The renormalization group and Fermi liquids

R.Shankar
Sloane Physics Lab
Yale University
New Haven CT 06520

Chapter 1

1.1 The RG: what, why and how

Imagine that you have some problem in the form of a partition function

$$Z(a,b) = \int dx \int dy e^{-a(x^2+y^2)} e^{-b(x+y)^4} \tag{1.1}$$

where a, b are the parameters.

First consider $b = 0$, *the gaussian model*. Suppose that you are just interested in x, say in its fluctuations. Then you have the option of integrating out y and working with the new partition function

$$Z(a) = N \int dx e^{-ax^2} \tag{1.2}$$

where N comes from doing the y-integration. We will ignore such an x-independent prefactor here and elsewhere since it will cancel in any averaging process.

Consider now the nongaussian case with $b \neq 0$. Here we have

$$\begin{aligned} Z(a', b'...) &= \int dx \left[\int dy e^{-a(x^2+y^2)} e^{-b(x+y)^4} \right] \\ &\equiv \int dx e^{-a' \ x^2} e^{-b'x^4 - c'x^6 + ...} \end{aligned} \tag{1.3}$$

where a', b' etc., define the parameters of the effective field theory for x. These parameters will reproduce exactly the same averages for x as the original ones. *This evolution of parameters with the elimination of uninteresting degrees of freedom, is what we mean these days by renormalization, and as such has nothing to do with infinities; you just saw it happen in a problem with just two variables.*

The parameters b, c etc., are called *couplings* and the monomials they multiply are called *interactions*. The x^2 term is called the *kinetic* or *free-field* term.

Notice that to get the effective theory we need to do a nongaussian integral. This can only be done perturbatively. At the simplest *tree Level*, we simply drop y and find $b' = b$. At higher orders, we bring down the nonquadratic exponential and integrate in y term by term and generate effective interactions for x. This procedure can be represented by Feynman graphs in which variables in the loop are limited to the ones being eliminated.

Why do we do this? Because certain tendencies of x are not so apparent when y is around, but surface to the top, as we zero in on x. For example, we are going to consider a problem in

which x stands for low-energy variables and y for high energy variables. Upon integrating out high energy variables a numerically small coupling can grow in size (or initially impressive one diminish into oblivion), as we zoom in on the low energy sector.

This notion can be made more precise as follows. Consider the gaussian model in which we have just $a \neq 0$. We have seen that this value does not change as y is eliminated since x and y do not talk to each other. This is called a *fixed point of the RG*. Now turn on new couplings or "interactions" (corresponding to higher powers of x, y etc.) with coefficients b, c and so on. Let a', b' etc., be the new couplings after y is eliminated. The mere fact that $b' > b$ does not mean b is more important for the physics of x. This is because a' could also be bigger than a. So we rescale x so that the kinetic part, x^2, has the same coefficient as before. If the quartic term still has a bigger coefficient, (still called b'), we say it is a *relevant* interaction. If $b' < b$ we say it is irrelevant. This is because in reality y stands for many variables, and as they are eliminated one by one, the coefficient of the quartic term will run to zero. If a coupling neither grows not shrinks it is called *marginal*.

There is another excellent reason for using the RG, and that is to understand the phenomenon of universality in critical phenomena. I must regretfully pass up the opportunity to explain this and refer you to Reference [1].

We will now see how this method is applied to interacting fermions in $d = 2$. Later we will apply these methods to quantum dots.

1.2 The problem of interacting fermions

Consider a system of nonrelativistic spinless fermions in two space dimensions. The one particle hamiltonian is

$$H = \frac{K^2}{2m} - \mu \tag{1.4}$$

where the chemical potential μ is introduced to make sure we have a finite density of particles in the ground state: all levels up the Fermi surface, a circle defined by

$$K_F^2/2m = \mu \tag{1.5}$$

are now occupied and occupying these levels lowers the ground-state energy.

Notice that this system has gapless excitations above the ground state. You can take an electron just below the Fermi surface and move it just above, and this costs as little energy as you please. Such a system will carry a dc current in response to a dc voltage. An important question one asks is if this will be true when interactions are turned on. For example the system could develop a gap and become an insulator. What really happens for the $d = 2$ electron gas?

We are going to answer this using the RG. Let us first learn how to do RG for noninteracting fermions. To understand the low energy physics, we take a band of of width Λ on either side of the Fermi surface. This is the first great difference between this problem and the usual ones in relativistic field theory and statistical mechanics. Whereas in the latter examples low energy means small momentum, here it means small deviations from the Fermi surface. Whereas in these older problems we zero in on the origin in momentum space, here we zero in on a surface. The low energy region is shown in Figure 1.1.

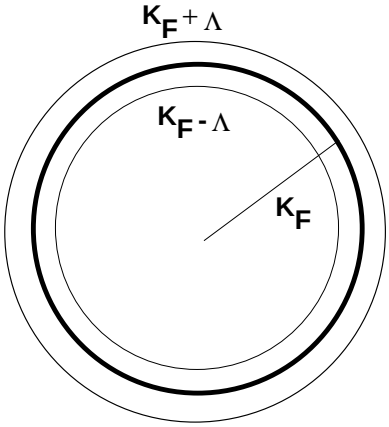

Figure 1.1: The low energy region for nonrelativistic fermions lies within the annulus concentric with the Fermi circle.

To apply our methods we need to cast the problem in the form of a path integral. Following any number of sources, say [2] we obtain the following expression for the partition function of free fermions:

$$Z_0 = \int d\psi d\overline{\psi} e^{S_0} \tag{1.6}$$

where

$$S_0 = \int d^2K \int_{-\infty}^{\infty} d\omega \overline{\psi}(\omega, \mathbf{K}) \left(i\omega - \frac{(K^2 - K_F^2)}{2m} \right) \psi(\omega, \mathbf{K}) \tag{1.7}$$

where ψ and $\overline{\psi}$ are called Grassmann variables. They are really weird objects one gets to love after some familiarity. Most readers can assume they are ordinary integration variables. The dedicated reader can learn more from Ref. [2].

We now adapt this general expression to the annulus to obtain

$$Z_0 = \int d\psi d\overline{\psi} e^{S_0} \tag{1.8}$$

where

$$S_0 = \int_0^{2\pi} d\theta \int_{-\infty}^{\infty} d\omega \int_{-\Lambda}^{\Lambda} dk \overline{\psi}(i\omega - v\ k)\psi. \tag{1.9}$$

To get here we have had to approximate as follows:

$$\frac{K^2 - K_F^2}{2m} \simeq \frac{K_F}{m} \cdot k = v_F\ k \tag{1.10}$$

where $k - K - K_F$ and v_F is the fermi velocity, hereafter set equal to unity. Thus Λ can be viewed as a momentum or energy cut-off measured from the Fermi circle. We have also replaced KdK by $K_F dk$ and absorbed K_F in ψ and $\overline{\psi}$. It will seen that neglecting k in relation to K_F is irrelevant in the technical sense.

Let us now perform mode elimination and reduce the cut-off by a factor s. Since this is a gaussian integral, mode elimination just leads to a multiplicative constant we are not

interested in. So the result is just the same action as above, but with $|k| \leq \Lambda/s$. Let us now do make the following additional transformations:

$$\begin{aligned} (\omega', k') &= s(\omega, k) \\ (\psi'(\omega', k'), \overline{\psi}'(\omega', k')) &= s^{-3/2}(\psi(\frac{\omega'}{s}, \frac{k'}{s}), \overline{\psi}(\frac{\omega'}{s}, \frac{k'}{s})). \end{aligned} \tag{1.11}$$

When we do this, the action and the phase space all return to their old values. So what? Recall that our plan is to evaluate the role of quartic interactions in low energy physics as we do mode elimination. Now what really matters is not the absolute size of the quartic term, but its size relative to the quadratic term. Keeping the quadratic term identical before and after the RG action makes the comparison easy: if the quartic coupling grows, it is relevant; if it decreases, it is irrelevant, and if it stays the same it is marginal. The system is clearly gapless if the quartic coupling is irrelevant. Even a marginal coupling implies no gap since any gap will grow under the various rescalings of the RG.

Let us now turn on a generic four-Fermi interaction in path-integral form:

$$S_4 = \int \overline{\psi}(4)\overline{\psi}(3)\psi(2)\psi(1)u(4,3,2,1) \tag{1.12}$$

where $\int$ is a shorthand:

$$\int \equiv \prod_{i=1}^{3} \int d\theta_i \int_{-\Lambda}^{\Lambda} dk_i \int_{-\infty}^{\infty} d\omega_i \tag{1.13}$$

At the tree level, we simply keep the modes within the new cut-off, rescale fields, frequencies and momenta , and read off the new coupling. We find

$$u'(k', \omega', \theta) = u(k'/s, \omega'/s, \theta) \tag{1.14}$$

This is the evolution of the coupling function. To deal with coupling constants with which we are more familiar, we expand the functions in a Taylor series (schematic)

$$u = u_o + ku_1 + k^2 u_2 ... \tag{1.15}$$

where k stands for all the k's and ω's. An expansion of this kind is possible since couplings in the Lagrangian are nonsingular in a problem with short range interactions. If we now make such an expansion and compare coefficients in Eqn. (1.14), we find that u_0 is marginal and the rest are irrelevant, as is any coupling of more than four fields. Now this is exactly what happens in ϕ_4^4, scalar field theory in four dimensions with a quartic interaction. The difference here is that we still have dependence on the angles on the Fermi surface:

$$u_0 = u(\theta_1, \theta_2, \theta_3, \theta_4)$$

Therefore in this theory we are going to get coupling functions and not a few coupling constants.

Let us analyze this function. Momentum conservation should allow us to eliminate one angle. Actually it allows us more because of the fact that these momenta do not come form

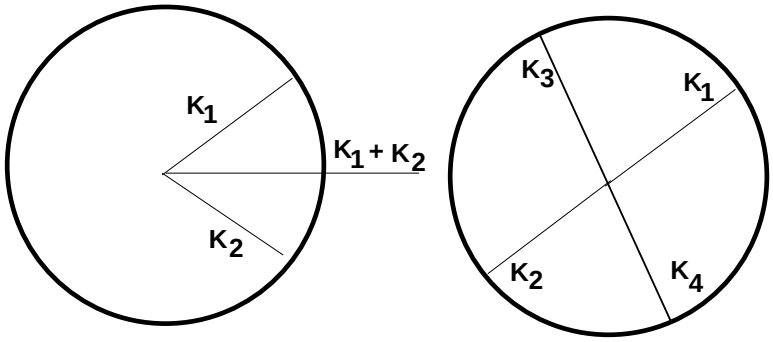

Figure 1.2: Kinematical reasons why momenta are either conserved pairwise or restricted to the BCS channel.

the entire plane, but a very thin annulus near K_F. Look at the left half of Figure 1.2. Assuming that the cutoff has been reduced to the thickness of the circle in the figure, it is clear that if two points 1 and 2 are chosen from it to represent the incoming lines in a four point coupling, the outgoing ones are forced to be equal to them (not in their sum, but individually) up to a permutation, which is irrelevant for spinless fermions. Thus we have in the end just one function of two angles, and by rotational invariance, their difference:

$$u(\theta_1, \theta_2, \theta_1, \theta_2) = F(\theta_1 - \theta_2) \equiv F(\theta). \tag{1.16}$$

About forty years ago Landau came to the very same conclusion[3] that a Fermi system at low energies would be described by one function defined on the Fermi surface. He did this without the benefit of the RG and for that reason, some of the leaps were hard to understand. Later detailed diagrammatic calculations justified this picture [4]. The RG provides yet another way to understand it. It also tells us other things, as we will now see.

The first thing is that the final angles are not slaved to the initial ones if the former are exactly opposite, as in the right half of Figure 1.2. In this case, the final ones can be anything, as long as they are opposite to each other. This leads to one more set of marginal couplings in the BCS channel, called

$$u(\theta_1, -\theta_1, \theta_3, -\theta_3) = V(\theta_3 - \theta_1) \equiv V(\theta). \tag{1.17}$$

The next point is that since F and V are marginal at tree level, we have to go to one loop to see if they are still so. So we draw the usual diagrams shown in Figure 3. We eliminate an infinitesimal momentum slice of thickness $d\Lambda$ at $k = \pm\Lambda$.

These diagrams are like the ones in any quartic field theory, but each one behaves differently from the others and its its traditional counterparts. Consider the first one (called ZS) for F. The external momenta have zero frequencies and lie of the Fermi surface since ω and k are irrelevant. The momentum transfer is exactly zero. So the integrand has the following schematic form:

$$\delta F \simeq \int d\theta \int dk d\omega \left(\frac{1}{(i\omega - \varepsilon(K))} \frac{1}{(i\omega - \varepsilon(K))} \right) \tag{1.18}$$

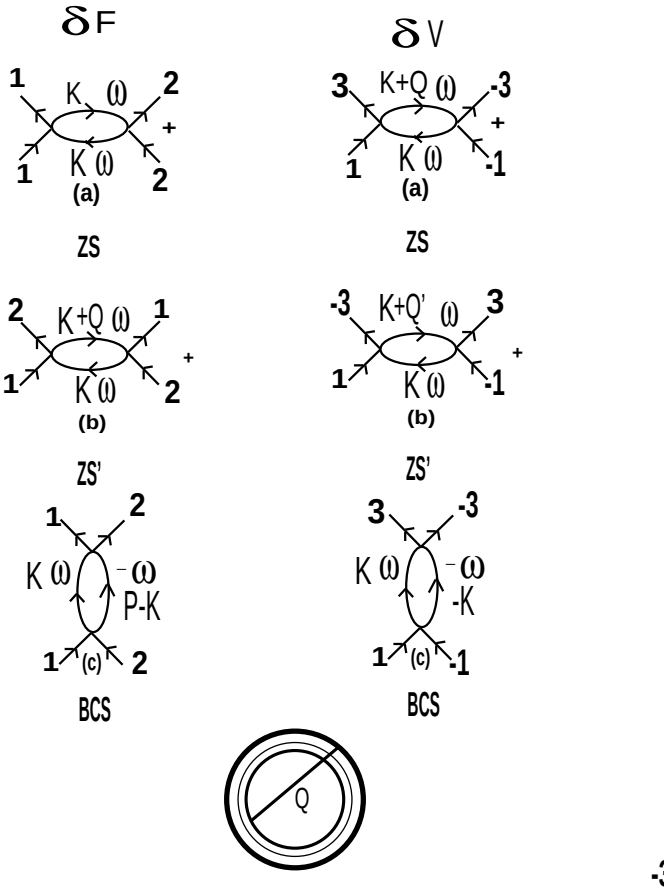

Figure 1.3: One loop diagrams for the flow of F and V. The last at the bottom shows that a large momentum Q can be absorbed only at two particular initial angles (only one of which is shown) if the final state is to lie in the shell being eliminated.

The loop momentum K lies in one of the two shells being eliminated. Since there is no energy difference between the two propagators, the poles in ω lie in the same half-plane and we get zero, upon closing the contour in the other half-plane. In other words, this diagram can contribute if it is a particle-hole diagram, but given zero momentum transfer we cannot convert a hole at $-\Lambda$ to a particle at $+\Lambda$. In the ZS' diagram, we have a large momentum transfer, called Q in the inset at the bottom. This is of order K_F and much bigger than the radial cut-off, a phenomenon unheard of in say ϕ^4 theory, where all momenta and transfers are bounded by Λ. This in turn means that the loop momentum is not only restricted in the direction to a sliver $d\Lambda$, but also in the angular direction in order to be able to absorb this huge momentum Q and land up in the other shell being eliminated (see bottom of Fig. (1.3). So we have $du \simeq dt^2$, where $dt = d\Lambda/\Lambda$. The same goes for the BCS diagram. Thus F does not flow at one loop.

Let us now turn to the renormalization of V. The first two diagrams are useless for the same reasons as before, but the last one is special. Since the total incoming momentum is zero, the loop momenta are equal and opposite and no matter what direction K has, $-K$ is guaranteed to lie in the same shell being eliminated. However the loop frequencies are now equal and opposite so that the poles in the two propagators now lie in opposite half-planes. We now get a flow (dropping constants)

$$\frac{dv(\theta_1-\theta_3)}{dt} = -\int d\theta v(\theta_1-\theta)\ v(\theta-\theta_3) \tag{1.19}$$

Here is an example of a flow equation for a coupling function. However by expanding in terms of angular momentum eigenfunctions we get an infinite number of flow equations

$$\frac{dv_m}{dt} = -v_m^2. \tag{1.20}$$

one for each coefficient. These equations tell us that if the potential in angular momentum channel m is repulsive, it will get renormalized down to zero (a result derived many years ago by Anderson and Morel) while if it is attractive, it will run off, causing the BCS instability. This is the reason the V's are not a part of Landau theory, which assumes we have no phase transitions. This is also a nice illustration of what was stated earlier: one could begin with a large positive coupling, say v_3 and a tiny negative coupling v_5. After much renormalization, v_3 would shrink to a tiny value and v_5 would dominate.

1.3 Large-N approach to Fermi liquids

Not only did Landau say we could describe Fermi liquids with an F function, he also managed to compute the response functions at small ω and q in terms of the F function even when it was large, say 10, in dimensionless units. Again the RG gives us one way to understand this. To this end we need to recall the the key ideas of "large-N" theories.

These theories involve interactions between N species of objects. The largeness of N renders fluctuations (thermal or quantum) small, and enables one to make approximations which are not perturbative in the coupling constant, but are controlled by the additional small parameter $1/N$.

As a specific example let us consider the Gross-Neveu model[5] which is one of the simplest fermionic large-N theories. This theory has N identical massless relativistic fermions interacting through a short-range interaction. The Lagrangian density is

$$\mathcal{L} = \sum_{i=1}^{N} \bar{\psi}_i\ \partial\!\!\!/ \psi_i - \frac{\lambda}{N}\Big(\sum_{i=1}^{N} \bar{\psi}_i\psi_i\Big)^2 \tag{1.21}$$

Note that the kinetic term conserves the internal index, as does the interaction term: any index that goes in comes out. You do not have to know much about the GN model to to follow this discussion, which is all about the internal indices.

Figure 1.4 shows the first few diagrams in the expression for the scattering amplitude of particle of isospin index i and j in the Gross-Neveu theory. The "bare" vertex comes with a factor λ/N. The one-loop diagrams all share a factor λ^2/N^2 from the two vertices. The first one-loop diagram has a free internal summation over the index k that runs over N values, with the contribution being identical for each value of k. Thus, this one-loop diagram acquires a compensating factor of N which makes its contribution of order λ^2/N, the *same order in* $1/N$ *as the bare vertex*. However, the other one-loop diagrams have no such free internal summation and their contribution is indeed of order $1/N^2$. Therefore, to leading order in $1/N$, one should keep only diagrams which have a free internal summation for every

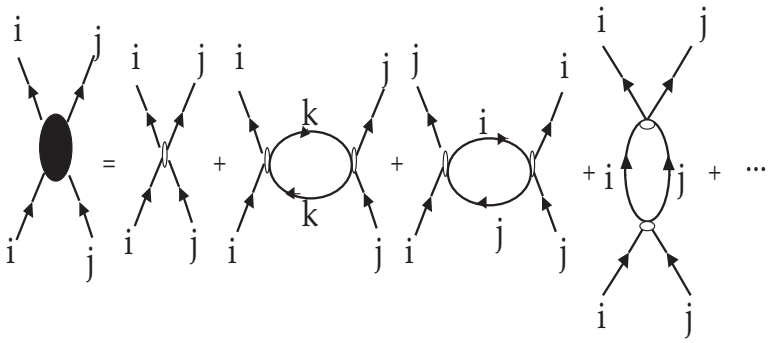

Figure 1.4: Some diagrams from large-N theory

vertex, that is, iterates of the leading one-loop diagram, which are called bubble graphs. *For later use remember that in the diagrams that survive (do not survive), the indices i and j of the incoming particles do not (do) enter the loops.* Let us assume that the momentum integral up to the cutoff Λ for one bubble gives a factor $-\Pi(\Lambda, q_{ext})$, where q_{ext} is the external momentum or frequency transfer at which the scattering amplitude is evaluated. To leading order in large-N the full expression for the scattering amplitude is

$$\Gamma(q_{ext}) = \frac{1}{N}\frac{\lambda}{1+\lambda\Pi(\Lambda, q_{ext})} \tag{1.22}$$

Once one has the full expression for the scattering amplitude (to leading order in $1/N$) one can ask for the RG flow of the "bare" vertex as the cutoff is reduced by demanding that the physical scattering amplitude Γ remain insensitive to the cutoff. One then finds (with $t = \ln(\Lambda_0/\Lambda)$)

$$\frac{d\Gamma(q_{ext})}{dt} = 0 \Rightarrow \frac{d\lambda}{dt} = -\lambda^2\frac{d\Pi(\Lambda, q_{ext})}{dt} \tag{1.23}$$

which is exactly the flow one would extract at one loop. Thus the one-loop RG flow is the exact answer to leading order in a large-N theory. All higher-order corrections must therefore be subleading in $1/N$.

1.3.1 Large-N applied to Fermi liquids

Consider now the $\bar{\psi}\psi - \bar{\psi}\psi$ correlation function (with vanishing values of external frequency and momentum transfer). Landau showed that it takes the form

$$\chi = \frac{\chi_0}{1+F_0}, \tag{1.24}$$

where F_0 is the angular average of $F(\theta)$ and χ_0 is the answer when $F = 0$. Note that the answer is not perturbative in F.

Landau got this result by working with the ground-state energy as a functional of Fermi surface deformations. The RG provides us with not just the ground-state energy, but an effective hamiltonian (operator) for all of low-energy physics. This operator problem can be solved using large N-techniques.

The value of N here is of order K_F/Λ, and here is how it enters the formalism. Imagine dividing the annulus in Fig. (1.2) into N patches of size (Λ) in the radial and angular directions. The momentum of each fermion $\mathbf{k}_i$ is a sum of a "large" part ($\mathcal{O}(k_F)$) centered on a patch labelled by a patch index $i = 1, ...N$ and a "small" momentum ($\mathcal{O}(\Lambda)$ within the patch[2].

Consider a four-fermion Green's function, as in Figure (1.4). The incoming momenta are labeled by the patch index (such as i) and the small momentum is not shown but implicit. We have seen that as $\Lambda \to 0$, the two outgoing momenta are equal to the two incoming momenta up to a permutation. At small but finite Λ this means the patch labels are same before and after. Thus the patch index plays the role of a conserved isospin index as in the Gross-Neveu model.

The electron-electron interaction terms, written in this notation, (with $\mathbf{k}$ integrals replaced by a sum over patch index and integration over small momenta) also come with a pre-factor of $1/N$ ($\simeq \Lambda/K_F$).

It can then be verified that in all Feynman diagrams of this cut-off theory the patch index plays the role of the conserved isospin index exactly as in a theory with N fermionic species. For example in Figure (1.4) in the first diagram, the external indices i and j do not enter the diagram (small momentum transfer only) and so the loop momentum is nearly same in both lines and integrated fully over the annulus, i.e., the patch index k runs over all N values. In the second diagram, the external label i enters the loop and there is a large momentum transfer ($\mathcal{O}(K_F)$). In order for both momenta in the loop to be within the annulus, and to differ by this large q, the angle of the loop momentum is limited to a range $\mathcal{O}(\Lambda/K_F)$. (This just means that if one momentum is from patch i the other has to be from patch j.) Similarly, in the last loop diagram, the angle of the loop momenta is restricted to one patch. In other words, the requirement that all loop momenta in this cut-off theory lie in the annulus singles out only diagrams that survive in the large N limit.

The sum of bubble diagrams, singled out by the usual large-N considerations, reproduces Landau's Fermi liquid theory. For example in the case of χ, one obtains a geometric series that sums to give $\chi = \frac{\chi_0}{1+F_0}$.

Since in the large N limit, the one-loop β-function for the fermion-fermion coupling is exact, it follows that the marginal nature of the Landau parameters F and the flow of V , Eqn. (1.20), are both exact as $\Lambda \to 0$.

A long paper of mine Ref. ([2]) explains all this, as well as how it is to be generalized to anisotropic Fermi surfaces and Fermi surfaces with additional special features and consequently additional instabilities. Polchinski [6] independently analyzed the isotropic Fermi liquid (though not in the same detail, since it was a just paradigm or toy model for an effective field theory for him).

This work was supported by NSF-DMR 0103639.

Bibliography

[1] M.E.Fisher *Critical Phenomena*, F. W. J. Hahne, Editor, Lecture Notes Number 186, Springer-Verlag, Berlin, (1983). These notes are from the Engelbrecht Summer School of 1983!

[2] R. Shankar, *Physica* **A177**, 530 (1991); R.Shankar, Rev. Mod. Phys. **66**, 129 (1994).

[3] L. D. Landau, Sov. Phys. JETP **3**, 920 (1956); Sov. Phys. JETP **5**, 101 (1957).

[4] A. A. Abrikosov, L. P. Gorkov, and I. E. Dzyaloshinski, *Methods of Quantum Field Theory in Statistical Physics*, Dover Publications, New York, 1963.

[5] D. J. Gross and A. Neveu, Phys. Rev.**D10**, 3235 (1974).

[6] P.Polchinski In: *TASI Elementary Particle Physics*, ed by J. Polchinski and J.Harvey (World Scientific, 1992.)