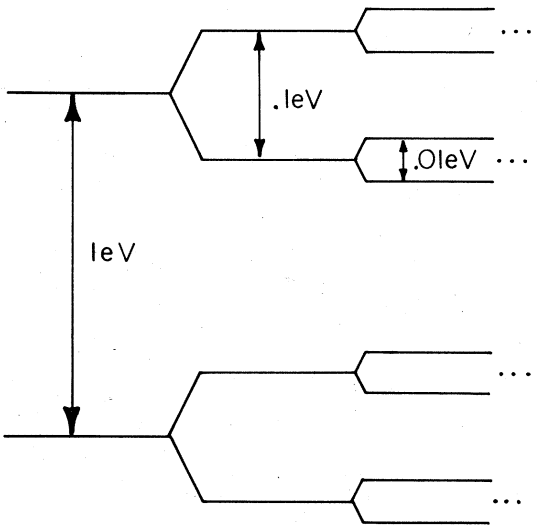

FIG. 12. Multiple energy scale structure of many-electron states of the conduction band.

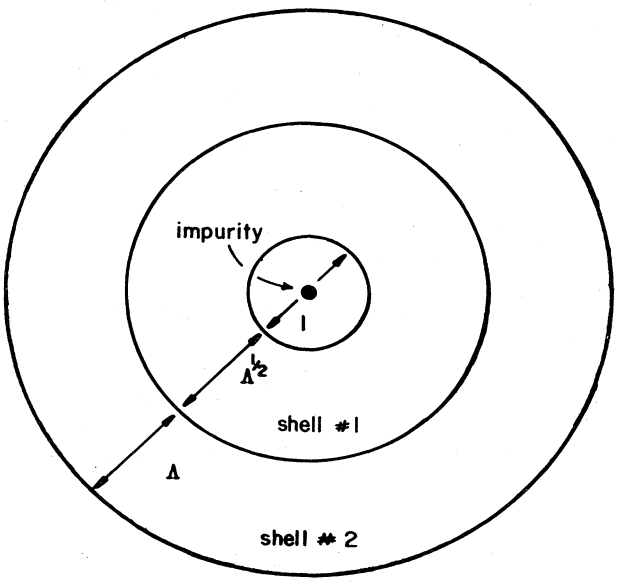

FIG. 13. Onion-like spherical shells giving the location of successive wave functions in the Kondo basis. The size of the smallest (inner) shell is a few Ångström units.

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 $\Lambda^{\frac{1}{2}}$. Here Λ is a parameter ≥ 1; it can be chosen arbitrarily. See below. For accidental reasons the width of the first layer is denoted $\Lambda^{\frac{1}{2}}$ rather than Λ. 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 nth state is concentrated closer and closer to the Fermi surface; thus the energy scale for these states decreases being $\sim\Lambda^{-n/2}$ for the nth 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 nth level is of order $\Lambda^{-n/2}$.

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 nth 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 nth tie represents the Hamiltonian containing n conduction band single electron states (that is, the nth 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 nth Hamiltonian corresponds to having the railroad car at the nth 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 nth 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.

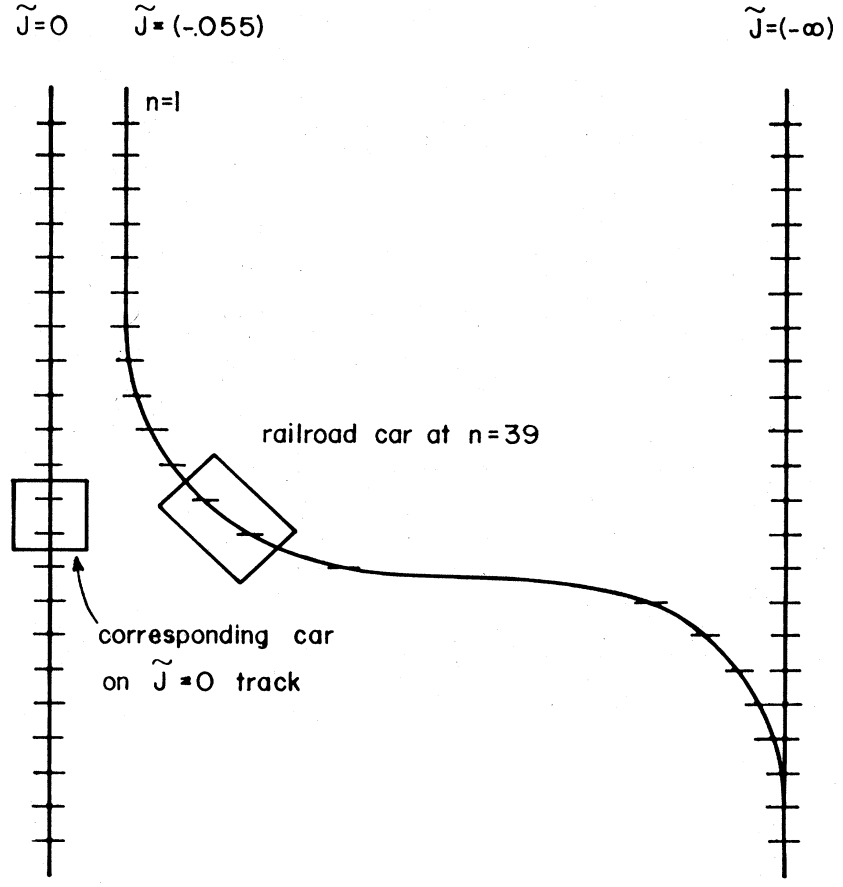

FIG. 14. Railroad track analogy for the Kondo calculation. Different tracks correspond to different initial values of $\tilde{J}$. A track from the top of the figure to the nth tie corresponds to the Kondo Hamiltonian with n electron states kept. The railroad cars illustrate the subset of energy levels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

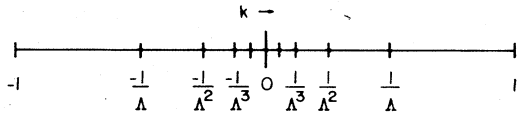

FIG. 15. Discretization on a logarithmic scale in k space.

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 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 nth 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 nth 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 calculations 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

$$H_K = \int_{-1}^{1} dk a_k^+ a_k - JA^+ \boldsymbol{\sigma} A \cdot \boldsymbol{\tau}, \tag{VII.4}$$

where

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

and

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

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 $a_k^+ a_k$ 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 $\Sigma_{\mu\nu} A_\mu^+ \sigma_{x\mu\nu} A_\nu$, 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)

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

(c is a constant). If one looks at only the J term it would appear that χ is proportional to T^{-1} for $T \to 0$. However, for $T \to 0$ 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"):

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

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 $T \to 0$; in this case $\chi(T)$ behaves as T^{-1} for $T \to 0$ 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

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

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., $f_n^+ f_n$) 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 $\Lambda \to 1$. 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 mth 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_{ml}. The first wave function $\psi_{m0}(k)$ we choose to be a constant; ater normalization one obtains

$$\psi_{m0}(k) = \Lambda^{m/2}(1 - \Lambda^{-1})^{-\frac{1}{2}} \quad (\Lambda^{-m-1} < k < \Lambda^{-m}); \quad \text{(VII.9)}$$

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

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

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

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

with

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

(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 $(1-\Lambda^{-1})^{\frac{1}{2}}\Lambda^{-m/2}$ 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 mth 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 corrollary 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 $H_0 + H_1 + H_2 + \cdots$ 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

$$\mathrm{Tr} S_z^2 \exp(-\beta H_K)/\mathrm{Tr} \exp(-\beta H_K),$$

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 $\Lambda \to 1$. 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:

$$H_K = \sum_{n=0}^{\infty} \epsilon_n (f_n^+ f_{n+1} + f_{n+1}^+ f_n) - 2J f_0^+ \boldsymbol{\sigma} f_0 \cdot \boldsymbol{\tau}, \quad \text{(VII.16)}$$

where the ϵ_n will be determined below. Consider now the general structure of an orthogonal transformation. One can write

$$f_m = \sum_m \{u_{nm} a_m + v_{nm} b_m\}. \quad \text{(VII.17)}$$

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

$$H_K = \frac{(1 + \Lambda^{-1})}{2} \sum_{m=0}^{\infty} \frac{1}{\sqrt{2}}(1 - \Lambda^{-1})^{\frac{1}{2}}\Lambda^{-3m/2}(a_m^+ - b_m^+)f_0$$

$$+\text{non-}f_0 \text{ terms}. \tag{VII.25}$$

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

$$f_1^+ = \frac{1}{\sqrt{2}}(1 - \Lambda^{-3})^{\frac{1}{2}} \sum_{m=0}^{\infty} \Lambda^{-3m/2}(a_m^+ - b_m^+), \tag{VII.27}$$

$$\epsilon_0 = \frac{(1 - \Lambda^{-1})^{\frac{1}{2}}}{(1 - \Lambda^{-3})^{\frac{1}{2}}} \frac{(1 + \Lambda^{-1})}{2}. \tag{VII.28}$$

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

$$u_{1m} = -v_m = (1/\sqrt{2})(1 - \Lambda^{-3})^{\frac{1}{2}}. \tag{VII.29}$$

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:

$$H_K = \sum_{m=0} \Lambda^{-5m/2} \frac{(1 - \Lambda^{-3})^{\frac{1}{2}}}{\sqrt{2}}(a_m^+ + b_m^+)f_1 \frac{(1 + \Lambda^{-1})}{2}$$

$$+\text{non-}f_1 \text{ terms}. \tag{VII.30}$$

Consider the coefficient of f_1. One knows that this coefficient can always be written in the form $af_0^+ + f_2^+$ 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

$$a = \tfrac{1}{2}(n \text{ even}) \text{ or } \tfrac{3}{2}(n \text{ odd}), \tag{VII.38}$$

$$r_M = \frac{n - \frac{3}{2} - a}{2}, \tag{VII.39}$$

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 $[(1+\Lambda^{-1})/2^{-n/2}]$ 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:

$$\sum_{n \text{ even}} u_{nm}u_{nm'} = \tfrac{1}{2}\delta_{mm'}, \tag{VII.46}$$

$$\sum_{n \text{ odd}} u_{nm}u_{nm'} = \tfrac{1}{2}\delta_{mm'}, \tag{VII.47}$$

$$\sum_{m} u_{nm}u_{n'm} = \tfrac{1}{2}\delta_{nn'} \text{ (if } n-n' \text{ is even)}. \tag{VII.48}$$

The last condition is equivalent to the requirement

$$\frac{1}{2\pi i}\oint \frac{dz}{z}U_n(z)U_{n'}(z^{-1}) = \tfrac{1}{2}\delta_{nn'} \tag{VII.49}$$

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

$$\sum_{n \text{ even}} U_n(z)U_n(z') = R_0(z,z') + R_0(z,z')R_2(z,z') + R_0(z,z')R_2(z,z')R_4(z,z') + \cdots \tag{VII.51}$$

where

$$R_0(z,z') = (1-\Lambda^{-1})/[2(1-\Lambda^{-\frac{1}{2}}z)(1-\Lambda^{-\frac{1}{2}}z')], \tag{VII.52}$$

and for $n \geq 2$

$$R_n(z,z') = \frac{[1-\Lambda^{(n-\frac{3}{2})}z][1-\Lambda^{(n-\frac{3}{2})}z']}{[1-\Lambda^{-(n+\frac{1}{2})}z][1-\Lambda^{-(n+\frac{1}{2})}z']} \times\frac{[1-\Lambda^{1-2n}]\Lambda^{3-2n}}{[1-\Lambda^{3-2n}]}. \tag{VII.53}$$

Write

$$\Sigma_n(z,z') = 1 + R_{n+2}(z,z') + R_{n+2}(z,z')R_{n+4}(z,z') + \cdots. \tag{VII.54}$$

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

$$\Sigma_n(z,z') = [1/(1-zz')][1-\Lambda^{-(n+\frac{1}{2})}z] \times[1-\Lambda^{(n+\frac{1}{2})}z][1-\Lambda^{-(2n+1)}]^{-1}. \tag{VII.56}$$

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

$$R_n(z,z') \simeq zz'. \tag{VII.57}$$

The product $R_0R_2\cdots R_n$, 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

$$\Sigma_0(z,z') = 1 + R_2 + R_2R_4{}^+\cdots R_2R_4R_6\cdots R_m\,\Sigma_m \tag{VII.58}$$

for any even m. For $m \to \infty$, the last term behaves as $(zz')^{m/2}$ and therefore vanishes. Thus for $m \to \infty$ 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:

$$H_N = \Lambda^{(N-1)/2}\{\sum_{n=0}^{N-1}\Lambda^{-n/2}(f_n{}^+f_{n+1} + f_{n+1}{}^+f_n) - \tilde{J}f_0{}^+\boldsymbol{\sigma}f_0\cdot\boldsymbol{\tau}\}. \tag{VIII.1}$$

Then the original Hamiltonian H of (VII.8) is

$$H = \lim_{N\to\infty}\Lambda^{-(N-1)/2}H_N. \tag{VIII.2}$$

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:

$$H_{N+1} = \Lambda^{\frac{1}{2}}H_N + f_{N+1}{}^+f_N + f_N{}^+f_{N+1} \tag{VIII.3}$$

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 $\Lambda^{-\frac{1}{2}}(f_1{}^+f_2 + f_2{}^+f_1)$, $\Lambda^{-1}(f_2{}^+f_3 + f_3{}^+f_2)$, etc., from the original Hamiltonian H, until for $N \to \infty$ 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 $f_{n\mu}(0 \le n \le N,\ \mu = \pm\frac{1}{2})$ 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.