The Effective Field Theory of Fermi Surfaces in the Vicinity of Van Hove Singularities

Thesis by
Tristan McKinney

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy in Physics

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2018
Defended May 29, 2018

Tristan McKinney
ORCID: 0000-0003-2365-1810

To Mom

ACKNOWLEDGEMENTS

The limitations of my memory and stamina make it impossible to list here all who have guided, cajoled, and comforted me on the way to completing this work. I apologize to everyone I wrongfully forget.

I would like to begin by thanking the generations of individuals who built and fought for a world where I can live and think freely. Without liberty, life is a hollow exercise in futility; with it, there is a chance for meaning.

More personally, I deeply and sincerely thank my advisor, Anton Kapustin. His mentorship and insight has helped me understand a small part of the universe in a way no one else has, and that is an incomparable gift. I also owe a debt of gratitude to our collaborator Ira Rothstein for teaching me to question all assumptions and drill down to the bedrock of truth. I thank the members of my committee: Maria Spiropulu, for always looking out for me as if I were her own student; Mark Wise, for his open door and willingness to answer any question; and Olexei Motrunich, for his flexibility and helpful input.

I was blessed with the opportunity to teach a number of delightfully intelligent students and to work with dedicated professors in our common goal of sharing knowledge. It is easy to forget that we are here not just to discover but also to communicate, but they always helped me remember.

The most wonderful part of being at Caltech is the unbelievable density of curious and amazingly intelligent people. I am especially thankful for the chance to have interacted with the physics students in my year, in particular Bogdan, Chia-Hsien, Kung Yi, Du, Min-Feng, and Elena. Discussing physics with the graduate students on my floor was one of the more satisfying parts of my work; I thank especially Murat, Hyungrok, Enrico, and Ashmeet in this regard. I would be remiss to omit my current officemate Pouria, whose example of tireless dedication helped me stay motivated through the last several months of preparing this work. I am also thankful to the younger graduate students, especially Olmo and Irene, who remind me of the excitement that comes with embarking on a new adventure.

My life at graduate school would have been empty without the friends I met here. I thank Krzysztof and Alex, for their fearlessness and their reminder that the value of family is beyond measure; Cristian, Javier, Dustin, and Belinda, both for their unwavering friendship and for years of lively repartee; Steve and Howard, for always

reminding me of the importance of balance in work and in life; and Vipul, for his honest and genuine nature, and most recently his efforts to keep me sane as this process came to a close.

I must make a special mention of Ingmar, my longtime officemate, roommate, and friend. His intelligence, humor, and thoughtful humanity kept me entertained and stimulated through good times and bad.

Finally, I am grateful to my family for their love and support. Thank you Arpana, my love, for your superhuman patience and capacity for kindness, and for your family close and far, who have given me another home; Brennen, for your courage and will, which has inspired me and lent me strength; William, Kathryn, and Erin, for bringing me more joy than anyone has a right to; Laura, for creating a home full of love and happiness; and Dad, for being an image of integrity that I will always strive to emulate.

ABSTRACT

The use of effective field theories to attack new and seemingly disparate problems has proliferated in the past several decades. In this thesis, we develop effective field theories for systems of fermionic quasiparticles possessing Fermi surfaces, with a particular focus on Fermi surfaces proximal to Van Hove singularities. Such systems are a fruitful source of complex and novel behavior in condensed matter physics. We begin with an overview of the renormalization group procedure at the heart of effective field theory by analyzing a simple example. We emphasize the concept that the RG relates the observables of one theory to those of another theory with precisely the same form but different numerical parameters. We also note the generality and extensibility of these concepts. We then apply this perspective to the study of quasiparticles with a round Fermi surface, employing the technique of binning the quasiparticle fields in momentum space to translate previous treatments into a more modern form. We next develop an effective field theory describing the excitations of modes around a Fermi surface with a Van Hove singularity. We resolve lingering questions about the presence of nonlocal interactions in similar models. We find a rich and complicated theory capable of describing deviations from typical Fermi liquid behavior that nonetheless displays some universal dependence on the interactions involving modes in the vicinity of the Van Hove point. We close with an analysis of the instabilities of this Van Hove effective field theory.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1] A. Kapustin, T. McKinney, and I. Z. Rothstein, "Wilsonian effective field theory of 2D van Hove singularities," submitted for publication (2018), arXiv:1804.01713 [cond-mat.str-el],
T. M. participated in the reworking of the project, performed analytical and numerical calculations, made figures, and participated in the writing of the manuscript.

CONTENTS

LIST OF FIGURES

Chapter 1

INTRODUCTION

This thesis chronicles an application of effective field theory (EFT) to the investigation of Fermi surfaces. In particular, it explores the behavior of systems of fermions in two spatial dimensions at zero temperature where the Fermi surface passes close to a Van Hove singularity, a point where the density of states diverges. The motivation is twofold. First, these systems serve as a novel testbed for understanding how we should modify an EFT when the interesting observables are situated along an extended manifold in momentum space. This is opposed to the setting of traditional EFTs, where we are interested in the neighborhood of a single point at zero energy and momentum. Second, a variety of materials with interesting properties have quasi-two-dimensional Fermi surfaces in the proximity of Van Hove singularities, including the cuprate superconductors [1]. Furthermore, previous treatments from the perspective of effective field theory have been afflicted by nonlocal divergences [2], a surprising result given the experience with similar EFTs in high-energy physics [3, 4]. The application of certain methods from these other EFTs clarifies the situation.

Chapter 2 reviews the Wilsonian view of effective field theory. We focus on the perspective that the renormalization group relates a subclass of observables from one theory to those of another, more useful theory and work through a simple example. We then briefly detail how it is possible to relate the low-energy theories that result from this procedure to the high-energy theories from which they descend, and we comment upon how to deal with the situation when the high-energy theory is not well understood. Finally, we emphasize that it is possible to generalize this strategy to focus on different regions of momentum space.

In Chapter 3, we begin with a review of the application of the methodology of effective field theory to nonsingular Fermi surfaces. In the process, we discover a complication involving the conservation of momentum that was not carefully addressed in the original, classic treatments [5, 6]. This leads us to introduce the concept of momentum space binning, a feature of several modern EFTs with multiple scales. In contrast to the conclusions of previous works, we find that within the binned formalism all interactions are irrelevant at the classical level. Despite this,

we determine that the large size of the Fermi surface can enhance various quantum effects, leading to nontrivial behavior in the low-energy theory. We conclude by recapitulating several famous results of Fermi liquid theory using the binned EFT.

Chapter 4 formulates the EFT for a Fermi surface with a Van Hove singularity. Resolving the issue of nonlocal divergences requires us to once again subdivide momentum space. The presence of the Van Hove singularity forces us to introduce a momentum-dependent coupling function, threatening the predictive power of the theory. Fortunately, the Van Hove singularity also simplifies the renormalization group flow, leading to universal predictions for the strong coupling scale. The presence of a large logarithm in the beta function leads to an enhancement of this scale compared to the analogous result for a round Fermi surface. This suggests the superconducting transition temperature is also enhanced, in agreement with previous findings [7–10]. We also discover that modes along the flat sections of the Fermi surface accompanying a van Hove singularity represent a microscopic realization of the marginal Fermi liquid, a phenomenological model describing the normal state of certain high-T_C superconductors [11].

Finally, in Chapter 5, we explore the instability of the Van Hove EFT to other ground states. The logarithmic enhancement found in Chapter 4 results in two poles in the two-particle Green's function at zero net momentum. One of these is physical and suggests that the system is unstable to Cooper pairing when the coupling is attractive, in agreement with the evidence provided by the beta function. Finally, we demonstrate the existence of another instability to a density-wave state, and we investigate how the presence of a finite chemical potential, which shifts the Fermi surface away from the Van Hove singularity, modifies this result.

Chapter 2

A PRIMER ON EFFECTIVE FIELD THEORIES

Effective field theories are a class of models used to describe systems in terms of the degrees of freedom (DOFs) relevant to some limited set of observations. Fundamentally, EFTs are useful because of the intuitive concept that if one is only interested in a small subset of all the potential behaviors of a system, it is often possible to simplify the description of that system by discarding irrelevant information. In the context in which the tools of effective field theory were first codified, this simplification involves focusing on the low-energy observables of a complicated quantum field theory, such as scattering amplitudes of low-energy particles. It is then possible to remove the high energy DOFs of the system, in this case the modes of the quantum field with short wavelengths, so these low-energy observables continue to be encoded in the theory. An extremely early example of such a theory is the Euler-Heisenberg Lagrangian, which describes the self-interaction of light below the mass of the electron and the positron. Another famous example is chiral perturbation theory, an EFT that describes the interactions between hadrons formed from the confined quarks and gluons of quantum chromodynamics [12]. More recent examples include effective field theories describing the low-energy dynamics of string theories [13]. We will discuss several more examples in Section 3.6 of the next chapter.

Despite this historical context, the set of systems amenable to description in terms of EFTs extends well beyond the low-energy limits of traditional quantum field theories. For example, the methodology has been applied to the study of the large-scale structure of the universe [14], hydrodynamics [15], and (of particular relevance to this thesis) Fermi liquids and superfluids [5, 6, 16]. In Chapters 3 and 4, we explore effective field theories of Fermi surfaces in two dimensions. This subject is particularly relevant to the observation of anomalous behavior in various quasi-two-dimensional materials that appear to possess a Fermi surface and display exotic behaviors such as high-temperature superconductivity [17, 18]. This chapter is dedicated to understanding effective field theory in a simpler context.

There are several technical requirements to describe a given physical system with an EFT. First, there should be something playing the role of spacetime in the system, a

domain on which the fields of the system will be defined. Second, we need fields, functions on this spacetime (it is field theory, after all). Note that the values of these fields are rarely the actual observable quantities of the theory. In the context of many-body quantum mechanical systems, these fields can often be thought of as being related to the wavefunctions of the particles in the system. Third, we must be able to formulate the dynamics of the fields in terms of an action principle. Fourth and finally, there should be some kind of transformation on the fields which allows us to "zoom in" on a particular set of DOFs. This will take the form of a coarse-graining operation where we eliminate the degrees of freedom that are not relevant to the observations we wish to make, absorbing their effects into other elements of the system. In the simplest examples of effective field theories, these eliminated DOFs corresponded to massive particles which cannot be produced at low energies.

The fact that the fields are governed by an action principle allows us to construct a mathematical object called a path integral. In quantum mechanics, this object is a weighted sum of the probability amplitudes for the various paths the system can take to move between states. We generally extract observables from the path integral by considering the correlation between the values of the fields at different points in spacetime. This information can then be used to compute particle scattering amplitudes, which can then be related to cross-sections for the scattering process [19]. From this, we can find the expected particle counts in a detector. These are more or less what we actually observe in real world experiments.

The definitive aspect of EFTs is the coarse-graining procedure. It is what allows us to describe the system in terms of a restricted, effective set of DOFs. In general, it can take many forms. In the context of quantum field theories, which we will focus on here, the procedure usually involves computing the part of the path integral involving the DOFs we wish to eliminate from the theory, leaving behind an expression which only describes the fields we are actually interested in. This is called integrating out the eliminated modes of the field. Let us turn now to an example.

2.1 A simple example

Following the example of Chapter 12 of [19], we consider an effective field theory involving real scalar fields in four dimensions. This analysis will closely follow theirs, which is in turn mostly based on [20]. Similar treatments can be found in [21] and, to a lesser degree, [5]. The focus of the presentation here is the motivation for rescaling momentum and the fact that we are constructing a **different**

theory. Special attention is paid to the simplification of the correlators in the new theory. This allows us to write down simple expressions for a subset of the correlators in the original theory using the new theory. In particular, our presentation makes it clear the renormalization group acts as a sort of generalized symmetry, demonstrating that fundamentally different theories produce the same correlators (at least perturbatively), up to rescalings.

Take a theory of a single real scalar field in four dimensions with an explicit momentum cutoff and a $\mathbb{Z}_2$ symmetry $\varphi \to -\varphi$. We assume the action can be expanded in powers of the field and its derivatives (i.e. local operators). Working in momentum space, the action is

$$S[\varphi, \{\lambda\}; \Lambda] = S_{\mathrm{k}}[\varphi; \Lambda] + \sum_{\{r\},n} \lambda_{\{r\},n} S_{\mathrm{mn}}[\varphi; \Lambda], \tag{2.1}$$

where

$$S_{\mathrm{k}}[\varphi; \Lambda] \equiv \int^{\Lambda} \mathrm{d}^4 k \frac{1}{2}(k^2 + m^2)\varphi(k)^2, \tag{2.2}$$

$$S_{\mathrm{mn}}[\varphi; \Lambda] \equiv \int^{\Lambda} \prod_i^n [\mathrm{d}^4 k_i] \delta^4\left(\sum_i k_i\right) k^{\{r\}} \varphi^n. \tag{2.3}$$

The momentum space arguments of the fields have been suppressed, and $k^{\{r\}}$ represents the different possible symmetrized combinations of m copies of the momenta of the fields. The UV cutoff is most conveniently defined after we have passed to the Euclideanized action, in which case we integrate over a ball of radius Λ in momentum space for each k_i. The delta function ensures the interactions are in fact local in position space. Finally, the sum over n includes only even numbers of the fields because of the aforementioned $\mathbb{Z}_2$ symmetry.

The time-ordered correlators of this theory can be found by analytically continuing from the Euclideanized correlators, which are given by the following path integral expression:

$$\langle \varphi(p_1)\varphi(p_2)\ldots\rangle = \frac{\int [\mathrm{D}\varphi(k)]^{\Lambda} \varphi(p_1)\varphi(p_2)\ldots e^{-S[\varphi,\{\lambda\};\Lambda]}}{\int [\mathrm{D}\varphi(k)]^{\Lambda} e^{-S[\varphi,\{\lambda\};\Lambda]}}. \tag{2.4}$$

The path integral is defined by assuming momentum space is a discrete lattice of finite size (governed by Λ) and spacing, treating the field at each point in momentum space to be a variable to be integrated over, then taking the limit of the resulting expression as we reduce the lattice spacing to zero size. The momenta then play the role of continuous labels on the integration variables. The path integral measure

is a product over the (infinitely many) differential elements associated to the field variable at each point in momentum space. We will ignore the technical difficulties in actually taking the continuum limit here. p_1 and p_2 are then labels of particular fields in momentum space.

The utility of effective field theories hinges on their ability to simplify the calculation of observable quantities by restricting the types of questions we ask. Here, that will mean only probing the low-momentum behavior of the theory, which means taking $p_i \ll \Lambda$ in Eq. (2.4). Our approach will be to relate the correlators in the given theory to those in another theory where we have *rescaled* all of the momenta. We call the original theory the UV theory and the final theory the IR theory.

To this end, consider a correlator where all of the external momenta are scaled by a factor ζ, where $\zeta < 1$:

$$\langle \varphi(\zeta p_1)\varphi(\zeta p_2)\ldots\rangle = \frac{\int [\mathrm{D}\varphi(k)]^\Lambda \varphi(\zeta p_1)\varphi(\zeta p_2)\ldots e^{-S[\varphi,\{\lambda\};\Lambda]}}{\int [\mathrm{D}\varphi(k)]^\Lambda e^{-S[\varphi,\{\lambda\};\Lambda]}}. \tag{2.5}$$

To relate this expression to a correlator with unscaled momenta, we perform a change of variables in the path integral (also known as a field redefinition), taking

$$\varphi(\zeta k) = \zeta^{-3}\varphi'(k). \tag{2.6}$$

The expression for the action in terms of the new field variables will be altered. The overall scaling factor of the fields is chosen so the kinetic term in the action remains the same after the change of variables:

$$\begin{aligned} S_\mathrm{k}[\varphi;\Lambda] &= \int^\Lambda \mathrm{d}^4k \frac{1}{2}k^2[\zeta^{-3}\varphi'(\zeta^{-1}k)]^2 \\ &= \int^{\zeta^{-1}\Lambda} \mathrm{d}^4k' \frac{1}{2}k'^2\varphi'(k')^2 = S_\mathrm{k}[\varphi';\zeta^{-1}\Lambda]. \end{aligned} \tag{2.7}$$

In the second equality, we have made a change of variables in momentum space, taking $k' = \zeta^{-1}k$. The third equality holds because k' is a dummy variable. Choosing the field redefinition in this way ensures the form of the free field propagator remains unchanged under rescaling.

S_mn transforms under the rescaling:

$$\begin{aligned} S_\mathrm{mn}[\varphi;\Lambda] &= \int^{\zeta^{-1}\Lambda} \prod_i^n [\mathrm{d}^4\zeta k_i'] \delta^4\left(\sum_i \zeta k_i'\right)(\zeta k)^{\{r\}}\zeta^{-3n}\varphi'^n \\ &= \zeta^{-4+(n+r)} S_\mathrm{mn}[\varphi';\zeta^{-1}\Lambda]. \end{aligned} \tag{2.8}$$

The exponent in the scaling factor is the naive scaling dimension of the interaction. It suggests the contributions from interactions with larger and larger values of r and n (more copies of the fields and derivatives of the fields) will make suppressed contributions to the correlators. We will see why this is the case shortly.

Consider the effect of the change of variables in the path integral on Eq. (2.5). From Eq. (2.6), we have

$$d\varphi = \zeta^{-3} d\varphi'. \tag{2.9}$$

This leads to a formally infinite Jacobian determinant in the path integral measure. However, because the correlators involve a ratio of two path integral expressions, this multiplicative factor cancels. Therefore

$$\langle \varphi(\zeta p_1)\varphi(\zeta p_2)\ldots\rangle = \zeta^{-3F} \frac{\int [\mathrm{D}\varphi']^{\zeta^{-1}\Lambda} \varphi'(p_1)\varphi'(p_2)\ldots e^{-S[\varphi',\{\zeta^{-4+(r+n)}\lambda\};\zeta^{-1}\Lambda]}}{\int [\mathrm{D}\varphi']^{\zeta^{-1}\Lambda} e^{-S[\varphi',\{\zeta^{-4+(r+n)}\lambda\};\zeta^{-1}\Lambda]}}, \tag{2.10}$$

where F is the number of fields in the correlator on the left. After this point, we will drop the primes on the fields because they are dummy variables that are integrated over.

The right-hand side of Eq. (2.10) looks nearly like a correlator of fields with unscaled momenta p_i, Eq. (2.4), but with modified values for the coupling constants of the theory. If they were exactly the same, we could show that the correlators involving scaled momenta are exactly equal to the correlators in another theory with unscaled momenta, but with different values for the couplings. Expressions in this theory would then be much simpler, because we should be able to ignore the effects of the higher dimensional operators. However, at this point we cannot yet make this claim, because the cutoffs in the two theories are not the same.

To resolve this, separate the path integral measure into modes with energy below Λ (the "slow" modes) and between Λ and $\zeta^{-1}\Lambda$ (the "fast" modes):

$$[\mathrm{D}\varphi]^{\zeta^{-1}\Lambda} = [\mathrm{D}\varphi_\mathrm{s}]^{\Lambda} [\mathrm{D}\varphi_\mathrm{f}]_{\Lambda}^{\zeta^{-1}\Lambda}. \tag{2.11}$$

The original field in the action can then be written as a sum of fast and slow modes, with the modes defined to have zero value outside their respective momentum ranges, so

$$\varphi = \varphi_\mathrm{s} + \varphi_\mathrm{f}. \tag{2.12}$$

Once we have made this division, the kinetic term (and the interaction term with $n = 2$, $r = 0$, which corresponds to the mass) in the action neatly divides into one

term for the slow modes and one term for the fast modes. On the other hand, the other interaction terms decompose into three pieces: one which involves only the slow modes, one which involves only the fast modes, and one which mixes the two. We write

$$S[\varphi, \{\zeta^{-4+(r+n)}\lambda\}; \zeta^{-1}\Lambda] = S_{\rm s} + S^{(\rm f)}_{\rm kin} + S_{\rm mix} + S_{\rm f}. \tag{2.13}$$

The part of the action involving the slow modes is simple,

$$S_{\rm s} = S[\varphi_{\rm s}, \{\zeta^{-4+(r+n)}\lambda\}; \Lambda], \tag{2.14}$$

and looks just like the original action. The kinetic term for the fast modes is also straightforward,

$$S^{(\rm f)}_{\rm kin} = \int_{\Lambda}^{\zeta^{-1}\Lambda} {\rm d}^4k \frac{1}{2}(k^2 + m^2){\varphi_{\rm f}}^2. \tag{2.15}$$

The "mixed" piece involves a binomial expansion of terms of the form $(\varphi_{\rm s} + \varphi_{\rm f})^n$. We can write it as

$$S_{\rm mix} = \sum_{\{r\},n} \zeta^{-4+(r+n)}\lambda_{\{r\},n} S^{(\rm mix)}_{\rm mn} \tag{2.16}$$

with

$$S^{(\rm mix)}_{\rm mn} = \sum_{q+p=n}^{q=n-1} C_{n,p} \int^{\zeta^{-1}\Lambda} \prod_i^n [{\rm d}^4k_i]\delta^4\left(\sum_i k_i\right) k^{\{r\}}\varphi_{\rm s}^p\varphi_{\rm f}^q. \tag{2.17}$$

Note that we have defined the fast modes to be zero when their momenta are less than Λ, and the slow modes to be zero when their momenta are greater than Λ, which restricts the actual range in the momentum integrals. Finally, the interactions between only fast modes take the form

$$S_{\rm f} = \sum_{\{r\},n} \zeta^{-4+(r+n)}\lambda_{\{r\},n} S^{(\rm f)}_{\rm mn} \tag{2.18}$$

with

$$S^{(\rm f)}_{\rm mn} = \int_{\Lambda}^{\zeta^{-1}\Lambda} \prod_j^n [{\rm d}^4k_j]\delta^4\left(\sum_j k_j\right) k^{\{r\}}\varphi_{\rm f}^n. \tag{2.19}$$

The numerator of Eq. (2.10) becomes

$$\int [{\rm D}\varphi_{\rm s}]^{\Lambda}\varphi_{\rm s}(p_1)\varphi_{\rm s}(p_2)\ldots e^{-S_{\rm s}} \int [{\rm D}\varphi_{\rm f}]_{\Lambda}^{\zeta^{-1}\Lambda} e^{-S^{(\rm f)}_{\rm kin}} e^{-S_{\rm mix}} e^{-S_{\rm f}}. \tag{2.20}$$

We have assumed here that the modes in the correlator are all slow; this is where we enforce the precept that we are only interested in observables involving small momenta.

We wish to construct a path integral which depends on the slow modes alone. To accomplish this, we must evaluate the path integral over the fast modes. At this point, we assume the perturbative approximation is valid for all interactions. Then we can expand the exponentials involving S_{mix} and S_{f} as a series in the coupling constants λ, and we can evaluate correlators between the fast modes in the resulting expression using Wick's theorem with a propagator defined by Eq. (2.15). The results are conveniently organized in terms of Feynman diagrams with internal lines corresponding to the fast modes and external lines corresponding to slow modes in a particular interaction term. The vacuum diagrams come from evaluating the correlators in S_{f} and cancel against the corresponding terms from the denominator of Eq. (2.10).

To get a feel for the results of this procedure, consider a term which comes from expanding the six-point interaction in S_{mix} with four slow modes and two fast modes:

$$\lambda_{0,6}(\varphi_{\text{s}} + \varphi_{\text{f}})^6 = C\lambda_{0,6}\varphi_{\text{s}}^4\varphi_{\text{f}}^2 + \dots \tag{2.21}$$

Integrating out the fast modes in this expression corresponds to evaluating the Feynman diagram corresponding in Fig. 2.1, where the double line corresponds to the propagator for the fast modes running in the loop. This diagram then represents a modification of the effective four-point coupling for the remaining slow modes.

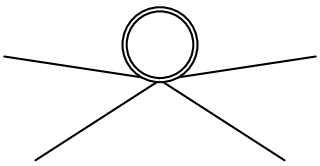


Figure 2.1: A contribution to the four point interaction in the effective action from integrating out fast modes.

Imagine we have made a small rescaling step, so $\zeta = 1 - \delta\Lambda/\Lambda$ and $\zeta^{-1} = 1 + \delta\Lambda/\Lambda$. This implies we can drop the rescaling factor for $\lambda_{0,6}$ from the expression, since it will only contribute at order $\delta\Lambda^2$. The change in the four point coupling due to integrating out the fast modes goes like

$$\delta\lambda_{0,4} \sim \lambda_{0,6}\int_{\Lambda}^{\Lambda+\delta\Lambda}\frac{\mathrm{d}^4k}{(2\pi)^4}\frac{1}{k^2+m^2}, \tag{2.22}$$

where we have used the usual result for the momentum space propagator for a real scalar field. The integration limits indicate we should perform the integral over the range of $|k|$ between Λ and $\Lambda + \delta\Lambda$, that is, the momenta at the edge of a

four-dimensional ball of radius $\Lambda + \delta\Lambda$. We have not been careful in keeping track of the numerical factor arising from the binomial expansion. Assuming $m \ll \lambda$, which should be true since we wish to describe particles which are not so massive that they should be integrated out of the theory, we have

$$\delta\lambda_{0,4} \sim \lambda_{0,6} \int \mathrm{d}\Omega_3 \frac{\Lambda^3 \delta\Lambda}{(2\pi)^4} \frac{1}{\Lambda^2} + O\left(\frac{m^2}{\Lambda^2}\right). \tag{2.23}$$

$\int \mathrm{d}\Omega_3 = 2\pi^2$ is the surface area of a four-dimensional ball. Therefore to leading order in m^2/Λ^2,

$$\delta\lambda_{0,4} \sim \Lambda\delta\Lambda \frac{\lambda_{0,6}}{8\pi^2}. \tag{2.24}$$

While this particular term was independent of the momentum of the slow modes, that will not always be the case. For example, the corrections arising from evaluating the diagram in Fig. 2.2 will depend on the momentum of the slow modes corresponding to the external lines. Since that dependence comes from the propagators of the fast modes, we will be able to expand these corrections in powers of q/Λ, where q is an external momentum. As a result, this diagram corrects the derivative interactions of the slow modes.

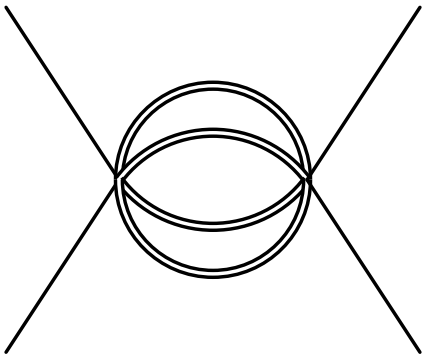


Figure 2.2: A diagram that depends on the momenta of the external modes and hence modifies the derivative couplings.

Fig 2.2 illustrates another important feature of the activity of integrating out the fast modes in that it involves two copies of the $\lambda_{0,6}$ coupling. Diagrams involving more than one vertex come from higher-order terms in the expansion of the exponential $\exp(-S_{\mathrm{mix}})$ in powers of the coupling constants. These involve copies of different vertices at different spacetime points, and the fast modes of the different vertices can be contracted against each other. These terms then re-exponentiate to modify the couplings in the effective action of the slow modes.

It is a very important point that the contributions of the interactions involving many powers of the fields and couplings are not at all small at the beginning of this procedure. For example, it is not true that the contribution of the six-point vertex to the new value of the effective four-point interaction between the slow modes is suppressed relative to the four-point interaction's contribution. The point here is that contributions to observables from the **effective** six-point coupling is small. This is because we are absorbing the important effects of the original six-point coupling into the effective four-point coupling. In other words, it is not true that the results in the IR theory are independent of the values of the irrelevant interactions in the UV theory. What is true is that the important effects of these irrelevant interactions are absorbed into the value of the marginal and relevant parameters of the theory.

Our results show that besides being rescaled, a generic interaction between the slow modes is modified by the effects of the integrated-out fast modes. Despite this fact, we see that the effective interactions between slow modes generated by this procedure will always take the form of local interactions. Thus, they serve to alter the values of the coupling constants that already exist in S_s. Also note that even if we had initially begun with a finite set of interactions between the modes of the theory, this procedure would (generically) generate all effective interactions consistent with the symmetries of the theory. This is a general feature of the Wilsonian approach to effective field theory.

Next, consider the diagram in Fig. 2.3. This diagram depends on the external momentum of the slow modes, and it involves two external slow lines. Hence, it should act to correct the kinetic term for the slow modes, $k^2\varphi_s$. However, we want the new theory we produce upon rescaling and integrating out the fast mode to have the same Feynman rules, with only the value of the couplings changing. This will not be the case if we modify the kinetic term for the soft modes. To rectify the situation, we introduce a second field redefinition which removes these effects:

$$\varphi' = \sqrt{Z_\varphi}\varphi_s. \tag{2.25}$$

Z_φ is referred to as the field strength renormalization. As before, the field redefinition yields an infinite multiplicative factor which cancels between the numerator and denominator of Eq. (2.4) and a number of leftover factors corresponding to the number of fields in the correlator. The final expression relating the correlators of the UV and IR theories is

$$\langle\varphi(\zeta p_1)\varphi(\zeta p_2)\dots\rangle_{S_0} = \zeta^{-3F} Z_\varphi^{-F/2}\langle\varphi(\zeta p_1)\varphi(\zeta p_2)\dots\rangle_{S'}. \tag{2.26}$$

The values of the couplings in the new IR theory defined by S' have been modified by the "classical" scaling of the fields, the quantum corrections due to integrating out the fast modes, and the field strength renormalization. This process as a whole is called the renormalization group (RG) flow.

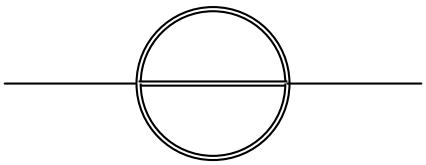


Figure 2.3: A diagram which corrects the kinetic energy of the slow modes and therefore induces a field strength renormalization.

2.2 Predictive power

The necessity of keeping an infinite set of interactions in the effective theory is disturbing. We want to make predictions. If there are an infinite number of tuning parameters corresponding to the values of the couplings, how can we possibly succeed? Fortunately, our prior results regarding the scaling dimensions of the couplings comes to the rescue. We found above that the couplings with n fields and r derivatives are scaled by a factor of ζ^{-4+r+n} if we ignore the quantum effects associated with loop diagrams. Since $z < 1$, this means the numerical value of these couplings in the IR correlator shrink as we iterate the process of scaling and integrating out modes. We have restricted ourself to the study of the perturbative features of the theory, which means only positive powers of the couplings can show up in the IR correlator. As a result, the parts of the correlators involving these higher-order interactions shrink to zero as we continue the RG, and we are left with only a finite number of couplings which do not shrink to zero in the IR. The couplings that do shrink are called the irrelevant couplings. The couplings that do not shrink are further subdivided into two classes: relevant couplings, which grow as we scale down, and marginal couplings, which remain the same.

If we do take into account the quantum effects described above, we generically find that the effective scaling dimension of the couplings is modified. The difference between the classical scaling dimension and the one with the quantum effects included is referred to as the anomalous dimension of the interaction associated with the coupling.

This definition allows us to state the requirement for our procedure to function correctly with a fair level of precision: the anomalous dimensions from integrating

out the high-energy modes should not overwhelm the naive scaling dimensions. In particular, the quantum effects should be small enough that no naively irrelevant interactions become relevant. An interaction for which this does occur is called a dangerously irrelevant interaction. Of course, this is not necessarily an insuperable obstacle, if we also continue to track the dangerously irrelevant interaction throughout the flow.

Let us summarize these results. We find the small-momentum correlators of our UV theory are equal to the correlators of a distinct IR theory with the **same** cutoff but modified values of the couplings. For all the interactions with a negative mass dimension (modulo issues of large anomalous dimensions), the effective couplings shrink throughout the RG flow. Since the functional form of the correlators is identical in each theory, and we assume the correlation functions are perturbative in the coupling constants, we may conclude the contributions from any particular irrelevant interaction will eventually be suppressed relative to the marginal/relevant interactions. Thus, if we are sufficiently far in the IR, the observables of the theory look independent of the effective irrelevant couplings. In the words of traditional QFT, we are left with a renormalizable effective theory in the IR.

2.3 Higher-derivative interactions

Higher-derivative terms in the effective action must be treated more carefully than those with higher powers of the fields. To see why, consider the contribution of the $\partial^2\varphi^4$ interaction to some correlation function. While the scaling argument we have made in the previous sections goes through without modification, we see the suppression of the coupling constant is offset by the simultaneous rescaling of the external momenta in the correlation function. As a result, the contribution of this interaction to the correlation functions of the IR theory is not suppressed relative to those of the UV theory.

To see the resolution to this quandary, notice that the contributions from these interactions must always be accompanied by powers of the external momenta. This follows immediately from dimensional analysis: the naive scaling dimension of an interaction with n derivatives must be n, so we can always write the coupling constant as $\lambda = \tilde{\lambda}/\Lambda^n$, where $\tilde{\lambda}$ is a dimensionless number and where Λ is the original UV cutoff. We are interested in this discussion only in instances where an external momentum shows up in a correlator involving this coupling. As a result, if we assume the correlators are analytic in λ, we must always end up with positive

powers of the external momenta in the correlators, so the contributions look like

$$\left(\frac{p}{\Lambda}\right)^r \tilde{\lambda} \tag{2.27}$$

for some positive r. Therefore the contributions from these interactions are guaranteed to be suppressed even in the UV theory.

2.4 Relevant operators and naturalness

We now know that any field theory looks renormalizable if we flow sufficiently far into the IR. However, notice that effective couplings with positive mass dimension, the relevant couplings, tend to quickly grow unless there is a precise cancellation due to the loop effects. Thus, if we imagine starting out in some random initial field theory in the UV, the effective values of these relevant couplings tend to become as large as the UV scale itself. If the coupling corresponds to an interaction, we should expect the perturbative approximation to quickly break down. In our example, there are no interactions which are classically relevant. However, the mass of the particle, the coefficient of φ^2, is relevant. This means that unless there is a precise cancellation in the theory, the mass should be too large to allow the particle to exist in the IR theory at all. Such a cancellation is referred to as a *fine tuning* of the model.

In the context of an effective theory describing a condensed matter system where we have control of the experimental parameters (such as temperature, pressure, or doping), it makes sense to consider fine-tuned models, because it may very well be possible to tune the necessary parameters in the lab. The same cannot be said of fundamental theories of nature. In that case, the existence of a relevant coupling or mass in the theory is surprising. For this reason, we call a theory with no relevant operators (which hence requires no fine tuning) natural.[1] In particular, a low mass fundamental scalar particle in four spacetime dimensions is unnatural.

Until recently [22, 23], no such particle had been observed. The observation of the Higgs boson changed this situation. The fact that the Higgs exists is one of the stronger pieces of theoretical evidence for the existence of a new UV theory at energy scales not too far above what is accessible with today's technology.

[1] This is often generalized to include theories with relevant couplings that restore a symmetry of the theory when the coupling is taken to zero. This is referred to as the t'Hooft naturalness condition.

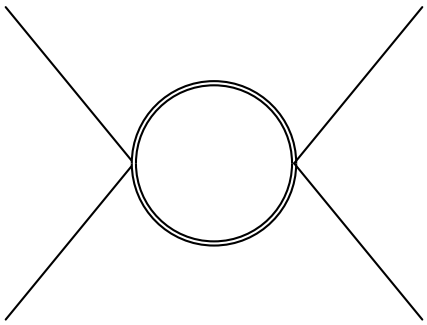


Figure 2.4: The leading contribution to the beta function for the four point coupling.

2.5 The beta function

Let us simply take the tuning to produce a small mass for the particle in our simple example as an axiom. Our analysis indicates that the behavior of the low-energy observables of our system depend most strongly on the value of the marginal (and relevant) couplings between the fields. Consider the model after we have already flowed into the IR for a bit, so the effective couplings of the irrelevant interactions are small. There is a single marginal interaction in the Lagrangian, the four point coupling. Call the coupling λ. As we continue the RG flow, the dominant effect on the flow of λ will come from λ itself. Furthermore, there is no classical scaling at all because the coupling is marginal. Finally, the field strength renormalization only affects the flow at higher orders in λ.

The lowest-order contribution to the flow is from diagrams like the one in Fig. 2.4.[2] Being cavalier about the numerical coefficient, this diagram results in a modification to the effective four-point coupling given by

$$\delta\lambda = C\lambda^2 \int \frac{\mathrm{d}^4 k}{(2\pi)^4} \frac{1}{k^2 + m^2} \frac{1}{(k+q)^2 + m^2}, \tag{2.28}$$

where q is the momentum flowing into the diagram. We can drop q from the expression since we assume the momenta of the soft modes are much smaller than Λ, and because this correction will modify a different interaction. Furthermore, we drop m^2 under our assumption that we have tuned the mass to a small value. The resulting expression is then

$$\delta\lambda = C'\lambda^2 \frac{\delta\Lambda}{\Lambda}, \tag{2.29}$$

where C' is some positive constant.

[2] In fact there are three such diagrams, each corresponding to ways to route different external momenta through the loop. They all give the same contribution.

We can take the limit where the iterative process we have described becomes a continuous transformation between theories. In that case, Eq. (2.29) becomes a differential equation governing the behavior of the effective coupling as we flow to the IR. The resulting expression, called the *beta function*, is

$$\Lambda \frac{\mathrm{d}\lambda}{\mathrm{d}\Lambda} = C'\lambda^2 + O(\lambda^3). \tag{2.30}$$

To order λ^2, the solution is

$$\lambda(\Lambda) = \frac{\lambda_0}{1 + C'\lambda_0 \log \frac{\Lambda}{\Lambda_0}}. \tag{2.31}$$

We see that the behavior of the effective coupling as we flow to the IR depends on the sign of λ_0.

2.6 Polchinski RG and matching

We have shown that the effects of the irrelevant terms in the UV action are absorbed into the effective marginal and relevant parameters and the observables in the IR theory are independent of the effective values of the nonrenormalizable couplings. However, notice that this observation only holds if we have flowed sufficiently far into the IR. If we want to consider external momenta and energies near the original cutoff, we are faced with a problem: the suppression we have found is not small, and therefore the entire infinite tower of irrelevant interactions contributes to the observables.

This is not an issue if we restrict ourselves to low-energy observables. However, we sometimes want to understand the low-energy effective field theory in terms of another theory living immediately above some physical cutoff, like the mass of a heavy particle. The Wilsonian approach described above demonstrates that whatever effective field theory we generate by integrating out this massive particle will eventually give useful descriptions in the IR. However, it does not give us a calculational method to relate these contributions to the original theory. We will see that it is possible to find a description of the low energy theory that can be related to the high energy completion. This process is often referred to as the construction of a top-down effective field theory [24, 25].

To understand the tools involved, it is useful to understand the traditional point of view of renormalization and renormalizability in quantum field theory. In that approach, the UV cutoff is regarded as an unphysical calculational crutch that we wish to remove at our earliest convenience. The Lagrangian is specified at the

outset to only include the interactions that we have termed marginal or relevant. The reason for this requirement is intimately tied to the technical details of loop computations. If we only include the renormalizable terms, we find it is possible to remove the divergences associated with the UV cutoff by allowing the values of these couplings to vary as we send the cutoff to infinity. However, this procedure fails if we include irrelevant interactions. In that case, we find we must include an infinite collection of interactions to remove all of the divergences generated by the loop diagrams (as opposed to the finite number when only marginal and irrelevant couplings are included). This then ruins the predictability of the theory [26, 27].

It is not too hard to understand why this happens from the point of view of effective field theory. The problem is that we are holding an irrelevant interaction constant while simultaneously sending the UV cutoff to an infinite value. From our discussion above, it is clear the irrelevant coupling should naturally be pushed to lower values by the RG as we flow further into the IR, with a limiting value of zero for all irrelevant couplings if we are infinitely far from the UV cutoff. Therefore we see there is an inherent contradiction in such a procedure from the EFT perspective.

In [28], Polchinski formalizes and extends the connection between EFT and the traditional view. He asks us to consider two different theories: one defined with all irrelevant couplings set to zero at the UV scale and another with nonzero values for at least some of these couplings. He then considers integrating out the high-energy modes (without the momentum rescaling we have discussed). He shows that the "distance" between the RG **trajectories** of these two theories becomes small in the IR.[3] In other words, we can always find a choice of values for the theory with only renormalizable couplings at the UV scale that will reproduce the observations of the theory with other, irrelevant couplings included, up to errors suppressed by powers of the final IR scale over the original UV scale.[4] This makes the connection with the traditional view clear: no matter what "true" theory we start with in the UV, it is always possible to find another "fake" theory containing only renormalizable interactions that will reproduce the observations we make in the IR.

If we wish to reduce the size of the errors, we must keep additional irrelevant interactions. However, we can stratify the space of interactions that we must keep

[3]Note that without the rescaling, the irrelevant parameters generated by the RG flow do not take on small values. However, they do approach the same low-dimensional manifold parameterized by the marginal (and relevant) couplings of the theory.

[4]The only exception to this is if the anomalous dimension of an operator (the part of the scaling due to the loop effects) overwhelms the classical scaling and promotes an irrelevant operator to relevancy.

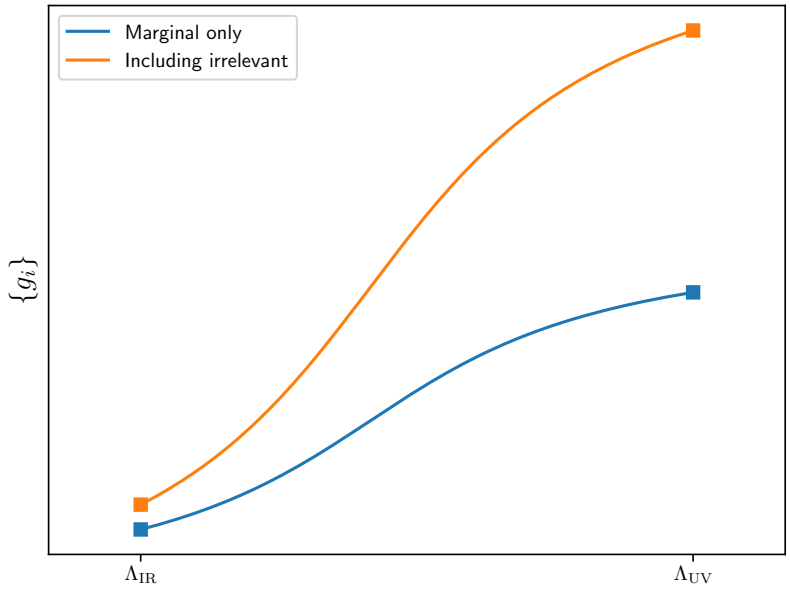


Figure 2.5: A representation of the RG trajectories of two theories. The vertical axis represents the space of couplings. As the RG scale is lowered, the two paths approach a submanifold parameterized by the marginal couplings alone.

to reproduce the predictions of the true effective field theory up to a given error. This stratification corresponds exactly to the power counting of the dimensions of the interactions described in the previous sections.

Besides connecting two disparate understandings of the renormalization group, [28] shows us how to relate the low-energy effective theory to a UV theory living above the cutoff. We will call this UV theory the full theory, although it is also often an EFT. To recap, the problem is that the true effective theory living directly below the physical cutoff, the one we would find by directly integrating out the modes of the full theory, is horribly complicated. It contains a huge number of unsuppressed interactions, and they generically all contribute to the RG flow. The key is to realize that by Polchinski's argument, we can replace this true theory by one containing only renormalizable interactions between the remaining degrees of freedom. This other theory will differ wildly from the true effective field theory in the vicinity of the cutoff, but it will be guaranteed to only make errors relative to the true EFT that are power suppressed once we consider IR observables.

The strategy, then, is to first calculate a set of observables in the fake theory

containing only the renormalizable interactions between the remaining degrees of freedom at a scale around the cutoff. Next, we calculate the same set of observables in the full theory living above the cutoff at the same scale. We then adjust the values of the couplings in the fake theory to reproduce the full theory results. Finally, we flow down to the scale we are actually interested in by using the RG, confident in our knowledge that the fake theory will reproduce the results of the true EFT up to small errors. To reduce the size of the errors, we simply increase the number of interactions in the fake theory to include the next rung of irrelevant couplings. Instead having to track the infinite set of couplings near the transition from one theory to the next, we have reduced the problem to finding an effective theory that will reproduce our results in the IR with a finite collection of couplings specified at the boundary. This whole procedure is known as matching [24, 25, 29, 30].

2.7 Bottom up EFTs

The top-down approach described in the previous section may be contrasted with a bottom-up approach, where the full theory describing the behavior above the cutoff is unknown, too complicated, or too strongly coupled to be useful in extracting information relevant to the low energy theory. Good examples of this include the chiral Lagrangian (strongly coupled full theory) and the standard model (unknown full theory). In these cases, we make no attempt to relate the values in the effective theory to the full theory. However, we can be confident that as long as there is some effective description of the system in terms of fields and a sufficiently large range of energies in which this description is useful, the dynamics should be described by the marginal and relevant interactions of the effective degrees of freedom.

This suggests a strategy for dealing with these scenarios: identify the possible forms the fields and interactions can take by enumerating the symmetries of the system involved, either by making use of some knowledge of the complicated full theory or through an appeal to experiment.[5] Then, write down all the relevant and marginal local interactions that satisfy the symmetry requirements and can be constructed from the fields. This theory should then be able to describe the dynamics of the system as long as we only consider observables within the domain of applicability of the theory.

[5] Finding the correct IR degrees of freedom can often be something of an art. An interesting example of the challenges involved in this process is given by the development of nonrelativistic quantum chromodynamics (NRQCD), the theory of nonrelativistic bound states of heavy quarks. In that case, it required several iterations before the correct description was found [31–33].

In the next chapter, we will meet an example of a thoroughly complicated UV theory, a high-density collection of strongly interacting electrons inside the atomic lattice of a material. There is not much hope of trying to understand the dynamics of this full theory. Instead, we will attempt to construct a bottom-up EFT to describe the behavior of the system for low excitation energies.

2.8 Generalizations of the RG

In our simple example, we chose to rescale all components of the momentum, and hence the spacetime coordinates of the fields, in a uniform manner. This made sense here, because we wished to preserve the Lorentz invariance of the system throughout the flow. However, nothing within the formalism of the RG necessitated that choice. In fact, we even could have chosen to rescale fields in different regions of momentum space differently. The only requirement for the procedure described above to function is that we should return the action to its original form (up to a change in the parameters of the theory). With this freedom, we can imagine focusing on different regions of momentum space to investigate particular classes of observables relevant to a given experimental question. We will meet examples of theories that do exactly this in the next chapter.

Chapter 3

THE ROUND FERMI SURFACE

3.1 The free electron model

The first theories of the behavior of electrons in the crystalline lattice of metals were developed at the beginning of the 20th century. In these models, the electrons were treated as nearly free classical particles bouncing off a lattice of stationary atomic nuclei. While this did a fairly good job explaining the electronic conductivity of metals, it failed in many respects, such as in predicting the heat capacity. The next major advance, pioneered by Arnold Sommerfeld, was to treat the electrons as a gas of nearly free particles obeying Fermi statistics.

In this picture, we imagine filling an empty box with noninteracting fermions. The energy eigenstates of the potential represented by the box correspond to sinusoidal wavefunctions in position space, which means they look like states of definite and discrete momentum in momentum space. The states with the longest wavelengths (and hence the smallest momentum) have the lowest energies. At low temperatures, the particles go into the lowest possible energy states. However, because these particles are fermions, the Pauli exclusion principle implies that it is impossible for more than one identical particle to occupy the same quantum state. If we take the particles to be spin-1/2, as electrons are, this means that we can place at most two into each energy eigenstate. As a result, as we place more and more particles into the box at zero temperature, the low momentum states are filled, and the new particles we introduce must go into higher and higher energy states.

If we plot the filled and unfilled states in momentum space, the filled states cluster around the origin at zero momentum. As a result, the shape of the space of filled states looks like a ball of the dimension appropriate to the system. Therefore, for example, the filled states at zero temperature look like a filled circle in two spatial dimensions, and they look like a filled sphere in three dimensions. The boundary between the filled and unfilled states in the system is called the Fermi surface. The region of filled states within the Fermi surface is called the Fermi sea.

The excitations of such a system involve adding energy to one of the particles in the Fermi sea, moving it into an unfilled state above the Fermi surface. While the total energy of the excited state is relatively large, the amount of energy to move a particle

from within the Fermi sea to just outside the Fermi surface can be very small — in the limit where we take the volume of our box to be infinite, the minimum energy required to produce this excited state actually goes to zero. Applying this model to a real metal explains why arbitrarily small electric fields can produce electric currents within the material: the small electric field puts the "conduction" electrons near the Fermi surface into a state with momentum in the direction of the field, leading to the transport of charge within the system.

Given this picture of the filled and unfilled states with the Fermi surface as the boundary, we can ask what what simultaneous amount of energy and momentum must be injected into the system to excite a state near the Fermi surface. To do this, we can linearize the energy of the electrons near the Fermi surface as a function of momentum. Since they are free electrons, the Fermi surface resides at a constant magnitude of momentum P_F, and we have

$$E(P) = \frac{P^2}{2m} = \frac{(P_F + \delta P)^2}{2m} = V_F \delta P + E_F + O(\delta P^2/m), \tag{3.1}$$

where the Fermi energy $E_F = P_F^2/2m$ is the energy of the electronic state living right at the Fermi surface and the Fermi velocity $V_F = P_F/m$ is the speed associated with these states. Then we see that we can excite the system by simultaneously adding an energy $E(P) - E_F = V_F \delta P$ and a momentum δP. In other words, the dispersion for the excitations of the system looks like

$$E = V_F P \tag{3.2}$$

at small momentum, where E and P refer to energy and momentum of the excitation relative to the ground state at zero temperature. This allows us to think of the excitations of the system as independent of the underlying electrons that make up the Fermi sea.

The process of exciting the electron into a state above the Fermi surface leaves behind an unfilled state below the Fermi surface. Compared to unexcited state, this looks like the absence of an electron. As a result, it takes on the character of a positively charged particle called a hole. From this point of view, the promotion of the fermionic particle into an excited state looks very similar to pair production in QED, where a sufficiently energetic photon can (in the presence of other particles, which is necessary to conserve momentum) produce an electron-positron pair. The major difference here is that while there is a minimum energy to produce the electron-positron pair, there is no minimum energy to produce the electron-hole pair in a

metal. This can be seen by observing that the energy associated with the excitations in Eq. (3.2) goes to zero as P goes to zero. Such a system is described as being gapless.

Amazingly, this treatment did an excellent job describing many properties of metals at low temperatures, despite the fact that the electrons in a metal are by no means nearly free. Indeed, they should constantly interact both with each other and the underlying lattice of atomic nuclei. Thus, it was a mystery why the model was so successful.

3.2 Landau Fermi liquid

In a series of papers in the late 1950s [34] [35] [36], Lev Landau greatly clarified the situation. His major idea was to consider starting with a gas of free electrons in a box at zero temperature and turning on the interactions between the electrons (and with the underlying lattice) adiabatically, that is, slowly enough that the eigenstates and energy levels of the system shift in a continuous matter and the higher energy states are not excited. Intuitively, this means that the boundary between the filled and unfilled states, the Fermi surface, should persist throughout the process of turning on the interaction. The excitations of this new system should then have a similar character to those in the free electron gas. The resulting state is called a Fermi liquid. The underlying particles are called quasiparticles to emphasize the fact that the process of turning on the interactions will modify some aspects of their behavior, such as their effective mass. Note that the quasiparticles will still display fermionic statistics. Landau's basic idea here was to elevate the importance of the Fermi surface, making it the defining characteristic of the new model. This picture is explained in a particularly straightforward manner in [37].

Within this model, we find an explanation for the approximation that the interactions between the quasiparticles can be regarded in some sense as small. In particular, the restriction of conservation of energy and momentum, combined with the Pauli exclusion principle, implies that pairs of quasiparticles with generic configurations of momenta can only scatter into states with nearly the same momenta. This then restricts the ability of the interaction to modify the behavior of the system. For example, it means that the probability of a given quasiparticle decaying into a pair of quasiparticles and a hole with lower energy goes like $(E - E_F)^2$ (at low temperatures) or T^2 (at higher temperatures) [37]. We will see that the methods of effective field theory reproduce this result, and place it in a more general context.

3.3 Effective field theory

Let us now postulate that we can describe the behavior of a system with a Fermi surface and quasiparticles in terms of an effective field theory, where fermionic fields will correspond to the quasiparticles. We will work directly in momentum space and in two spatial dimensions, and we will mostly ignore the issues of the spins of the quasiparticles by making the assumption of spin-singlet interactions. This treatment will mostly follow [5] and [6]. The first works to revisit the Fermi liquid theory from the perspective of renormalization and effective field theory were [38] and [39]. Historically, the increased interest in this point of view stemmed from an attempt to understand the possible ways in which Fermi liquid theory could fail. This was due to the discovery of high temperature superconductivity [40], and the observation that the materials displaying high temperature superconductivity do indeed seem to posses a Fermi surface [17, 41], but nonetheless appear to have properties qualitatively and quantitatively different from what is expect within Fermi liquid theory [42].

In analogy with the dispersion relation Eq. (3.2) for the free Fermi gas, we take the kinetic term in the Lagrangian for the quasiparticles to be

$$\int \mathrm{d}\omega \, \mathrm{d}^2 k \psi^\dagger [\omega - \varepsilon(k)] \psi, \tag{3.3}$$

where the dispersions will be assumed to only depend on the perpendicular distance from the Fermi surface:

$$\varepsilon(k) = V_F k_\perp + O(k_\perp^2). \tag{3.4}$$

This will be justified momentarily by the observation that within the effective field theory, higher order terms in this expansion will be irrelevant in the RG sense. For simplicity, we will focus on a circular Fermi surface with a constant Fermi velocity, that is, a rotationally invariant theory. The conclusions of the analysis will remain the same in the case of a general round Fermi surface where the Fermi velocity is always nonzero, given that we use a system of coordinates where $k_\perp$ is the component of the momentum locally perpendicular to the Fermi surface.

Our previous discussion of the Fermi liquid detailed how the low-energy excitations of the system represent states in the vicinity of the Fermi surface. Therefore in order to construct a low-energy effective theory of this system, the high energy states (fast modes) that we integrate out during the renormalization process should be those far from the Fermi surface. We therefore cut off the theory such that the momenta lie in a band around the Fermi surface of width Λ/V_F. The renormalization process will

involve rescaling in the component of momentum perpendicular to this surface, then integrating out the modes at the edge of the allowed region of momentum space in order to return the cutoff to its original numerical value. As in our example system in Chapter 2, we will be relating the low-energy correlators (in this case, corresponding to correlators in the vicinity of the Fermi surface) to correlators of another theory. The new theory should then produce a simpler description of the system in terms of a small number of effective operators appearing in the Lagrangian.

The UV theory of this particular example is in fact not a quantum field theory at all. Instead, the degrees of freedom are the atoms inside the lattice of the material. This emphasizes the versatility of effective field theory, in that we may describe the low-energy dynamics of a system with a completely different character in terms of the effective fields of the quasiparticles.

3.4 Classical scaling

To understand the behavior of this theory under the renormalization procedure described above, consider a particular correlator with an action S_0 that is perturbative but otherwise arbitrary,

$$\langle \psi(\zeta^{-1}\omega, \zeta^{-1}p_\perp, \theta) \dots \rangle_{S_0}. \tag{3.5}$$

As in the example of the scalar φ^4 theory in the introductory chapter, $z > 1$ is a scale factor emphasizing that we are interested in the low energy modes of the system. The choice of the scaling, while arbitrary, is motivated by the observation that it will preserve the form of the kinetic term, allowing us to use the same propagator in the scaled and unscaled theories. Furthermore, it is physically motivated by the observation that if we wish to understand the low energy behavior of the system, we should focus on modes near the Fermi surface.

We perform an analogous rescaling of the modes in the path integral. The path integral expression for the correlator is

$$\frac{1}{Z}\int [\mathrm{D}\psi^\dagger\, \mathrm{D}\psi]_{-\Lambda}^{\Lambda} \psi(\zeta^{-1}\omega, \zeta^{-1}p_\perp, \theta) \dots \exp(\mathrm{i}S_0), \tag{3.6}$$

where

$$Z_= \int [\mathrm{D}\psi^\dagger\, \mathrm{D}\psi]_{-\Lambda}^{\Lambda} \exp(\mathrm{i}S_0). \tag{3.7}$$

The limits on the path integral measure emphasize that we must include momenta both above and below the Fermi surface to properly account for the presence of

holes. Also, we place limits on $\varepsilon(k) = V_F k_\perp$, so we include only the modes in momentum space with $-\Lambda < k_\perp < \Lambda$, where

$$k_\perp = |\vec{k} - K_F \hat{k}|/V_F. \tag{3.8}$$

The integral in the kinetic part of the action in terms of $k_\perp$ then looks like

$$\int \mathrm{d}\omega \int |k|\,\mathrm{d}|k|\,\mathrm{d}\theta = \int \mathrm{d}\omega \int_{-\Lambda/V_F}^{\Lambda/V_F} (k_\perp + K_F)\,\mathrm{d}k_\perp\,\mathrm{d}\theta. \tag{3.9}$$

The term that is linear in $k_\perp$ will scale more quickly than that involving K_F, so it will be irrelevant in the sense of the renormalization group. Hence we can drop it from the theory. We will choose to define the theory without a cutoff on energy because it is convenient and does not lead to additional divergences.

Performing a field redefinition

$$\psi(\zeta^{-1}\omega, \zeta^{-1}k_\perp) = \zeta\psi'(\omega, k_\perp), \tag{3.10}$$

$$\psi(\omega, k_\perp) = \zeta^{[\psi]}\psi'(\zeta\omega, \zeta k_\perp) \tag{3.11}$$

transforms the path integral measure and the kinetic term in the action:

$$[\mathrm{D}\psi^\dagger\,\mathrm{D}\psi]_{-\Lambda}^{\Lambda} = [\mathrm{D}\psi'^\dagger\,\mathrm{D}\psi']_{-\zeta\Lambda}^{\zeta\Lambda}, \tag{3.12}$$

$$S_\mathrm{k} = \int \mathrm{d}\omega\,\mathrm{d}^2k\,\zeta^{2[\psi]}\psi'(\zeta\omega, \zeta k_\perp)[\omega - V_F k_\perp]\psi(\zeta\omega, \zeta k_\perp). \tag{3.13}$$

Changing variables in the kinetic term to $k'_\perp = \zeta k_\perp$ and $\omega' = \zeta\omega$ and dropping the primes yields

$$S_\mathrm{k} = \zeta^{2[\psi]-3} \int \mathrm{d}\omega\,\mathrm{d}^2k\,\psi^\dagger[\omega - \varepsilon(k)]\psi, \tag{3.14}$$

which implies that we should choose the scaling dimension $[\psi] = 3/2$ for the fields in momentum space in order to preserve the kinetic term throughout the RG.

Let us analyze the scaling interactions in the theory. The general form is

$$\int (\mathrm{d}\omega\,\mathrm{d}^2k)^n \delta\left(\sum \omega\right) \delta^2\left(\sum k\right) g_{m,n,p} k_\perp^m \omega^p (\psi^\dagger\psi)^{n/2}. \tag{3.15}$$

The delta functions enforce conservation of energy and momentum. Crucially, we also define $g_{m,n,p}$ to be an arbitrary function of the angular coordinates of the particles; we will comment on this further below. The number of powers of ζ that multiply the interaction is

$$\begin{aligned}[g_{m,n,p}] &= -2n + 1 - m - p + \left[\delta^2\left(\sum k\right)\right] + \frac{3}{2}n \\ &= -\frac{n}{2} - (m+p) + 1 + \left[\delta^2\left(\sum k\right)\right].\end{aligned} \tag{3.16}$$

We have left the scaling of the momentum conserving delta function ambiguous for now. However, we should expect $0 \leq [\delta^2(k)] \leq 1$, because only the perpendicular component of each momentum is rescaled. The ambiguity of the scaling of this delta function will be the topic of the next section.

In the limit where the coupling constants are all small, we may neglect the effects arising from the integrated-out "fast" modes of the field, just as in Chapter 2. Eq. (3.16) then implies that in this approximation, the correlators of the scaled (IR) theory will then take exactly the same functional form as those in unscaled (UV) theory, but with $g_{m,n,p} \to \zeta^{[g_{m,n,p}]} g_{m,n,p}$. Given the constraint on the scaling of the delta function, all interactions with six or more copies of the field and any number of powers of $k_\perp$ or ω should be irrelevant under the rescaling, that is, $g_{m,n,p}$ is replaced by smaller and smaller values as we scale further into the IR. Furthermore, any interaction with four fields and one or more powers of $k_\perp$ or ω should be irrelevant, as is any interaction with two fields and two or more powers of $k_\perp$ or ω.

If we assume for now that the interaction involving four fields is marginal, we arrive at the conclusion that the most general action containing only relevant or marginal interactions (i.e. containing only terms that are not scaled to zero) is given by

$$S = S_{\mathrm{k}} + S_{\mathrm{i}}, \tag{3.17}$$

where

$$S_{\mathrm{k}} = \int \mathrm{d}\omega \, \mathrm{d}^2 k \psi^\dagger [\omega - V_F(\theta) k_\perp] \psi \tag{3.18}$$

and

$$S_{\mathrm{i}} = \int (\mathrm{d}\omega \, \mathrm{d}^2 k)^4 \delta\left(\sum \omega\right) \delta^2\left(\sum k\right) g(\theta_1, \theta_2, \theta_3, \theta_4) (\psi^\dagger \psi)^2. \tag{3.19}$$

If we enforce rotational invariance, $V_F(\theta)$ becomes independent of the angle of the particle. However, the coupling function will still depend on the three independent differences of the angular coordinates of the scattered particles.

A careful reader will notice that we have omitted a term from the Lagrangian: a momentum- and energy-independent constant multiplying $\psi^\dagger \psi$. Call the constant μ, so that the term in the Lagrangian is $\mu \psi^\dagger \psi$. From the perspective of an effective field theory, this term is particularly worrisome, because it is a relevant operator. In a "normal" relativistic field theory, this term would correspond to a mass for the particles, and its relevance would imply that it is unnatural for it to have a small value in the IR theory. However, the situation is different here.

Including this in the kinetic term of the Lagrangian yields

$$\psi^\dagger[\omega - V_F(\theta)k_\perp + \mu]\psi. \tag{3.20}$$

Consider the following field redefinition:

$$\psi(k_\perp) = \psi(k'_\perp - \mu/V_F). \tag{3.21}$$

The kinetic term becomes

$$\psi'^\dagger(k_\perp - \mu/V_F)[\omega - V_F(\theta)k_\perp + \mu]\psi'(k_\perp - \mu/V_F). \tag{3.22}$$

Changing variables to $k'_\perp = k_\perp - \mu/V_F$ then restores the original form of the kinetic term, but with the primed fields and perpendicular components of the momentum.

Thus, we can eliminate the constant term in the quadratic part of the action by making a field redefinition. What has happened here is that we have "shifted" the location of the true Fermi surface. In the vocabulary of condensed matter physics, μ corresponds to a shift in the chemical potential of the particles. The important point here is that the excitations of the system with the shifted chemical potential continue to remain gapless.

As we iterate the process of integrating out the high energy modes, we will have to continually shift the location of the Fermi surface via field redefinitions. However, the modifications to the location of the Fermi surface will always be smaller than the width of the band around the Fermi surface defined by the cutoff as long as the theory remains perturbative. This is because the corrections from loops to the chemical potential will always take the form

$$\delta\mu \sim g^n(\Lambda_0 - \Lambda). \tag{3.23}$$

Fig 3.1 demonstrates this shifting of the position of the Fermi surface.

It is important to observe that the kind of field redefinition used here would not be possible in the case of a massive relativistic particle. The simplest way to see why this could not work is to notice that if we tried to define a new momentum k' such that $k'^2 = k^2 + m^2$ (working in Euclideanized momentum space), the topology of the allowed modes of the field would change. Instead of looking like a ball in momentum space centered at the origin, the allowed space of modes would look like a hollow ball, with inner radius $k' = m$. As a result, the structure of the correlators of the redefined theory would not look massless at all. The point in this case is that the topology already looks like a hollow ball, so the correlators of the redefined fields will look like those of the original fields.

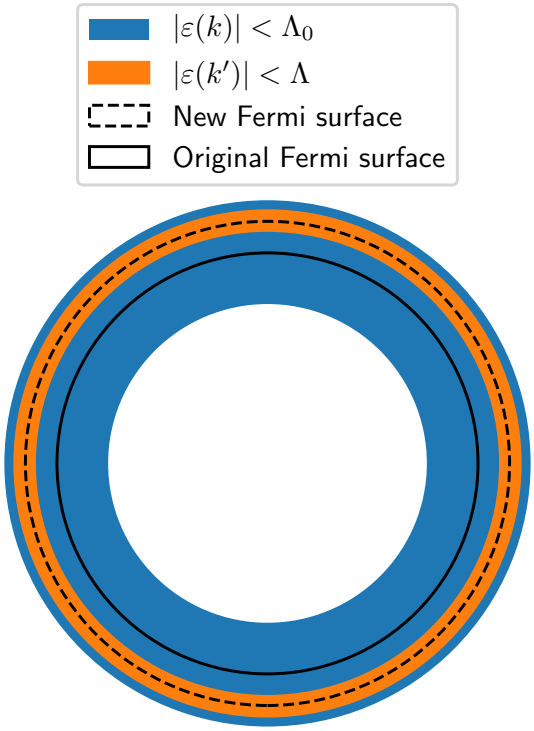


Figure 3.1: Shifting of the position of the Fermi surface due to renormalization of the chemical potential.

3.5 Issues with the scaling

Now we must return to the issue of the scaling of the momentum-conserving delta function. [6] and [5] have two different approaches to this problem. Polchinski claims that the scaling of the delta function differs depending on the relative orientation of the momenta of the interacting particles. To understand this point of view, let us write out the form of the two dimensional momentum-conserving delta function. Writing it in terms of $|k|$ and θ, the delta function in the k_x direction is

$$\delta(|k_1|\cos\theta_1 + |k_2|\cos\theta_2 - |k_3|\cos\theta_3 - |k_4|\cos\theta_4). \tag{3.24}$$

Replacing $|k| = k_\perp K_F$, the argument of this becomes

$$k_{1\perp}\cos\theta_1 + k_{2\perp}\cos\theta_2 - k_{3\perp}\cos\theta_3 - k_{4\perp}\cos\theta_4 + K_F(\cos\theta_1 + \cos\theta_2 - \cos\theta_3 - \cos\theta_4); \tag{3.25}$$

the k_y-conserving delta function has the same argument with sin $\to$ cos.

His reasoning runs as follows: we may assume that the $k_\perp$ are small compared to K_F since we are interested in the behavior of the theory deep in the IR, where only

a thin band of modes around the Fermi surface remains. Then we may drop the $k_\perp$ from Eq. (3.25), presumably in some sort of multipole expansion. In this case the delta function no longer depends on the $k_\perp$ at all, and therefore it will not scale, so $[\delta^2(k)] = 0$ and the interaction is irrelevant.

However, exceptional behavior occurs when we take two of the particles to be approximately back-to-back, so that (for example) $\theta_1 = \theta_2 + \pi$. In that case, the argument k_x delta function becomes

$$(k_{1\perp} - k_{2\perp}) \cos\theta_1 - k_{3\perp} \cos\theta_3 - k_{4\perp} \cos\theta_4 - K_F(\cos\theta_3 + \cos\theta_4). \tag{3.26}$$

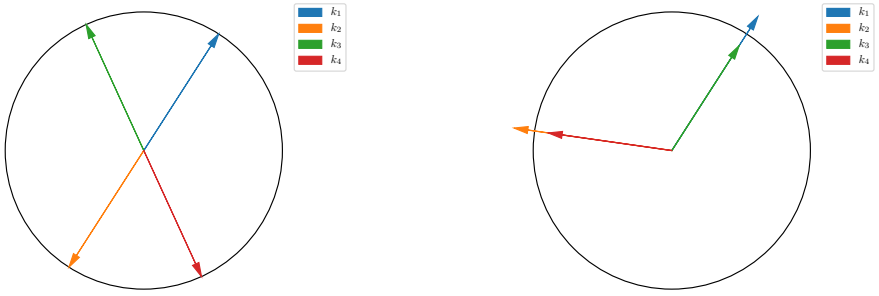
Once again dropping the $k_\perp$ parts of this expression relative to K_F, we see that the delta function constrains $\cos\theta_3 = -\cos\theta_4$. This is possible if either $\theta_3 = \theta_4 + \pi$ or $\theta_3 = \pi - \theta_4$. It is generically impossible to satisfy the second condition if all of the momenta lie in a thin band around the Fermi surface. Therefore the k_x delta function alone constrains the momenta of particles 3 and 4 to point in opposite directions, given that the momenta of 1 and 2 are opposite. The argument of the k_y delta function then becomes

$$(k_{1\perp} - k_{2\perp}) \sin\theta_1 - (k_{3\perp} - k_{4\perp}) \sin\theta_4. \tag{3.27}$$

In this expression, the dependence on K_F has entirely dropped out. Therefore the k_y delta function must scale with $k_\perp$, and we have $[\delta^2(k)] = 1$. Note that if we had first used the k_y delta function, the roles of the two delta functions would have been reversed, but the result would have been the same.

In general, this argument implies that the interaction is marginal only in certain kinematic configurations. In particular, the two incoming particles must be back-to-back, so that $k_1 = -k_2$, or we must have forward scattering, where $k_1 = k_3$ or $k_1 = k_4$.

While the logic here is initially convincing, trying to make the argument more rigorous reveals several problems. First, what precisely is the meaning of "taking" $\theta_1 = \theta_2 + \pi$ in the action? Second, what is the nature of the expansion of the delta function that allows us to neglect the perpendicular components of the momenta? Finally, what happens if we simultaneously "take" $\theta_1 = \theta_2 + \pi$ and $\theta_3 = \theta_4 + \pi$? This final issue is particularly troublesome, because under those circumstances, it appears that both the k_x and k_y delta functions will scale with $k_\perp$. This would mean $[\delta^2(k)] = 2$, resulting in a **relevant** interaction. It turns out that this issue is



(a) Back-to-back scattering (b) Forward scattering

Figure 3.2: The two possible configurations of momentum that allow for the conservation of momentum in the limit where the cutoff is taken to zero. The arrows in the figure on the right are of different lengths only to make them visible in the figure.

particularly subtle for a theory with a Fermi surface with a Van Hove singularity, the topic of the next chapter.

In light of these difficulties, let us turn to the point of view put forward in [5]. In that work, the delta function is used to evaluate the integral over the fourth particle's momentum before the scaling analysis. If we do this, the interaction becomes

$$\int (\mathrm{d}\omega\, \mathrm{d}^2k)^3 g(\theta_i)\theta(\Lambda - V_F|K_F - |k_1 + k_2 - k_3||) \times \psi^\dagger(k_1)\psi^\dagger(k_2)\psi(k_3)\psi(k_1 + k_2 - k_3). \quad (3.28)$$

The step function is a result of the constraint that the fourth momentum must also reside within a small band around the Fermi surface; it is a residual effect of the cutoffs in momentum space for the integrated variables.

Consider generic values of θ_1 and θ_2 when they are not back-to-back. Unless θ_3 is in the vicinity of θ_1 or θ_2, it will be impossible to satisfy the constraints in the step function. This is most easily seen by taking the limit $\Lambda \to 0$, so that the only allowed momenta lie on the Fermi surface itself. The we see that if we add three arbitrary momenta lying on a circle (corresponding here to k_1, k_2, and $-k_3$), the resulting vector $-k_4$ will not lie on the circle at all. This means that the step function forces k_3 to be approximately equal to k_1 or k_2 — the other interactions are simply not allowed. The only exception occurs when k_1 and k_2 are back-to-back. In that case, the step function will be automatically satisfied as long as k_3 is within the

small band around the Fermi surface. The remaining allowed interactions (Fig. 3.2) are the same as the marginal interactions found by Polchinski.

From Shankar's point of view, then, it is simply not possible to ask about low energy scattering events between particles at arbitrary positions on the Fermi surface, because momentum conservation will always force one of the particles to be very energetic unless they are in one of the special kinematic configurations described above. What about the scaling? The kinetic term remains the same in this version of the theory, but the measure in the interaction Eq. (3.28) has changed, and the delta function is gone. If we simply naively count count the scaling dimension of the remaining terms in the action and the fields, we find that the four point interaction goes like

$$[\mathrm{d}\omega\, \mathrm{d}k_\perp] \times 3 + [\psi] \times 4 = -2 \times 3 + \frac{3}{2} \times 4 = 0. \tag{3.29}$$

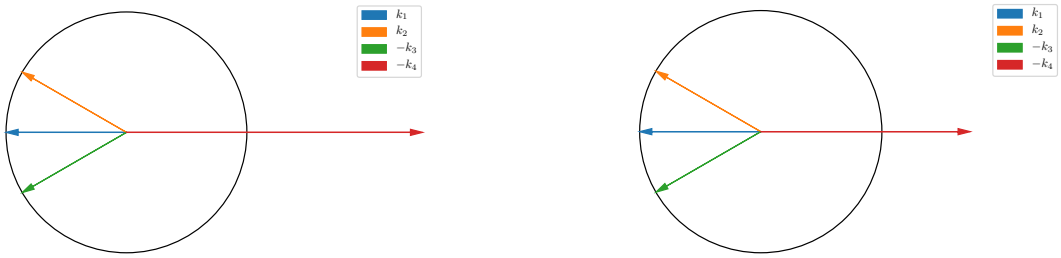
Thus, it appears to be marginal. According to this, we should simply regard any interaction that satisfies the momentum constraints as marginal. From this point of view, all other interactions have already been integrated out of the theory.

Unfortunately, there is also a problem here. Instead of choosing to evaluate both the θ_4 and $k_{4\perp}$ integrals using the two delta functions in Eq. (3.19), imagine choosing to perform the θ_3 and θ_4 integrals. Then there are no remaining delta functions in the action, but the measure involves $\mathrm{d}^4 k_\perp$. Thus, if we follow the same logic as in the argument above, we see that the scaling dimension of the interaction should be

$$[\mathrm{d}\omega] \times 3 + [\mathrm{d}k_\perp] \times 4 + [\psi] \times 4 = -1. \tag{3.30}$$

We get a different answer for the scaling just because we evaluated the delta functions differently! Furthermore, we could instead choose to evaluate two of the integrals under the perpendicular components of the momentum. This would imply that all interactions are relevant, which would make our effective theory useless.

Clearly, something has gone wrong here. It should not be possible to alter the scaling dimension simply by choosing to evaluate the expression for the action in a different way. The resolution to this issue is to realize that we have forgotten a step in the RG process: we must perform a field redefinition to return the action to its original form. For the first three fields in the action, with momenta k_1, k_2, and k_3, this is perfectly fine — the field redefinition necessary to return the kinetic term in the action to its original form acts correctly on those fields. However, the argument of the fourth field does not scale correctly for generic values of k_1, k_2, and k_3.



(a) Before rescaling (b) After rescaling

Figure 3.3: In general, an inhomogeneous scaling of the momenta is not compatible with conservation of momentum. In this figure, we see the effect of scaling momenta towards the Fermi surface.

It is simplest to illustrate this with a picture. For illustrative purposes, consider a theory where there is no UV cutoff, but we still scale the momenta towards the Fermi surface. Within this theory, consider the interaction between the four fields in Fig. 3.3a. k_1, k_2, and $-k_3$ are on the Fermi surface but point in approximately the same direction. This implies $-k_4$ must be in nearly the opposite direction and be far from the Fermi surface. Upon rescaling (Fig. 3.3b), the first three momenta remain on the Fermi surface, since $k_\perp = 0$ for them already. However, $k_{4\perp}$ is rescaled drastically. Therefore the act of scaling momentum towards the Fermi surface is incompatible with conserving momentum for general scattering events.

While this is an extreme example, the conclusion holds more generally. In fact, the only interactions for which momentum is still conserved after the rescaling described above are back-to-back and forward scattering.

To understand more precisely what is going on, let us look directly at the scaling of the tree-level corrections to the four point correlator in a momentum conserving theory at a fixed scale. This is analogous to performing the expansion of the Feynman rules for the propagator in more standard EFTs, such as in HQET [43], to determine the leading order terms in the propagators for the effective Lagrangian. We can think of this as inserting the interaction vertex and four propagators inside some larger Feynman diagram, and then seeing how this subdiagram scales as we reduce the momentum of the system.

To follow this plan, consider the scaling of the product of four propagators in momentum space, with the fourth propagator's momentum k_4 set equal to the $k_1 + k_2 - k_3$. Call this quantity G_4/g. This is basically the **unamputated** tree-level

contribution to the Green's function of the theory. It will scale like

$$G_4/g \sim \frac{1}{k_{1\perp}}\frac{1}{k_{2\perp}}\frac{1}{k_{3\perp}}\frac{1}{|k_1+k_2-k_3|-K_F}. \tag{3.31}$$

The denominator of the final factor is the interesting one. Define

$$k_{4\perp} = \sqrt{k_1^2+k_1^2+k_1^2+2k_1\cdot k_2-2k_1\cdot k_3-2k_2\cdot k_3}. \tag{3.32}$$

Note that

$$|k_1| = K_F + k_{1\perp}. \tag{3.33}$$

To understand the scaling of Eq. (3.32), we may take $k_{1\perp} = k_{2\perp} = k_{3\perp} = k$. Then

$$k_{4\perp} \sim (K_F + k)\sqrt{3+2\chi}, \tag{3.34}$$

where

$$\chi = \cos\theta_{12} - \cos\theta_{13} - \cos\theta_{23}. \tag{3.35}$$

Note that

$$\theta_{23} = \theta_{13} - \theta_{12}. \tag{3.36}$$

We can write

$$k_{4\perp} = K_F(\sqrt{3+2\chi}-1) + k\sqrt{3+2\chi}. \tag{3.37}$$

From this it is clear that for generic values of χ, the first term dominates and $k_{4\perp}$ does not scale with k at all. On the other hand, if

$$\chi(\theta_{12},\theta_{13}) = -1, \tag{3.38}$$

the first term will drop out of the expression. Plotting χ in terms of θ_{12} and θ_{13}, Fig. 3.4, immediately yields the solutions

$$\theta_{12} = \theta_{13}, \tag{3.39}$$

$$\theta_{13} = 0, \tag{3.40}$$

with θ_{12} any value, and

$$\theta_{12} = \pi, \tag{3.41}$$

with θ_{13} any value. Eq. (3.39) and Eq. (3.40) correspond to forward scattering and Eq. (3.41) corresponds to back-to-back scattering.

The upshot of this discussion is the following: the contributions of scattering events that do not satisfy one of the conditions in Eqs. [(3.39)-(3.41)] are suppressed relative

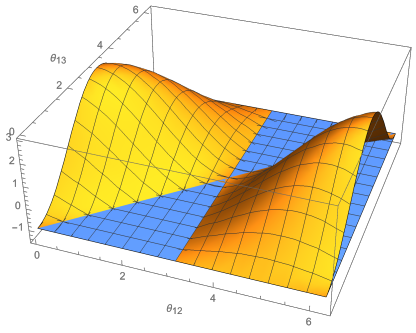


Figure 3.4: Graphical solution to the condition of Eq. (3.38). The points where more than one solution holds are saddle points.

to those that do. Thus, to reproduce the behavior of the theory at sufficiently low energies, we may simply set the coupling function to zero outside a small region in θ-space around the solutions to Eq. (3.38).

To find the size of this region and understand the suppressed corrections to the theory, we should expand Eq. (3.37) around the solutions. Consider first the case $\theta_{12} = \theta_{13} + \delta$. Then

$$\chi = \cos\theta_{12} - \cos(\theta_{12} + \delta) - \cos\delta. \tag{3.42}$$

Expanding to second order in δ,

$$\cos(\theta_{12} + \delta) = \cos\theta_{12}\cos\delta - \sin\theta_{12}\sin\delta \approx \left(1 - \frac{\delta^2}{2}\right)\cos\theta_{12} - \delta\sin\theta_{12}. \tag{3.43}$$

Therefore

$$\chi \approx -1 + \delta\sin\theta_{12} + \frac{1}{2}\delta^2(1 + \cos\theta_{12}) \tag{3.44}$$

and

$$\sqrt{3 + 2\chi} \approx 1 + \delta\sin\theta_{12} + \frac{1}{2}\delta^2(1 + \cos\theta_{12} - \sin^2\theta_{12}). \tag{3.45}$$

Keeping only the first order terms in k and δ for now,

$$k_{4\perp} \sim K_F\delta\sin\theta_{12} + k. \tag{3.46}$$

This implies that $k_{4\perp}$ scales like k under the condition

$$|\delta| \lesssim \frac{k}{K_F\sin\theta_{12}}. \tag{3.47}$$

The case $\theta_{13} = \delta'$ yields

$$\chi = \cos\theta_{12} - \cos\delta' - \cos\theta_{12} - \delta'. \tag{3.48}$$

This is identical to Eq. (3.42) up to a minus sign, and therefore Eq. (3.47) holds for the other type of forward scattering also.

Taking $\theta_{12} = \pi + \delta_{12}$ gives us

$$\chi = -\cos\delta_{12} - \cos\theta_{13} + \cos(\theta_{13} - \delta_{12}). \tag{3.49}$$

Then

$$\sqrt{3 + 2\chi} \approx 1 + \delta_{12}\sin\theta_{13} + \frac{1}{2}\delta_{12}^2(1 - \cos\theta_{13} - \sin^2\theta_{13}), \tag{3.50}$$

and the expression analogous to Eq. (3.46) is

$$k_{4\perp} \sim K_F\delta_{12}\sin\theta_{13} + k. \tag{3.51}$$

We are trying to find the effects of setting the coupling function to zero outside a small range parameterized by δ, δ', and δ_{12}. Take, for example, the case of near-forward scattering. The ratio of the unamputated Green's function for small δ to the unamputated Green's function at $\delta = 0$ exactly is

$$\frac{1/(K_F\delta\sin\theta_{12} + k)}{1/k} = \frac{k}{K_F\delta\sin\theta_{12} + k}. \tag{3.52}$$

If we take $\delta > \Lambda/K_F$ (which is still a small number), we have

$$\frac{k}{\Lambda\sin\theta_{12} + k} \sim \frac{k}{\Lambda} \tag{3.53}$$

as long as $\sin\theta_{12}$ is not too small. This means that we only omit terms of order k/Λ from the theory if we set the coupling function to zero for $\delta > \Lambda/K_F$. An analogous statement holds for back-to-back scattering.

The preceding logic demonstrates that we can capture the leading order behavior of the theory by setting the coupling function to zero for any interaction that is not within Λ/K_F of a special kinematic configuration. We would also like to craft an effective theory that naturally handles conservation of momentum correctly. One possible way of formulating this is to use an exact RG, which works by constructing differential equations relating the flows of the exact vertices directly [44]. We instead appeal to the example of EFTs that deal with these issues by binning modes in momentum space.

3.6 Binning

In an EFT that implements binning, momentum space is divided into a large number of discrete cells or bins. The position of the center of each such bin is denoted by

a large momentum known as the label momentum, and the positions in momentum space within the bin are written as a sum of the large label momentum and a small residual momentum. Generally, it is this small residual momentum that is scaled in the theory, although in some instances the label momenta are also scaled.

The the idea of this binning in momentum space was first advocated in [45] and [46] as a way to analyze the decays of hadrons with a single heavy quark. The motivation for these effective field theories was the heavy quark approximation, pioneered in such papers as [47, 48], in which the heavy constituent quark was treated as a nonrelativistic field interacting with the "brown muck" of QCD excitations bubbling around it [25]. The expansion parameter in such a theory is the mass of the heavy quark, and the fundamental observation allowing for the simplification of the system is that small transfers of momentum cannot alter the velocity of that quark to leading order. This, then, implies that as long as there is a large separation of scales between Λ_{QCD}, the momentum scale describing the light quark and gluon fields interacting with the heavy quark, and m, the mass of the heavy quark, we can regard the heavy quark as moving along at a constant velocity. See [43, 49] for a comprehensive overview of the implications.

Georgi [46] in particular emphasized that we can construct this effective theory through the idea of binning. In this picture, the size of the each small bin in momentum space is set by Λ_{QCD}. These bins are then labeled by the large momentum of the constituent heavy quark of the system. We can then remove this large momentum with a field redefinition. As a result, the antiquark field decouples from the dynamics of the system, and we may integrate it out. The remaining degrees of freedom all live within the small box in momentum space, and the resulting effective field theory allows for calculation of various properties of the system.

HQET suffered from the inability to simultaneously treat bound states of multiple heavy quarks. Another effective field theory, NRQCD, was formulated [50, 51] to address this situation. Unfortunately, this formulation of the effective field theory did not make the power-counting of the fields explicit at the level of the Lagrangian. [31] showed a way to count powers consistently in a small parameter v, the velocity of the bound state of the system. However, there are issues with treating on-shell gluons using this method. Another approach, pioneered in [33], advocated using a different coordinate rescaling for the transverse components of the gluon field in Coulomb gauge to deal with these issues. This then calls for a multipole expansion of the transverse components, and in turn momentum is not conserved at the vertices

of the theory, much in the way that we encountered in the case of the Fermi surface. However, this treatment fails to correctly account for potential scattering. [32] resolved the conflict by including a second gluon field in the effective field theory.

Eventually, the idea to divide momentum space into bins was applied to NRQCD [52]. By implementing the binning, it was possible to recast the Coulomb potential as a local interaction depending only on the label momenta of the fields. This demonstrated that the behavior of an effective theory can in the end depend nonanalytically on the label momenta of the fields while remaining analytic (and hence local) in the residual momentum. This guarantees that we can make predictions within this framework, despite a more complicated functional dependence of the interactions. We will see an example of this phenomenon in the case of Fermi surfaces with Van Hove singularities. Soft-collinear effective theory (SCET), used to describing situations with a light QCD particle boosted to a very large momentum, such as the large three-momentum of a light constituent of a decaying heavy quark or an energetic jet in a collider experiments, is yet another example of a theory that organizes the modes of the field into bins in momentum space [3, 53]. A novel feature of that theory is the necessity of introducing additional cutoffs separating the various modes in momentum space [4, 54]. We will see a similar feature arise in the next chapter.

3.7 Binning for the round Fermi surface EFT

The idea of dividing the momentum space around the Fermi surface into bins was already implied in Shankar's original treatment [5]. This is apparent from his reference to the behavior of the theory as a sort of large-N limit, where the N in question corresponds to the inverse of the size of the bins. [55] is another example of a similar concept, where the Fermi surface is "sectorized" and the renormalization group procedure is discretized in order to prove rigorously that the diagrammatic expansions for a system with a Fermi surface have nonzero radii of convergence, under certain circumstances. Other works analyzed the problem of scaling the momentum towards a single point on the Fermi surface, with an eye towards understanding other potential instabilities of the Fermi liquid, such as coupling to an emergent gauge field [56–58]. In [59], Chapter 18, this scaling towards a single point is also used to demonstrate the small but nonzero width of the quasiparticles. However, these treatments do not accurately capture the effects associated with back-to-back scattering. Here, we wish to treat the subject of the binned effective theory with slightly more precision.

3.8 Momentum scaling

We will break up the momentum a la SCET:

$$\vec{k} = \vec{K} + \vec{k}_{\parallel} + \vec{k}_{\perp}, \tag{3.54}$$

where

$$\vec{K} = K_F(\hat{x}\cos\theta_{\vec{K}} + \hat{y}\sin\theta_{\vec{K}}) \tag{3.55}$$

is a label momentum on the Fermi surface, $\vec{k}_{\parallel}$ is tangent to the Fermi surface, and $\vec{k}_{\perp}$ is perpendicular to the Fermi surface. We can think of $\vec{K}$ as taking on a discrete set of values denoting the centers of boxes in momentum space lying on the Fermi surface. In analogy to HQET/NRQCD/SCET, we write the quasiparticle field as a sum of fields with the large component of the momenta on the Fermi surface explicitly factored out:

$$\psi = \sum_{\vec{K}} \exp\left(-\mathrm{i}\vec{K}\cdot\vec{x}\right)\psi_{\vec{K}}. \tag{3.56}$$

In principle, we should eventually take $\vec{K}$ to be continuous and then mod out the extra dimensions of momentum space as in SCET. Practically, this corresponds to resumming the bins into an integral when performing loop calculations while being careful to avoid overcounting, which can be rather subtle due to the presence of so-called zero bins [60]. Fortunately, there will not be any such complications for us.

Drawing upon our experience in the Section 3.4, the kinetic term in the action is

$$S_{\mathrm{k}} = \sum_{\vec{K}} \int \mathrm{d}\omega\, \mathrm{d}k_{\perp}\, \mathrm{d}k_{\parallel} \mathcal{L}_{\vec{K}}, \tag{3.57}$$

$$\mathcal{L}_{\vec{K}} = \psi^{\dagger}_{\vec{K}}\left(\omega + V_F k_{\perp}\right)\psi_{\vec{K}}. \tag{3.58}$$

Note the sum over the large label momenta. We will not carefully track the factors of π appearing in the action in these sections, since we are focused only on the scaling behavior of the fields and interactions.

Following the same procedure as our toy example in Chapter 2, we state that we wish to investigate the behavior of correlators of fields near the Fermi surface. This means taking $k_{\perp}$ in particular to be small. In order to conserve the residual momentum of interactions on the Fermi surface, we are also forced to take the residual momentum $k_{\parallel}$ to scale in the same manner. This is a consequence of the fact that the direction perpendicular to the Fermi surface changes as we move along it. We also want to

maintain the form of the propagator, so we scale the energy of the fields in the same way as the residual momenta. The correlators we are interested in look like

$$G_N(\zeta k) = \langle \psi_{\vec{K}}(\zeta\omega, \zeta k_\perp, \zeta k_\perp) \ldots \rangle_{S_0}, \tag{3.59}$$

where $\zeta < 1$ is a small scale factor.

The path integral expression for this then looks like

$$G_N(\zeta k) = \frac{1}{Z} \int \mathrm{D} \prod_{\vec{K}_i} [\psi^\dagger_{\vec{K}_i} \psi_{\vec{K}_i}] \psi_{\vec{K}}(\zeta k) \ldots \exp(\mathrm{i}S_0), \tag{3.60}$$

where

$$Z = \int \mathrm{D} \prod_{\vec{K}_i} [\psi^\dagger_{\vec{K}_i} \psi_{\vec{K}_i}] \exp(\mathrm{i}S_0) \tag{3.61}$$

and we have suppressed various labels. Our goal is to relate this correlator to that of the IR theory between modes with momenta k. Therefore we perform a field redefinition,

$$\psi(\zeta k) = \zeta^{[\psi]} \psi'(k). \tag{3.62}$$

The infinite product in the measure resulting from the field redefinition cancels, but we are left an overall scaling of the Green's function due to the extra copies of the fields in the numerator. We will see after the following manipulations that

$$G_N(\zeta k) = \zeta^{N[\psi]} \langle \psi_{\vec{K}}(\omega, k_\perp, k_\perp) \ldots \rangle_S. \tag{3.63}$$

In other words, the correlators of fields near the Fermi surface in the original theory will be given by the correlators of a theory with rescaled momenta and coupling constants. We ignore (for now) the effects of the field strength renormalization. This is handled via another field redefinition. We will choose $[\psi]$ in order to make the kinetic term invariant under the scaling procedure.

As we reduce the size of the bins in momentum space, we must simultaneously increase the number of bins in order to continue to cover the entire Fermi surface. More precisely, upon rescaling the residual momenta, the maximum value of both $k_\perp$ and $k_\parallel$ within an individual bin is increased. We only integrate out the high frequency modes with large values of $k_\perp$ and not those with large value of $k_\parallel$. This is because we do not wish to integrate out the soft particles from the theory, which would in general produce nonlocality at the scale of the residual momenta. However, we wish to keep each bin approximately square so that the form of the theory remains invariant under the rescaling procedure. This means that we must repartition the

modes into a larger number of bins as the scaling procedure progresses, effectively increasing the value of K_F. The width of the bins along the direction of the Fermi surface is set by the scaling of the momentum tangent to the Fermi surface, so we must take

$$[\Sigma] = -[k_\parallel] = -1. \tag{3.64}$$

The brackets here represent the power of ζ that accompanies the enclosed term upon the RG transformation in the action, so for example

$$k_\parallel \to \zeta^1 k_\parallel. \tag{3.65}$$

We choose $[\psi]$ to render the kinetic term in the action invariant under the scaling, which will allow us to use the same propagator in the calculation of observables. Counting the powers of the scale factor ζ, we have

$$[\Sigma] + [\omega] + [k_\perp] + [k_\parallel] + [\omega] + 2[\psi] = 0, \tag{3.66}$$

which implies

$$[\psi] = -\frac{3}{2}. \tag{3.67}$$

3.9 Interaction terms

Now let us turn to the scaling of the four-point interaction term. We first investigate an interaction that can depend arbitrarily on the four label momenta of the fields, but has no dependence on the residual momenta or energies. This is given by

$$S_{\mathrm{i}} = \sum_{\vec{K}_1,\vec{K}_2,\vec{K}_3,\vec{K}_4} \delta(\vec{K}_1+\vec{K}_2-\vec{K}_3-\vec{K}_4) \int (\mathrm{d}\omega\, \mathrm{d}k_\perp\, \mathrm{d}k_\parallel)^3 g(\vec{K}_i) \psi^\dagger_{\vec{K}_1} \psi^\dagger_{\vec{K}_2} \psi_{\vec{K}_3} \psi_{\vec{K}_4}. \tag{3.68}$$

We have already evaluated a delta function ensuring conservation of energy and residual momentum. The remaining delta function in Eq. (3.68) enforces conservation of the large label momentum, and is therefore not literally a Dirac delta function.

The shape of the Fermi surface has a profound effect on the conservation of label momentum. In particular, instead of the sum over label momentum reducing to a sum over three arbitrary momenta with the fourth set by conservation, we end up with three sums over two momenta. This is because the sum of three generic momenta on the Fermi surface will not lie on the Fermi surface — there is an additional constraint. This is just Shankar's observation that only back-to-back and forward scattering are allowed. The three sums correspond to taking three different

pairs of particles to be back-to-back: 1 and 2, 1 and 3, and 1 and 4. In order to avoid overcounting it is convenient to isolate the repeated terms in these sums. These represent interactions in which all four particles are collinear. In other words, in these interactions, two pairs of particles share the same label momenta, and the pairs are also back-to-back. There are then two sums with a single free label momentum:

$$\begin{aligned}\sum_{\vec{K}_1,\vec{K}_2,\vec{K}_3,\vec{K}_4} &\delta(\vec{K}_1+\vec{K}_2-\vec{K}_3-\vec{K}_4)\psi^\dagger_{\vec{K}_1}\psi^\dagger_{\vec{K}_2}\psi_{\vec{K}_3}\psi_{\vec{K}_4} \\ &= \sum_{\vec{K}_1,\vec{K}_2\neq\vec{K}_1}\psi^\dagger_{\vec{K}_1}\psi^\dagger_{-\vec{K}_1}\psi_{\vec{K}_2}\psi_{-\vec{K}_2} + \sum_{\vec{K}_1,\vec{K}_2\neq\vec{K}_1}\psi^\dagger_{\vec{K}_1}\psi^\dagger_{\vec{K}_2}\psi_{\vec{K}_1}\psi_{\vec{K}_2} \\ &\quad + \sum_{\vec{K}_1,\vec{K}_2\neq\vec{K}_1}\psi^\dagger_{\vec{K}_1}\psi^\dagger_{\vec{K}_2}\psi_{\vec{K}_2}\psi_{\vec{K}_1} + \sum_{\vec{K}_1}\psi^\dagger_{\vec{K}_1}\psi^\dagger_{-\vec{K}_1}\psi_{\vec{K}_1}\psi_{-\vec{K}_1} \\ &\quad + \sum_{\vec{K}_1}\psi^\dagger_{\vec{K}_1}\psi^\dagger_{\vec{K}_1}\psi_{-\vec{K}_1}\psi_{-\vec{K}_1}. \end{aligned} \tag{3.69}$$

We will consider the case of a spin-singlet interaction here, which allows us to combine the second, third, and fifth terms in the above sum into a generic forward-scattering sum,

$$\sum_{\vec{K}_1,\vec{K}_2\neq-\vec{K}_1}\psi^\dagger_{\vec{K}_1}\psi^\dagger_{\vec{K}_2}\psi_{\vec{K}_1}\psi_{\vec{K}_2}. \tag{3.70}$$

Note that we have separated out the overlap between forward and back-to-back scattering in the fourth term in Eq. (3.69), which includes only a single sum over momenta.

The scaling of the double-sum interactions is

$$2[\Sigma] + 3([\omega] + [k_\perp] + [k_\parallel]) + 4[\psi] = -2 + 9 - 6 = 1, \tag{3.71}$$

and the scaling of the single-sum interaction is

$$[\Sigma] + 3([\omega] + [k_\perp] + [k_\parallel]) + 4[\psi] = 2. \tag{3.72}$$

This implies that within the label-momentum formalism described here, the four-point interactions are all irrelevant.

Interactions with more fields will be even more irrelevant. To see this, note that label-momentum conservation implies that all fields must come in pairs with either the same or opposite label momentum. This is due to the curvature of the Fermi surface and the fact that the residual momentum cannot alter the label momentum.

The power counting for a generic interaction of $2n$ fields with the maximal number of sums is

$$n[\Sigma] + (2n-1)([\omega] + [k_\perp] + [k_\parallel]) + 2n[\psi] = -n + 6n - 3 - 3n = 2n - 3. \quad (3.73)$$

More powers of ω or $k_\perp$ or fewer sums (i.e. more restricted kinematic configurations) only makes the interaction more irrelevant.

Including only the least irrelevant interactions, we have for the action of the IR theory

$$S_{\rm IR} = S_{\rm k} + S_{\rm B} + S_{\rm F} + S_{\rm BF}, \quad (3.74)$$

where

$$S_{\rm B} = \sum_{K_1, K_2 \neq K_1} \int (\mathrm{d}\omega\, \mathrm{d}^2 k)^3 V(K_1, K_2) \psi^\dagger_{\vec{K}_1} \psi^\dagger_{-\vec{K}_1} \psi_{\vec{K}_2} \psi_{-\vec{K}_2}, \quad (3.75)$$

$$S_{\rm F} = \sum_{K_1, K_2 \neq -K_1} \int (\mathrm{d}\omega\, \mathrm{d}^2 k)^3 F(K_1, K_2) \psi^\dagger_{\vec{K}_1} \psi^\dagger_{\vec{K}_2} \psi_{\vec{K}_1} \psi_{\vec{K}_2}, \quad (3.76)$$

$$S_{\rm BF} = \sum_{K_1} \int (\mathrm{d}\omega\, \mathrm{d}^2 k)^3 g(K_1) \psi^\dagger_{\vec{K}_1} \psi^\dagger_{-\vec{K}_1} \psi_{\vec{K}_1} \psi_{-\vec{K}_1}. \quad (3.77)$$

Technically Eq. (3.77) is power-suppressed relative to the other interactions. We include it here because the discussion of loop effects would be less clear without it. Note that we have adopted the naming convention of [5] for the back-to-back and forward interactions.

An important feature of effective theories with binning is that the couplings can be generic (even nonanalytic) functions of the label momentum. An example of this occurs in NRQCD, where the potential interaction between quarks is nonanalytic in the transfer label momentum [52]. Therefore there is no a priori constraint on the forms of F and V in the above expression. This will have important ramifications when we turn to the analysis of systems with a Van Hove singularity.

3.10 A puzzle

Eq. (3.72) and Eq. (3.71) indicate that all possible interactions in our theory are irrelevant. What has happened? Our earlier analysis suggested that we should still recover unusual behavior in the back-to-back and forward scattering amplitudes, and we have found a result that disagrees with the two most famous analyses of the problem. We do indeed have different behavior for those interactions: we have input by hand the condition that these are the only interactions allowed by conservation of label momentum. However, in the process of resolving the issue of homogeneously

scaling the delta functions in the interaction (and ensuring locality and translation invariance in the in the residual, IR coordinates), we have smoothed out the singular behavior that arose in [6].

3.11 The back-to-back interaction

To understand the resolution of this issue, we must calculate the one-loop results. Let us start first with the correction to the back-to-back scattering amplitude. We will make the assumption of a constant coupling V and a spin-singlet interaction; this will not affect the reasoning that follows. Consider first the one-loop s-channel amplitude with two copies of the back-to-back vertices. After computing the ω integral via contours, the amplitude is

$$\mathcal{A}_{\mathrm{S,B}} = -\frac{V^2}{4\pi^2}\sum_P \int d^2k \frac{\theta(\varepsilon_P(k))\theta(\varepsilon_{-P}(K-k)) - \theta(-\varepsilon_P(k))\theta(-\varepsilon_{-P}(K-k))}{\varepsilon_P(k) + \varepsilon_{-P}(K-k) - E - i\epsilon\,\mathrm{sign}\,\varepsilon_P(k)}; \tag{3.78}$$

see Appendix B. K in the above equation is the sum of the residual momenta of the external particles and E is the net energy. P is the label momentum of the particle in the loop. Substituting the appropriate expression for $\varepsilon_K(k)$ for the loop particles and remembering that they are in the vicinity of opposite sides of the Fermi surface, we have

$$\mathcal{A}_{\mathrm{S,B}} = -\frac{V^2}{4\pi^2}\sum_P \int \mathrm{d}k_\parallel I_\perp, \tag{3.79}$$

where

$$I_\perp = \int_{-\Lambda/V_F}^{\Lambda/V_F} \mathrm{d}k_\perp \frac{\theta(k_\perp)\theta((k-K)_\perp) - \theta(-k_\perp)\theta(-(k-K)_\perp)}{V_F(2k-K)_\perp - E - i\epsilon\,\mathrm{sign}\,\varepsilon_P(k)}. \tag{3.80}$$

The key to our quandary is the label momentum sum that appears in Eq. (3.78). Under rescaling, this must grow by a factor of ζ^{-1}, because the effective field theory must contain a larger number of bins to cover the fixed Fermi surface in momentum space. Another way to say this is that the circumference of the Fermi surface must grow larger when measured in the units of the residual momentum, so

$$K_F \to K_F' = \zeta^{-1} K_F \tag{3.81}$$

as we rescale.

Continuing the calculation, we recombine the label momentum sum with the parallel momentum integral as in NRQCD [52] to find

$$\mathcal{A}_{\mathrm{S,B}} = -\frac{V^2}{4\pi^2}\int K_F\,\mathrm{d}\theta I_\perp(E,K). \tag{3.82}$$

The omission of the single term in the sum corresponding to S_{BF} does not affect this result to leading order, since in the deep IR limit this corresponds to removing a single point from the angular integration measure. Take $K = 0$ and change variables to $k = 2V_F k_\perp$. Then

$$I_\perp(E, 0) = \frac{1}{2V_F} \int_0^{2\Lambda} \frac{\mathrm{d}k}{k - E - \mathrm{i}\epsilon} - \frac{1}{2V_F} \int_{-2\Lambda}^{0} \frac{\mathrm{d}k}{k - E + \mathrm{i}\epsilon}. \tag{3.83}$$

For scattering between particles with $E > 0$, there is a pole at $E + \mathrm{i}\epsilon$. The integral is equivalent to putting the pole on the real axis, avoiding it by integrating around a semicircle of vanishing radius at the position of the pole, and taking the principle value of the integral for the real part. The result is

$$I_\perp(E, 0) = \frac{1}{V_F} \left(\log \frac{2\Lambda}{E} + \frac{\mathrm{i}\pi}{2} \right) \tag{3.84}$$

up to power-suppressed terms. Then

$$\mathcal{A}_{\mathrm{S,B}} = -\frac{V^2}{2\pi} \frac{K_F}{V_F} \left(\log \frac{2\Lambda}{E} + \frac{\mathrm{i}\pi}{2} \right) = -\frac{m^* V^2}{2\pi} \left(\log \frac{2\Lambda}{E} + \frac{\mathrm{i}\pi}{2} \right), \tag{3.85}$$

where we have used the definition of the effective mass of the quasiparticles on the Fermi surface,

$$m^* = \frac{K_F}{V_F}. \tag{3.86}$$

Consider the scaling of this result under the leading-order RG (that is, neglecting the anomalous dimension introduced by this and other loop corrections). The coupling goes like $V \to \zeta V$, but as stated above $K_F \to \zeta^{-1} K_F$ and hence $m^* \to \zeta^{-1} m^*$, so the overall "classical" scaling of this result is ζ^1. As a result, it has the same scaling as the tree level result due to the enhancement from loop label momentum sum. This is very important, and we will investigate the implications shortly.

Now consider the contribution from the s-channel amplitude with two copies of the forward-scattering vertex. We see immediately that there can be no sum over the label momenta in the loop because the form of the forward scattering interaction constrains the incoming and outgoing label momenta to match. As a result, there is no loop enhancement in this diagram and the contributions are therefore suppressed within our power-counting scheme.

We must next address the t-channel diagram (the u-channel is just given by exchanging the two outgoing particles). Inserting a back-to-back interaction at either vertex constrains the label momentum in the loop to a single value, so there can be

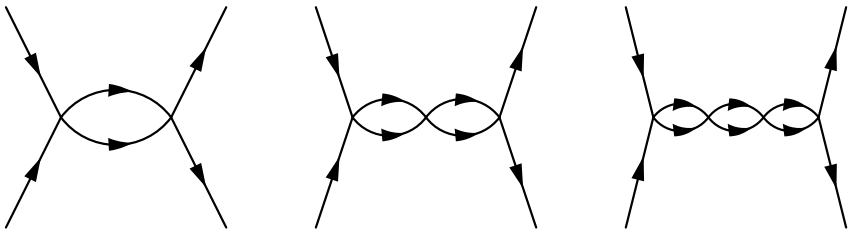


Figure 3.5: The leading-power contributions to the back-to-back scattering for the first several loop orders.

no compensating loop sum. We can only insert forward-scattering vertices if the interaction is both forward and back-to-back simultaneously. However, our power-counting demonstrates that the coupling for this interaction must scale as ζ^2, so the enhancement from the label momentum loop sum is not sufficient to promote this to a leading power effect.

In fact, we can extend the above reasoning to higher orders in the number of loops. At each loop order, the only diagram with a loop enhancement to go along with every factor of V in the vertex is the series of iterated bubble diagrams in Fig. 3.5. This can be thought of as being due to a large-N limit, where $N = K_F/\Lambda$ depends on the renormalization scale [5]. The fact that the leading-order contribution is from these diagrams is not new at all, of course — [61] makes exactly the same observation, for example. This presentation is only meant to continue the program of [5, 6, 39] and put the reasoning in the terms of a more modern perspective on effective field theories while simplifying the power counting.

3.12 Superconducting instability

The following treatment is very similar to that in [61], with appropriate modifications related to the bin summation implied by our effective theory.

Consider the following Green's function in our effective theory:

$$\langle \psi^\dagger \psi^\dagger(x)\psi\psi(y)\rangle. \tag{3.87}$$

We will see that the Fourier transform of Eq. (3.87) has a pole in the upper half-plane in complex frequency space, indicating an instability to Cooper paring. We can use Eq. (3.56) to write this in terms of a sum over the label fields:

$$\sum_{\vec{K}_i} e^{\mathrm{i}(\vec{K}_1+\vec{K}_2)x} e^{-\mathrm{i}(\vec{K}_3+\vec{K}_4)y} \langle \psi^\dagger_{\vec{K}_1} \psi^\dagger_{\vec{K}_2}(x)\psi_{\vec{K}_3}\psi_{\vec{K}_4}(y)\rangle. \tag{3.88}$$

Since $|K_i| = K_F$, the leading contribution comes from the terms in the sum where $\vec{K}_2 = -\vec{K}_1$ and $\vec{K}_4 = -\vec{K}_3$. Making a perturbative expansion in the couplings of the theory, we see that the resulting sum of terms at leading power looks just like insertions of the iterated bubble diagrams for the four-point amplitude. Thus, we can investigate the poles in E of the four-point coupling to understand the stability of the ground state.

The leading-power all-loop four-point diagram is given by a sum over the iterated bubbles in Fig. 3.5. For the simple case of a constant back-to-back coupling considered here, each additional bubble simply contributes a factor of

$$\frac{m^* V}{2\pi}\left(\log\frac{2\Lambda}{E} + \frac{i\pi}{2}\right). \tag{3.89}$$

Therefore the sum is a geometric series, and (ignoring the momentum dependence) we have

$$\mathcal{A}(E) = \frac{V}{1 + \frac{m^* V}{2\pi}\left(\log\frac{2\Lambda}{E} + \frac{i\pi}{2}\right)}. \tag{3.90}$$

For an attractive coupling $V < 0$, this has a pole at

$$E_0 = 2i\Lambda \exp\left(-\frac{2\pi}{m^*|V|}\right). \tag{3.91}$$

This implies the ground state is unstable to the formation of Cooper pairs. The distance of the pole from the real axis is inversely proportional to the relaxation time of the ground state and therefore gives an estimate of the binding energy of the Cooper pairs. This then implies the existence of a gap in the quasiparticle spectrum. A more detailed analysis, as in [61], shows that the size of the gap is given by the the absolute value $|E_0|$ exactly, and the critical temperature above which the gap closes is given by

$$T_c = \frac{e^\gamma}{\pi}|E_0|, \tag{3.92}$$

where γ is the Euler-Mascheroni constant.

In the context of our effective theory, the important feature of Eq. (3.90) is the fact that the product $m^* V$ appears in the denominator. This means that despite the fact that our interaction (and all of the bubble diagrams contributing to $\mathcal{A}(E)$) have positive scaling dimension and hence scale to zero in the IR, there will always be a pole in the back-to-back scattering amplitude. This is entirely due to the enhancement from the sum over the large label momenta. As a result, the instability of the ground state of the theory persists despite the presence of only formally irrelevant couplings.

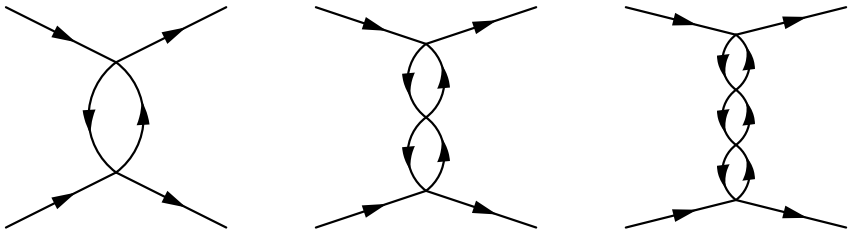


Figure 3.6: The leading-power contributions to forward scattering for the first several loop orders.

For a detailed treatment of the superconducting gap in a system without rotational invariance and a nonconstant coupling, see [5] and particularly [16].

3.13 The forward interaction and zero sound

We turn now to forward scattering. The s-channel diagram includes no label sum and is therefore power-suppressed. There can be no insertions of the back-to-back interaction for the t-channel diagram. Therefore the only possible contribution is from forward-scattering interactions in the t-channel diagram. Furthermore, this implies that the leading contributions in our power-counting involving forward scattering will simply be the sum of the iterated t-channel diagrams, such as in Fig. 3.6.

The most striking effect of this simplification to a series of iterated diagrams is the phenomenon of zero sound [36]. We have now seen within our effective theory that the leading diagrams in the power counting consist of the iterated s- and t-channel diagrams, because these diagrams are the only ones that will have a label sum enhancement for every instance of the couplings. This means in particular that collisions which change the momenta of the scattered particles are strongly suppressed in the IR as long as Cooper pairing has not become the dominant effect. As a result, it becomes impossible for normal hydrodynamic sound to propagate in Fermi liquids at low temperatures [61]. However, another type of sound, corresponding to oscillations of the shape of the Fermi surface that do not change the enclosed volume [35], continues to propagate.

To see this within our model, consider the density-density correlator:

$$\langle \psi^\dagger \psi(x) \psi^\dagger \psi(y) \rangle. \tag{3.93}$$

The analytic structure of the Fourier transform of this object tells us about the density oscillations of the system in response to an external perturbation [5]. Following that

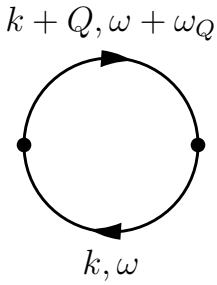


Figure 3.7: The first term in the expansion for the compressibility.

treatment, we define the compressibility χ by

$$(2\pi)^3\delta^3(0)\chi(Q,\omega_Q) = -\langle\rho(Q,\omega_Q)\rho(-Q,-\omega_Q)\rangle. \tag{3.94}$$

The quantity on the right is computed via the series of bubble diagrams with propagators that have momenta k and $k+Q$ and energy ω and $\omega+\omega_Q$. From this it is clear that the singularities of the expression, which determine the spectrum of sound-like excitations in the Fermi liquid, must be very similar to the singularities of the forward-scattering amplitude with transfer momentum Q and transfer energy ω.

Taking the forward-scattering coupling to be a constant, assuming a spin-singlet interaction, and working with a rotationally invariant Fermi surface is sufficient to demonstrate the existence of the singularity associated with zero sound. The leading term is given by the susceptibility $I(Q,\omega_Q)$, given by i times the Feynman diagram in Fig. 3.7. This is

$$I(Q,\omega_Q) = \sum_P \int \frac{\mathrm{d}^2k}{4\pi^2} \frac{\theta(\varepsilon_P(k))\theta(-\varepsilon_P(k+Q)) - \theta(-\varepsilon_P(k))\theta(\varepsilon_P(k+Q))}{\varepsilon_P(k) - \varepsilon_P(k+Q) + \omega - \mathrm{i}\epsilon\ \mathrm{sign}\,\varepsilon_P(k)}. \tag{3.95}$$

We have already performed the contour integral for the loop energy; see Appendix A for a similar calculation.

We may combine the sum over label momenta with the integral over the parallel component of the residual momenta as usual. The resulting expression is

$$I(Q,\omega_Q) = \frac{1}{4\pi^2}(I_{F+} + I_{F-}), \tag{3.96}$$

where

$$I_{F+} = \int K_F\,\mathrm{d}\theta\,\mathrm{d}k_\perp \frac{\theta(k_\perp)\theta(-(k+Q)_\perp)}{V_F k_\perp - V_F(k+Q)_\perp + \omega - \mathrm{i}\epsilon}, \tag{3.97}$$

$$I_{F-} = -\int K_F \, \mathrm{d}\theta \, \mathrm{d}k_\perp \frac{\theta(-k_\perp)\theta((k+Q)_\perp)}{V_F k_\perp - V_F(k+Q)_\perp + \omega + \mathrm{i}\epsilon}. \tag{3.98}$$

The step functions constrain the loop momentum to live in a small region in the vicinity of the Fermi surface when Q is small. To see this, note that the step functions in Eq. (3.98) constrain $k_\perp < 0$ while $(k+Q)_\perp > 0$. The second condition can be written

$$\sqrt{k^2 + 2kQ\cos\theta + Q^2} - K_F > 0. \tag{3.99}$$

Expanding this for small Q, we have

$$k_\perp > -Q\cos\theta. \tag{3.100}$$

Then

$$I_{F-} = -m^* \int \mathrm{d}\theta \int_{-Q\cos\theta}^{0} \frac{\mathrm{d}k_\perp \theta(\cos\theta)}{-Q\cos\theta + \omega/V_F + \mathrm{i}\epsilon}. \tag{3.101}$$

The $k_\perp$ integral is now trivial, and we have

$$I_{F-} = -m^* \int_{-\pi/2}^{\pi/2} \mathrm{d}\theta \frac{Q\cos\theta}{\omega/V_F - Q\cos\theta + \mathrm{i}\epsilon}. \tag{3.102}$$

Partially evaluating I_{F+} gives the same integrand but the opposite constraint on the sign of $\cos\theta$, so

$$I(Q,\omega_Q) = -\frac{m^*}{4\pi^2} \int \mathrm{d}\theta \frac{\cos\theta}{s - \cos\theta + \mathrm{i}\epsilon\cos\theta}, \tag{3.103}$$

where

$$s = \frac{\omega}{V_F Q}. \tag{3.104}$$

Only the iterated forward scattering diagrams will be loop enhanced. Thereore the leading-power contribution to the compressibility comes from a geometric series. Including the signs from the Wick contractions and spin sums, we find

$$\mathrm{i}\chi(Q,\omega_Q) = \mathrm{i}I(Q,\omega_Q) + \mathrm{i}F[\mathrm{i}I(Q,\omega_Q)]^2 + \cdots = \frac{\mathrm{i}I(Q,\omega_Q)}{1 - \frac{m^*F}{4\pi^2} I_\theta}, \tag{3.105}$$

where

$$I_\theta = \int \mathrm{d}\theta \frac{\cos\theta}{s - \cos\theta + \mathrm{i}\epsilon\cos\theta}. \tag{3.106}$$

Since we are interested in stable propagating modes of sound, we look for poles in Eq. (3.105) that lie on the positive real axis for complex ω. F is real, so this can only occur if I_θ is also real. Inspection of Eq. (3.106) demonstrates that this only occurs for $s > 1$. Evaluating the integral under this condition yields

$$I_\theta = 2\pi\left(\frac{s}{\sqrt{s^2-1}} - 1\right). \tag{3.107}$$

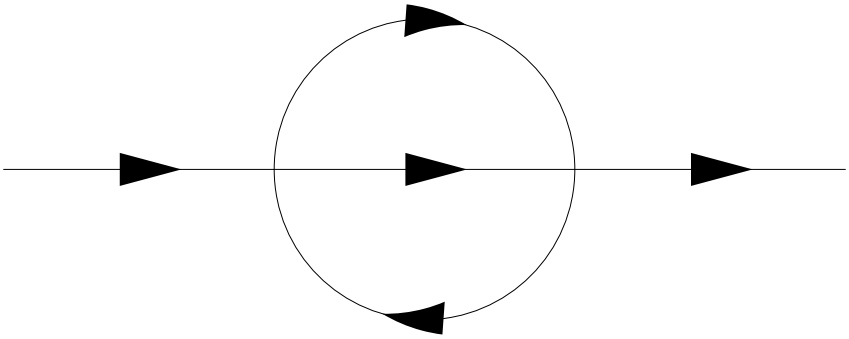


Figure 3.8: The nonzero result for the imaginary part of this diagram implies that the quasiparticles are unstable. However, the diagram is power suppressed as we flow to the IR. The result is that low energy quasiparticles have long lifetimes.

The condition for there to be a pole Eq. (3.105) is then

$$\frac{2\pi}{m^* F} = \left(\frac{s}{\sqrt{s^2 - 1}} - 1 \right). \tag{3.108}$$

Eq. (3.108) has solutions for any positive value of the coupling F, so we see that there is indeed a propagating sound mode at small Q and ω. The condition $s > 1$ implies that the speed associated with these modes, ω/Q, is always greater than the Fermi velocity V_F. In the limit of weak coupling, the left hand side of Eq. (3.108) goes to infinity, which implies that s must approach one. Expanding in $\delta s = s - 1$, we find

$$s = 1 + \left(\frac{m^* F}{2\pi} \right)^2 \tag{3.109}$$

in the weak coupling limit.

3.14 Stability of the quasiparticles

We found above that all higher loop diagrams in the theory are power-suppressed, with the exception of the iterated s- and t-channel diagrams. This means that in particular all diagrams that can contribute to a finite width for the quasiparticles must be suppressed. Furthermore, the quasiparticle width tells us about the lifetime of the quasiparticles and hence provides an estimate of the resistivity of the Fermi liquid. The sunrise diagram in Fig. 3.8 is the leading-order contribution to the width. Based on the fact that there are two vertices and no loop enhancement, this diagram must go like ζ^2 as we lower the energy. This implies that the width goes like T^2 at low temperatures [6].

3.15 Renormalization of the couplings

The result for the one-loop s-channel diagram with back-to-back external momenta in Eq. (3.85) depends on the UV cutoff Λ. This means that as we integrate out the UV modes from the theory, the back-to-back interaction is renormalized. Following the standard method of determining the change in the value of the effective coupling due to integrating out the high-energy modes, imagine rescaling the residual momenta by $\zeta^{-1} = (1 + \delta\Lambda/\Lambda)$. Then the path integral over the high-energy modes leads to a leading-order contribution to the effective back-to-back coupling that looks just like the one-loop amplitude in Eq. (3.78), but with $k_\perp$ restricted to lie in the small range Λ/V_F to $(\Lambda + \delta\Lambda)/V_F$. To understand the flow in terms of a continuous RG transformation, we simply take the limit of infinitesimal $\delta\Lambda$ and compute the corresponding infinitesimal change in the coupling.

From the description of this procedure, it is clear that the contribution of the infinitesimal change can be performed by simply taking the derivative of the amplitude with respect to Λ. In other words, we can write

$$\delta\tilde{V} = \delta\Lambda \frac{\mathrm{d}\mathcal{A}_{\mathrm{S,B}}}{\mathrm{d}\Lambda}, \tag{3.110}$$

where $\tilde{V} = \zeta V$. We have chosen to separate out the classical scaling here in light of our previous observation that the higher-loop results in the bubble diagram all have the same classical scaling as the tree level result (thanks to the label sum enhancement), and because all other contributions to the beta function are power-suppressed.

Taking the derivative of the one-loop result in Eq. (3.85) yields

$$\Lambda \frac{\mathrm{d}\mathcal{A}_{\mathrm{S,B}}}{\mathrm{d}\Lambda} = -\frac{m^*\tilde{V}^2}{2\pi}. \tag{3.111}$$

The beta function then reads

$$\Lambda \frac{\mathrm{d}\tilde{V}}{\mathrm{d}\Lambda} = \frac{m^*\tilde{V}^2}{2\pi}, \tag{3.112}$$

and the solution is

$$\tilde{V}(\Lambda) = \frac{\tilde{V}_0}{1 + \frac{\tilde{V}_0 m^*}{2\pi} \log \frac{\Lambda_0}{\Lambda}}. \tag{3.113}$$

When V_0 is negative, corresponding to an attractive potential, this implies that the coupling becomes strong at exactly the scale of the Cooper pair binding energy we found above. This is the indicator within the renormalization group that our effective field theory will fail to be applicable below the energies associated with the Cooper pairs.

In the more general case of a back-to-back coupling function that depends on the label momentum, the RG equation takes the form of an integral equation, and the flow depends on the value of the coupling between various different bins [5]. The details of that analysis demonstrate that the instability to a superconducting state is fairly robust, in that if any eigenvalue of the eigenmodes of the Hermitian kernel of the integral equation is negative, the system will tend toward a superconducting ground state [16].

Our calculation of the one-loop forward-scattering amplitude for small transfer momentum and energy in Section 3.13 demonstrated that there was no dependence on the UV cutoff. This implies that the forward coupling has no anomalous dimension at leading order in the power counting in our theory. However, the next section explores a subtle issue with this reasoning.

3.16 Subtleties in the binned theory

As an alternative to the binning we have performed above, we can instead simply regulate the theory with a hard cutoff, allow the coupling function (including the effects of both the forward and back-to-back interactions) to be an arbitrary function of the external momenta, and calculate the one-loop beta function by taking the derivative of the sum of the tree level and one-loop amplitudes and setting it equal to zero. This is the "old-fashioned" renormalization procedure first pioneered in [62].

Appendix B contains the calculations described above, and Fig. 4.4 shows the numerical results for the contributions to the beta function if we allow arbitrary incoming net and transfer momenta for the s- and t-channel coupling assuming constant, equal couplings for back-to-back and forward scattering. The s-channel diagram produces unsurprising results: we see that whenever the net momentum is less than Λ/K_F, the coupling is strongly renormalized. For values of the net momentum greater Λ/K_F, the beta function rapidly falls to zero. This demonstrates the consistency of our assumption that residual momentum and label momentum must be conserved separately for the back-to-back interactions.

On the other hand, the renormalization due to the t-channel diagram is much stranger. We see that for small values of the transfer momentum there is no contribution from the one-loop diagram. This is entirely consistent with our observation in the previous section that the exact forward scattering amplitude does not depend on the UV cutoff. However, when the transfer momentum is within the range $\Lambda/V_F < Q < 2\Lambda/V_F$, the beta function increases to a nonzero and unsuppressed value. This situation

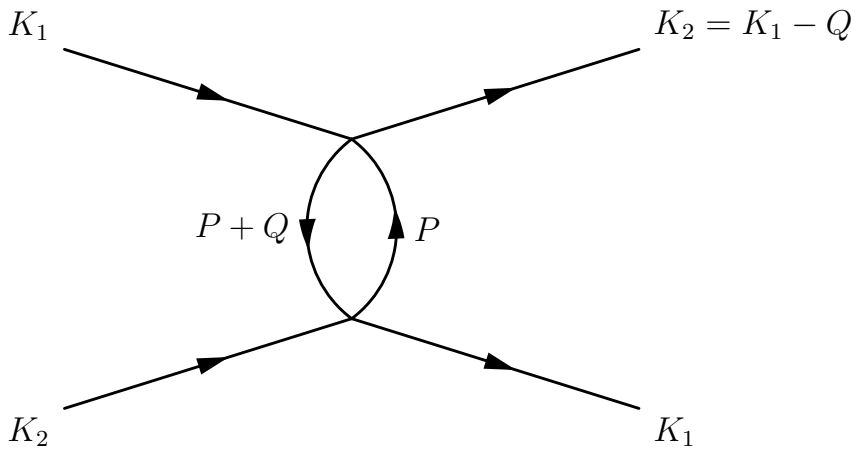


Figure 3.9: The diagram leading to a strange contribution to the forward coupling Gell-Mann and Low–style beta function. $V_F|Q| \sim \Lambda$ here.

corresponds to Fig. B.2b in Appendix B. Note that it is possible to have a nonzero transfer momentum flowing through the loop even with the external particles on the Fermi surface. The topology of the associated diagram is shown in Fig. 3.9. The situation described above represents forward scattering with a small opening angle.

The cause of this unusual behavior is the presence of an "anomalous" loop enhancement in the diagram of Fig. 3.9. Within our effective theory, it is only possible to satisfy conservation of label momentum at the two vertices if $P = K_2$ or $-K_1$. However, within the Gell-Mann and Low–style scheme, the loop particles go far off shell, allowing the residual momentum to mix with the label momentum. In the terminology of the binned effective theory, this allows for the label P to take on multiple values and enhances the loop contribution.

Fortunately, this phenomenon does not actually invalidate our effective theory. Remember that as we perform the RG, the Green's functions in the UV theory are mapped to Green's functions in IR theories with larger and larger values of the momentum. In other words, the momentum Q in the worrisome diagram becomes larger when measured in units of the residual momentum. Imagine starting with a dangerous momentum in the range $\Lambda < V_F Q < 2\Lambda$. Initially the forward coupling is strongly renormalized for this particular configuration of momenta. As we continue to run the RG, $V_F Q$ eventually exceeds 2Λ. We see in Fig. 4.4 that the beta function quickly drops off in this situation. Hence, the interaction in question is only renormalized for a small period of RG time. The result is equivalent to a power-suppressed contribution to the beta function for the entire flow. Note that

the same phenomenon does not occur for the back-to-back coupling because an arbitrarily small net momentum leads to a nonvanishing beta function, unlike the forward-scattering case.

The subtleties we have seen here become even more important in the case of a Fermi surface with long, flat sections. In that case, a large range of transfer and net momentum can allow for loop enhancements. The Fermi surface we consider in the next chapter is an example of such a case. To analyze it, we will forgo the formulation of a binned effective theory, simply using the behavior of the one-loop diagrams under the change of the UV cutoff to understand the theory.

3.17 Conclusion

The reinterpretation of Fermi liquid theory in terms of effective field theory does not in itself yield new predictive power. In fact, the vast majority of quantitative and qualitative predictions of the behaviors of Fermi liquids had already been worked out by Landau in his original works on the theory; most everything else was known by around 1965. The goal was actually to understand the diversity of possible low-energy fixed points of a theory of fermions in the presence of a Fermi surface. This was directly motivated by the observation of a Fermi surface in materials with exotic properties, like high-T_C superconductors. Therefore what was being sought was either a new fixed point, not already understood by Landau and company, or some kind of no-go theorem ruling out such a possibility. Indeed, the works cited above did produce something like a no-go theorem. [6] in particular claims that the only natural perturbative theories with a Fermi surface and only fermionic quasiparticles take the form of Landau's Fermi liquid at low energies. As with most no-go theorems, we should be careful in interpreting its domain of applicability. In the next chapter, we find that it is possible to construct a nontrivial theory that reproduces some non-Fermi liquid behavior by tuning a single parameter.

Chapter 4

EFFECTIVE FIELD THEORY OF 2D VAN HOVE SINGULARITIES

4.1 Abstract

We study 2D fermions with a short-range interaction in the presence of a Van Hove singularity. It is shown that this system can be consistently described by an effective field theory whose Fermi surface is subdivided into regions as defined by a factorization scale, and that the theory is renormalizable in the sense that all of the counterterms are well defined in the IR limit. The theory has the unusual feature that the renormalization group equation for the coupling has an explicit dependence on the renormalization scale, much as in theories of Wilson lines. In contrast to the case of a round Fermi surface, there are multiple marginal interactions with nontrivial RG flow. The Cooper instability remains strongest in the BCS channel. We also show that the marginal Fermi liquid scenario for the quasiparticle width is a robust consequence of the Van Hove singularity. Our results are universal in the sense that they do not depend on the detailed properties of the Fermi surface away from the singularity.

4.2 Introduction

In the 1990s and early 2000s, extensive theoretical work was devoted to the study of systems of fermions in 2D with the Fermi level close to a Van Hove singularity [9, 10, 63–72]. In such a system, the Fermi velocity vanishes at isolated points on the Fermi surface. We will refer to these points as Van Hove points. The existence of saddle points in the phonon and quasiparticle dispersions was first shown to be a generic feature of systems with a periodic lattice in [73]. We make the assumption that the chemical potential of the system has been set such that the Fermi surface passes through these points. From a theoretical standpoint, the Van Hove singularity is one of the simplest situations in which deviations from standard Fermi liquid theory are expected [74]. For example, the leading order computation

This chapter is an expanded and modified version of A. Kapustin, T. McKinney, and I. Z. Rothstein, "Wilsonian effective field theory of 2D van Hove singularities," submitted for publication (2018), arXiv:1804.01713 [cond-mat.str-el].

of the self-energy [63, 64] shows that with a short-range interaction the width of the quasiparticles is linear in the energy, a characteristic behavior of the Marginal Fermi Liquid (MFL) [11]. Since the MFL paradigm has been proposed to explain some peculiar properties of the normal state of high-T_c superconductors, it was speculated that high-T_c superconductors are special due to their proximity to a Van Hove singularity [1, 8, 64, 75–77]. While this scenario has fallen out of favor, understanding the effect of Van Hove singularities on the Fermi liquid remains an important problem.

Most of the studies cited above focus on the 2D Hubbard model on a square lattice at or near half-filling because of its relevance to cuprate superconductors. In this model, the Fermi surface is diamond-shaped and features two inequivalent Van Hove points (i.e. points where the Fermi velocity vanishes) as well as nesting. These features complicate the analysis, and it is hard to disentangle the effects of Van Hove points and nesting. In this paper we study in detail the case of a single Van Hove point from the point of view of Effective Field Theory (EFT). When applied to the case of a nonsingular Fermi surface, the EFT approach explains the ubiquity of both the Fermi liquid and BCS-type superconductivity [5, 6, 16, 38, 39].

As was previously noticed in [9, 67], the hyperbolic dispersion law characteristic of electrons near a 2D Van Hove point leads to additional divergences not regulated by the Wilsonian cutoff Λ, and necessitates the introduction of an additional regulator, which we take to be a Fermi velocity cutoff Υ. Υ also plays the role of a factorization scale, splitting the Fermi surface into two regions, $v_F > \Upsilon$ and $v_F < \Upsilon$, where two different power counting schemes apply. In each region momenta are split into large "label" momenta and small "residual" components. Previous work on the 2D Van Hove singularity has been plagued by nonlocal divergences, and a recent detailed study [2] concluded that the Van Hove EFT is not renormalizable when the Fermi level is exactly at the Van Hove singularity and has a very narrow range of applicability when the Fermi level is close to it. However, as we show, when momenta are properly power counted, all of the counterterms are independent of the residual momenta in each respective region and therefore should be considered local. Furthermore, the coupling in each region can only depend upon the label momenta. The coupling can be assumed to be independent of momenta only when all components of the momenta are smaller than Λ/Υ.

In the BCS channel, the RG equation for the coupling function explicitly depends on the logarithm of the ratio of the Wilsonian cutoff Λ to the bandwidth W and leads

to the well known double logarithmic running [9, 10, 67]. This dependence on the UV scale W represents a form of UV/IR mixing and has interesting consequences discussed below.

The situation is reminiscent of high energy scattering processes, such as the Sudakov form factor, where the phase space of gauge bosons is split into two regions which dominate the IR behavior. This splitting leads to additional (rapidity) divergences that necessitate a new regulator [4, 54] to distinguish between soft and collinear modes. Summing contributions from these two sectors leads to a cancellation of the regulator but, as in the present case, the cancellation leaves behind a Cheshire log in the beta function. This in turn leads to double logs in the renormalization group flow.

We utilize our results to study how a Van Hove singularity modifies the low energy behavior. In particular, we discuss the Cooper instability and the range of applicability of the Marginal Fermi Liquid scenario. We show that the Cooper instability is the strongest in the BCS channel, as in the case of the circular Fermi surface, but is also present for other kinematic configurations. This happens because of additional marginal interactions that lead to a breakdown of the Fermi liquid picture. We also show that a certain generalization of the MFL scenario is a robust consequence of the Van Hove EFT.

4.3 A toy model of a Van Hove singularity

In the 2D Hubbard model on a square lattice, there are two VH points in the Brillouin zone: $p_{VH} = (0, \pi)$ and $p_{VH} = (\pi, 0)$. When the hopping parameters in the x and y directions are not equal, the energies of these two points are different. If the Fermi level is much closer to one than the other, the effective field theory of a single VH singularity should apply. At both of the VH points, $2p_{VH} \sim 0$. We assume there is a unique VH point in the Brillouin zone and time-reversal (T) symmetry, which takes $p \mapsto -p$, is present. Therefore the singularity sits at the origin, a fixed point under T.

Such a scenario is realized by expanding the nearest-neighbor Hubbard model Hamiltonian around the point $p = 0$ to lowest order in momentum components and assuming a zero-range interaction. The resulting action is

$$S = \int dt\, d^2x \left[\psi^\dagger (i\partial_t - \varepsilon(-i\nabla) + \mu)\psi - \frac{g}{2}(\psi^\dagger\psi)^2\right], \tag{4.1}$$

where the dispersion relation is

$$\varepsilon(p) = p^2 \equiv t_x p_x^2 - t_y p_y^2 \tag{4.2}$$

and is unbounded from below. p^2 denotes the square of the 2D vector p with respect to the indefinite metric $\mathrm{diag}(t_x, -t_y)$. It is convenient to set $t_x = t_y = 1$ by rescaling p_x and p_y, such that metric becomes $\mathrm{diag}(1, -1)$, and absorbing a factor of $1/\sqrt{t_x t_y}$ into g. If we regard p_x, p_y as periodic with period of order k_B, then t_x, t_y are of order W/k_B^2, where W is the bandwidth.

As usual, all states with $\varepsilon(p) < \mu$ are assumed to be occupied, so in the free ($g = 0$) limit the excitations of the system are particles and holes, both with nonnegative energy. When the Fermi level μ vanishes, the system has a discrete symmetry, $\psi \leftrightarrow \psi^\dagger, x \leftrightarrow y$, that exchanges particles and holes. Furthermore, the quadratic dispersion relation has $O(1, 1)$ invariance, and the short-range interaction preserves this symmetry. Also, for $\mu = 0$, the action (4.1) is invariant under dilatations

$$\psi(t, x) \to \lambda^{-1}\psi(\lambda^2 t, \lambda x). \tag{4.3}$$

Invariance with respect to Galilean boosts is spontaneously broken by the Fermi sea for all values of μ. As usual, the dilatation symmetry is anomalous on the quantum level. Internal symmetries include $U(1)$ particle-number symmetry and $SU(2)$ spin symmetry.

The interaction term in (4.1) has zero range, and in momentum space corresponds to a four-point vertex with no momentum dependence. A naive justification for this simple ansatz is that any vertex with more than four fermionic fields or polynomial momentum dependence is irrelevant in the RG sense. Here we assume that under the RG transformations the momenta scale as

$$p_x \to \lambda p_x, \; p_y \to \lambda p_y, \tag{4.4}$$

so energy has scaling dimension 2 and ψ has scaling dimension 1. Then the chemical potential μ is relevant, the coupling g is marginal, and more complicated interactions are irrelevant.

This naive argument is, as well known, incorrect, due to the fact that momenta tangent to the Fermi surface should not scale under RG flow. Moreover, the theory defined with a contact interaction, eq. (4.1), is not consistent, as corrections to the four-point function include nonlocal divergences that cannot be absorbed into a renormalization of g [9, 67]. These divergences can be traced to the noncompactness of the Fermi surface.

4.4 Setting up the Van Hove EFT

To make the theory (4.1) well defined, one must impose a cutoff on momenta to render the Fermi surface compact. This cutoff is imposed in addition to the Wilsonian cutoff $|\varepsilon(p)| \leq \Lambda$. We assume Λ is much smaller than the bandwidth $W \sim k_B^2$. We also assume that $|\mu| \ll \Lambda$, so the modes near the Fermi surface are not integrated out.

Let Υ denote this momentum cutoff. The largest possible value for Υ is of order k_B, the size of the Brillouin zone, and thus it is natural to assume that $\Lambda \ll \Upsilon^2$. The region

$$|p_\pm| \leq \Upsilon, \quad |p_+ p_-| \leq \Lambda, \tag{4.5}$$

where $p_+ = p_x + p_y$ and $p_- = p_x - p_y$, will be called the VH region.[2] Within this region, the dispersion law is

$$\varepsilon = p_+ p_-. \tag{4.6}$$

We are using Λ and Υ as both explicit regulators and factorization scales. Υ has a natural value of order V_F, the typical value of the Fermi velocity away from the VH points, but it can also be chosen parametrically smaller. In any physical result the dependence on Υ should cancel, since its role is merely to separate the VH and NVH regions. On the other hand, in any physical prediction Λ is a placeholder for the RG scale.

The VH region is the part of the Λ-neighborhood of the Fermi surface that is close to the saddle point. In this region, the dispersion relation (4.6) implies that if both components of momentum are of the same order, then $p_\pm \sim \sqrt{\Lambda} \ll \Upsilon$. In addition to these "soft modes," the VH region is populated by collinear and anticollinear modes whose momenta scale as $(\Upsilon, \Lambda/\Upsilon)$ and $(\Lambda/\Upsilon, \Upsilon)$ respectively.

The rest of the Λ-neighborhood of the Fermi surface will be called the NVH region. Within this region, the dispersion law is the standard

$$\varepsilon(p) = p_\perp v_F(p_\parallel), \tag{4.7}$$

where $p_\perp / p_\parallel$ are normal/tangential to the Fermi surface. We assume that the NVH region is "featureless," in the sense that the Fermi velocity does not change too much there, and that it is free of nesting. The first assumption simply means that there are no other Van Hove singularities nearby, while the importance of the second assumption will be discussed in Section 4.11. Fig. (4.1) illustrates the division of a representative Fermi surface into the VH and NVH regions.

[2] Note that Υ breaks the $O(1,1)$ symmetry but preserves the particle-hole symmetry.

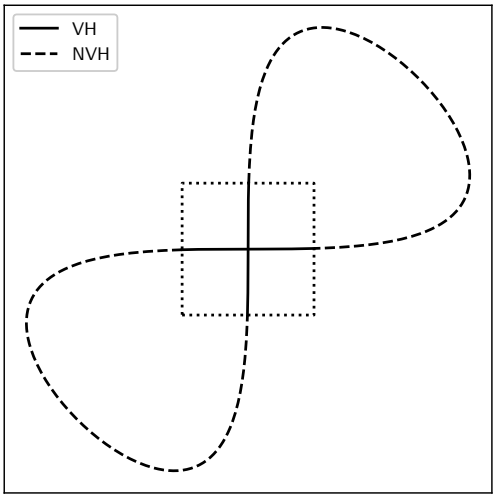


Figure 4.1: An example of the division of the Fermi surface into Van Hove and non-Van Hove regions.

In general, loop calculations involving modes from the VH region alone will depend on Υ in such a way that the $\Upsilon \to \infty$ limit leads to additional divergences. Thus, a sensible EFT must include both the VH region and the NVH region. We use the term "full theory" for such an EFT. We make no assumptions about the shape of the Fermi surface in the NVH region. As we will show below, our results are universal to leading log accuracy in the sense that they only depend upon V_F, the typical Fermi velocity in the NVH region, and not the detailed shape of the Fermi surface. Therefore, our results apply to any system with a VH singularity near the Fermi surface that is weakly coupled at energies of order the bandwidth.

We will denote the fields annihilating electrons in the VH and NVH regions ψ_V and ψ_N respectively. The interaction part of the action is

$$S_{int} = \int dt \prod_{i=1}^{4} d^2 p_i \sum_{\alpha\beta\gamma\delta} g_{\alpha\beta\gamma\delta} \psi^\dagger_\alpha(p_1)\psi^\dagger_\beta(p_2)\psi_\gamma(p_3)\psi_\delta(p_4), \tag{4.8}$$

where the indices $\alpha, \beta, \gamma, \delta$ take values V and N. In general, $g_{\alpha\beta\gamma\delta}$ can depend on the momenta p_i of the interacting modes. The couplings must match onto each other as the VH modes approach the NVH region and vice versa. For example, g_{NNVV} must match onto g_{VVVV} as p_1 and p_2 approach the VH region.

Naively, in light of the dispersion laws (4.6) and (4.7), one might think that the

coupling functions in (4.8) should only depend on the $p_\parallel$ of the NVH modes and that the only marginal interaction between the VH modes should be a momentum-independent constant. We will see in the next section that this is not self-consistent: one-loop calculations imply that the couplings must depend on momentum in a more generic manner. This is because when both the rapidity cutoff Υ and the Wilsonian cutoff Λ are present, a low momentum scale Λ/Υ also plays a role. We will call Λ/Υ the ultrasoft scale.

We can achieve some simplification by recalling that momentum and energy conservation limits the interactions of the NVH modes to special kinematic configurations [5]. These configurations correspond to forward scattering and back-to-back (BCS) scattering. This implies that interactions between NVH modes and VH modes are of two kinds: (1) forward scattering between a VH mode and an NVH mode and (2) scattering of nearly back-to-back VH modes to nearly back-to-back NVH modes and vice versa. As a result, the numbers of VH and NVH particles are separately conserved.

4.5 Specific heat

As an example of the anamalous behavior expected from systems with a van Hove singularity, consider the specific heat. In the case of a normal Fermi liquid, the specific heat is linear in temperature to leading order in the interactions between the quasiparticles [37]. Note that it can be rather complicated to go beyond this leading-order result [78]. In any case, we will find that even the leading order result is modified in the van Hove case, with a logarithmic enhancement at low temperatures.

The average energy density of the a Fermi liquid at finite temperature T in two dimensions can be computed using the Fermi distribution:

$$\langle E\rangle = 2\int \frac{\mathrm{d}^2k}{(2\pi)^2}\frac{\varepsilon(k)}{e^{-\beta\varepsilon(k)}+1}, \tag{4.9}$$

where $\beta = 1/k_\mathrm{B}T$ (k_B is the Boltzmann constant) and

$$\varepsilon(k) = t_x k_x^2 - t_y k_y^2. \tag{4.10}$$

We include a factor of two for the spin degeneracy. Changing variables to

$$k_\pm = \sqrt{t_x}k_x \pm \sqrt{t_y}k_y, \tag{4.11}$$

this becomes

$$\langle E\rangle = \frac{1}{\sqrt{t_x t_y}}\int \frac{\mathrm{d}^2k}{(2\pi)^2}\frac{\varepsilon(k)}{e^{-\beta\varepsilon(k)}+1}, \tag{4.12}$$

where $\varepsilon(k) = k_+ k_-$. The derivative with respect to temperature gives us the specific heat:

$$C_V = \left(\frac{\partial \langle E \rangle}{\partial T}\right)_N = -\beta^2 \frac{\partial \langle E \rangle}{\partial \beta}. \tag{4.13}$$

This yields

$$C_V = \frac{1}{\sqrt{t_x t_y}} \int \frac{\mathrm{d}^2 k}{(2\pi)^2} \frac{\beta^2 \varepsilon(k)^2 e^{-\beta\varepsilon(k)}}{\left(e^{-\beta\varepsilon(k)} + 1\right)^2}. \tag{4.14}$$

We now institute the Fermi velocity cutoff $|k_\pm| < \Upsilon$ and change variables via

$$k_\pm = k e^{\pm\eta}, \tag{4.15}$$

with $k > 0$. The Jacobian is

$$\left| \frac{\partial k_+}{\partial k} \frac{\partial k_-}{\partial \eta} - \frac{\partial k_+}{\partial \eta} \frac{\partial k_-}{\partial k} \right| = 2k. \tag{4.16}$$

Note that this only works for the first quadrant in $k_\pm$ space. We will need to multiply our final result by four to take into account the other quadrants. After changing variables, Eq. (4.14) is

$$C_V = \frac{2}{\sqrt{t_x t_y}} \int \frac{k\,\mathrm{d}k}{(2\pi)^2} \int_{-\log \Upsilon/k}^{\log \Upsilon/k} \mathrm{d}\eta \frac{(\beta k^2)^2 e^{-\beta k^2}}{\left(e^{-\beta k^2} + 1\right)^2}. \tag{4.17}$$

Evaluating the integral over η yields

$$C_V = \frac{2}{\sqrt{t_x t_y}} \int \frac{k\,\mathrm{d}k}{(2\pi)^2} \frac{(\beta k^2)^2 e^{-\beta k^2}}{\left(e^{-\beta k^2} + 1\right)^2} \log \frac{\Upsilon^2}{k^2}. \tag{4.18}$$

Changing variables yet again to $u = \sqrt{\beta} k$ gives us

$$C_V = \frac{2}{\beta \sqrt{t_x t_y}} (I_1 \log \beta \Upsilon^2 - I_2), \tag{4.19}$$

where

$$I_1 = \int_0^{\Upsilon\sqrt{\beta}} \frac{u\,\mathrm{d}u}{(2\pi)^2} \frac{u^4 e^{-u^2}}{\left(e^{-u^2} + 1\right)^2}, \tag{4.20}$$

$$I_2 = \int_0^{\Upsilon\sqrt{\beta}} \frac{u\,\mathrm{d}u}{(2\pi)^2} \frac{u^4 e^{-u^2}}{\left(e^{-u^2} + 1\right)^2} \log u^2. \tag{4.21}$$

For temperatures well below Υ^2 these functions do not differ appreciably from their value with the upper cutoff taken to infinity, and we have

$$I_1 = \frac{1}{48}, \tag{4.22}$$

$$I_2 \approx 0.022. \tag{4.23}$$

In the low temperature limit the term involving I_2 is negligible. Including the factor of four mentioned above, we have

$$C_V = \frac{k_B T}{6\sqrt{t_x t_y}} \log \frac{\Upsilon^2}{k_B T}. \tag{4.24}$$

As mentioned previously, the Fermi velocity cutoff can by approximated by the typical Fermi velocity in the non-van Hove region, V_F. Therefore

$$C_V = \frac{k_B T}{6\sqrt{t_x t_y}} \log \frac{V_F^2}{k_B T} \tag{4.25}$$

at low temperatures. This characteristic $-T \log T$ dependence has been observed in cuprate superconductors under conditions where the superconducting transition has been suppressed with strong magnetic fields [79], although the authors of that work ascribe this to the existence of a quantum critical point instead of to the van Hove singularity.

4.6 The one-loop beta function

Generic kinematic configuration

Consider the scattering of VH modes in a generic kinematic configuration. Conservation of momentum implies the NVH modes will not contribute. Thus tree-level interactions are described by a single coupling function of three independent VH momenta. We would like to determine how this function is renormalized. It is enlightening to first assume that the coupling is a momentum-independent constant, as naive power counting suggests. The manner in which this assumption fails will show us how to appropriately modify the theory.

We subdivide the VH region into three parts: the soft region, where $p_\pm \sim \sqrt{\Lambda}$; the collinear region, where $p_+ \sim \Upsilon$ and $p_- \sim \frac{\Lambda}{\Upsilon}$; and the anticollinear region, where $p_- \sim \Upsilon$, and $p_+ \sim \frac{\Lambda}{\Upsilon}$. Fig. 4.2 illustrates the location of these subregions. This separation is useful for categorizing the contributions to the beta function. Since in this subsection the kinematic configuration is assumed to be generic, the differences and sums of external momenta are of the same order as the momenta themselves.

As usual, we have three diagrams at one-loop level, which we refer to as s-channel ($\mathcal{A}_S$), t-channel ($\mathcal{A}_T$), and u-channel ($\mathcal{A}_U$); see Fig. 4.3. These three diagrams depend on $K = p_1 + p_2$, $Q = p_1 - p_3$, and $Q' = p_1 - p_4$ respectively, and each contributes independently to the beta function. The u-channel diagram is identical

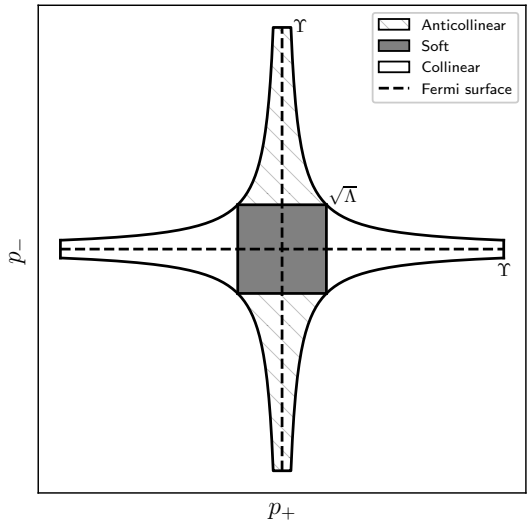
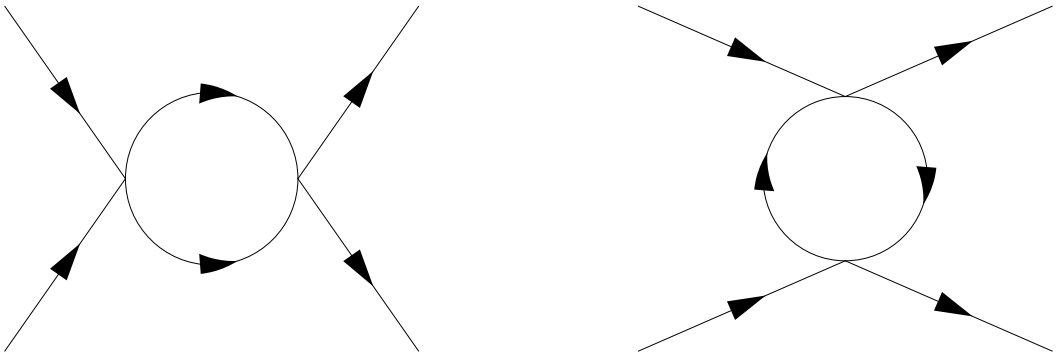


Figure 4.2: Subdivision of the VH region.



(a) s-channel (b) t-channel

Figure 4.3: The diagrams contributing to the renormalization of the coupling at one loop. Not shown is the u-channel diagram, which is given by interchanging the final state particles in the t-channel diagram.

to the t-channel diagram if we take $Q \leftrightarrow Q'$, so we focus on the t- and s-channel diagrams.

We find (see Appendix A) that any one-loop diagram where a collinear external mode and an anticollinear external mode meet at a vertex leads to a power-suppressed contribution to the beta function. This is because the K or Q involved in the interaction always sets a large energy scale that acts to suppress the associated diagram.

Generic t-channel diagrams that do not involve collinear-anticollinear vertices make order-one contributions to the beta function. For example, for a generic interaction between soft modes,

$$\Lambda \frac{d\mathcal{A}_{\mathrm{T}}}{d\Lambda} = \frac{g^2}{4\pi^2} \tag{4.26}$$

plus power-suppressed terms. There are exceptions in certain special kinematic configurations; see Section 4.6.

The behavior of the s-channel diagrams is more complicated. Defining

$$\varepsilon_K = K_+ K_-, \tag{4.27}$$

we find that generic s-channel diagrams that do not involve collinear-anticollinear vertices interpolate between being log enhanced when $\varepsilon_K \ll \Lambda$ and order one when $\varepsilon_K \sim \Lambda$. As an example, for generic interactions between soft modes,

$$\Lambda \frac{d\mathcal{A}_{\mathrm{S}}}{d\Lambda} = -\frac{g^2}{4\pi^2} \log\left(\frac{\Lambda}{\varepsilon_K}\right) \tag{4.28}$$

plus suppressed terms. To avoid confusion, we note that ε_K is not the net energy of the incoming particles.

Special kinematic configurations

Eq. (4.28) appears to imply that the beta function diverges as ε_K approaches zero, thus necessitating the existence of a nonlocal counterterm, which would mean the formalism lacked a systematic power-counting scheme. However, (4.28) does not apply in the $\varepsilon_K \to 0$ limit. The divergent behavior is an unphysical artifact of taking the Van Hove region to be infinite in extent. If we take the rapidity cutoff Υ into account, we find that when one component of K, say K_-, satisfies

$$|K_-| < \frac{\Lambda}{\Upsilon}, \tag{4.29}$$

such as for an interaction between only collinear modes, then

$$\Lambda \frac{d\mathcal{A}_{\mathrm{S}}}{d\Lambda} = -\frac{g^2}{4\pi^2} \log\left(\frac{\Upsilon}{K_+}\right) \tag{4.30}$$

plus order-one terms. If both components of K are ultrasoft (i.e. smaller in magnitude than Λ/Υ), we find to leading log order

$$\Lambda \frac{d\mathcal{A}_{\mathrm{S}}}{d\Lambda} = -\frac{g^2}{4\pi^2} \log\left(\frac{\Upsilon^2}{\Lambda}\right). \tag{4.31}$$

We can summarize the detailed behavior of the s-channel contribution to the beta function in the following manner:[3]

$$\Lambda \frac{d\mathcal{A}_S}{d\Lambda} = \begin{cases} -\frac{g^2}{4\pi^2} \log\left(\frac{\Lambda}{\max(K_+, \Lambda/\Upsilon) \max(K_-, \Lambda/\Upsilon)}\right), & \varepsilon_K \lesssim \Lambda \\ O(1) \times \frac{\Lambda}{\varepsilon_K} g^2, & \varepsilon_K \gtrsim \Lambda. \end{cases} \tag{4.32}$$

If $K_+ \sim \Upsilon$, the log in (4.30) will not be large, and hence the order-one "corrections" cannot be ignored. As a result, the dependence on K_+ becomes complicated. Similarly, if one component of Q is large while the other is ultrasoft, the t-channel diagram has a complicated dependence on the large component (though unlike the s-channel diagram, it can never become log enhanced). These cases are discussed in more detail in Section 4.11. Finally, the t-channel contribution to the beta function vanishes if both components of Q are ultrasoft.

Binning and leading-log behavior

At first glance, the behavior of the beta function implied by the above results is rather odd. The contribution from the s-channel diagram in Eq. (4.32) sometimes depends nonanalytically on the momentum, and the functional forms of the results change when the components of K or Q pass a particular threshold (around the scale Λ/Υ). Previous authors [2] have particularly regarded the behavior of the t-channel diagram as a sign of unavoidable nonlocality in the theory. However, as discussed in the next section, similar behavior appears already for a circular Fermi surface, and is dealt with using bins in momentum space of size Λ/K_F. This notion of binning allows for a clear separation between large and small momenta, and was previously used in the context of the theory of nonrelativistic heavy quarks [52]. Binning is also implicit in the standard Fermi-surface RG [5]. We apply the same method here.

We divide momentum space into bins of size Λ/Υ, each with a label momentum corresponding to the center of the bin and a residual momentum, of order Λ/Υ, corresponding to the position within the bin. The couplings are then indexed by the discrete label momenta, and we can Taylor expand in the residual momenta. The beta function then depends at leading order on the label momenta alone, and all results are analytic in the residual momenta. The theory is therefore renormalizable, although the couplings depend in an arbitrary way on the label momenta. The same is true for a circular Fermi surface (see the next section).

[3] Note that for the s-channel diagram, taking $\varepsilon_K \gtrsim \Lambda$ is equivalent to injecting a large virtuality into the loop, which is formally outside the range of validity of the effective theory. The effects of such modes in intermediate states are properly accounted for in higher dimensional, power-suppressed, operators. This is consistent with the result in (4.32).

The nonanalytic dependence on the net momentum implies that our assumption of a momentum-independent coupling was inconsistent, and the RG flow will generate dependence on the label momenta even for modes within the soft region. While this complicated behavior threatens the predictive power of the theory, we will see in Section 4.8 that the enhancement of the beta function for modes with small net momentum allows for several important simplifications.

4.7 Revisiting the round Fermi surface

Let us revisit some old results involving a round Fermi surface. In that context, previous authors [5, 6, 16, 38, 39] found that only certain coupling functions are present in the IR theory. In particular, only forward scattering and interactions between back-to-back particles (the BCS channel) are marginal, in the language of effective field theory. Furthermore, these authors found that only the BCS coupling is renormalized (that is, corrections to forward scattering are power suppressed), and that the beta function for the BCS interaction is in fact one-loop exact for generic round Fermi surfaces [5, 16].

These results hold in the limit where the Wilsonian cutoff Λ on the energy of the modes included in the theory (or, in other words, the "width" of modes around the Fermi surface) is taken to zero while the size of Fermi surface itself is held fixed. For nonzero Λ, near-forward and near-BCS scattering continue to be present in the theory. To understand their role more precisely, let us consider their contributions to the one-loop beta function.

We may parameterize a generic coupling function in terms K, Q, and Q', the same functions of the external momenta defined in Section (4.6). As before, the s-, t-, and u-channel diagrams ([5] calls these the BCS, ZS, and ZS' diagrams), depend on K, Q, and Q' respectively, and the t- and u-channel diagrams are exchanged under $Q \leftrightarrow Q'$. BCS scattering occurs for $K = 0$ and forward scattering occurs when either Q or Q' is zero.

It is straightforward to show that when any of these momenta are order K_F (the radius of the Fermi surface), the presence of the Fermi surface forces the contribution from the corresponding one-loop diagram to the beta function to be suppressed. For example, the log derivative of the one-loop s-channel diagram is

$$\Lambda \frac{d\mathcal{A}_S}{d\Lambda} \sim \frac{\Lambda}{v_F |K|} g^2 \tag{4.33}$$

when $|K| \approx K_F$. A similar statement holds for the t-channel and u-channel diagrams.

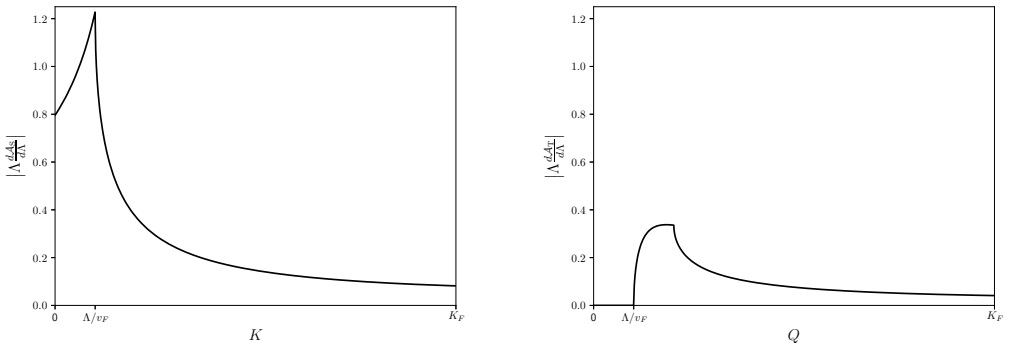


Figure 4.4: The log derivatives of the s- and t-channel diagrams measured in units of g^2 for a circular Fermi surface. We assume a constant coupling.

From this point of view, the one-loop contributions are generically power suppressed. The exceptional behavior occurs when K (or Q or Q') is of order Λ/v_F. Unlike the case for large K or Q, the behavior qualitatively differs between the s and t channels.

For the t-channel diagram to make a nonsuppressed contribution to the beta function, the following must hold:

$$\frac{\Lambda}{v_F} < |Q| < O(1) \times \frac{\Lambda}{v_F}. \tag{4.34}$$

Thus, there is a window of values where the contribution is nonzero, and the position of the edges of this widow depend on Λ. On the other hand, for the s-channel diagram to make an unsuppressed contribution to the beta function, K must satisfy

$$|K| < O(1) \times \frac{\Lambda}{v_F}. \tag{4.35}$$

In particular, $K = 0$ gives an order-one contribution while $Q = 0$ does not. Fig. 4.4 demonstrates the behavior of the log derivatives assuming a constant coupling.

This difference has a profound effect. In the course of the RG flow, the condition that K or Q is order Λ/K_F changes, since we take Λ to scale down. If K is actually zero from the beginning, there will always be an order-one contribution to the beta function, and this condition is stable throughout the RG flow. This allows attractive couplings with $K = 0$ (the BCS channel) to become strong at small Λ. On the other hand, the condition for the t-channel diagram to give an unsuppressed contribution to the beta function is not stable under the RG flow. Hence for *any* fixed Q, the t channel only contributes to the beta function for a small period of RG time.

In summary, the contribution to the beta function is power-suppressed throughout the RG flow for generic (large) K and Q. If K or Q is small enough, there are

order-one contributions to the beta function, but only for a short RG time. The only exception is "true" BCS scattering, where $|K| < \Lambda/v_F$ throughout the flow. If we assume the UV coupling is weak, this means the only coupling that can be relevant to the ground state instability involves the BCS configuration.

With this context, the startling results for the one-loop VH beta function [Eqs. (4.26)-(4.32)] are less surprising. Even with a round Fermi surface, the beta function, and therefore the coupling, depends on K and Q. This is even true for the BCS coupling, which is generically a function of two angular coordinates [5, 16] (playing the role of label momenta) for noncircular Fermi surfaces.

Finally, the transition from zero contribution to the beta function from the t-channel diagram to a finite contribution as we increase Q from zero is also present for the circular Fermi surface. The major difference in the VH case is the long, flat section of the Fermi surface, which guarantees that the window in Q for which the t channel is not power suppressed is larger than for a circular Fermi surface. Fortunately, we will see that we may once again neglect the contribution from the t channel relative to the s channel, at least for certain observables.

4.8 The leading contribution at one loop

Section 4.6 demonstrates that only the s-channel diagram contains a logarithmic enhancement at one-loop order. Furthermore, the largest possible contribution to the beta function occurs when $K \simeq 0$. This indicates that the kinematic configuration of near-zero net momentum, the BCS channel,[4] dominates the low-energy behavior of the theory.

With this in mind, assume the UV the dependence on the external momenta is analytic. This condition will not be preserved under the RG, because the s-channel introduces a nonanalytic dependence on the net momentum K in the four-point coupling. However, if we focus on the BCS configuration we may ignore any nonanalytic dependence on the other momenta to leading-log order.

In the following calculations, we sum the leading VH and NVH contributions. While the precise form of the full results generically depends on the detailed shape of the NVH portion of the Fermi surface, the leading contribution is independent of these details. Instead, this summing procedure turns out to be identical to taking the VH results and replacing the cutoff Υ with V_F, its natural value.

[4] In this context, the term "BCS" means back-to-back up to an ultrasoft momentum. Generic configurations of only ultrasoft modes therefore qualify as BCS.

Parameterize the BCS coupling as $g_{\rm B}(p_1, p_3)$, where p_1 is the label momentum of one of the incoming pair of particles (the other has label momentum $-p_1$) and p_3 is the label momentum of one of the pair of outgoing particles. We find with logarithmic accuracy (see Appendix A)

$$\Lambda \frac{dg_{\rm B}(p_1, p_3)}{d\Lambda} = \frac{1}{4\pi^2} g_{\rm B}(p_1, 0) g_{\rm B}(0, p_3) \log \frac{V_F^2}{\Lambda}. \tag{4.36}$$

An unusual feature of this equation is that the beta function has an explicit dependence on Λ, as well as V_F^2. The latter can be regarded as an energy scale of order of the bandwidth, $V_F^2 \sim W$. Thus the IR physics retains some information about the UV scale W.

The solution to (4.36) is

$$\begin{aligned} g_{\rm B}(p_1, p_3; \Lambda) &= g_{\rm B}(p_1, p_3; \Lambda_0) \\ &\quad - \left(\frac{1}{8\pi^2}\right) \frac{g_{\rm B}(p_1, 0; \Lambda_0) g_{\rm B}(0, p_3; \Lambda_0) \left(\log^2 \frac{V_F^2}{\Lambda} - \log^2 \frac{V_F^2}{\Lambda_0}\right)}{1 + \frac{g_{\rm B}(0,0;\Lambda_0)}{8\pi^2} \left(\log^2 \frac{V_F^2}{\Lambda} - \log^2 \frac{V_F^2}{\Lambda_0}\right)}. \end{aligned} \tag{4.37}$$

The coupling in the vicinity of the Van Hove singularity, $g_{\rm B}(0, 0)$, plays a special role: it "drives" the RG for the other couplings, and when it is attractive at the scale Λ_0, it sets the one-loop estimate of the strong-coupling scale,

$$\Lambda^* = V_F^2 \exp\left(-\sqrt{\log^2 \frac{{V_F}^2}{\Lambda_0} + \frac{8\pi^2}{|g_{\rm B}(0, 0; \Lambda_0)|}}\right). \tag{4.38}$$

As in the ordinary BCS theory [80] the strong-coupling scale is non-perturbative in $g(\Lambda_0)$. However, the usual dependence of this scale on the microscopic parameters differs from (4.38). While (4.38) simplifies considerably if we set $\Lambda_0 = V_F^2 \sim W$, this choice may be unphysical if the Van Hove EFT is obtained by integrating out some other degrees of freedom at a scale below W. For example, if the short-range interaction arises both from the screened Coulomb repulsion and the phonon-mediated attraction, the Van Hove EFT applies only up to energy scales of the order of the Debye frequency ω_D, which is usually much smaller than the bandwidth W. Then the natural choice for Λ_0 is ω_D, and we have a hierarchy of scales $V_F^2 \simeq W \gg \omega_D$.

To understand some of the limitations of this formalism, consider the amplitude (as opposed to the beta function) in the BCS configuration. If we assume a momentum-independent BCS coupling, it is straightforward to evaluate the one-loop amplitude

with logarithmic accuracy:

$$\mathcal{A}_{\rm BCS}(E) = \frac{g_{\rm B}^2}{8\pi^2}\left(\log^2\frac{V_F^2}{\Lambda} - \log^2\frac{V_F^2}{E} - i\pi\log\frac{V_F^2}{E}\right), \tag{4.39}$$

where we have kept only the leading terms in the real and imaginary parts. Taking the log derivative of equation (4.39) with respect to Λ reproduces the beta function (4.36) for $g_{\rm B}(0,0)$. However, the imaginary part of the amplitude depends on $\log\frac{V_F^2}{E}$. This large log is not resummed by the standard beta function and indicates that something akin to the rapidity renormalization group introduced in [4, 54] would be necessary to resum subleading logs.

In the special case $\Upsilon^2 = \Lambda$ our scheme in the VH region resembles that of Ref. [66]. In that work it is implicitly assumed that g is repulsive, and that Λ can be taken as high as the bandwidth, so that the NVH region is effectively absorbed into the VH region. However, lowering Λ then results in integrating some low-energy modes and requires nonlocal counterterms.

4.9 Higher-order renormalization

Let us discuss how higher-order corrections modify Eq. (4.39). This is particularly important for the kinematic configuration with zero net momentum, which controls the Cooper instability. Since the beta function at zero net momentum contains a logarithm of a large ratio, $\log(V_F^2/\Lambda)$, one may wonder if the one-loop computation is reliable in this kinematic configuration, or if one needs to resum the logs in the beta function itself. We will call logs containing V_F^2, such as $\log(V_F^2/E)$ or $\log(V_F^2/\Lambda)$, rapidity logs. We want to estimate the contribution of higher rapidity logs to the beta function at zero net momentum.

We will limit ourselves to the analysis of 2-loop diagrams. We take $\Upsilon \sim V_F$, in which case there are no large rapidity logs in non-VH loops. The renormalized coupling g is related to the bare coupling g_b by

$$g_b = g Z_4 Z_2^{-2}, \tag{4.40}$$

where Z_4 is the renormalization factor for the particle-particle four-point amplitude, and Z_2 is the wave function renormalization. Z_2 is finite at one loop, and at two-loop order is determined from the on-shell behavior of the self-energy diagram, Fig. 4.5, whose imaginary part is finite even without the rapidity cut-off [63, 64], and therefore does not contain rapidity logs.

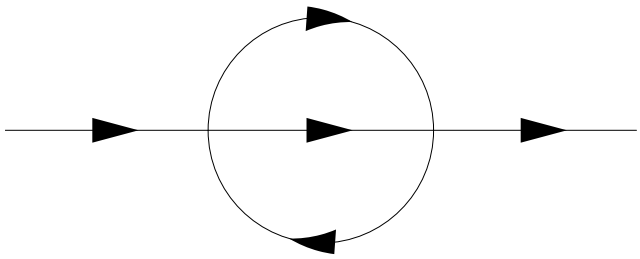


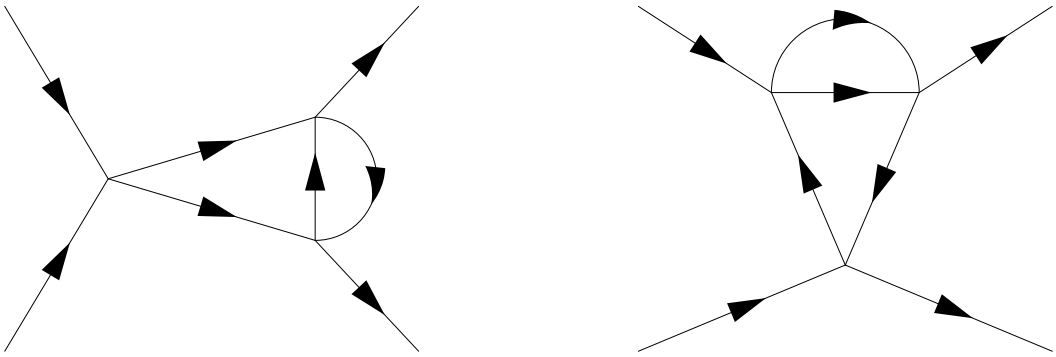
Figure 4.5: The two-loop self-energy with finite imaginary part.

Two-loop contributions to Z_4 arise from diagrams such as in Fig. 4.6. (Iterations of one-loop diagrams do not contribute since their infinities are removed by one-loop counter-terms.) Their contributions to the beta function can be estimated using what we already know about the one-loop diagrams. For example, the diagram Fig. 4.6a is obtained from the one-loop s-channel diagram by replacing one of the vertices with the one-loop t-channel diagram. The latter does not contain rapidity logs, so the contribution of the whole diagram to the beta function should behave in the same way as that of the one-loop s-channel diagram. In particular, it contains at most a single $\log(V_F^2/\Lambda)$ at zero net momentum. The diagram Fig. 4.6b can be regarded as a one-loop t-channel diagram with one vertex replaced with a one-loop s-channel diagram. The latter amplitude contains at most two rapidity logs, so the contribution of Fig. 4.6b to the beta function contains at most $\log^2(V_F^2/\Lambda)$. We conclude that with logarithmic accuracy the two-loop beta function at zero net momentum has the form

$$\beta(g) = \frac{1}{4\pi^2}(g^2 + Cg^3)\log\frac{V_F^2}{\Lambda} + C'g^3\log^2\frac{V_F^2}{\Lambda}, \tag{4.41}$$

where C and C' are constants.

Now we can see if the resummation of rapidity logs in the beta function is necessary. Eq. (4.37) indicates that the one-loop RG equations resum logs of the form $g\log^2(V_F^2/\Lambda)$. Thus we are assuming $g\log^2(V_F^2/\Lambda) \lesssim 1$, while $g\log(V_F^2/\Lambda) \ll 1$. This implies $g^3\log^2(V_F^2/\Lambda)$ is parametrically suppressed relative to $g^2\log(V_F^2/\Lambda)$. We conjecture that this behavior persists at higher loops, in the sense that every extra power of g is accompanied by at most a single rapidity log. If this is true, then resumming the rapidity logs in the beta function will not will not change qualitative conclusions regarding the RG flow and the Cooper instability.



(a) (b)

Figure 4.6: Examples of two-loop contributions to the beta function. The diagrams with iterated loops are not shown. The diagram on the right can contribute a double rapidity log to the beta function.

4.10 Kinematic requirements for s-channel dominance

We have seen that the s-channel diagram dominates the renormalization of the coupling function in the BCS configuration, where the net momentum of the incoming particles is very small. It is natural to ask to what degree this result holds for other configurations of the momentum, that is, under what circumstances the s-channel diagram alone will determine the leading log behavior. To answer this question, note that whenever the s-channel diagram's contribution to the beta function is enhanced by a logarithm of the ultrasoft scale relative to the t- or u-channel diagrams, we can drop those contributions. If we lower Λ using the RG such that the t- and u-channel diagrams make only power-suppressed contributions to the beta function, we can drop them in the computation of the leading log amplitudes. Therefore if the net momentum K is ultrasoft in terms of the varying value of Λ whenever the t- and u-channel diagrams make nonsuppressed contributions to the beta function, we can ignore these diagrams without making an error at leading log.

The t- and u-channel diagrams make suppressed contributions once the transfer momenta satisfy $\varepsilon(Q) > \Lambda$ and $\varepsilon(Q') > \Lambda$, respectively. The s-channel diagram is maximally log enhanced (K is still ultrasoft) as long as both components of K satisfy $K_\pm < \Lambda/\Upsilon$. Then as long as $\Lambda > K_\pm \Upsilon$, we can neglect the effects of the other diagrams. Therefore from the perspective of the UV theory, we can find "universal" results for the amplitudes (in that they depend only on the s-channel, and we can drop any nonanalytic dependence on Q and Q') when

$$\max(K_+, K_-) < \frac{\min(\varepsilon(Q), \varepsilon(Q'))}{\Upsilon}. \tag{4.42}$$

Hence we can make universal predictions whenever the interaction is more back-to-back than forward, in the sense of Eq. (4.42).

4.11 The collinear region as a marginal Fermi liquid

By definition, the collinear region is the part of the VH region where $|p_+|$ is of order of the rapidity cutoff Υ, while $|p_-|$ is less or equal than Λ/Υ. The anticollinear region is defined similarly, but with p_+ and p_- exchanged. Each of the following statements regarding the collinear region also applies to the anticollinear region.

Everywhere in the collinear region, the Fermi velocity is nonzero. Naively, one might conclude that this region is no different from the NVH region. In particular, one might think that the usual Fermi surface EFT [5] applies both in the NVH and the collinear region, but this is incorrect. To see why, recall that canonical Fermi surface EFT predicts that all interactions (apart from forward and BCS scattering) are irrelevant, and thus the quasiparticle width scales like $E^2/v_F k_B$ for small E. In the NVH region, v_F is of order W/k_B, thus the Fermi liquid theory applies for E much smaller than W. But it is well known [5] that additional marginal interactions arise when a portion of the Fermi surface is related to another portion of the Fermi surface by a translation in momentum space (nesting). The translation vector Q is called the nesting vector. The collinear region is an extreme example of this, since the Fermi surface is approximately invariant with respect to arbitrary shifts with $Q = (Q_+, 0)$. Following Wilczek and Nayak [58], we will refer to such a Fermi surface as flat.

Wilczek and Nayak emphasized the failure of the Fermi liquid theory for flat Fermi surfaces and proposed that the correct EFT for flat Fermi surfaces is quasi-1D, with the component of momentum parallel to the Fermi surface playing the role of a continuous label. In particular, the four-fermion interaction is marginal for generic combinations of momenta rather than irrelevant.

But there is also an important difference between the collinear region and the model of interacting 1D fermions (the Luttinger model). In the Luttinger model, the coupling is exactly marginal (has vanishing beta function). This is most easily seen using bozonization, which turns the Luttinger model into a free boson with a linear dispersion law. The vanishing of the beta function does not apply to the EFT describing the collinear region. The reason is that, unlike in the 1D case, the Fermi velocity varies along the Fermi surface. For definiteness, let us consider the collinear region and set $\mu = 0$. Then the "small" component of momentum is p_-,

while the "large" one is p_+. If we treat p_+ as a continuous label, the Fermi velocity is

$$v_F(p_+) = p_+. \tag{4.43}$$

As long as we consider generic scattering events between particles for which p_+ is $O(\Upsilon)$, the four-fermion coupling can be Taylor expanded in p_-, but not in p_+. Thus the leading interaction term

$$S_{int} = \int dt \int d^2p_1 d^2p_2 d^2p_3 \frac{1}{4} g(p_{1+}, p_{2+}, p_{3+}) \epsilon^{\alpha_1\alpha_2} \epsilon^{\alpha_3\alpha_4} \psi_{\alpha_1} \psi_{\alpha_2} \psi^\dagger_{\alpha_3} \psi^\dagger_{\alpha_4} \tag{4.44}$$

depends on a function of three real variables $g(p_{1+}, p_{2+}, p_{3+})$ that we take to be spin independent. This choice of spin structure for the interaction corresponds to the spin-singlet coupling, which we will focus on here. Furthermore, we take g to be symmetric under $p_1 \leftrightarrow p_2$ and $p_3 \leftrightarrow p_1 + p_2 - p_3$ independently, so the vertex factor is

$$i(\delta_{\alpha_1\alpha_3}\delta_{\alpha_2\alpha_4} - \delta_{\alpha_1\alpha_4}\delta_{\alpha_2\alpha_3}) g(p_1, p_2, p_3). \tag{4.45}$$

It is straightforward to compute the beta function for g. We find:

$$\begin{aligned} \frac{dg(p_{1+}, p_{2+}, p_{3+})}{d\log\mu} &= \frac{1}{2\pi^2} \int_K^\Upsilon dq \, \frac{g(p_{1+}, p_{2+}, q) g(q, K-q, p_{3+})}{2q - K} \\ &+ \frac{1}{8\pi^2 Q} \int_{Q - \min(Q,\Upsilon)}^{\min(Q,\Upsilon)} dq \, g(p_{1+}, q, p_{3+}) g(p_{2_+}, q+Q, p_{4+}) \\ &+ \frac{1}{8\pi^2 Q'} \int_{Q' - \min(Q',\Upsilon)}^{\min(Q',\Upsilon)} dq \, g(p_{1+}, q, p_{4+}) g(p_{2_+}, q+Q', p_{3+}) \end{aligned} \tag{4.46}$$

plus terms suppressed by $\varepsilon(p)/\Lambda$, where p is one of the external momenta. Here $K = p_{1+} + p_{2+}$ and $Q = p_{1+} - p_{3+}$ are assumed to be positive, for definiteness. Even if we take g to be independent of the "large" components of momenta at some scale, the RG evolution is nontrivial and introduces momentum dependence. At higher orders we will also have to take into account the renormalization of the Fermi velocity function $v_F(p_+)$. Finally, we neglected the spin-triplet coupling. Even if it is set to zero in the UV, it will be generated by radiative corrections, and thus a renormalizable theory should have both couplings. The above computation, which takes into account only the spin-singlet coupling, merely illustrates our point that the beta functions are nonzero in the collinear region.

The EFT that includes only the collinear region is sufficient to compute the width of the quasiparticle whose momentum is in the collinear region, where $|p_+|$ is of order Υ. If one formally takes the limit $\Upsilon \to \infty$ and assumes that the coupling g

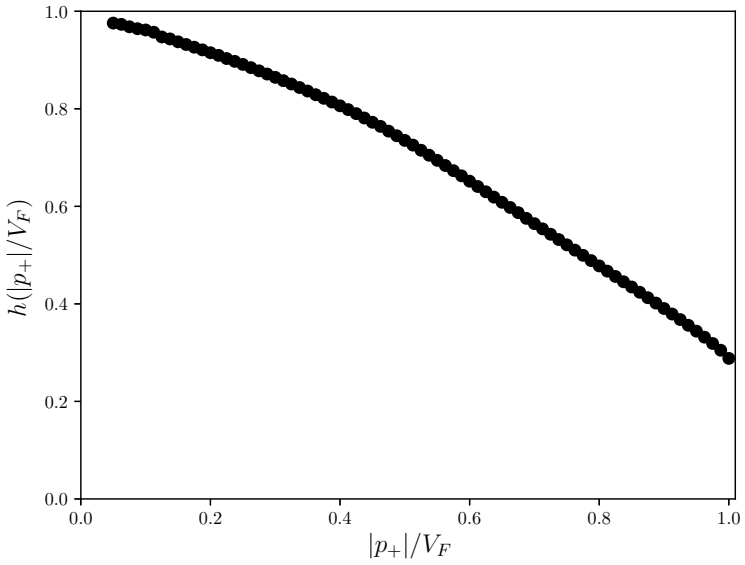


Figure 4.7: Numerical results for the dependence of h on $|p_+|/V_F$ in units of g^2 assuming a constant coupling. h is normalized to g^2 for $V_F \to \infty$.

is independent of momenta, the leading-order computation can be performed in the toy model and gives [63, 64]:

$$\Gamma(E) \sim g^2 E. \tag{4.47}$$

The linear dependence on E follows from dimensional analysis and is a hallmark of the marginal Fermi liquid [11]. The computation in the toy model cannot be extended to higher orders, since it is not a renormalizable theory. However, if we include the NVH region by introducing the rapidity cutoff $\Upsilon = V_F$, dimensional analysis gives a similar result:

$$\Gamma(E) \sim h(|p_+|/V_F)E, \tag{4.48}$$

where V_F is the typical Fermi velocity in the NVH region. At leading (two-loop) order the function $h(x)$ is of order g^2, but is not a constant even if one assumes, for simplicity, that g is a constant. Evaluating the imaginary part of the self-energy diagram (Fig. 4.5) numerically, we find the result in Fig. 4.7. Appendix C describes the calculations involved in determining this function.

Eq. (4.48) is valid provided we can neglect the chemical potential μ, a relevant coupling. Thus it holds in the range $|\mu| \ll E \ll W$. The corrections are of several sorts. The NVH region contribution is of order E^2/W, as usual. The corrections

from a nonzero μ are of order μ^2/E. Finally, higher orders in perturbation theory will give the function h a weak (logarithmic) dependence on E.

One of the defining properties of the MFL is that the quasiparticle width, defined via the imaginary part of the on-shell self-energy, is proportional to energy. The above arguments show that the Marginal Fermi Liquid behavior [11] is a robust consequence of the proximity to a Van Hove singularity. On the other hand, the dependence of the width on the "large" component of momentum can be nontrivial, unlike in the simplest models of Marginal Fermi Liquids.

4.12 Conclusions

We have presented a systematic effective field theory description of systems with a Van Hove singularity. The formalism is valid to leading power in an expansion in E/W and generalizes the classic results in [5, 6]. We have shown that the theory is renormalizable with all counterterms being local in the sense that they are finite in the zero energy limit. That such a formalism exists had to be the case given that any well-defined microscopic local theory must yield a renormalizable description, if it is properly formulated. A crucial ingredient in generating such a theory is the inclusion of all the relevant modes on the Fermi surface. Given that the entire surface is necessarily part of the IR description of the theory, it is not surprising that focusing solely on one region leads to nonlocalities.

The EFT that we constructed depends on a coupling function $g(k_1, k_2, k_3, -k_1 - k_2, -k_3)$ that cannot be expanded in powers of momenta (except when all momenta are ultrasoft). The appearance of an arbitrary function of six variables makes the theory much less predictive than the usual Fermi surface RG, which has two marginal couplings that depend on two variables each (for a 2D Fermi liquid). Nevertheless, we showed that in the BCS channel the EFT can be greatly simplified, provided we keep only logarithmically-enhanced terms. In this channel, one is left with a single function of two variables that satisfies a simple RG equation.

We have utilized our formalism to show that generic theories with Van Hove singularities will lead to Marginal Fermi Liquid behavior as previously anticipated using toy models [63, 64]. This behavior arises in both the soft and collinear subsectors of the VH region, the latter of which can constitute a considerable fraction of the Fermi surface. Thus our conclusions disagree with [66], where it was argued that for $E \gg \mu$ the Fermi liquid picture is valid. Our treatment of the collinear region clarifies the physics of Fermi surfaces with flat regions as discussed in [58]. We

also show that the running of the coupling in the BCS channel is logarithmically enhanced, and the coupling itself runs double logarithmically, in agreement with [9, 65].

Chapter 5

INSTABILITIES OF THE VAN HOVE EFT

In the previous chapter, we constructed an effective field theory to describe the normal state of two dimensional fermionic systems with Van Hove singularities on their Fermi surfaces. Understanding the instabilities of this theory is a particularly important topic because of the rich structure of the phase diagram of materials possessing van Hove singularities [77] and the observation of non-Fermi liquid behavior in the normal state [11]. We began to address this issue by studying the beta function, and we determined that the flow for the coupling between particles in a nearly back-to-back configuration suggests an instability to Cooper pairing.

In this chapter, we address the question of the instabilities of the system using another approach. In general, the analytic structure of the correlators in a field theory provide a great deal of information about the behavior of the theory itself, including the spectrum of excitations [19], bound states [81], and resonances. As an example, we found in Chapter 3 that it is possible to deduce the existence of the zero sound mode in Fermi liquids from the properties of the density-density correlator. The analytic structures of these correlators also give us clues about instabilities through their relationship to the response of the system to external perturbations. For example, the existence of a pole in frequency space along the positive imaginary axis for the retarded two-particle Green's function indicates the presence of an exponentially growing response and hence instability [61]. Analyzing these correlators in standard Fermi liquid theory reveals the onset of charge-density waves [82–85], spin-density waves [59, 86], and superconductivity [61], for example. We will apply the same reasoning to our system, beginning with an investigation the structure of the retarded two-particle Green's function.

5.1 Poles in the two-particle Green's function

From the arguments in Chapter 4, the strongest interactions in the IR should be due to the BCS coupling between back-to-back particles. Therefore we consider the analytic properties of the retarded two-particle Green's function at zero net momentum. This directly parallels our earlier analysis for the round Fermi surface in Section 3.12. As in that discussion, we closely follow [61].

We ignore the dependence of the coupling function on the transverse components of the momentum. This will be justified if we focus on states in the vicinity of the Van Hove singularity: the coupling between these modes drives the RG, so we can resum logs associated with that configuration alone. Call the coupling g.

We have shown that the leading diagrams are the iterated s-channel bubbles, as in the case of a round Fermi surface. Unlike in the round Fermi surface, where other diagrams are power suppressed, these diagrams are only suppressed by powers of the large logarithm $\log(V_F^2/\Lambda)$. Despite this fact, resumming the leading order diagrams should at least allow us to derive approximate results for the ground-state instability.

The two-particle Green's function at zero net momentum will have the same analytic structure as the four-particle scattering amplitude at zero net momentum [61]. Call the leading log contribution to the connected contribution to the four-particle scattering amplitude at zero net momentum Γ. This is the sum of the iterated s-channel diagrams, and it is given by a geometric series:

$$\Gamma(\omega) = \frac{g}{1 - \frac{g}{8\pi^2}\left(\log^2 \frac{V_F^2}{\Lambda} - \log^2 \frac{V_F^2}{\omega/2} - i\pi \log \frac{V_F^2}{\omega/2} + \frac{\pi^2}{6}\right)}. \tag{5.1}$$

We have dropped power-suppressed corrections but kept the order-one terms. At this point, we analytically continue Eq. (5.1) and look for singularities. Poles in the lower half-plane correspond to unstable excitations (i.e. those with a finite width), those on the real axis to stable states (with branch cuts corresponding to a continuum of multiparticle states), and poles in the upper half-plane to instabilities of the system. Furthermore, the location of the pole in this final scenario determines both the relaxation time of the unstable ground state and, via the uncertainty principle, the binding energy of the Cooper pairs [61].

From Eq. (5.1),

$$g\Gamma^{-1} = 1 - \tilde{g}\left(\lambda^2 - \log^2 \frac{V_F^2}{\omega/2} - i\pi \log \frac{V_F^2}{\omega/2} + \frac{\pi^2}{6}\right), \tag{5.2}$$

where

$$\tilde{g} \equiv \frac{g}{8\pi^2} \tag{5.3}$$

and

$$\lambda \equiv \log \frac{V_F^2}{\Lambda}. \tag{5.4}$$

We can write

$$\log^2 \frac{V_F^2}{\omega/2} + \mathrm{i}\pi \log \frac{V_F^2}{\omega/2} = \left(\log \frac{V_F^2}{\omega/2} + \frac{\mathrm{i}\pi}{2} \right)^2 + \frac{\pi^2}{4}. \tag{5.5}$$

Complexifying $\omega = |\omega| e^{\mathrm{i}\varphi}$ yields the analytic continuation of Eq. (5.2). The zeros Eq. (5.2) are the locations of the singularities of Γ, and are given by

$$\left(\log \frac{2V_F^2}{|\omega|} - \mathrm{i}\varphi + \frac{\mathrm{i}\pi}{2} \right)^2 = \gamma, \tag{5.6}$$

where

$$\gamma \equiv \lambda^2 - \frac{1}{\tilde{g}} - \frac{\pi^2}{12}. \tag{5.7}$$

We should also note that the presence of the logarithm implies that there is a branch cut in Γ, which we will take to run along the negative real axis.

Since γ is real, the expression in the parentheses in Eq. (5.6) must be along either the real or imaginary axis. Therefore either $\varphi = \pi/2$ or $\log \frac{2V_F^2}{|\omega|} = 0$. If $\varphi = \pi$, we must have

$$\log \frac{2V_F^2}{|\omega|} = \pm\sqrt{\gamma}. \tag{5.8}$$

Since the expression on the left is real, this can only hold if $\gamma > 0$. Solving for ω, we find two poles:

$$\omega = 2iV_F^2 e^{\pm\gamma}. \tag{5.9}$$

If we instead take $\log \frac{2V_F^2}{|\omega|} = 0$, Eq. (5.6) implies

$$-\left(\varphi + \frac{\pi}{2}\right)^2 = \gamma. \tag{5.10}$$

Since $(\varphi + \frac{\pi}{2})^2 > 0$, it is clear that this condition can be satisfied only if $\gamma < 0$. Solving for the phase yields

$$\varphi = \frac{\pi}{2} \pm \sqrt{-\gamma}, \tag{5.11}$$

and hence

$$\omega = 2\mathrm{i}V_F^2 e^{\pm\mathrm{i}\sqrt{-\gamma}}. \tag{5.12}$$

This implies that once again there should be two poles. However, instead of being along the imaginary axis, these poles lie on the circle $|\omega| = 2V_F^2$. If we take the branch cut to be along the negative real axis, one of the poles will reach the branch cut once $\gamma < -(\pi/2)^2$.

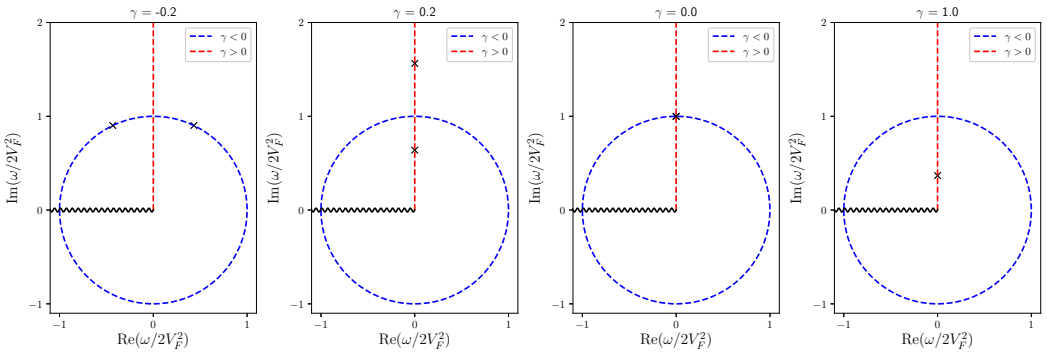
When $\gamma = 0$ the two poles in Eq. (5.9) and Eq. (5.12) merge, and we are left with

$$\omega = 2iV_F^2. \tag{5.13}$$

We can summarize the preceding results by writing the location of the poles in $\Gamma(\omega)$ as

$$\omega_\pm = 2iV_F^2 e^{\pm\sqrt{\gamma}}, \tag{5.14}$$

with $\sqrt{\gamma}$ understood to mean $i\sqrt{|\gamma|}$ when $\gamma < 0$. Fig. 5.1 demonstrates the positions of the poles and branch cut for several values of γ.



(a) (b) (c) (d)

Figure 5.1: Locations of the poles in $\Gamma(\omega)$ for different values of γ.

Let us find the form of $\Gamma(\omega)$ in the vicinity of the poles. We will work under the assumption that $\gamma > -(\pi/2)^2$ for now to avoid any issues with the branch cut. Also, note that taking $\gamma < -(\pi/2)^2$ push the pole moving clockwise around the circle as we decrease γ below the real axis, which would correspond to an unstable resonance. Define

$$z = \omega - \omega_\pm. \tag{5.15}$$

Then

$$g\Gamma^{-1} = 1 - \tilde{g}\left[\lambda^2 - \left(\log\frac{2V_F^2}{z+\omega_\pm} + \frac{i\pi}{2}\right)^2 - \frac{\pi^2}{12}\right]. \tag{5.16}$$

Since we assume that we are in the vicinity of the poles, $|z| \ll |\omega_\pm|$. Expanding the logarithm, we see

$$\log\frac{2V_F^2}{z+\omega_\pm} + \frac{i\pi}{2} = \log\frac{2V_F^2}{-i\omega_\pm} + \frac{z}{\omega_\pm} = \log e^{\mp\sqrt{\gamma}} - \frac{z}{\omega_\pm} = \mp\sqrt{\gamma} - \frac{z}{\omega_\pm}. \tag{5.17}$$

As long as $\gamma \neq 0$, the poles remain separate. When $\gamma \to 0$, they pinch together. Assume $\gamma \neq 0$ for now. Substituting Eq. (5.17) in Eq. (5.16) then yields

$$g\Gamma^{-1} = 1 - \tilde{g}\left(\frac{1}{\tilde{g}} \mp 2\sqrt{\gamma}\right) = \pm 2\sqrt{\gamma}\tilde{g}\frac{z}{\omega_\pm}, \tag{5.18}$$

or

$$\Gamma(\omega) = \frac{\pm 4\pi^2 \omega_\pm / \sqrt{\gamma}}{\omega - \omega_\pm} \tag{5.19}$$

in the vicinity of the poles at $\omega_\pm$. Then the residues at ω_+ and ω_- are

$$\mathrm{Res}(\Gamma, \omega_\pm) = \frac{\pm 4\pi^2 \omega_\pm}{\sqrt{\gamma}} = \frac{\pm 8\pi^2 V_F^2 \mathrm{i} e^{\pm\sqrt{\gamma}}}{\sqrt{\gamma}}. \tag{5.20}$$

If $\gamma = 0$, we instead have

$$g\Gamma^{-1} = 1 - \tilde{g}\left(\lambda^2 - \frac{z^2}{\omega_\pm} - \frac{\pi^2}{12}\right) \tag{5.21}$$

with $\omega_+ = \omega_- = 2\mathrm{i}V_F^2$. $\gamma = 0$ implies

$$\lambda^2 - \frac{\pi^2}{12} = \frac{1}{\tilde{g}}, \tag{5.22}$$

so Eq. (5.21) becomes

$$g\Gamma^{-1} = \frac{z^2}{\omega_\pm^2} \tag{5.23}$$

or

$$\Gamma(\omega) = \frac{-4V_F^4 g}{(\omega - 2\mathrm{i}V_F^2)^2} \tag{5.24}$$

in the vicinity of the double pole.

For $\gamma > 0$, the two poles lie on the imaginary axis. As γ increases, the poles move away from $2\mathrm{i}V_F^2$. ω_- moves towards the real axis, asymptotically approaching it as γ becomes large, and ω_+ moves to larger imaginary values at an exponential rate. In both cases the residue is purely imaginary, similar to the situation for the single pole in the BCS model with a constant coupling. As mentioned above, the position of the pole on the positive imaginary axis sets the relaxation time of the unstable Fermi liquid "ground state" and the binding energy of the Cooper pairs, with larger values for the position of the pole corresponding to shorter relaxation times and greater binding energies. This would naively imply that the pole at ω_+ should be the determining factor in the system.

However, the magnitude of the imaginary frequency at which that pole is located is exponentially larger than $2V_F^2$ for large γ. Since V_F^2 is by assumption greater than or equal to the UV cutoff for our effective theory, it is reasonable to assume that ω_+ represents an unphysical pole in this scenario.

Now consider $\gamma < 0$. At $\gamma = 0$ both poles merge together at $2iV_F^2$. As we lower γ, $\sqrt{\gamma}$ becomes imaginary and $e^{\pm\sqrt{\gamma}}$ is a pure phase with opposite signs for the two poles. Thus, the poles separate and move along the circle $|\omega| = 2V_F^2$. Eventually it seems that ω_- will merge into the branch cut along the negative real axis, while ω_+ will go below the real axis. In any case, for values of γ slightly greater than $-(\pi/2)^2$, the two poles are close to the real axis, and hence the relaxation time associated with these instabilities should be long (and the binding energies small). However, given that the magnitude of the complex frequency for both of these poles is $2V_F^2$, it is likely that they are also unphysical. In that case, the only physically relevant situation is when ω_- approaches the real axis along the imaginary axis for large, positive values of γ.

Remember that a negative coupling corresponds to an attractive potential. In that case, γ will be greater than zero whenever

$$\log \frac{V_F^2}{\Lambda} + \frac{8\pi^2}{|g|} > \frac{\pi^2}{12}. \tag{5.25}$$

The precise value of the order-one constant on the right-hand side of this equation may be modified by higher-loop effects and the behavior of the coupling away from the Van Hove point. Eq. (5.25) means that any theory with a negative coupling that satisfies

$$|g| > \frac{8\pi^2}{\frac{\pi^2}{12} - \log^2 \frac{V_F^2}{\Lambda}} \tag{5.26}$$

will have a Cooper pairing instability, at least in the iterated one-loop approximation considered here. The binding energy of the Cooper pairs is approximated by

$$\Delta E \sim 2V_F^2 \exp\left(-\sqrt{\log^2 \frac{V_F^2}{\Lambda} - \frac{8\pi^2}{g} - \frac{\pi^2}{12}}\right). \tag{5.27}$$

This is of course closely related to the strong coupling scale predicted by the RG flow in Section 4.8. Note that as long as the effective coupling is negative when the UV scale obeys

$$\Lambda < e^{-\pi/\sqrt{12}} V_F^2 \sim 0.40 V_F^2, \tag{5.28}$$

Eq. (5.26) is automatically satisfied.

Interestingly, the reasoning above implies it is also possible for a positive coupling to produce a pairing instability. This occurs whenever

$$|g| > \frac{8\pi^2}{\log^2 \frac{V_F^2}{\Lambda} - \frac{\pi^2}{12}}. \tag{5.29}$$

If the UV cutoff is order V_F^2, Eq. (5.29) would necessitate a nonperturbative coupling, and hence we should not trust the result in that limit. On the other hand, if there is a large separation of scales between the Fermi velocity and the cutoff of the effective theory, g could be relatively small and still satisfy the condition for ω_- to lie on the imaginary axis. However, in that case we can write

$$\sqrt{\lambda^2 - \frac{1}{\tilde{g}} - \frac{\pi^2}{12}} \approx \lambda - \frac{1}{2\lambda}\left(\frac{1}{\tilde{g}} + \frac{\pi^2}{12}\right), \tag{5.30}$$

and therefore the binding energy of the Cooper pairs is approximately

$$\Delta E \sim 2\Lambda \exp\left[\frac{1}{2\lambda}\left(\frac{1}{\tilde{g}} + \frac{\pi^2}{12}\right)\right]. \tag{5.31}$$

The argument of the exponential is positive. This implies the binding energy must be outside domain of validity of our theory, which suggests that positive values of the coupling likely do not actually result in a pairing instability.

In summary, we have found that the only physical pole in the two-particle Green's function occurs at

$$\omega_- = 2\mathrm{i}V_F^2 e^{-\sqrt{\gamma}}, \tag{5.32}$$

which implies a binding energy of

$$\Delta E \sim 2V_F^2 \exp\left(-\sqrt{\log^2 \frac{V_F^2}{\Lambda} - \frac{8\pi^2}{g} - \frac{\pi^2}{12}}\right) \tag{5.33}$$

for the Cooper pairs. This is consistent with our renormalization group analysis. Higher loop effects are likely to modify the order-one constant inside the exponential. There are two interesting features of Eq. (5.33). First, the argument of the exponential goes like $1/\sqrt{g}$ when g is small, unlike the behavior in a regular Fermi liquid, where the binding energy goes like $1/g$. Second, the scale being multiplied by the exponential is V_F^2, as opposed to the usual UV cutoff, which acts as a stand-in for either the bandwidth of the Debye frequency. Both of these effects imply a larger binding energy, which is also suggestive of a higher superconducting transition temperature.

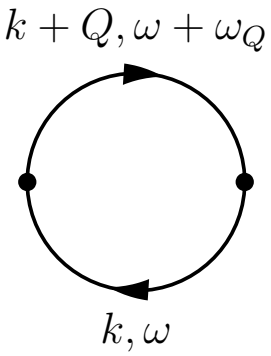


Figure 5.2: The diagram corresponding to the susceptibility.

5.2 The susceptibility

Let us turn now to the calculation of the particle-hole susceptibility represented by the diagram in Fig. 5.2. Within standard Fermi liquid theory, this quantity is omnipresent. This can be traced to its role as the leading contribution in perturbation theory to the density-density correlator. The density-density correlator in turn describes the response of various attributes of the system to an external forcing. The susceptibility also clearly corresponds to the loop integral in the one-loop t-channel diagram in a theory with constant coupling. Furthermore, as we noted in Section 3.13, Landau showed that the analytic structure of this diagram at vanishing energy and momentum relates the forward scattering amplitude to the phenomenon of zero sound [36]. In one spatial dimension, the presence of a divergence in the susceptibility at $Q = 2K_F$ and zero frequency indicates that the system is unstable to various density wave states [59, 83–85, 87]. We will see that analogous behavior appears for Fermi surfaces with Van Hove singularities.

As the response of the system depends on the density-density correlator, we will focus on a related quantity, the compressibility. From [5], the compressibility $\chi(Q, \omega_Q)$ is given by

$$(2\pi)^3 \delta^3(0) \chi(Q, \omega_Q) \equiv -\langle \rho(Q, \omega_Q) \rho(-Q, -\omega_Q) \rangle, \tag{5.34}$$

where ρ is the Fourier transform of the fermion density $\psi^\dagger \psi(x)$. As an example of the importance of this quantity, acoustic excitations correspond to poles in the compressibility at finite ω_Q and Q [88]. We will focus on the case where ω_Q is small in this analysis because we are interested in low energy dynamics of the system. We calculated this diagram in Section 3.13 in the case of a round Fermi surface. We will follow a similar approach here.

We will start by assuming a constant spin-singlet interaction. This is certainly not a natural assumption (in the sense of effective field theory), since we have already seen that quantum corrections are guaranteed to alter this condition upon renormalization. Despite that, it will help us understand the nature of the interaction when we add the complication of a momentum-dependent coupling. Call the regular part g. In this approximation, each additional bubble contributes a single factor of $gI(Q,\omega_Q)$ to a given diagram, where

$$I(Q,\omega_Q) = \mathrm{i}\int \frac{\mathrm{d}\omega\,\mathrm{d}^2k}{(2\pi)^3}\frac{\mathrm{i}}{\omega-\varepsilon(k)+\mathrm{i}\epsilon\operatorname{sign}\varepsilon(k)} \times \frac{\mathrm{i}}{\omega+\omega_Q-\varepsilon(k+Q)+\mathrm{i}\epsilon\operatorname{sign}\varepsilon(k+Q)} \tag{5.35}$$

is the susceptibility. All cutoffs have been suppressed for now.

The sum of all the contributing diagrams is a geometric series, so

$$\chi = \frac{I}{1-gI}. \tag{5.36}$$

Therefore the poles in the compressibility occur when

$$I(Q,\omega_Q) = \frac{1}{g}. \tag{5.37}$$

I is basically the same as the one-loop t-channel amplitude. Performing the energy contour integral and changing variables to $k_\pm$ yields

$$I(Q,\omega_Q) = \frac{1}{8\pi^2}\int \mathrm{d}^2k \frac{\theta(-\varepsilon(k))\theta(\varepsilon(k+Q)) - \theta(\varepsilon(k))\theta(-\varepsilon(k+Q))}{\varepsilon(k+Q)-\varepsilon(k)-\omega_Q+\mathrm{i}\epsilon\operatorname{sign}\varepsilon(k)}. \tag{5.38}$$

We have suppressed the cutoff dependence in the above expression. We will take the Fermi velocity cutoff to be of the order of the typical Fermi velocity in the non-Van Hove part of the Fermi surface. Assume $\omega_Q > 0$ and $Q_\pm > 0$. Then the region allowed by the step functions looks like rectangular strips of width Q_+ and Q_- along the Fermi surface as long as both components of Q are ultrasoft. If either is larger than the ultrasoft scale Λ/V_F, the strips will intersect with the energy cutoffs on the two particles, leading to a complicated shape.

We can perform the same rescaling used in Appendix A to deal with the situation where the components of Q are not both ultrasoft. Choose $Q_+ > Q_-$. By changing variables to

$$k_+ = \eta k'_+,\ k_- = \frac{k'_-}{\eta}, \tag{5.39}$$

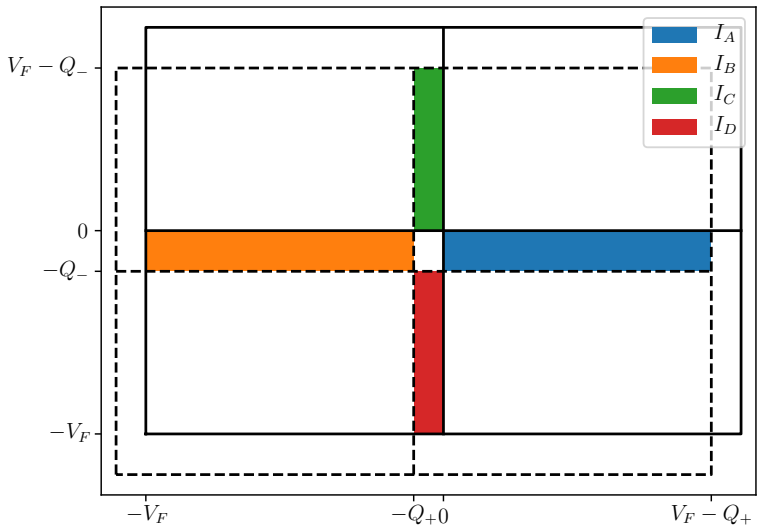


Figure 5.3: The integration region for calculating the contribution of bubbles in the compressibility. The dotted lines show the location of the Fermi surface and Fermi velocity cutoffs for the propagator with momentum $k + Q$. The solid lines are for the propagator with momentum k. Note that the energy cutoffs and the restrictions they impose on the integration regions are not shown.

where

$$\eta = \frac{\varepsilon(Q)}{\sqrt{Q_-}} = \sqrt{\frac{Q_+}{Q_-}}, \tag{5.40}$$

we can effectively set $Q_+ = Q_- = \sqrt{\varepsilon(Q)} \equiv Q$ within the integral. This is due to the $O(1,1)$ invariance of the problem in the absence of V_F. Note that η is greater than one. Under this change of variables, the step functions enforcing the rapidity cutoff transform. For example, the one constraining k_+ becomes

$$\theta(V_F - |k_+|) = \theta\left(\frac{V_F}{\eta} - |k'_+|\right), \tag{5.41}$$

the one constraining k_- becomes

$$\theta(V_F - |k_-|) = \theta(\eta V_F - |k'_-|), \tag{5.42}$$

and similarly for those constraining $k + Q$.

Writing out the integrand of Eq. (5.38),

$$\frac{1}{k_+Q + k_-Q + \varepsilon(Q) - \omega_Q + \mathrm{i}\epsilon\,\mathrm{sign}\,\varepsilon(k)} = \left(\frac{1}{\varepsilon(Q)}\right)\frac{1}{k_+/Q + k_-/Q + 1 - \omega_Q^* + \mathrm{i}\epsilon\,\mathrm{sign}\,\varepsilon(k)}, \tag{5.43}$$

where $\omega_Q^* \equiv \omega_Q/\varepsilon(Q)$, suggests another change of variables to $\tilde{k}_\pm = k_\pm/Q$. Then Eq. (5.38) is

$$I(Q,\omega_Q) = \frac{1}{8\pi^2}(I_A - I_B + I_C - I_D), \tag{5.44}$$

with each integral corresponding to one of the strips labeled in Fig. 5.3. Each of the these integrals can be written in the form

$$I_A = \int_0^{V_F^*/\eta - 1} \mathrm{d}\tilde{k}_+ \int_{-1}^{0} \mathrm{d}\tilde{k}_- \frac{f_A}{\tilde{k}_+ + \tilde{k}_- + 1 - \omega_Q^* \mp \mathrm{i}\epsilon}, \tag{5.45}$$

where

$$V_F^* \equiv \frac{V_F}{Q}, \tag{5.46}$$

the sign of $\mathrm{i}\epsilon$ matches the sign in of the corresponding term in Eq. (5.44), and f_A contains the information about the energy cutoffs. Note that the sign in the pole prescription is reversed in I_B and I_D compared to I_A. The limits come from measuring the strips in Fig. 5.3 in units where $Q_+ = Q_- = 1$ and remembering that the cutoff has been scaled to V_F/η in the k_+ direction and ηV_F in the k_- direction. Walking through the various changes of variables shows us

$$f_A = f_C = \theta(\Lambda^* + \tilde{k}_+\tilde{k}_-)\theta(\Lambda^* - (\tilde{k}_+ + 1)(\tilde{k}_- + 1)), \tag{5.47}$$

$$f_B = f_D = \theta(\Lambda^* - \tilde{k}_+\tilde{k}_-)\theta(\Lambda^* + (\tilde{k}_+ + 1)(\tilde{k}_- + 1)), \tag{5.48}$$

where

$$\Lambda^* \equiv \frac{\Lambda}{\varepsilon(Q)}. \tag{5.49}$$

We should change variables again to

$$x = \tilde{k}_+ + \tilde{k}_-, \tag{5.50}$$

$$y = \tilde{k}_+ - \tilde{k}_-, \tag{5.51}$$

so

$$k_\pm = \frac{1}{2}(x \pm y). \tag{5.52}$$

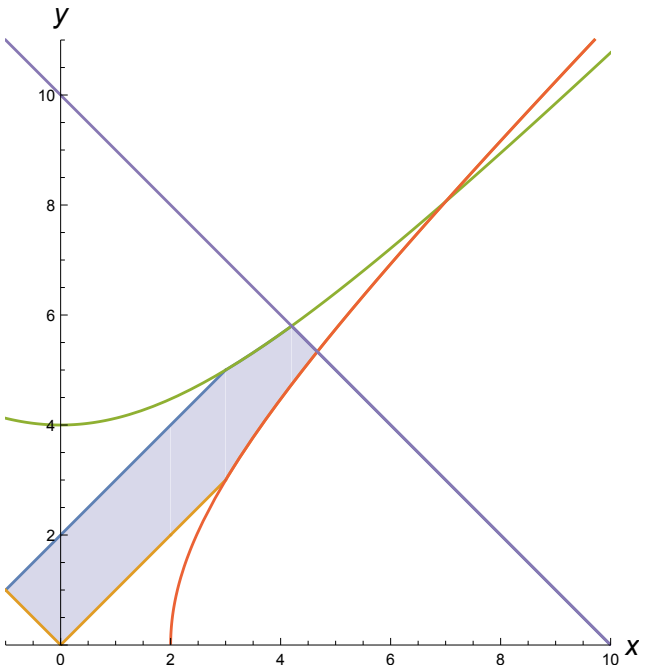


Figure 5.4: An example of the integration region for I_A in the rotated coordinates x and y. $\Lambda^* = 4$, $V_F^* = 12$, and $\eta = 2$. The red and green hyperbolas are the energy cutoffs, while the purple line is the rapidity cutoff.

Then

$$I_A = \frac{1}{2}\int \mathrm{d}y\,\mathrm{d}x \frac{1}{x + 1 - \omega_Q^* - \mathrm{i}\epsilon}. \tag{5.53}$$

This change of variables is a rotation, flip, and rescaling, and it moves the strip to a diagonal orientation. The energy cutoffs in terms of the new variables are

$$f_A = f_C = \theta(4\Lambda^* + x^2 - y^2)\theta(4\Lambda^* - (x+2)^2 + y^2), \tag{5.54}$$

$$f_B = f_D = \theta(4\Lambda^* - x^2 + y^2)\theta(4\Lambda^* + (x+2)^2 - y^2). \tag{5.55}$$

Λ^* is the cutoff measured in units of $\varepsilon(Q)$. Therefore we always have $\Lambda^* > 1$ for momenta within the domain of applicability of the effective theory. Fig. 5.4 shows the effects of the energy cutoffs on the rotated integration region.

In these coordinates, it is not hard to see that the hyperbolas describing the energy cutoffs intersect at $x = 2\Lambda^* - 1$ for I_A. Furthermore, the hyperbolas both intersect the straight edges of the strip at $x = \Lambda^* - 1$. The rapidity cutoff is the line

$$y = 2\left(\frac{V_F^*}{\eta} - 1\right) - x, \tag{5.56}$$

where

$$V_\eta \equiv \frac{V_F^*}{\eta} = \frac{V_F}{Q_+}. \tag{5.57}$$

The pole in the integrand occurs at the vertical line $x = \omega_Q^* - 1$ and must therefore always occur before the point at which the hyperbolas intersect the edges of the strip.

The integrand approaches its pole at $x = \omega_Q^* - 1$. Performing the integration with the $\mathrm{i}\epsilon$ prescription is equivalent to choosing an integration contour along the real axis except for a positively oriented semicircle of vanishing radius passing below the real axis at the location of the pole. Therefore

$$\frac{1}{x + 1 - \omega_Q^* - \mathrm{i}\epsilon} = \mathrm{P.V.}\frac{1}{x + 1 - \omega_Q^*} + \mathrm{i}\pi\delta(x + 1 - \omega_Q^*). \tag{5.58}$$

In fact, it is straightforward to see that the imaginary part of I_A is proportional to the height in the y direction of the rotated strip for constant $x = \omega_Q^* - 1$. The fact that the pole always occurs before the intersection with the hyperbola means we can ignore the energy cutoffs. Then

$$\mathrm{Im}\, I_A = \begin{cases} \pi \min(\omega_Q^*, 1), & \omega_Q^* < V_\eta - 1 \\ \pi\left(V_\eta - 1 - \omega_Q^*\right), & V_\eta - 2 \le \omega_Q^* < V_\eta - 1 \\ 0, & \omega_Q^* \ge V_\eta - 1, \end{cases} \tag{5.59}$$

where we have remembered to divide by 2 due to the change of variables to x and y. An identical result holds for $\mathrm{Im}\, I_C$ with $V_\eta \to \eta V_F^*$. I_B and I_D do not make contributions to the imaginary part because we have assumed that $\omega_Q > 0$ and $Q_\pm > 0$. The roles of these integrals would have been interchanged with I_A and I_C if we had assumed the signs of ω_Q and $\varepsilon(Q)$ differed.

We can recover the results of [63] by letting the Fermi velocity cutoff go to infinity. Then $\mathrm{Im}\, I_A = \mathrm{Im}\, I_C = \pi \min(\omega_Q^*, 1)$, implying

$$\mathrm{Im}\, I(Q, \omega_Q) = \frac{1}{4\pi} \min\left(\frac{\omega_Q}{\varepsilon(Q)}, 1\right), \tag{5.60}$$

which agrees with Eq. (14) in [63] up to a difference in normalization.

To determine the real part, we must consider the different possible cases for the location of the rapidity cutoff relative to the onsets of the effects of the energy cutoffs. The rapidity cutoff matters whenever V_η is small enough that the line

impinges on the integration region. Using Eq. (5.56), the requirements for the rapidity cutoff to modify the integration region will take the form

$$V_\eta < \frac{x+y}{2} + 1, \tag{5.61}$$

where x and y will be the coordinates of some point on the boundary of the strip with the hyperbolic energy cutoffs included.

The position of the intersection of the hyperbolas is $x = 2\Lambda^* - 1$, $y = \sqrt{4\Lambda^{*2}+1}$. Therefore the rapidity cutoff first matters when

$$V_\eta < \frac{1}{2}\left(\sqrt{4\Lambda^{*2}+1} + 2\Lambda^* + 1\right) \equiv 2\bar{\Lambda} \approx 2\Lambda^*. \tag{5.62}$$

Similarly, when

$$V_\eta < \Lambda^* + 1, \tag{5.63}$$

the rapidity cutoff prevents the upper hyperbola from having any effect. Finally, when

$$V_\eta < \Lambda^*, \tag{5.64}$$

neither of the energy cutoffs play a role at all for I_A and I_B. The role of V_η and ηV_F^* are reversed for the strips along the k_- axis, I_C and I_D. Therefore the energy cutoffs are guarantee to drop out of the calculation when

$$\eta V_F^* < \Lambda^*, \tag{5.65}$$

which corresponds to the situation when both components of the momentum are ultrasoft.

We analyze three cases. First, we find the susceptibility in the limit where the energy cutoff Λ is low enough that the rapidity cutoff has no effect. This corresponds to the inverse of the condition in Eq. (5.62) above. Second, we analyze the situation where the rapidity cutoff affects the strips along one axis and not the other. This corresponds to

$$\frac{\Lambda^*+1}{\eta} < V_F^* < \eta\Lambda^*. \tag{5.66}$$

This is only possible when Q_+ and Q_- are sufficiently different. Finally, we consider the scenario where Eq. (5.65) holds. We will not focus on the "transition zones" where a given strip is partially affected by the rapidity cutoff and partially by the energy cutoffs. These situations correspond to artefacts of the hard cutoffs we have implemented and will not give us insight into the behavior of the system.

Neither component ultrasoft, $V_F^* > 2\eta\bar{\Lambda}$

In this case, the rapidity cutoff has no effect. We will cut the integration region into three pieces,

$$I_A = \frac{1}{2}(I_{A1} + I_{A2} + I_{A3}), \tag{5.67}$$

and we will take the principal value in each case, since we have already determined the imaginary part.

$$I_{A1} = \int_{-1}^{0} \mathrm{d}x \int_{-x}^{x+2} \mathrm{d}y \frac{1}{x+1-\omega_Q^*}, \tag{5.68}$$

$$I_{A2} = \int_{0}^{\Lambda^*-1} \mathrm{d}x \int_{x}^{x+2} \mathrm{d}y \frac{1}{x+1-\omega_Q^*}, \tag{5.69}$$

$$I_{A3} = \int_{\Lambda^*-1}^{2\Lambda^*-1} \mathrm{d}x \int_{\sqrt{(x+2)^2-4\Lambda^*}}^{\sqrt{x^2+4\Lambda^*}} \mathrm{d}y \frac{1}{x+1-\omega_Q^*}. \tag{5.70}$$

I_{A1} and I_{A2} are straightforward:

$$I_{A1} = \int_{-1}^{0} \mathrm{d}x \frac{2(x+1)}{x+1-\omega_Q^*} = 2\left(1 + \omega_Q^* \log\left|\frac{\omega_Q^*-1}{\omega_Q^*}\right|\right), \tag{5.71}$$

$$I_{A2} = \int_{0}^{\Lambda^*-1} \mathrm{d}x \frac{x+2-x}{x+1-\omega_Q^*} = 2(\log|\Lambda^*-\omega_Q^*| - \log|\omega_Q^*-1|). \tag{5.72}$$

For I_{A3}, we have

$$I_{A3} = \int_{\Lambda^*-1}^{2\Lambda^*-1} \mathrm{d}x \frac{\sqrt{x^2+4\Lambda^*} - \sqrt{(x+2)^2-4\Lambda^*}}{x+1-\omega_Q^*}. \tag{5.73}$$

The indefinite integral of the first term is

$$\begin{aligned}\int \mathrm{d}x \frac{\sqrt{x^2+4\Lambda^*}}{x+1-\omega_Q^*} &= x_1 + (\omega_Q^*-1)\log(x_1+x) \\ &\quad + 2\Lambda_1\{\log(x+1-\omega_Q^*) - \log[4\Lambda^* + (\omega_Q^*-1)x + 2x_1\Lambda_1]\},\end{aligned} \tag{5.74}$$

where

$$x_1 \equiv \sqrt{x^2+4\Lambda^*} \tag{5.75}$$

and

$$\Lambda_1 \equiv \frac{1}{2}\sqrt{4\Lambda^* + (\omega_Q^*-1)^2}. \tag{5.76}$$

The indefinite integral of the second term is

$$\int \mathrm{d}x \frac{\sqrt{(x+2)^2 - 4\Lambda^*}}{x + 1 - \omega_Q^*} = x_2 - 2\Lambda_2 \arctan\left(\frac{2 - 4\Lambda^* + 2\omega_Q^* + (1+\omega_Q^*)x}{2\Lambda_2 x_2}\right) + (1+\omega_Q^*)\log\left(\frac{2 + x + x_2}{\omega_Q^*}\right), \quad (5.77)$$

where

$$x_2 \equiv \sqrt{(x+2)^2 - 4\Lambda^*} \quad (5.78)$$

and

$$\Lambda_2 \equiv \frac{1}{2}\sqrt{4\Lambda^* - (\omega_Q^* + 1)^2}. \quad (5.79)$$

With this, we can substitute the appropriate limits from Eq. (5.73) and find the precise result for I_A. However, this is not particularly enlightening. Instead, we can simply take the limit of the various expressions as $\omega_Q^* \to 0$, corresponding to the condition for small values of the transfer energy, and $\Lambda^* \to \infty$, corresponding to small values of the transfer momentum in the model without a rapidity cutoff. In the $\omega/\varepsilon(Q) \to 0$ limit, we have

$$I_A \to 2 + \log\frac{\Lambda^*}{2} = 2 + \log\frac{\Lambda}{2\varepsilon(Q)}. \quad (5.80)$$

Since I_C is identical to I_A under the exchange $k_+ \leftrightarrow k_-$, it must have the same value (the rapidity cutoff does not come into play). I_B requires a bit of work. After changing variables to $\tilde{k}_\pm$, we have

$$I_B = \int_{-V_F^*/\eta}^{-1} \mathrm{d}\tilde{k}_+ \int_{-1}^{0} \mathrm{d}\tilde{k}_- \frac{f_B}{\tilde{k}_+ + \tilde{k}_- + 1 - \omega_Q^* + \mathrm{i}\epsilon}. \quad (5.81)$$

After changing variables to x and y, the rotated strip has corners at $(-2, 0)$, $(-1, -1)$, $(-V_\eta, -V_\eta)$, and $(-V_\eta - 1, -V_\eta + 1)$. The energy hyperbolas are given by

$$y = -\sqrt{x^2 - 4\Lambda^*} \quad (5.82)$$

and

$$y = -\sqrt{(x+2)^2 + 4\Lambda^*}. \quad (5.83)$$

The intercepts with the edges of the strip occur at $x = -\Lambda^* - 1$, and the hyperbolas intercept each other at $x = -2\Lambda^* - 1$. Then we have

$$I_B = \frac{1}{2}(I_{B1} + I_{B2} + I_{B3}), \quad (5.84)$$

$$I_{B1} = \int_{-2\Lambda^*-1}^{-\Lambda^*-1} \mathrm{d}x \int_{-\sqrt{(x+2)^2+4\Lambda^*}}^{-\sqrt{x^2-4\Lambda^*}} \mathrm{d}y \frac{1}{x+1-\omega_Q^*}, \tag{5.85}$$

$$I_{B2} = \int_{-\Lambda^*-1}^{-2} \mathrm{d}x \int_{x}^{x+2} \mathrm{d}y \frac{1}{x+1-\omega_Q^*}, \tag{5.86}$$

$$I_{B3} = \int_{-2}^{-1} \mathrm{d}x \int_{x}^{-2-x} \mathrm{d}y \frac{1}{x+1-\omega_Q^*}. \tag{5.87}$$

Evaluating I_{B2} and I_{B3} is straightforward:

$$I_{B2} = \int_{-\Lambda^*-1}^{-2} \mathrm{d}x \frac{2}{x+1-\omega_Q^*} = 2(\log|\omega_Q^*+1| - \log|\Lambda^*+\omega_Q^*|), \tag{5.88}$$

$$I_{B3} = \int_{-2}^{-1} \mathrm{d}x \frac{-2-2x}{x+1-\omega_Q^*} = -2\left(1+\omega_Q^* \log\left|\frac{\omega_Q^*}{\omega_Q^*+1}\right|\right). \tag{5.89}$$

I_{B1} is ugly,

$$I_{B1} = \int_{-2\Lambda^*-1}^{-\Lambda^*-1} \mathrm{d}x \frac{\sqrt{(x+2)^2+4\Lambda^*} - \sqrt{x^2-4\Lambda^*}}{x+1-\omega_Q^*}, \tag{5.90}$$

but if we change variables to $x' = -x-2$, we see that it is actually given by

$$I_{B1} = \int_{\Lambda^*-1}^{2\Lambda^*-1} \mathrm{d}x \frac{\sqrt{(x+2)^2-4\Lambda^*} - \sqrt{x^2+4\Lambda^*}}{x+1+\omega_Q^*}, \tag{5.91}$$

which is identical to $-I_{A3}$ with $\omega_Q^* \to -\omega_Q^*$.

Under the current circumstances, the rapidity cutoffs do not matter, which means that I_D has the same value as I_B. Then

$$\mathrm{Re}\, I(Q,\omega_Q) = \frac{1}{8\pi^2}(I_{A1}+I_{A2}+I_{A3}-I_{B1}-I_{B2}-I_{B3}), \tag{5.92}$$

where the factors of one half in Eq. (5.67) and Eq. (5.84) have been canceled by I_C and I_D having the same values as I_A and I_B. Taking the limit $\omega \to 0$, $\Lambda^* \to \infty$ of the expressions found above using a computer algebra system, we find

$$I(Q,\omega_Q) = \frac{1}{2\pi^2}\left(2+\log\frac{\Lambda}{2\varepsilon(Q)} + \frac{\mathrm{i}\pi}{2}\frac{\omega}{\varepsilon(Q)}\right), \tag{5.93}$$

plus terms of order $\varepsilon(Q)/\Lambda$ or ω/Λ, in the $\omega/\varepsilon(Q) \to 0$ limit for small $\varepsilon(Q)$ when both components of Q are not ultrasoft. There is no cancellation between the positive and negative strips, and they each contribute half of the real part of Eq. (5.93).

One component ultrasoft

The condition in Eq. (5.66) implies that the strip along the original k_+ axis will be affected by the rapidity cutoff. Thus, the values of the real parts of I_A and I_B will be modified from those found in the previous section. Fortunately, the strips maintain a rectangular shape under the currently considered circumstances, greatly simplifying the analysis.

We again cut I_A into three pieces,

$$I_A = \frac{1}{2}(I_{A1} + I_{A2} + I_{A3}), \tag{5.94}$$

with the same I_{A1} from before, Eq. (5.68), and

$$I_{A2} = \int_0^{V_\eta - 2} \mathrm{d}x \int_x^{x+2} \mathrm{d}y \frac{1}{x + 1 - \omega_Q^*}, \tag{5.95}$$

$$I_{A3} = \int_{V_\eta - 2}^{V_\eta - 1} \mathrm{d}x \int_x^{2V_\eta - 2 - x} \mathrm{d}y \frac{1}{x + 1 - \omega_Q^*}, \tag{5.96}$$

where as before we take the principal value. The expression for I_{A2} here is the same as Eq. (5.69) with $\Lambda^* \to V_\eta - 1$, so

$$I_{A2} = 2(\log|V_\eta - \omega_Q^* - 1| - \log|\omega_Q^* - 1|). \tag{5.97}$$

For I_{A3}, we have

$$I_{A3} = -2\int_{V_\eta - 2}^{V_\eta - 1} \mathrm{d}x \frac{x + 1 - V_\eta}{x + 1 - \omega_Q^*} = -2\left(1 + (\omega_Q^* - V_\eta)\log\left|\frac{V_\eta - \omega_Q^*}{V_\eta - \omega_Q^* - 1}\right|\right). \tag{5.98}$$

Summing these gives us

$$I_A = \omega_Q^* \log\left|\frac{\omega_Q^* - 1}{\omega_Q^*}\right| + \log\left|\frac{V_\eta - \omega_Q^* - 1}{\omega_Q^* - 1}\right| + (V_\eta - \omega_Q^*)\log\left|\frac{V_\eta - \omega_Q^*}{V_\eta - \omega_Q^* - 1}\right|. \tag{5.99}$$

The calculation of I_B is similar. We write

$$I_B = \frac{1}{2}(I_{B1} + I_{B2} + I_{B3}), \tag{5.100}$$

with

$$I_{B1} = \int_{-V_\eta - 1}^{-V_\eta} \mathrm{d}x \int_{-2V_\eta - x}^{x+2} \mathrm{d}y \frac{1}{x + 1 - \omega_Q^*}, \tag{5.101}$$

$$I_{B2} = \int_{-V_\eta}^{-2} \mathrm{d}x \int_x^{x+2} \mathrm{d}y \frac{1}{x + 1 - \omega_Q^*}, \tag{5.102}$$

and I_{B3} from Eq. (5.89). We find

$$I_{B1} = 2 + 2(V_\eta + \omega_Q^*) \log \left| \frac{V_\eta + \omega_Q^* - 1}{V_\eta + \omega_Q^*} \right| \tag{5.103}$$

and

$$I_{B2} = 2 \log \left| \frac{\omega_Q^* + 1}{V_\eta + \omega_Q^* - 1} \right|. \tag{5.104}$$

Summing the contributions gives us

$$I_B = (V_\eta + \omega_Q^*) \log \left| \frac{V_\eta + \omega_Q^* - 1}{V_\eta + \omega_Q^*} \right| + \log \left| \frac{\omega_Q^* + 1}{V_\eta + \omega_Q^* - 1} \right| - \omega_Q^* \log \left| \frac{\omega_Q^*}{\omega_Q^* + 1} \right|. \tag{5.105}$$

Taking $\omega \to 0$ yields

$$I_A - I_B = 2[V_\eta \log V_\eta - (V_\eta - 1) \log (V_\eta - 1)]. \tag{5.106}$$

We are considering the limit where Q_- is ultrasoft and Q_+ is not. There is nothing inconsistent with taking the $\Lambda^* \to \infty$ limit with this condition in place (we simply decrease Q_- at fixed Q_+), which allows us to compare directly to the scenario described by Eq. (5.93). As mentioned above, the contributions from I_C and I_D are one half the result with no rapidity cutoff. Realizing that the imaginary part should not be affected because we are taking the $\omega \to 0$ limit, we have

$$I(Q, \omega_Q) = \frac{1}{4\pi^2} \left[\frac{V_F}{Q_+} \log \frac{V_F}{Q_+} - \left(\frac{V_F}{Q_+} - 1 \right) \log \left(\frac{V_F}{Q_+} - 1 \right) \right. \\ \left. +2 + \log \frac{\Lambda}{2\varepsilon(Q)} + i\pi \frac{\omega}{\varepsilon(Q)} \right], \tag{5.107}$$

plus terms of order $\varepsilon(Q)/\Lambda$ or ω/Λ, in the $\omega/\varepsilon(Q) \to 0$ limit for small $\varepsilon(Q)$ when only one component of Q (in this case, Q_-) is ultrasoft.

Both components ultrasoft

When both components are ultrasoft, I_C and I_D are given by expressions identical to I_A and I_B in the last section but with $Q_+ \leftrightarrow Q_-$. The solution is therefore

$$I(Q, \omega_Q) = \frac{1}{4\pi^2} \left[\frac{V_F}{Q_+} \log \frac{V_F}{Q_+} - \left(\frac{V_F}{Q_+} - 1 \right) \log \left(\frac{V_F}{Q_+} - 1 \right) \right. \\ \left. + \frac{V_F}{Q_-} \log \frac{V_F}{Q_-} - \left(\frac{V_F}{Q_-} - 1 \right) \log \left(\frac{V_F}{Q_-} - 1 \right) + i\pi \frac{\omega}{\varepsilon(Q)} \right], \tag{5.108}$$

plus terms of order $\varepsilon(Q)/\Lambda$ or ω/Λ, in the $\omega/\varepsilon(Q) \to 0$ limit for ultrasoft Q. Dropping terms of order $Q_\pm/V_F$ gives us

$$\frac{V_F}{Q_+}\log\frac{V_F}{Q_+} - \left(\frac{V_F}{Q_+} - 1\right)\log\left(\frac{V_F}{Q_+} - 1\right) \to \log\frac{V_F}{Q_+} + 1, \tag{5.109}$$

so Eq. (5.108) becomes

$$I(Q, \omega_Q) = \frac{1}{4\pi^2}\left[\log\frac{V_F^2}{\varepsilon(Q)} + 2 + \mathrm{i}\pi\frac{\omega}{\varepsilon(Q)}\right]. \tag{5.110}$$

We should regard the imaginary part as vanishingly small but positive in this limit.

5.3 Discussion of susceptibility

Eq. (5.93), Eq. (5.107), and Eq. (5.108) are the results for the susceptibility when zero, one, or two of the components of the momentum are ultrasoft. Perhaps unsurprisingly, we see that the rapidity cutoff and the UV cutoff effectively exchange roles as we vary the size of the components of the momentum. The real part of the susceptibility diverges for small values of the transfer momentum with $\omega = 0$ in each case. We may summarize the leading log results:

$$I(Q, \omega_Q) = \frac{1}{4\pi^2}\left(\log\frac{\Lambda}{Q_+ \max(Q_-, \Lambda/V_F)} + \log\frac{\Lambda}{Q_- \max(Q_+, \Lambda/V_F)}\right) \tag{5.111}$$

as $\omega_Q \to 0$.

In fact, the form we have found for the susceptibility is basically the only choice consistent with the t-channel beta function calculation in Appendix A. We found in that calculation that the log derivative of the susceptibility is nonzero for Q outside the ultrasoft region. This must persist in the soft region if we take the rapidity cutoff to infinity. Therefore the only energy scale that can show up in the logarithm once we take $\omega \to 0$ is $\varepsilon(Q)$, and the divergent part of the susceptibility must be proportional to $\log(\Lambda/\varepsilon(Q))$. As we reduce the momentum into the ultrasoft region, Λ should be eliminated in favor of the energy scale set by the rapidity cutoff, V_F^2. We see that this is exactly what happens.

It is straightforward to see how this occurs within the calculation. If we ignore the complicated curved region of the strip along the k_+ direction that depends on the energy cutoffs, we see that the remaining rectangular region extends out to $\Lambda^* - 1$ in the $x - y$ coordinate system. The integral over this long strip (when $\varepsilon(Q)$ is small) is the origin of the logarithmic divergence. As we reduce the size of Q_-, the end of the rectangle is replaced with $V_\eta - 1$. Therefore we see that the effect of the rapidity

cutoff is precisely to substitute V_F/Q_+ for $\Lambda/\varepsilon(Q)$ when Q_- is small for one of the strips, reducing the strength of the divergent log.

The fact that we can trace the appearance of $\varepsilon(Q)$ in the divergent logarithm to a paucity of alternative energy scales strongly suggests that introducing another scale, namely the chemical potential, will alter our results. μ will act to regulate the divergences at small values of $\varepsilon(Q)$, so each instance of $\varepsilon(Q)$ in the $\omega \to 0$ limit of the susceptibility found above will be replaced with $\max(\varepsilon(Q), \mu)$. This is actually also straightforward to see from the calculation. If we include a chemical potential, the step functions constraining the signs of ε for each of the loop momenta will take the form $\theta(\varepsilon(k) - \mu)$. Note that the integrand will remain unchanged, because the μ in the denominators of the propagators will cancel against each other when we perform the contour integration over the energy. The form of these step functions is exactly the same as those which appear in the step functions related to the UV cutoff, but with μ taking the place of Λ and with the signs reversed. These will therefore set the lower limits on the strip integrals. The fact that μ shows up in exactly the same manner as Λ (up to a sign) means that we can approximately write down the lower limit that shows up in the strip integrals in the transformed coordinates; it must basically be $\mu/\varepsilon(Q)$. The shape of the strip will also be slightly deformed, but this will be a lower order effect. Therefore we immediately see that the large logarithm from the strip will be proportional to

$$\int_{\mu^*}^{\Lambda^*} \mathrm{d}x \frac{1}{x + 1 - \omega_Q^*} \sim \log \frac{\Lambda^*}{\mu^*} = \log \frac{\Lambda}{\mu}, \tag{5.112}$$

where $\mu^* \equiv \mu/\varepsilon(Q)$, when μ is large compared with $\varepsilon(Q)$.

This is consistent with the intuition that once the scale μ becomes comparable to the UV cutoff, the Fermi surface will start to behave in a manner more similar to that of a round Fermi surface. In that case, the replacement $\varepsilon(Q) \to \mu$ in Eq. (5.93) (the theory without the rapidity cutoff) will eliminate the logarithmic divergence in the $\omega \to 0$ limit of the susceptibility entirely. There will be a residual finite part, but that is to be expected — the fact that the $\omega \to 0$ limit yields a nonzero result for the round Fermi surface is intimately connected to the phenomenon of zero sound.

This argument also allows us to see what happens to the value of the large log for an ultrasoft transfer momentum with a chemical potential. The integral analogous to Eq. (5.112) will look like

$$\int_{\mu^*}^{V_\eta} \mathrm{d}x \frac{1}{x + 1 - \omega_Q^*} \sim \log \frac{V_\eta}{\mu^*} = \log \frac{V_F Q_-}{\mu} \tag{5.113}$$

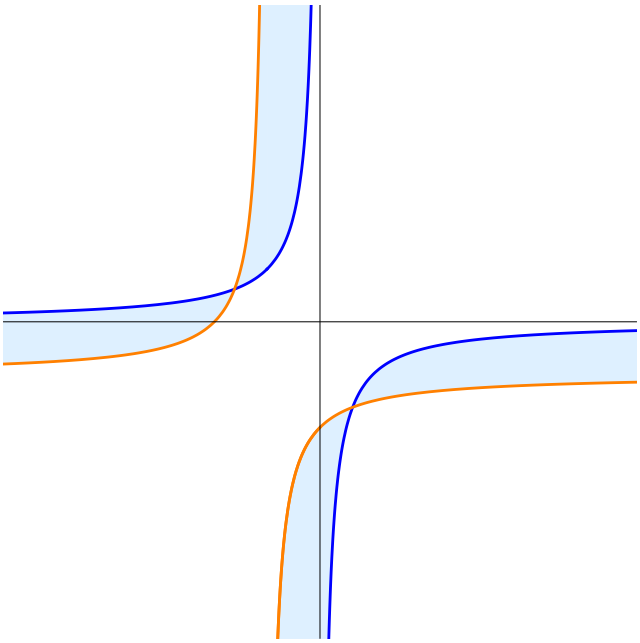


Figure 5.5: The integration region for the strips is modified in the presence of finite chemical potential.

for one of the strips and

$$\log \frac{V_F Q_+}{\mu} \tag{5.114}$$

for the other, so summing them will yield

$$\log \frac{V_F^2 \varepsilon(Q)}{\mu^2}. \tag{5.115}$$

Since by definition ultrasoft momenta satisfy $Q_\pm < \Lambda/V_F$, the large log from Eq. (5.115) is strictly less than that of Eq. (5.112) and disappears in the $\varepsilon(Q) \to 0$ limit, confirming the intuition that the system should act like a round Fermi surface in the limit of large μ.

5.4 Density wave instability

When we take the chemical potential to zero, there is a divergence in the real part of the susceptibility at zero frequency. Due to the relationship with the density-density correlator and the response of the system to external perturbations, this behavior is usually taken as a sign of the instability of the ground state to a density wave state. For example, the divergence of the susceptibility in one spatial dimension for zero frequency and $Q = 2K_F$ is the hallmark of the charge-density wave in one dimension [84].

In the more physical case of a small but finite chemical potential, we see that the divergence is regulated. However, we still find a pole in the sum of the iterated t-channel bubble diagrams if there is a sufficiently large separation of scales between the UV cutoff and the chemical potential. Including the effects of the chemical potential, we may write

$$I(Q, \omega_Q) \sim \frac{C(Q)}{4\pi^2} \log \frac{\Lambda}{\min\left(\varepsilon(Q), \mu\right)} \tag{5.116}$$

where $C(Q)$ varies between approximately one for generic collinear Q_+ and two when the large log approaches its maximum value as Q_+ approaches the ultrasoft region. The condition for a pole to exist in the summed bubble diagram, Eq. (5.37), then implies that the system should be unstable to a density-wave state when

$$\mu \lesssim \exp\left(-\frac{2\pi^2}{g}\right). \tag{5.117}$$

Note that the sign in this expression implies that the instability occurs for positive values of the coupling, unlike the instability to a superconducting state.

There are several caveats to our conclusion regarding the instability of the system. First, we have totally ignored the momentum dependence of the couplings. Second, and relatedly, there is no reason to expect that the bubble sum gives a particularly good approximation to the true compressibility in this situation. Unlike the case of the round Fermi surface, there is no power-suppression of the higher-loop diagrams for generic collinear scattering. We may say that this one-loop result provides evidence for the existence an instability but not conclusive evidence. The nature of the order parameter describing this state, and features of the resulting theory (such as the opening of a gap in the spectrum of the Fermi liquid), will be left to future work. Possible techniques for that analysis include constructing a mean field theory for the order parameter describing the broken symmetry of the new phase [59, 85] or creating a new effective theory for the order parameter field by integrating out the quasiparticle fields as in Hertz's treatment of quantum critical phenomena [59, 89].

Our results are broadly consistent with the conclusion in [85] that the half-filled Hubbard model, the prototypical example of a system with Van Hove singularities, is unstable to the generation of a charge- or spin-density wave, and with the treatment in [90], where divergent susceptibilities are also found in a system with several Van Hove points. In those treatments, it is stated that the instability is due to the existence of a momentum space nesting vector in the half-filled Hubbard model. While we do not have a single nesting vector for our Fermi surface, there is a continuous form

of nesting: translations in momentum space along the Fermi surface take one leg of the Fermi surface into itself (up to issues involving the Fermi velocity cutoff, which should be small if the translations are sufficiently small). This is indeed partially responsible for the phenomenon observed here.

We mention in passing that various authors find divergent susceptibilities from the spin correlators in the Hubbard model [7, 9, 91] and more generally in the presence of Van Hove singularities [92]. These divergent susceptibilities are then indicative of the existence of ferromagnetic or antiferromagnetic order [93]. We have not made an attempt to analyze the spin correlators here, so we cannot comment on whether this behavior is to be expected within our effective theory, or what role it may have in competing with superconductivity.

Chapter 6

CONCLUSIONS

This thesis reviews the tools of effective field theory and applies them to systems with Fermi surfaces. It clarifies several issues in the traditional approach to this subject by using some modern methods from EFTs describing elementary particles. It also constructs a rich and complicated effective theory for Fermi surfaces with a Van Hove singularity and explores some of the implications.

In Chapter 2, we walk through a toy model of an effective field theory to see the renormalization group in action. We find that it is possible to relate certain observables of our original UV theory to those in another IR theory with different values of the coupling constants. As a result of the changing values of the coupling constants, the calculation of the low energy observables of the original theory become simple in terms of the new description. We find that creating the effective theory in general generates all interactions between the effective fields, but that the effects of most of these interactions are suppressed in a way that preserves the predictive power of the theory to any given order in the ratio of the IR scale to the UV scale. We saw that this expansion of the interactions corresponds to the scaling dimension we assigned to the fields, but that the scaling is modified by the effects of quantum corrections from integrating out high energy modes. We discuss how it is possible to relate the predictions of the effective theory to knowledge of the full theory from which it descends via the matching procedure in terms of Polchinski's argument about the RG. Finally, we note how the procedure can be generalized by altering the scaling procedure and scaling fields in various regions of momentum space differently.

Chapter 3 applies the ideas of Chapter 2 to a system of fermionic quasiparticles at finite density with a circular Fermi surface in two spatial dimensions. We find that previous work on the subject tacitly ignores the issue of momentum conservation when the allowed range of momenta around the Fermi surface is nonvanishing, and that the appropriate way to address this is to bin the modes in momentum space in analogy to theories like NRQCD. In this formalism, we find that all interactions are technically irrelevant, but that the theory does not look free in the IR because of anomalous loop enhancements stemming from the large size of the Fermi surface

from the perspective of the IR theory. We comment on the instability of the resulting theory to a superconducting ground state and explore the phenomenon of zero sound. Finally, we comment upon several subtleties related to the nonvanishing contributions to the beta function for small transfer momentum.

Building upon this intuition, we construct an effective field theory in Chapter 4 for Fermi surfaces in two spatial dimensions that pass through a Van Hove singularity, a point where the Fermi velocity vanishes. This situation is particularly relevant to the cuprate superconductors and the understanding of the strange metal phase above the superconducting transition. We find that the subtlety discussed in Chapter 3 becomes a full-blown complication that vastly modifies the theory, allowing for a much greater range of allowed interactions. Despite this complication, we find that double logs associated with the back-to-back interactions with quasiparticles localized in momentum space around the Van Hove singularity result in certain universal behaviors. In the process, we address lingering questions about the nature of these double logs and nonlocal operators generated by the RG flow. Finally, we comment on the deviations from standard Fermi liquid behavior for the modes along the long, flat stretches of the Fermi surface proximal to the Van Hove singularity.

Chapter 5 explores in more detail the possible instabilities of the Van Hove EFT by analyzing the analytic structure of several correlators computable using perturbation theory in the EFT. We find that double log behavior generically leads to the existence of two poles in the two-particle Green's function, but that these poles are often unphysical and outside the domain of validity of the effective theory. However, when the coupling is attractive in the vicinity of the Van Hove singularity, we find that one of the poles is physical and represents the instability of the system to a superconducting ground state, in agreement with the RG analysis in Chapter 4. We also discover a divergence in the susceptibility, a hallmark of the instability to a density-wave state. We find that this divergence is regulated by the presence of a finite chemical potential moving the Fermi surface slightly away from the Van Hove singularity, but that the instability is likely to persist for relatively weak repulsive couplings as long as the separation of scales between the chemical potential and the UV cutoff of the theory is large.

BIBLIOGRAPHY

[1]R. Markiewicz, "A survey of the van hove scenario for high-tc superconductivity with special emphasis on pseudogaps and striped phases," Journal of Physics and Chemistry of Solids **58**, 1179–1310 (1997).

[2]S. Ghamari, S.-S. Lee, and C. Kallin, "Renormalization group analysis of a neck-narrowing lifshitz transition in the presence of weak short-range interactions in two dimensions," Phys. Rev. B **92**, 085112 (2015).

[3]C. W. Bauer, S. Fleming, D. Pirjol, and I. W. Stewart, "An effective field theory for collinear and soft gluons: heavy to light decays," Phys. Rev. D **63**, 114020 (2001).

[4]J.-Y. Chiu, A. Jain, D. Neill, and I. Z. Rothstein, "A formalism for the systematic treatment of rapidity logarithms in quantum field theory," Journal of High Energy Physics **2012**, 84 (2012).

[5]R. Shankar, "Renormalization group approach to interacting fermions," Rev. Mod. Phys. **66**, 129–192 (1994).

[6]J. Polchinski, "Effective field theory and the Fermi surface," in Proceedings, Theoretical Advanced Study Institute (TASI 92): From Black Holes and Strings to Particles: Boulder, USA, June 1-26, 1992 (1992), pp. 0235–276, arXiv:hep-th/9210046 [hep-th].

[7]H. Q. Lin, and J. E. Hirsch, "Two-dimensional hubbard model with nearest- and next-nearest-neighbor hopping," Phys. Rev. B **35**, 3359–3368 (1987).

[8]I. Dzyaloshinskii, "Superconducting transitions due to van hove singularities in the electron spectrum," Sov. Phys. JETP **93**, 1487–98 (1987).

[9]J. V. Alvarez, J. González, F. Guinea, and M. A. H. Vozmediano, "Superconducting, ferromagnetic and antiferromagnetic phases in the t- t' hubbard model," Journal of the Physical Society of Japan **67**, 1868–1871 (1998), eprint: https://doi.org/10.1143/JPSJ.67.1868.

[10]C. J. Halboth, and W. Metzner, "Renormalization-group analysis of the two-dimensional hubbard model," Phys. Rev. B **61**, 7364–7377 (2000).

[11]C. M. Varma, P. B. Littlewood, S. Schmitt-Rink, E. Abrahams, and A. E. Ruckenstein, "Phenomenology of the normal state of cu-o high-temperature superconductors," Phys. Rev. Lett. **63**, 1996–1999 (1989).

[12]S. Weinberg, "Phenomenological Lagrangians," Physica **A96**, 327–340 (1979).

[13]K. Becker, M. Becker, and J. H. Schwarz, *String theory and m-theory: a modern introduction* (Cambridge University Press, 2006).

[14]J. J. M. Carrasco, M. P. Hertzberg, and L. Senatore, "The Effective Field Theory of Cosmological Large Scale Structures," JHEP **09**, 082 (2012), arXiv:1206.2926 [astro-ph.CO].

[15]S. Dubovsky, L. Hui, A. Nicolis, and D. T. Son, "Effective field theory for hydrodynamics: thermodynamics, and the derivative expansion," Phys. Rev. **D85**, 085029 (2012), arXiv:1107.0731 [hep-th].

[16]S. Weinberg, "Effective action and renormalization group flow of anisotropic superconductors," Nucl. Phys. **B413**, 567–578 (1994), arXiv:cond-mat/9306055 [cond-mat].

[17]W. E. Pickett, D. J. Singh, H. Krakauer, and R. E. Cohen, "Fermi surfaces, fermi liquids, and high-temperature superconductors," Science **255**, 46–54 (1992), eprint: http://science.sciencemag.org/content/255/5040/46.full.pdf.

[18]P. W. Anderson, "New physics of metals: fermi surfaces without fermi liquids," Proceedings of the National Academy of Sciences **92**, 6668–6674 (1995).

[19]M. E. Peskin, and D. V. Schroeder, *An Introduction to quantum field theory* (Addison-Wesley, Reading, USA, 1995).

[20]K. G. Wilson, and J. B. Kogut, "The Renormalization group and the epsilon expansion," Phys. Rept. **12**, 75–200 (1974).

[21]A. V. Manohar, "Effective field theories," Lect. Notes Phys. **479**, 311–362 (1997), arXiv:hep-ph/9606222 [hep-ph].

[22]S. Chatrchyan, et al., "Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc," Physics Letters B **716**, 30–61 (2012).

[23]G. Aad, et al., "Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc," Physics Letters B **716**, 1–29 (2012).

[24]I. Z. Rothstein, "TASI lectures on effective field theories," in (2003), arXiv:hep-ph/0308266 [hep-ph].

[25]H. Georgi, "Effective field theory," Ann. Rev. Nucl. Part. Sci. **43**, 209–252 (1993).

[26]M. Srednicki, *Quantum field theory* (Cambridge University Press, 2007).

[27]S. Weinberg, *The Quantum theory of fields. Vol. 1: Foundations* (Cambridge University Press, 2005).

[28]J. Polchinski, "Renormalization and Effective Lagrangians," Nucl. Phys. **B231**, 269–295 (1984).

[29]H. Georgi, "On-shell effective field theory," Nucl. Phys. **B361**, 339–350 (1991).

[30]D. B. Kaplan, "Effective field theories," in Beyond the standard model 5. Proceedings, 5th Conference, Balholm, Norway, April 29-May 4, 1997 (1995), arXiv:nucl-th/9506035 [nucl-th].

[31]M. Luke, and A. V. Manohar, "Bound states and power counting in effective field theories," Phys. Rev. D **55**, 4129–4140 (1997).

[32]M. Luke, and M. J. Savage, "Power counting in dimensionally regularized nonrelativistic qcd," Phys. Rev. D **57**, 413–423 (1998).

[33]B. Grinstein, and I. Z. Rothstein, "Effective field theory and matching in nonrelativistic gauge theories," Phys. Rev. D **57**, 78–82 (1998).

[34]L. Landau, "The theory of a fermi liquid," Sov. Phys. JETP **3**, 920–925 (1957).

[35]L. Landau, "Oscillations in a fermi liquid," Sov. Phys. JETP **5**, 101–108 (1957).

[36]L. Landau, "On the theory of the fermi liquid," Sov. Phys. JETP **8**, 70 (1959).

[37]P. Nozières, and D. Pines, *The theory of quantum liquids* (Addison-Wesley, 1989).

[38]G. Benfatto, and G. Gallavotti, "Perturbation theory of the fermi surface in a quantum liquid. a general quasiparticle formalism and one-dimensional systems," Journal of Statistical Physics **59**, 541–664 (1990).

[39]G. Benfatto, and G. Gallavotti, "Renormalization-group approach to the theory of the fermi surface," Phys. Rev. B **42**, 9967–9972 (1990).

[40]J. G. Bednorz, and K. A. Müller, "Possible hightc superconductivity in the ba-la-cu-o system," Zeitschrift für Physik B Condensed Matter **64**, 189–193 (1986).

[41]J. C. Campuzano, G. Jennings, M. Faiz, L. Beaulaigue, B. W. Veal, J. Z. Liu, A. P. Paulikas, K. Vandervoort, H. Claus, R. S. List, A. J. Arko, and R. J. Bartlett, "Fermi surfaces of $YBa_2Cu_3O_{6.9}$ as seen by angle-resolved photoemission," Phys. Rev. Lett. **64**, 2308–2311 (1990).

[42]P. W. Anderson, "Experimental constraints on the theory of high-tc superconductivity," Science **256**, 1526–1531 (1992), eprint: http://science.sciencemag.org/content/256/5063/1526.full.pdf.

[43]A. V. Manohar, and M. B. Wise, "Heavy quark physics," Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol. **10**, 1–191 (2000).

[44]P. Kopietz, and T. Busche, "Exact renormalization group flow equations for nonrelativistic fermions: scaling toward the fermi surface," Phys. Rev. B **64**, 155101 (2001).

[45]E. Eichten, and B. Hill, "An effective field theory for the calculation of matrix elements involving heavy quarks," Physics Letters B **234**, 511–516 (1990).

[46]H. Georgi, "An effective field theory for heavy quarks at low energies," Physics Letters B **240**, 447–450 (1990).

[47]N. Isgur, and M. B. Wise, "Weak decays of heavy mesons in the static quark approximation," Physics Letters B **232**, 113–117 (1989).

[48]N. Isgur, and M. B. Wise, "Weak transition form factors between heavy mesons," Physics Letters B **237**, 527–530 (1990).

[49]M. B. Wise, "Heavy quark physics: Course," in Probing the standard model of particle interactions. Proceedings, Summer School in Theoretical Physics, NATO Advanced Study Institute, 68th session, Les Houches, France, July 28-September 5, 1997. Pt. 1, 2 (1997), pp. 1051–1089, arXiv:hep-ph/9805468 [hep-ph].

[50]W. Caswell, and G. Lepage, "Effective lagrangians for bound state problems in qed, qcd, and other field theories," Physics Letters B **167**, 437–442 (1986).

[51]G. T. Bodwin, E. Braaten, and G. P. Lepage, "Rigorous qcd analysis of inclusive annihilation and production of heavy quarkonium," Phys. Rev. D **51**, 1125–1171 (1995).

[52]M. E. Luke, A. V. Manohar, and I. Z. Rothstein, "Renormalization group scaling in nonrelativistic qcd," Phys. Rev. D **61**, 074025 (2000).

[53]C. W. Bauer, S. Fleming, and M. Luke, "Summing sudakov logarithms in $\vec{B} \; X_s\gamma$ in effective field theory," Phys. Rev. D **63**, 014006 (2000).

[54]J.-y. Chiu, A. Jain, D. Neill, and I. Z. Rothstein, "Rapidity renormalization group," Phys. Rev. Lett. **108**, 151601 (2012).

[55]J. Feldman, H. Knörrer, and E. Trubowitz, "A two dimensional fermi liquid. part 1: overview," Communications in mathematical physics **247**, 1–47 (2004).

[56]J. Polchinski, "Low-energy dynamics of the spinon-gauge system," Nuclear Physics B **422**, 617–633 (1994).

[57]C. Nayak, and F. Wilczek, "Non-fermi liquid fixed point in 2 + 1 dimensions," Nuclear Physics B **417**, 359–373 (1994).

[58]C. Nayak, and F. Wilczek, "Renormalization group approach to low temperature properties of a non-Fermi liquid metal," Nucl. Phys. **B430**, 534–562 (1994), arXiv:cond-mat/9408016 [cond-mat].

[59]S. Sachdev, *Quantum phase transitions* (Cambridge university press, 2011).

[60]A. V. Manohar, and I. W. Stewart, "The zero-bin and mode factorization in quantum field theory," Phys. Rev. D **76**, 074002 (2007).

[61]A. Abrikosov, L. Gorkov, and I. Dzyaloshinski, *Methods of quantum field theory in statistical physics*, Dover Books on Physics Series (Dover Publications, 1963).

[62]M. Gell-Mann, and F. E. Low, "Quantum electrodynamics at small distances," Phys. Rev. **95**, 1300–1312 (1954).

[63]S. Gopalan, O. Gunnarsson, and O. K. Andersen, "Effects of saddle-point singularities on the electron lifetime," Phys. Rev. B **46**, 11798–11806 (1992).

[64]P. C. Pattnaik, C. L. Kane, D. M. Newns, and C. C. Tsuei, "Evidence for the van hove scenario in high-temperature superconductivity from quasiparticle-lifetime broadening," Phys. Rev. B **45**, 5714–5717 (1992).

[65] Igor Dzyaloshinskii, "Extended van-hove singularity and related non-fermi liquids," J. Phys. I France **6**, 119–135 (1996).

[66]D. Menashe, and B. Laikhtman, "Fermi-liquid properties of a two-dimensional electron system with the fermi level near a van hove singularity," Phys. Rev. B **59**, 13592–13595 (1999).

[67]C. Honerkamp, M. Salmhofer, N. Furukawa, and T. M. Rice, "Breakdown of the landau-fermi liquid in two dimensions due to umklapp scattering," Phys. Rev. B **63**, 035109 (2001).

[68]D. Zanchi, and H. J. Schulz, "Weakly correlated electrons on a square lattice: renormalization-group theory," Phys. Rev. B **61**, 13609–13632 (2000).

[69]V. Y. Irkhin, A. A. Katanin, and M. I. Katsnelson, "Effects of van hove singularities on magnetism and superconductivity in the $t - t'$ hubbard model: a parquet approach," Phys. Rev. B **64**, 165107 (2001).

[70]G. Kastrinakis, "Quasiparticle scattering rate in overdoped superconducting cuprates," Phys. Rev. B **71**, 014520 (2005).

[71]G. Kastrinakis, "A fermi liquid model for the overdoped and optimally doped cuprate superconductors: scattering rate, susceptibility, spin resonance peak and superconducting transition," Physica C: Superconductivity **340**, 119–132 (2000).

[72]R. Nandkishore, L. Levitov, and A. Chubukov, "Chiral superconductivity from repulsive interactions in doped graphene," Nature Physics **8**, 158 (2012).

[73]L. Van Hove, "The occurrence of singularities in the elastic frequency distribution of a crystal," Phys. Rev. **89**, 1189–1193 (1953).

[74]S. Schmitt, "Non-fermi-liquid signatures in the hubbard model due to van hove singularities," Phys. Rev. B **82**, 155126 (2010).

[75]J. E. Hirsch, and D. J. Scalapino, "Enhanced superconductivity in quasi two-dimensional systems," Phys. Rev. Lett. **56**, 2732–2735 (1986).

[76]I. Dzyaloshinskii, "Maximal increase of the superconducting transition temperature due to the presence of van't hoff singularities," JETP Lett **46** (1987).

[77]J. Bok, and J. Bouvier, "Superconductivity and the van hove scenario," Journal of Superconductivity and Novel Magnetism **25**, 657–667 (2012).

[78]A. V. Chubukov, D. L. Maslov, S. Gangadharaiah, and L. I. Glazman, "Singular perturbation theory for interacting fermions in two dimensions," Phys. Rev. B **71**, 205112 (2005).

[79]B. Michon, C. Girod, S. Badoux, J. Kačmarčík, Q. Ma, M. Dragomir, H. A. Dabkowska, B. D. Gaulin, J.-S. Zhou, S. Pyon, T. Takayama, H. Takagi, S. Verret, N. Doiron-Leyraud, C. Marcenat, L. Taillefer, and T. Klein, "Thermodynamic signatures of quantum criticality in cuprates," ArXiv e-prints (2018), arXiv:1804.08502 [cond-mat.supr-con].

[80]J. R. Schrieffer, *Theory of superconductivity* (CRC Press, 2018).

[81]E. E. Salpeter, and H. A. Bethe, "A relativistic equation for bound-state problems," Phys. Rev. **84**, 1232–1242 (1951).

[82]H. Fröhlich, "On the theory of superconductivity: the one-dimensional case," Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences **223**, 296–305 (1954).

[83]G. Gruner, *Density waves in solids* (CRC Press, 2018).

[84]G. Grüner, "The dynamics of charge-density waves," Rev. Mod. Phys. **60**, 1129–1181 (1988).

[85]X.-G. Wen, *Quantum field theory of many-body systems: from the origin of sound to an origin of light and electrons* (Oxford University Press on Demand, 2004).

[86]A. W. Overhauser, "Spin density waves in an electron gas," Phys. Rev. **128**, 1437–1452 (1962).

[87]C. P. Enz, *A course on many-body theory applied to solid-state physics*, Vol. 11 (World Scientific Publishing Company, 1992).

[88]E. M. Lifshitz, and L. P. Pitaevskii, *Statistical physics: theory of the condensed state*, Vol. 9 (Elsevier, 2013).

[89]J. A. Hertz, "Quantum critical phenomena," Phys. Rev. B **14**, 1165–1184 (1976).

[90]T. M. Rice, and G. K. Scott, "New mechanism for a charge-density-wave instability," Phys. Rev. Lett. **35**, 120–123 (1975).

[91]R. Hlubina, S. Sorella, and F. Guinea, "Ferromagnetism in the two dimensional $t - t^{'}$ hubbard model at the van hove density," Phys. Rev. Lett. **78**, 1343–1346 (1997).

[92]M. Vozmediano, J. González, F. Guinea, J. Alvarez, and B. Valenzuela, "Properties of electrons near a van hove singularity," Journal of Physics and Chemistry of Solids **63**, Proceedings of the Conference on Spectroscopies in Novel Superconductors, 2295–2297 (2002).

[93]D. R. Penn, "Stability theory of the magnetic phases for a simple model of the transition metals," Phys. Rev. **142**, 350–365 (1966).

Appendix A

THE ONE-LOOP BETA FUNCTION FOR THE VAN HOVE CASE

We consider only the spin-singlet interaction. The interaction part of the Lagrangian is

$$\frac{g}{4}\epsilon^{ab}\epsilon^{cd}\psi_a^\dagger\psi_b^\dagger\psi_c\psi_d. \tag{A.1}$$

The tree-level four-point amplitude is

$$S_{abcd}g, \tag{A.2}$$

where

$$S_{abcd} = \epsilon_{ab}\epsilon_{cd} \tag{A.3}$$

is the spin structure of the amplitude. Momentum conservation means the coupling is a function of three momenta (six real variables) $g(p_1, p_2, p_3)$ which we take to be symmetric under $p_1 \leftrightarrow p_2$ and $p_3 \leftrightarrow p_1 + p_2 - p_3$.

The one-loop four-point amplitude has contributions from the s, t, and u channels. To understand the spin dependence and symmetry factors, let us momentarily take the coupling to be a constant and analyze the four-point correlation function in position space at tree level and one-loop. From this, we will be able to extract the contributions to the scattering amplitude. The tree level result is the connected part of

$$G_{\text{tree}} = \sum_{\alpha,\beta,\gamma,\delta}\frac{1}{4}\mathrm{i}g\epsilon_{\alpha\beta}\epsilon_{\gamma\delta}\langle ab\alpha^\dagger\beta^\dagger\gamma\delta c^\dagger d^\dagger\rangle, \tag{A.4}$$

where we have labeled the fields by their spin indices alone, reserving Greek indices for terms corresponding to the insertion of the interaction vertex, and suppressed the position-space arguments of the fields. The brackets should be interpreted to denote a time-ordered product of the fields in the interaction picture.

Using Wick's theorem, we can right the time ordered product as a sum of products of normal-ordered terms (whose expectation vanishes) and contractions of fields corresponding to free-field propagators [19]. Only the fully contracted terms survive. We use the convention

$$N(\psi_1\psi_2\psi_3^\dagger\psi_4^\dagger) = -\overbrace{\psi_1\psi_3^\dagger}N(\psi_2\psi_4^\dagger) \tag{A.5}$$

to properly deal with the exchanges of the anticommuting fields. Also note that we are only interested in the fully connected topology of contractions, as the others represent vacuum bubbles whose effects are canceled by the normalization of the interacting Green's function.

The propagator just involves a Kronecker delta for the spins, so

$$\langle ab\alpha^\dagger\beta^\dagger\gamma\delta c^\dagger d^\dagger\rangle = (-1^2)(\delta_{a\alpha}\delta_{b\beta}\delta_{c\gamma}\delta_{d\delta} - \delta_{a\alpha}\delta_{b\beta}\delta_{c\gamma}\delta_{d\delta} - \delta_{a\alpha}\delta_{b\beta}\delta_{c\gamma}\delta_{d\delta} + \delta_{a\alpha}\delta_{b\beta}\delta_{c\gamma}\delta_{d\delta}) \times (\text{position space}). \quad \text{(A.6)}$$

We have explicitly written out the antisymmetrization of the spin labels here that occurs due to the fact that we have found the time-ordered product of fermionic fields. "Position space" represents the part of the function involving the position space propagators. Since we are looking at the tree-level contribution here, there are no loop integrals involved. Therefore we discard the position space from the external propagators to find the contribution to the amplitude. Finally, the delta functions in Eq. (A.6) contract against the antisymmetric symbols from the interaction in Eq. (A.4) to yield

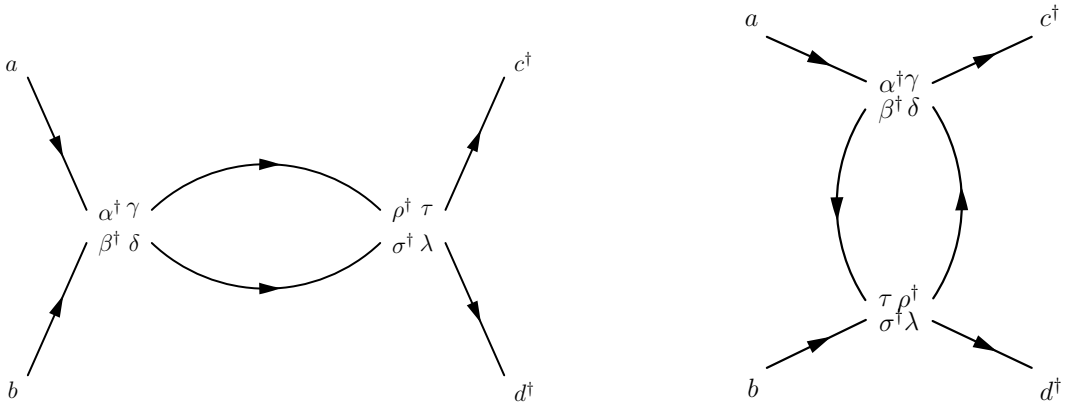
$$\mathrm{i}\mathcal{A}_{\text{tree}} = 4 \times \frac{1}{4}\mathrm{i}\epsilon_{ab}\epsilon_{cd}g = \mathrm{i}\epsilon_{ab}\epsilon_{cd}g. \quad \text{(A.7)}$$

Let us turn now to the one-loop corrections. The contribution to the Green's function is

$$G_{\mathrm{S}} = \sum \frac{1}{2!}\left(\frac{1}{4}\mathrm{i}g\right)^2 \epsilon_{\alpha\beta}\epsilon_{\gamma\delta}\epsilon_{\rho\sigma}\epsilon_{\tau\lambda}\langle ab\alpha^\dagger\beta^\dagger\gamma\delta\rho^\dagger\sigma^\dagger\tau\lambda c^\dagger d^\dagger\rangle_{\mathrm{S}}, \quad \text{(A.8)}$$

where the sum is over the Greek indices and the subscript S denotes that we will only include the contractions with the correct topology for the s-channel diagram. This means that we only include diagrams where a and b are contracted against fields from a single vertex (at a single spacetime point) and $c^\dagger$ and $d^\dagger$ are contracted against fields at the other vertex, with a pair of contractions spanning the two vertices. Fig. A.1a demonstrates an example of one such contraction.

The antisymmetry of the contractions in the normal ordering and the antisymmetric symbols ensure that each contraction with the same topology contributes with the same sign. It requires three "swaps" for the daggered and undaggered fields to be next to each other: $b \leftrightarrow \alpha$, $\delta \leftrightarrow \rho^\dagger$, and $\lambda \leftrightarrow c^\dagger$. Therefore there is an overall sign of $(-1)^3$ from the normal ordering process. Now we can count contractions. There are four ways to contract a against the daggered vertex fields. Once this is chosen, there is only one way to contract b against the other daggered field in the



(a) s-channel (b) t-channel

Figure A.1: Examples of topologies of contractions for different diagrams.

vertex. Then there are two ways to contract $d^\dagger$ against the undaggered fields on the remaining vertex, and one way to contract $c^\dagger$ once that is chosen. Finally, there are two ways to contract the remaining fields in the vertices with each other in the loop, so we have a total of

$$4 \times 2 \times 2 = 16 \tag{A.9}$$

possible contractions for this topology.

To compute the sum over the spin indices, consider the following representative contraction:

$$\langle \overbrace{ab\alpha^\dagger}\beta^\dagger \overbrace{\gamma\delta\rho^\dagger}\sigma^\dagger \overbrace{\tau\lambda c^\dagger}\, d^\dagger\rangle. \tag{A.10}$$

The spin sum becomes

$$\epsilon_{ab}\epsilon_{cd} \sum \epsilon_{\rho\sigma}\epsilon_{\tau\lambda}\delta_{\gamma\rho}\delta_{\delta\sigma} = 2\epsilon_{ab}\epsilon_{cd}. \tag{A.11}$$

Note that the s channel has the same spin-singlet structure as the tree-level amplitude. Putting this together, we have

$$G_\mathrm{S} = 16 \times \frac{1}{2!}\left(\frac{1}{4}\mathrm{i}g\right)^2 \times 2\epsilon_{ab}\epsilon_{cd} \times (-1)^3 \times (\text{position space}). \tag{A.12}$$

Amputating and going to momentum space yields the following loop integral:

$$I_{\text{loop,S}} = \int \frac{\mathrm{d}\omega\, \mathrm{d}^2k}{(2\pi)^3} \frac{\mathrm{i}}{\omega - \varepsilon(k) + \mathrm{i}\epsilon \operatorname{sign}\varepsilon(k)} \times \frac{\mathrm{i}}{E - \omega - \varepsilon(K-k) + \mathrm{i}\epsilon \operatorname{sign}\varepsilon(K-k)}, \tag{A.13}$$

where $E = E_1 + E_2$ is the net incoming energy and $K = p_1 + p_2$ is the net incoming momentum. We can write this as

$$I_{\text{loop,S}} = -\mathrm{i}^2 \int \frac{\mathrm{d}^2 k}{(2\pi)^2} I_{\omega,\text{S}}, \tag{A.14}$$

where

$$I_{\omega,\text{S}} = \int \frac{\mathrm{d}\omega}{2\pi} \frac{1}{\omega - \omega_A} \frac{1}{\omega - \omega_B} \tag{A.15}$$

and

$$\omega_A = \varepsilon(k) - \mathrm{i}\epsilon \operatorname{sign} \varepsilon(k), \tag{A.16}$$

$$\omega_B = -\varepsilon(K - k) + E + \mathrm{i}\epsilon \operatorname{sign} \varepsilon(K - k). \tag{A.17}$$

For $I_{\omega,\text{S}}$ to yield a nonzero result, the poles must lie on either side of the real axis, because otherwise we could simply close the contour on the opposite side from the poles. Choosing to close the contour above,

$$I_{\omega,\text{S}} = \frac{2\pi\mathrm{i}}{2\pi} \left(\frac{\theta(\operatorname{Im} \omega_A)\theta(-\operatorname{Im} \omega_B)}{\omega_A - \omega_B} + \frac{\theta(\operatorname{Im} \omega_B)\theta(-\operatorname{Im} \omega_A)}{\omega_B - \omega_A} \right) \tag{A.18}$$

because $1/(\omega - \omega_B)$ is analytic in the upper half-plane when $\operatorname{Im} \omega_B < 0$ and similarly for ω_A. Then

$$I_{\omega,\text{S}} = -\mathrm{i} \left(\frac{\theta(\varepsilon(k))\theta(\varepsilon(K - k)) - \theta(-\varepsilon(k))\theta(-\varepsilon(K - k))}{\omega_A - \omega_B} \right). \tag{A.19}$$

Next, it is convenient to change variables to $k_\pm = k_x \pm k_y$, so

$$k_x = \frac{1}{2}(k_+ + k_-), \tag{A.20}$$

$$k_y = \frac{1}{2}(k_+ - k_-). \tag{A.21}$$

The Jacobian determinant associated with this is

$$\left| \frac{\partial k_x}{\partial k_+} \frac{\partial k_y}{\partial k_-} - \frac{\partial k_x}{\partial k_-} \frac{\partial k_y}{\partial k_+} \right| = \frac{1}{2}. \tag{A.22}$$

Combining the various expressions together, we find

$$\begin{aligned} \mathrm{i}\mathcal{A}_\text{S} &= (-1)\mathrm{i}\epsilon_{ab}\epsilon_{cd} g^2 \\ &\quad \times \frac{1}{8\pi^2} \int \mathrm{d}k_+ \, \mathrm{d}k_- \frac{\theta(\varepsilon(k))\theta(\varepsilon(K - k)) - \theta(-\varepsilon(k))\theta(-\varepsilon(K - k))}{\varepsilon(k) + \varepsilon(K - k) - E - \mathrm{i}\epsilon \operatorname{sign} \varepsilon(k)}. \end{aligned} \tag{A.23}$$

Thus as long as the external energy E is negligible the integral is positive, and therefore the s-channel result is negative compared to tree level amplitude. This will then imply a positive contribution to the beta function.

The t-channel diagram differs only in the topology of contractions. We will use as our "reference" the contraction given in Fig. (A.1b). It requires 12 swaps to place the contracted fields next to each other, so there is no negative sign from the Wick contractions. Let us count the number of topologically equivalent contractions. There are four ways to contract a against either of the daggered fields in either vertex. Then there are two ways to contract b against the daggered fields in the other vertex. There are two ways to contract $c^\dagger$ against the undaggered fields in the vertex that a is contracted against, and two ways to contract $d^\dagger$ against the vertex b is contracted against. Finally, there is only one way to contract the remaining vertices in a loop. This gives us

$$4 \times 2 \times 2 \times 2 = 32 \tag{A.24}$$

ways to contract the fields and match the topology of the diagram. As before, the antisymmetry of the ϵ in the interaction guarantees that all contractions give the same sign.

The spin sum for the t-channel diagram is

$$\sum \epsilon_{a\beta}\epsilon_{c\delta}\epsilon_{\delta b}\epsilon_{\beta d} = \sum_{\beta} \epsilon_{a\beta}\epsilon_{\beta d} \sum_{\delta} \epsilon_{c\delta}\epsilon_{\delta b} = (-\delta_{ad})(-\delta_{cb}). \tag{A.25}$$

We can write this as

$$\begin{aligned}\delta_{ad}\delta_{bc} &= \frac{1}{2}(\delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc}) - \frac{1}{2}(\delta_{ac}\delta_{bd} - \delta_{ad}\delta_{bc}) \\ &= \frac{1}{2}(\delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc}) - \frac{1}{2}\epsilon_{ab}\epsilon_{cd}.\end{aligned} \tag{A.26}$$

The first term is a contribution to the spin-triplet interaction and the second is a contribution to the spin-singlet interaction. We will see below that the u-channel diagram will produce a contribution to the spin-triplet interaction with the opposite sign. This implies that in the case of a momentum-independent coupling, contributions to the beta function from the t- and u-channels for the spin-triplet interaction cancel if we constrain our initial Lagrangian to only include a spin-singlet interaction. In other words, the assumption of a purely spin-singlet interaction is stable under the RG flow *if the coupling can be regarded as a momentum-independent constant*. As we will see, this assumption must fail in the case at hand, so in principle a spin-triplet interaction should be radiatively generated even if it is not present in the UV Lagrangian. Nonetheless, to simplify this analysis, we ignore the spin-triplet contributions entirely.

Isolating the singlet contribution, amputating, and going to momentum space yields

$$\mathrm{i}\mathcal{A}_{\mathrm{T}} = (-1)^{12} \times 32 \times \frac{1}{2!}\left(\frac{1}{4}\mathrm{i}g\right)^2 \times (-1) \times \frac{1}{2}\epsilon_{ab}\epsilon_{cd} \times I_{\mathrm{loop,T}}. \tag{A.27}$$

We can write

$$I_{\mathrm{loop,T}} = \mathrm{i}^2 \int \frac{\mathrm{d}^2k}{(2\pi)^2} I_{\omega,\mathrm{T}}, \tag{A.28}$$

where

$$I_{\omega,\mathrm{T}} = \int \frac{\mathrm{d}\omega}{2\pi} \frac{1}{\omega - \omega_C} \frac{1}{\omega - \omega_D}, \tag{A.29}$$

$$\omega_C = \varepsilon(k) - \mathrm{i}\epsilon \mathrm{sign}\, \varepsilon(k), \tag{A.30}$$

$$\omega_D = \varepsilon(k+Q) - E_{\mathrm{T}} - \mathrm{i}\epsilon \mathrm{sign}\, \varepsilon(k+Q). \tag{A.31}$$

$Q = p_1 - p_3$ is the transfer momentum and $E_{\mathrm{T}} = E_1 - E_3$ is the transfer energy. This is the same form as $I_{\omega,\mathrm{S}}$, and we find

$$I_{\omega,\mathrm{T}} = (-\mathrm{i}) \left(\frac{\theta(\varepsilon(k))\theta(-\varepsilon(k+Q)) - \theta(-\varepsilon(k))\theta(\varepsilon(k+Q))}{\omega_C - \omega_D} \right). \tag{A.32}$$

As before, there is a final factor of $1/2$ from changing variables to $k_\pm$. Combining the expressions, we find the contribution to the spin-singlet interaction is

$$\begin{aligned} \mathrm{i}\mathcal{A}_{\mathrm{T}} &= \mathrm{i}\epsilon_{ab}\epsilon_{cd}g^2 \\ &\quad \times \frac{1}{16\pi^2} \int \mathrm{d}k_+ \, \mathrm{d}k_- \frac{\theta(\varepsilon(k))\theta(-\varepsilon(k+Q)) - \theta(-\varepsilon(k))\theta(\varepsilon(k+Q))}{\varepsilon(k) - \varepsilon(k+Q) + E_{\mathrm{T}} - \mathrm{i}\epsilon \mathrm{sign}\, \varepsilon(k)}. \end{aligned} \tag{A.33}$$

As long as the transfer energy $E_{\mathrm{T}} = E_1 - E_3$ is negligible, the t-channel is positive compared to the tree level contribution.

For the u-channel, there is a factor of -1 from swapping $c \leftrightarrow d$ in the Green's function. However, the spin sum becomes

$$\sum_{\beta\delta} = \epsilon_{a\beta}\epsilon_{d\delta}\epsilon_{\delta b}\epsilon_{\beta c} = \delta_{ac}\delta_{bd} = \frac{1}{2}(\delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc}) + \frac{1}{2}\epsilon_{ab}\epsilon_{cd}. \tag{A.34}$$

This implies that the overall contribution to the singlet is the same sign as for the t-channel diagram while the contribution to triplet has the opposite sign as promised. Isolating the singlet then gives us

$$\begin{aligned} \mathrm{i}\mathcal{A}_{\mathrm{U}} &= \mathrm{i}\epsilon_{ab}\epsilon_{cd}g^2 \\ &\quad \times \frac{1}{16\pi^2} \int \mathrm{d}k_+ \, \mathrm{d}k_- \frac{\theta(\varepsilon(k))\theta(-\varepsilon(k+Q')) - \theta(-\varepsilon(k))\theta(\varepsilon(k+Q'))}{\varepsilon(k) - \varepsilon(k+Q') + E_{\mathrm{U}} - \mathrm{i}\epsilon \mathrm{sign}\, \varepsilon(k)}, \end{aligned} \tag{A.35}$$

where $Q' = p_3 - p_2$ and $E_{\mathrm{U}} = E_3 - E_2$.

Since we have isolated only the spin-singlet contributions from the t- and u-channel diagrams and the s-channel diagram only makes a spin-singlet contribution, we have found that the full one-loop amplitude takes the form

$$\mathcal{A}_{abcd} = S_{abcd}(\mathcal{A}_{\mathrm{S}} + \mathcal{A}_{\mathrm{T}} + \mathcal{A}_{\mathrm{U}}). \tag{A.36}$$

We will focus on the s and t channels because the u channel is identical to the t channel if we replace $Q \to Q'$ and $E_{\mathrm{T}} \to E_{\mathrm{U}}$. Besides the momentum dependence of the coupling, the s- and t-channel diagrams only depend on the external momenta through $K = p_1 + p_2$ and $Q = p_1 - p_3$, respectively.

Restoring the momentum dependence of the coupling yields

$$\mathcal{A}_{\mathrm{S}} = -\frac{1}{8\pi^2}\int d^2k \frac{\theta(\varepsilon(k))\theta(\varepsilon(K-k)) - \theta(-\varepsilon(k))\theta(-\varepsilon(K-k))}{\varepsilon(k) + \varepsilon(K-k) - E - i\epsilon\,\operatorname{sign}\varepsilon(k)} \times g(p_1, p_2, k)g(k, K-k, p_3)f(k)f(K-k). \tag{A.37}$$

f contains all information regarding the cutoffs:

$$f(k) = \theta(\Lambda - |\varepsilon(k)|)\theta(\Upsilon - |k_+|)\theta(\Upsilon - |k_-|). \tag{A.38}$$

Similarly, the t-channel amplitude is

$$\mathcal{A}_{\mathrm{T}} = \frac{1}{16\pi^2}\int d^2k \frac{\theta(\varepsilon(k))\theta(-\varepsilon(k+Q)) - \theta(-\varepsilon(k))\theta(\varepsilon(k+Q))}{\varepsilon(k) - \varepsilon(k+Q) + E_{\mathrm{T}} - i\epsilon\,\operatorname{sign}\varepsilon(k)} \times g(p_1, k, p_3)g(k+Q, p_2, k)f(k)f(k+Q). \tag{A.39}$$

To compute the contribution to the beta function, we only need the logarithmic derivative of the these expressions with respect to Λ. Note that

$$\frac{d}{d\Lambda}\theta(\Lambda - |\varepsilon(k)|) = \delta(\Lambda - \varepsilon(k)) + \delta(\Lambda + \varepsilon(k)). \tag{A.40}$$

The delta function implies that each term in the beta function reduces to a one-dimensional integral. Next,

$$\frac{d}{d\Lambda}f(k)f(K-k) = [\delta(\Lambda - \varepsilon(k)) + \delta(\Lambda + \varepsilon(k))]\theta(\Upsilon - |k_+|)\theta(\Upsilon - |k_-|)f(K-k) + (k \leftrightarrow K-k). \tag{A.41}$$

These two terms are exchanged under the change of variables $k' = k - K$, and the rest of the integrand in (A.37) (including the coupling functions, due to their exchange symmetry) is invariant under this exchange. Then

$$\Lambda \frac{d\mathcal{A}_{\rm S}}{d\Lambda} = -\frac{1}{4\pi^2}(I_{S+} + I_{S-}), \tag{A.42}$$

$$\Lambda \frac{d\mathcal{A}_{\rm T}}{d\Lambda} = \frac{1}{8\pi^2}(I_{T+} + I_{T-}), \tag{A.43}$$

where

$$I_{S\pm} \equiv \pm\Lambda \int d^2k \frac{\delta(\Lambda \mp \varepsilon(k))\theta(\pm\varepsilon(K-k))\theta(\Upsilon - |k_+|)\theta(\Upsilon - |k_-|) f(k-K)}{\varepsilon(k) + \varepsilon(K-k)} \times g(p_1, p_2, k) g(k, K-k, p_3), \tag{A.44}$$

$$I_{T\pm} \equiv \pm\frac{1}{2}\Lambda \int d^2k \frac{\delta(\Lambda \mp \varepsilon(k))\theta(\mp\varepsilon(k+Q))\theta(\Upsilon - |k_+|)\theta(\Upsilon - |k_-|) f(k+Q)}{\varepsilon(k) - \varepsilon(k+Q)} \times [g(p_1, k, p_3) g(k+Q, p_2, k) + g(p_1, -k-Q, p_3) g(-k, p_2, -k-Q)]. \tag{A.45}$$

We have dropped E and $E_{\rm T}$ because they lead to power-suppressed terms in the beta function, and the $i\epsilon$ term because the delta functions ensure the denominator never approaches zero. The remaining integrals are similar to each other. They involve integrating over the one-dimensional space where one of the particles in the loop has $\varepsilon = \pm\Lambda$ and the other has either the same sign for ε (for the s channel) or the opposite sign (for the t channel).

Define P to be equal to K for the s-channel diagram and $-Q$ for the t-channel diagram. We exploit the $O(1,1)$ invariance of the dispersion to replace P by $\tilde{P} = \sqrt{|\varepsilon(P)|}(\text{sign}\, P_+, \text{sign}\, P_-)$ in each of the integrals by changing variables:

$$k_+ = k'_+/\eta, \tag{A.46}$$

$$k_- = \eta k'_-, \tag{A.47}$$

with

$$\eta \equiv \frac{\sqrt{|\varepsilon(P)|}}{|P_+|} = \sqrt{\left|\frac{P_-}{P_+}\right|}. \tag{A.48}$$

We may take η to be less than one by exchanging k_+ and k_- if necessary. The step functions involving the rapidity cutoff Υ are not invariant under this change of variables. In particular,

$$\theta(\Upsilon - |k_+|) \to \theta(\eta\Upsilon - |k_+|), \tag{A.49}$$

$$\theta(\Upsilon - |k_-|) \to \theta(\Upsilon/\eta - |k_-|), \tag{A.50}$$

$$\theta(\Upsilon - |k_+ - P_+|) \to \theta(\eta\Upsilon - |k_+ - \sqrt{|\varepsilon(P)|}\,\mathrm{sign}\,P_+|), \tag{A.51}$$

$$\theta(\Upsilon - |k_- - P_-|) \to \theta(\Upsilon/\eta - |k_- - \sqrt{|\varepsilon(P)|}\,\mathrm{sign}\,P_-|). \tag{A.52}$$

These set the limits of integration on the remaining k_+ integrals (once we have performed the k_- integrals with the delta function) if the energy constraints do not set stricter limits. With that in mind, let us first analyze the limits in the absence of a rapidity cutoff.

Integration limits

We can write generic expressions for the various possible integration limits in each of the four remaining integrals. As before, take P to be either K or $-Q$. Define

$$s_\pm = \mathrm{sign}\,P_\pm, \tag{A.53}$$

$$s_k = \begin{cases} 1 & \text{for } I_{S+}, I_{T+}, \\ -1 & \text{for } I_{S-}, I_{T-}, \end{cases} \tag{A.54}$$

$$s_p = \begin{cases} 1 & \text{for } I_{S+}, I_{T-}, \\ -1 & \text{for } I_{S-}, I_{T+}. \end{cases} \tag{A.55}$$

In terms of these quantities, the four integrals all contain the following factor in the integrand:

$$\delta(\varepsilon(k) - s_k\Lambda)\theta(s_p\varepsilon(k-P))\theta(\Lambda - s_p\varepsilon(k-P)). \tag{A.56}$$

We have used the fact that $\varepsilon(p) = \varepsilon(-p)$ here. This factor restricts the values of $\varepsilon(k)$ and $\varepsilon(k-P)$. It implies that once we use the delta function to perform the k_- integral, the remaining limits of the k_+ integral must come from the intersection of the hyperbola $\varepsilon(k) = s_k\Lambda$ with either the hyperbola $\varepsilon(k-P) = s_p\Lambda$ or the degenerate hyperbola $\varepsilon(k-P) = 0$, which is equivalent to the pair of lines $k_\pm = P_\pm$. Therefore we should simply find the locations of these intersections.

The intersection with the lines occur at

$$\lambda_A \equiv s_+\sqrt{|\varepsilon(P)|}, \tag{A.57}$$

and

$$\lambda_B \equiv s_- s_k \frac{\Lambda}{\sqrt{|\varepsilon(P)|}}. \tag{A.58}$$

These two values will then always serve as limits for the remaining k_+ integral. Note that

$$\lambda_A = \frac{s_+ s_- s_k \Lambda}{\lambda_B}. \tag{A.59}$$

This is a result of the $k_+ \leftrightarrow k_-$ symmetry of the integrals (without a rapidity cutoff) combined with the fact that the delta function constrains $\varepsilon(k) = s_k\Lambda$. Whenever the quantities

$$\lambda_\pm \equiv \frac{1}{2}\sqrt{|\varepsilon(P)|}\left(s_+ + s_-(s_k - s_p)\frac{\Lambda}{|\varepsilon(P)|} \pm\sqrt{1 - 2s_+s_-(s_k + s_p)\frac{\Lambda}{|\varepsilon(P)|} + (s_k - s_p)^2\frac{\Lambda^2}{\varepsilon(P)^2}}\right) \tag{A.60}$$

are purely real, the remaining integrals also have limits at $\lambda_\pm$ as a result of intersections of the nondegenerate hyperbolas. In that case the integration region splits into two disjoint pieces.

For I_{S+} and I_{S-}, $s_k = s_p$ and Eq. (A.60) simplifies to

$$\lambda_\pm = \frac{1}{2}\sqrt{|\varepsilon(P)|}\left(s_+ \pm \sqrt{1 - 4s_+s_-s_k\frac{\Lambda}{|\varepsilon(P)|}}\right). \tag{A.61}$$

For I_{T+} and I_{T-}, $s_p = -s_k$ and this simplifies to

$$\lambda_\pm = \frac{1}{2}\sqrt{|\varepsilon(P)|}\left(s_+ + 2s_-s_k\frac{\Lambda}{|\varepsilon(P)|} \pm \sqrt{1 + \frac{4\Lambda^2}{\varepsilon(P)^2}}\right). \tag{A.62}$$

Therefore the t-channel integrals always split into two pieces. The s-channel integrals split unless

$$|\varepsilon(K)| \leq 4\Lambda, \tag{A.63}$$

in which case $I_{S\operatorname{sign}\varepsilon(K)}$ is over a single contiguous region bounded by λ_A and λ_B. In that case,

$$I_{S\operatorname{sign}\varepsilon(K)} = \int_{\min(\lambda_A,\lambda_B)}^{\max(\lambda_A,\lambda_B)} \ldots \tag{A.64}$$

We will see that this integral (and only this one) is generally divergent as $\varepsilon(K) \to 0$, and that this divergence is cured by the rapidity cutoff.

For the remaining three integrals (and for $I_{S\operatorname{sign}\varepsilon(K)}$ if (A.63) is not satisfied), the integration regions are bounded on one side by either λ_A or λ_B and on the other by either λ_+ or λ_-. The remaining integrals take the form

$$I = \left(\int_{\lambda_1}^{\lambda_2} + \int_{\lambda_3}^{\lambda_4}\right) dk_+ \ldots \tag{A.65}$$

where λ_1 through λ_4 are the limits sorted in ascending order.

Rapidity limits

At this point, let us simplify the discussion by taking g to be a momentum-independent constant. We will find that this assumption is not consistent, because the beta function depends on the momentum. Suppressing the integration limits, the remaining integrals take the following form:

$$I(P_+, P_-, s_k, s_p) = g^2 \int d^2k \frac{\delta(\varepsilon(k) - s_k\Lambda)\theta(s_p\varepsilon(k-P))F(k,P)}{1 + s_p\frac{\varepsilon(k-P)}{\Lambda}}, \tag{A.66}$$

where

$$F(k,P) = \theta(\Upsilon - |k_+|)\theta(\Upsilon - |k_-|)f(k-P). \tag{A.67}$$

Note that the step function constrains the value of the denominator to be between 1 and 2 throughout the integration region, so all of the integrals are nonnegative.

Under $P_+ \to -P_+$ or $P_- \to -P_-$, $I_{S+} \leftrightarrow I_{S-}$ and $I_{T+} \leftrightarrow I_{T-}$. Thus we can take both components of K and Q to be positive without loss of generality. This simplification would have held earlier if we had assumed the coupling function obeys particle-hole symmetry, but this symmetry is generically broken by the NVH region.

We can now find the effect of the rapidity cutoff on the integration limits for the various integrals. The lower limit on k_+ imposed by the rapidity cutoff for I_{S+} and I_{S-} is

$$\lambda_{R1} = \eta \max\left(\frac{\Lambda}{\Upsilon}, P_+ - \Upsilon\right), \tag{A.68}$$

where $\eta = \sqrt{|P_-/P_+|}$. The upper limit is

$$\lambda_{R2} = \begin{cases} \eta\Upsilon, & P_- \le \Upsilon \\ \eta \min\left(\Upsilon, \frac{\Lambda}{P_- - \Upsilon},\right), & P_- > \Upsilon. \end{cases} \tag{A.69}$$

$\lambda_{R1}/\lambda_{R2}$ replaces the lower/upper limits in (A.64) or (A.65) when it is within either integration region. Alternatively, if it is less than/greater than both limits in one of the integrals, the integral is set to zero.

For I_{S-} and I_{T-}, one of the integration regions has negative k_+ and the other has positive k_+. There are four possible rapidity limits:

$$\lambda_{R3} = \eta(P_+ - \Upsilon), \tag{A.70}$$

$$\lambda_{R4} = -\frac{\eta\Lambda}{\Upsilon}, \tag{A.71}$$

$$\lambda_{R5} = \frac{\eta\Lambda}{\Upsilon - P_-}, \tag{A.72}$$

$$\lambda_{R6} = \eta\Upsilon. \tag{A.73}$$

$\lambda_{R3}/\lambda_{R4}$ replaces the lower/upper limit for the negative integration region and $\lambda_{R5}/\lambda_{R6}$ replaces the lower/upper limit for the positive region, or they set the appropriate integrals to zero, acting in a manner analogous to that described above for λ_{R1} and λ_{R2}.

Indefinite integrals

Evaluating the delta function in (A.66) and changing variables to $x = \frac{k_+}{\sqrt{\Lambda}}$ yields

$$I = g^2 \int \frac{dx}{|x|} \frac{1}{1 + s_k - s_p \alpha(x + s_k/x - \alpha)}, \tag{A.74}$$

where

$$\alpha \equiv \sqrt{\frac{|\varepsilon(P)|}{\Lambda}}. \tag{A.75}$$

We can directly compute the indefinite integrals for $I_{S\pm}$ and $I_{T\pm}$ as long as we make use of the restrictions on the integration limits implied by Section A. We find

$$I_{S+}(x) = \frac{g^2}{\sqrt{\alpha^4+4}} \log \left| \frac{-2 - \alpha^2 + \sqrt{\alpha^4+4} + 2\alpha x}{2 + \alpha^2 + \sqrt{\alpha^4+4} - 2\alpha x} \right|, \tag{A.76}$$

$$I_{S-}(x) = \frac{g^2 \operatorname{sign} x}{\sqrt{\alpha^4+4}} \log \left| \frac{-2 + \alpha^2 + \sqrt{\alpha^4+4} - 2\alpha x}{2 - \alpha^2 + \sqrt{\alpha^4+4} + 2\alpha x} \right|. \tag{A.77}$$

For I_{T+}, the appropriate indefinite integral depends on the magnitude of α, or in other words on the relative size of $|\varepsilon(Q)|$ and Λ:

$$I_{T+}(x) = \begin{cases} -\frac{2g^2}{\alpha\sqrt{4-\alpha^2}} \arctan\left(\frac{\alpha - 2x}{\sqrt{4-\alpha^2}}\right), & |\varepsilon(Q)| < 4\Lambda, \\ \frac{g^2}{\alpha\sqrt{\alpha^2-4}} \log \left| \frac{\alpha + \sqrt{\alpha^2-4} - 2x}{-\alpha + \sqrt{\alpha^2-4} + 2x} \right|, & |\varepsilon(Q)| > 4\Lambda. \end{cases} \tag{A.78}$$

Finally,

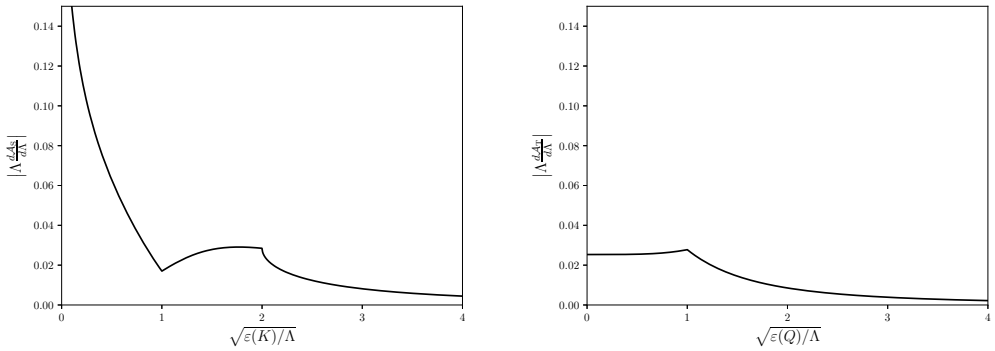
$$I_{T-}(x) = \frac{g^2 \operatorname{sign} x}{\alpha\sqrt{\alpha^2+4}} \log \left| \frac{-\alpha + \sqrt{\alpha^2-4} + 2x}{\alpha + \sqrt{\alpha^2-4} - 2x} \right|. \tag{A.79}$$

Collinear-anticollinear limit

Consider scattering between generic collinear and anticollinear particles. In this case, $|\varepsilon(P)| \gg 4\Lambda$, so $\alpha \gg 2$. In the $\alpha \to \infty$ limit,

$$\frac{1}{\alpha\sqrt{\alpha^2+4}} \approx \frac{1}{\alpha\sqrt{\alpha^2-4}} \approx \frac{1}{\sqrt{\alpha^4+4}} \to \frac{1}{\alpha^2} = \frac{\Lambda}{|\varepsilon(P)|}. \tag{A.80}$$

This suggests the beta function is suppressed by $\Lambda/|\varepsilon(P)|$ for collinear-anticollinear scattering, although we must also check the behavior of the log functions.



(a) (b)

Figure A.2: The log derivatives of the s- and t-channel diagrams in units of g^2 with no rapidity cutoff for a Fermi surface with a Van Hove singularity. We assume a constant coupling.

When $a > 2$, there are always two disjoint integration regions. Exchanging k_+ and k_- exchanges the two regions, so when we ignore the rapidity cutoff they must have the same value. Evaluating the integral between

$$\frac{\lambda_+}{\sqrt{\Lambda}} = \frac{1}{2}\alpha\left(1 + (s_k - s_p)\frac{1}{\alpha^2} + \sqrt{1 - 2(s_k + s_p)\frac{1}{\alpha^2} + (s_k - s_p)^2\frac{1}{\alpha^4}}\right) \tag{A.81}$$

and

$$\frac{\lambda_A}{\sqrt{\Lambda}} = \alpha, \tag{A.82}$$

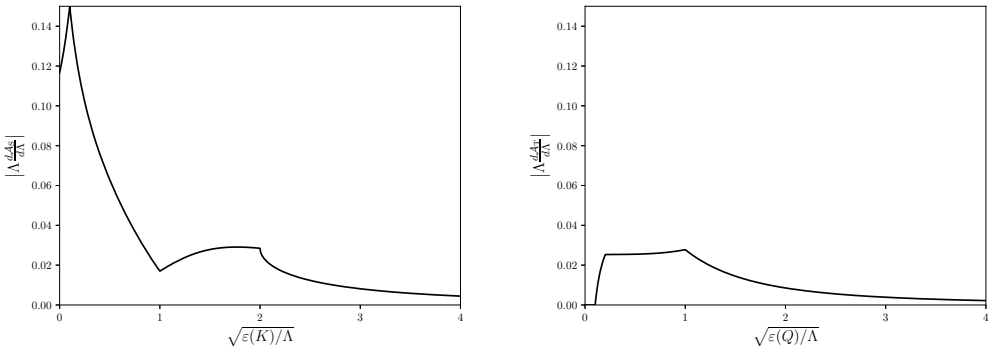
reversing the order if $\lambda_+ > \lambda_A$, and taking the $\alpha \to \infty$ limit yields the same result for each integral:

$$I = (2\log 2)\frac{\Lambda}{|\varepsilon(P)|}g^2 + O\left(\frac{\Lambda^2}{\varepsilon(P)^2}\right) \tag{A.83}$$

for collinear-anticollinear scattering. All one-loop contributions to the beta function are therefore power suppressed in this limit. This remains true when we include the rapidity cutoff, since it can only reduce the size of the integration region. Furthermore, such interactions continue to be power suppressed after we drop the assumption of a momentum-independent coupling, since the integration region always shrinks to zero size as $\varepsilon(P)$ becomes large.

Collinear limit

Consider the scenario where all scattered particles are restricted to the collinear region and assume the external momenta p_{i+} and their sums/differences (K_+, Q_+, and Q'_+) are all order Υ and much larger than $\sqrt{\Lambda}$. Furthermore, assume the scattered



(a) (b)

Figure A.3: Plots demonstrating how the rapidity cutoff modifies Fig. (A.2).

particles have energies well below the cutoff, so $\varepsilon(p) \ll \Lambda$. Together, these imply that the perpendicular components of the momenta are small:

$$p_{i-} = \frac{\varepsilon(p_i)}{p_{i+}} = O\left(\frac{\varepsilon(p_i)}{\Upsilon}\right) \ll \frac{\Lambda}{\Upsilon}. \tag{A.84}$$

The following results also hold, with appropriate modifications, if all momenta lie in the anticollinear region.

In this limit, only the rapidity cutoff on the collinear components of momenta comes into play. Furthermore,

$$\varepsilon(K) = \varepsilon(p_1)\left(1 + \frac{p_{2+}}{p_{1+}}\right) + \varepsilon(p_2)\left(1 + \frac{p_{1+}}{p_{2+}}\right) \ll \Lambda \tag{A.85}$$

since we have assumed that the collinear components of the incoming particles are all of the same order. As a result,

$$\alpha = \sqrt{\frac{\varepsilon(K)}{\Lambda}} \ll 1. \tag{A.86}$$

Similar statements hold for the t- and u-channel contributions.

After a change of variables, the integration limits for both the I_{S+} and I_{S-} integrals are $\sqrt{|\varepsilon(K)/\Lambda|} = \alpha$ and $\eta\Upsilon/\sqrt{\Lambda} = \alpha\Upsilon/|K_+|$. Only one of the integration regions for I_{S-} remains after we impose the rapidity cutoff. Substituting these limits into the indefinite integrals and taking the small-α limit yields

$$I_{S+} \approx I_{S-} \to \frac{g^2}{2}\log\left(\frac{2\Upsilon}{K_+} - 1\right). \tag{A.87}$$

Figure A.4: Integration region for collinear external momentum.

For the I_{T+} and I_{T-} integrals, only one of the two integration regions remains. The limits are

$$\frac{\lambda_{\mp}}{\sqrt{\Lambda}} = \frac{1}{2}\alpha\left(1 \pm \frac{2}{\alpha^2} \mp \sqrt{1+\frac{4}{\alpha^4}}\right), \tag{A.88}$$

with λ_- for I_{T+} and λ_+ for I_{T-}, and $\lambda_A/\sqrt{\Lambda} = \alpha$. Substituting these into the appropriate indefinite integrals gives

$$I_{T+} \approx I_{T-} \to \frac{g^2}{2} \tag{A.89}$$

in the collinear limit.

There are two important features of (A.87) and (A.89). First, the contributions from the s, t, and u channels will all be order g^2. Second, the integrals are independent of the small (anticollinear) components of the external momenta. These conclusions do not depend on our assumption of a momentum-independent coupling. Backtracking through our derivation and restoring the momentum dependence yields (4.46).

Forward scattering

If both components of Q are smaller than Λ/Υ, both integration regions for I_{T+} and I_{T-} shrink to zero size. Thus, the t-channel contribution to the beta function disappears in the forward-scattering limit in the presence of a rapidity cutoff; see Fig. A.3. This is analogous to the situation discussed in [5], where the forward scattering function makes no contribution to the beta functions for a round Fermi surface. As in the case of a round Fermi surface, there is a sharp change in the contribution to the beta function once $\varepsilon(Q)$ exceeds a threshold; compare Fig. 4.4 and Fig. A.3.

BCS limit

Consider I_{S+} in the $K_\pm \to 0$ limit. Since $\varepsilon(K) < 4\Lambda$, there is a single contiguous integration region, bounded by $\lambda_A = \sqrt{|\varepsilon(K)|}$ and $\lambda_B = \sqrt{\Lambda/|\varepsilon(K)|}$. The extent of this region diverges as we lower $\varepsilon(K)$. We find

$$I_{S+} = \frac{g^2}{\sqrt{\alpha^4+4}} \log\left(\frac{(-2+\alpha^2+\sqrt{\alpha^4+4})(\alpha^2+\sqrt{\alpha^4+4})}{(-\alpha^2+\sqrt{\alpha^4+4})(2-\alpha^2+\sqrt{\alpha^4+4})}\right). \tag{A.90}$$

Taking the small α limit yields

$$I_{S+} \to \frac{1}{2}g^2 \log\frac{4\Lambda}{\varepsilon(K)}, \tag{A.91}$$

which diverges at $\varepsilon(K) = 0$. This is the divergence that forced us to introduce the rapidity regulator. I_{S-} has the same value as I_{S+} in the small α limit.

Introducing the rapidity cutoff regulates the divergence. The rapidity cutoff restricts the I_{S+} integral to run from $\eta\sqrt{\Lambda}/\Upsilon$ to $\eta\Upsilon/\sqrt{\Lambda}$ and the I_{S-} integral to run from η to $\eta\Upsilon/\sqrt{\Lambda}$. Plugging these into (A.76) and (A.77) and taking the $\alpha \to 0$ limit yields

$$I_{\pm} \to \frac{1}{2} g^2 \log \frac{\Upsilon^2}{\Lambda}, \tag{A.92}$$

so

$$\Lambda \frac{d\mathcal{A}_{\rm S}}{d\Lambda} = -\frac{g}{4\pi^2} \log \frac{\Upsilon^2}{\Lambda} \tag{A.93}$$

for back-to-back interactions.

Generic BCS beta function

The previous results indicate that we may take the coupling for fixed ultrasoft net momentum (the BCS configuration) to be analytic in the other momenta to leading-log order. Furthermore, we may drop all but the s-channel diagram to this order. Parameterize the BCS coupling $g_{\rm B}(p_1, p_3)$ in terms of one of the incoming momenta p_1 and one of the outgoing momenta p_3 at fixed ultrasoft K. The log derivative of the amplitude is (A.42), with

$$I_{S\pm} \equiv \pm\Lambda \int d^2k \frac{\delta(\Lambda \mp \varepsilon(k))\theta(\pm\varepsilon(K-k))F(k,K)}{2\varepsilon(k) - E} g_{\rm B}(p_1, k) g_{\rm B}(k, p_3) \tag{A.94}$$

and $F(k, K)$ from (A.67). Take the components of K to be positive but infinitesimal to avoid ambiguity from the definition of the step functions. Eq. (A.94) receives contributions from several one-dimensional regions of momentum space; see Fig. A.5.

Call I_{++} the contribution from the region with $k_+ > k_- > 0$. Evaluating the k_- integral with the delta function yields

$$I_{++} = \frac{1}{2} \int_{\sqrt{\Lambda}}^{\Upsilon} \frac{dk}{k} g_{\rm B}(p_1, k) g_{\rm B}(k, p_3) \tag{A.95}$$

up to power-suppressed terms. Assume it is possible to expand the coupling function in k_+ and k_-. The resulting expression for I_{++} will include terms of the form

$$\int_{\sqrt{\Lambda}}^{\Upsilon} \frac{dk_+}{k_+} k_+^m \left(\frac{\Lambda}{k_+}\right)^n \partial_{k_+}^m \partial_{k_-}^n [g_{\rm B}(p_1, 0) g_{\rm B}(0, p_3)]. \tag{A.96}$$

The natural scale for the derivatives is $(1/V_F)^{m+n}$. As a result, terms with $m \neq n$ give at most order-one contributions to the beta function. When $m = n$, terms in (A.96)

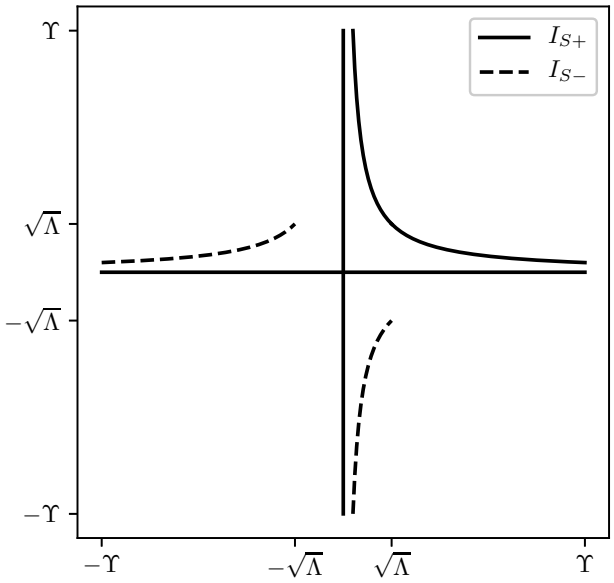


Figure A.5: The integration regions for $I_{S\pm}$.

take the form

$$\left(\frac{\Lambda}{V_F^2}\right)^n \int_{\sqrt{\Lambda}}^{\Upsilon} \frac{dk_+}{k_+} = \frac{1}{2}\left(\frac{\Lambda}{V_F^2}\right)^n \log\frac{\Upsilon^2}{\Lambda}. \tag{A.97}$$

Since we assume $\Lambda \ll V_F^2$, these are suppressed unless $n = 0$. The $n = m = 0$ term is log enhanced, and the leading-log result is therefore

$$I_{++} = \frac{1}{4} g_{\mathrm{B}}(p_1, 0) g_{\mathrm{B}}(0, p_3) \log\frac{\Upsilon^2}{\Lambda}. \tag{A.98}$$

A similar analysis holds for each of the the terms in $I_{S+} + I_{S-}$. Adding the NVH region cancels the Υ dependence. Finally, setting the log derivative with respect to Λ of the sum of the tree-level amplitude $g_{\mathrm{B}}(p_1, p_3)$ and the one-loop amplitude equal to zero implies

$$\Lambda \frac{d g_{\mathrm{B}}(p_1, p_3)}{d\Lambda} = \frac{1}{4\pi^2} g_{\mathrm{B}}(p_1, 0) g_{\mathrm{B}}(0, p_3) \log\frac{V_F^2}{\Lambda}. \tag{A.99}$$

The beta function for the the coupling between modes in the vicinity of the VH point, $g_{\mathrm{B}}(0, 0)$, is independent of the other couplings, and the solution is

$$g_{\mathrm{B}}(0, 0; \Lambda) = \frac{g_{\mathrm{B}}(0, 0, \Lambda_0)}{1 + \frac{g_{\mathrm{B}}(0,0,\Lambda_0)}{8\pi^2}\left(\log^2\frac{V_F^2}{\Lambda} - \log^2\frac{V_F^2}{\Lambda_0}\right)}. \tag{A.100}$$

Using this, the beta function for $g_{\mathrm{B}}(p_1, 0)$ becomes

$$\Lambda \frac{dg_{\mathrm{B}}(p_1,0)}{d\Lambda} = \frac{1}{4\pi^2} \frac{g_{\mathrm{B}}(p_1,0) g_{\mathrm{B}}(0,0,\Lambda_0)}{1 + \frac{g_{\mathrm{B}}(0,0,\Lambda_0)}{8\pi^2}\left(\log^2 \frac{V_F^2}{\Lambda} - \log^2 \frac{V_F^2}{\Lambda_0}\right)} \log \frac{V_F^2}{\Lambda}, \tag{A.101}$$

with solution

$$g_{\mathrm{B}}(p_1, 0; \Lambda) = \frac{g_{\mathrm{B}}(p_1, 0; \Lambda_0)}{1 + \frac{g_{\mathrm{B}}(0,0,\Lambda_0)}{8\pi^2}\left(\log^2 \frac{V_F^2}{\Lambda} - \log^2 \frac{V_F^2}{\Lambda_0}\right)}. \tag{A.102}$$

An analogous result holds for $g_{\mathrm{B}}(0, p_3; \Lambda)$. Substituting these into the beta function for $g_{\mathrm{B}}(p_1, p_3)$ and solving yields

$$\begin{aligned} g_{\mathrm{B}}(p_1, p_3; \Lambda) &= g_{\mathrm{B}}(p_1, p_3; \Lambda_0) \\ &\quad - \left(\frac{1}{8\pi^2}\right) \frac{g_{\mathrm{B}}(p_1, 0; \Lambda_0) g_{\mathrm{B}}(0, p_3; \Lambda_0) \left(\log^2 \frac{V_F^2}{\Lambda} - \log^2 \frac{V_F^2}{\Lambda_0}\right)}{1 + \frac{g_{\mathrm{B}}(0,0;\Lambda_0)}{8\pi^2}\left(\log^2 \frac{V_F^2}{\Lambda} - \log^2 \frac{V_F^2}{\Lambda_0}\right)}. \end{aligned} \tag{A.103}$$

We see that the expressions for $g_{\mathrm{B}}(0, 0; \Lambda)$, $g_{\mathrm{B}}(p_1, 0; \Lambda)$, and $g_{\mathrm{B}}(0, p_3; \Lambda)$ are in fact special cases of this general result.

Appendix B

THE ONE-LOOP BETA FUNCTION FOR A ROUND FERMI SURFACE

We make the same assumption of a spin-singlet interaction as in the calculation of the beta functions for the Van Hove case in Appendix A. We also choose to combine the forward and back-to-back couplings into a single coupling function g. As in the Van Hove calculation, the one-loop diagrams will depend on the net momenta K and the transfer momenta Q and Q'. We absorb the sum over label momenta into the integral over the parallel component of momentum. It will be sufficient for our purposes to assume an initial condition of a constant coupling g and a circular Fermi surface. The initial part of the calculation closely follows that of the Van Hove case, and we find

$$\Lambda \frac{d\mathcal{A}_{\mathrm{S}}}{d\Lambda} = -\frac{1}{2\pi^2}(I_{S+} + I_{S-}), \tag{B.1}$$

$$\Lambda \frac{d\mathcal{A}_{\mathrm{T}}}{d\Lambda} = \frac{1}{4\pi^2}(I_{T+} + I_{T-}), \tag{B.2}$$

where

$$I_{S\pm} \equiv \pm\Lambda g^2 \int d^2k \frac{\delta(\Lambda \mp \varepsilon(k))\theta(\pm\varepsilon(K-k))}{\varepsilon(k) + \varepsilon(K-k)}, \tag{B.3}$$

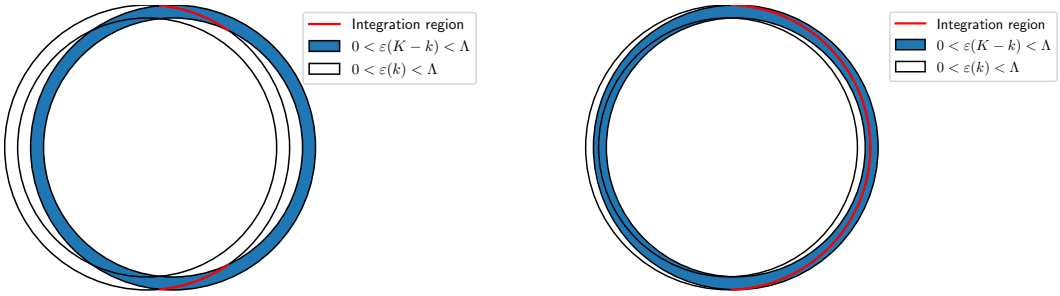
$$I_{T\pm} \equiv \pm\Lambda g^2 \int d^2k \frac{\delta(\Lambda \mp \varepsilon(k))\theta(\mp\varepsilon(k+Q))}{\varepsilon(k) - \varepsilon(k+Q)}. \tag{B.4}$$

There is a factor of two difference because we do not change variables to $k_{\pm}$.

The interesting behavior is due to the fact that the net and transfer momenta crucially affect the step functions in the above expressions. In Eq. (B.3), the momentum k is constrained to lie at $k_{\perp} = \pm\Lambda$, while simultaneously the step function demands that $\varepsilon(K-k)$ have the same sign as $\varepsilon(k)$. This means that for I_{S+} the integral is over the parts of a circle of radius $|k| = K_F + \Lambda/V_F$ that intersects an annulus with inner and outer radii Λ/V_F and $\Lambda/V_F + K_F$ and which is displaced from the origin by a distance K; see Fig. B.1.

Using the step function to evaluate the $k_{\perp}$ integral and realizing that the step function acts to constrain the angular integral, we have for example

$$I_{S+} = \frac{g^2}{V_F}\left(\int_{\theta_{\min}}^{\theta_{\max}} + \int_{-\theta_{\max}}^{-\theta_{\min}}\right) \frac{K_F \, \mathrm{d}\theta}{1 + \varepsilon^*(\theta)/\Lambda}. \tag{B.5}$$



(a) $V_F|K| > \Lambda$ (b) $V_F|K| < \Lambda$

Figure B.1: The integration region for I_{S+}. The two copies of the Fermi surface are displaced by the net momentum K.

The energy of the loop particle with momentum $K - k$ is given by

$$\varepsilon^*(\theta) = V_F \left[\sqrt{(K_F^2 + \Lambda)^2 - 2(K_F + \Lambda)|K| \cos\theta + K^2} - K_F \right]. \tag{B.6}$$

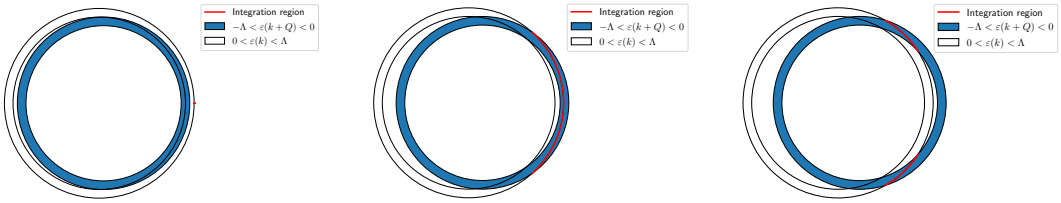
The minimum and maximum values in the angular integral can be found by straightforward geometry to be

$$\theta_{\min} = \mathrm{Re}\left(\arccos \frac{K^2 + 2K_F\Lambda + \Lambda^2}{2|K|(K_F + \Lambda)} \right), \tag{B.7}$$

$$\theta_{\max} = \mathrm{Re}\left(\arccos \frac{|K|}{2(K_F + \Lambda)} \right), \tag{B.8}$$

where θ is measured relative to the direction of K. When $V_F|K| < \Lambda$, the minimum value is $\theta = 0$ and the two integrals in Eq. (B.5) combine into a single integral from $-\theta_{\max}$ to $\theta_{\max}$.

The analysis for I_{S-} and $I_{T\pm}$ is similar. However, the integrals $I_{T\pm}$ differ in an important manner. Whenever $V_F|Q| < \Lambda$, they are identically zero. Fig. B.2a shows that the integration region is only of nonzero size when $V_F|Q| > \Lambda$. Once $V_F|Q| > 2\Lambda$, it divides into two small integration regions that shrink as Q is increased. Fig. 4.4 in Chapter 4 demonstrates the results of numerically integrating the resulting expressions.



(a) $V_F|Q| < \Lambda$ (b) $\Lambda < V_F|Q| < 2\Lambda$ (c) $V_F|Q| > 2\Lambda$

Figure B.2: The integration region for I_{T+}. The two copies of the Fermi surface are displaced by the transfer momentum Q.

Appendix C

THE IMAGINARY PART OF THE SELF-ENERGY

We begin with the imaginary part of the susceptibility in the presence of a Fermi velocity cutoff. We found the value of this quantity in Chapter 5 for zero chemical potential, but we extend the analysis to nonzero chemical potential here. Starting with Eq. (1) from [63] with $\hbar = 1$ and $\varepsilon_F = \mu$, perform the trivial k_z integral:

$$\text{Im } \tilde{\chi} = \frac{1}{2}\int \mathrm{d}k_- \,\mathrm{d}k_+ \theta(\varepsilon(k+q)-\mu)\theta(\mu-\varepsilon(k)) \times \delta(q_+k_- + q_-k_+ + \varepsilon(q) - \omega). \quad \text{(C.1)}$$

The hard cutoff on the Fermi velocity shows up in this calculation as a cutoff Υ on $|k_+|$, $|k_-|$, $|k_+ + q_+|$, and $|k_- + q_-|$. Therefore Eq. (C.1) is modified:

$$\text{Im } \tilde{\chi} = \frac{1}{2}\int \mathrm{d}k_- \,\mathrm{d}k_+ \theta(\varepsilon(k+q)-\mu)\theta(\mu-\varepsilon(k)) \times \delta(q_+k_- + q_-k_+ + \varepsilon(q) - \omega)\theta(\Upsilon - |k_i|)\theta(\Upsilon - |k_i+q_i|). \quad \text{(C.2)}$$

This is the correct expression for $\omega \geq 0$ only. As usual, q corresponds to the transfer momentum injected into the loop.

C.1 Transfer momentum outside the first quadrant

Change variables in Eq. (C.2) to $k_i \to -k_i$:

$$\text{Im } \tilde{\chi} = \frac{1}{2}\int \mathrm{d}k_- \,\mathrm{d}k_+ \theta(\varepsilon(-k+q)-\mu)\theta(\mu-\varepsilon(-k)) \times \delta(-q_+k_- - q_-k_+ + \varepsilon(q) - \omega)\theta(\Upsilon - |k_i|)\theta(\Upsilon - |-k_i+q_i|). \quad \text{(C.3)}$$

Since $\varepsilon(p) = p_+p_- = \varepsilon(-p)$,

$$\text{Im } \tilde{\chi} = \frac{1}{2}\int \mathrm{d}k_- \,\mathrm{d}k_+ \theta(\varepsilon(k-q)-\mu)\theta(\mu-\varepsilon(k)) \times \delta(-q_+k_- - q_-k_+ + \varepsilon(-q) - \omega)\theta(\Upsilon - |k_i|)\theta(\Upsilon - |k_i-q_i|). \quad \text{(C.4)}$$

Therefore

$$\text{Im } \tilde{\chi}(-q,\omega) = \text{Im } \tilde{\chi}(q,\omega). \quad \text{(C.5)}$$

From [63], Im $\tilde{\chi}$ for $\omega < 0$ is given by

$$\text{Im } \tilde{\chi}(q,-\omega) = -\,\text{Im } \tilde{\chi}(-q,\omega). \quad \text{(C.6)}$$

(C.5) then implies

$$\text{Im } \tilde{\chi}(q,\omega) = \text{sign}\,\omega \,\text{Im } \tilde{\chi}(q,|\omega|). \tag{C.7}$$

Change variables again:

$$k'_+ = k_+ + q_+, \tag{C.8}$$

$$k'_- = -k_- - q_-. \tag{C.9}$$

Then

$$\varepsilon(k+q) = -\varepsilon(k') \tag{C.10}$$

and

$$\varepsilon(k) = -(k'_+ + q_+)(k'_- - q_-) = -\varepsilon(k+\tilde{q}) \tag{C.11}$$

where $\tilde{q} = (q_+, -q_-)$. The argument of the delta function in Eq. (C.2) can be written

$$\varepsilon(k+q) - \varepsilon(k) = -\varepsilon(k') + \varepsilon(k'+\tilde{q}). \tag{C.12}$$

Dropping the primes gives us

$$\text{Im } \tilde{\chi} = \frac{1}{2}\int \mathrm{d}k_- \,\mathrm{d}k_+ \theta(\varepsilon(k+\tilde{q})+\mu)\theta(-\mu-\varepsilon(k)) \times \delta(\varepsilon(k+\tilde{q}) - \varepsilon(k) - \omega)\theta(\Upsilon - |k_i|)\theta(\Upsilon - |k_i + \tilde{q}_i|). \tag{C.13}$$

Along with Eq. (C.7), this implies

$$\text{Im } \tilde{\chi}(q_+, q_-, \omega, \mu) = \text{sign}\,\omega \,\text{Im } \tilde{\chi}(|q_+|, |q_-|, |\omega|, \text{sign}\, q_+ \text{sign}\, q_- \mu). \tag{C.14}$$

Thus we need only find Im $\tilde{\chi}$ for q in the first quadrant for positive ω and arbitrary values of μ to determine its value everywhere. In fact, we can also exchange $k_+ \to k_-$ and vice versa to see that Im $\tilde{\chi}(q_+, q_-) =$ Im $\tilde{\chi}(q_-, q_+)$, so we only need the value of Im $\tilde{\chi}$ in the first quadrant for $q_- > q_+$ to reproduce everything.

C.2 The integration region for zero chemical potential

Call D_1 the region where $\varepsilon(k) < 0$. For $\mu = 0$, this is the union of the second and fourth quadrants of k-space. Call D_2 the region where $\varepsilon(k+q) > 0$. For $\mu = 0$, this is the union of the first and third quadrants in k-space displaced by $-q$. Finally, call L the line where

$$k_+ q_- + k_- q_+ = \omega - \varepsilon(q). \tag{C.15}$$

Imagine doing the k_+ integral with the delta function. Then Im $\tilde{\chi}$ is given by one half of the width in k_- space of the intersection of D_1, D_2, and L divided by q_-, as

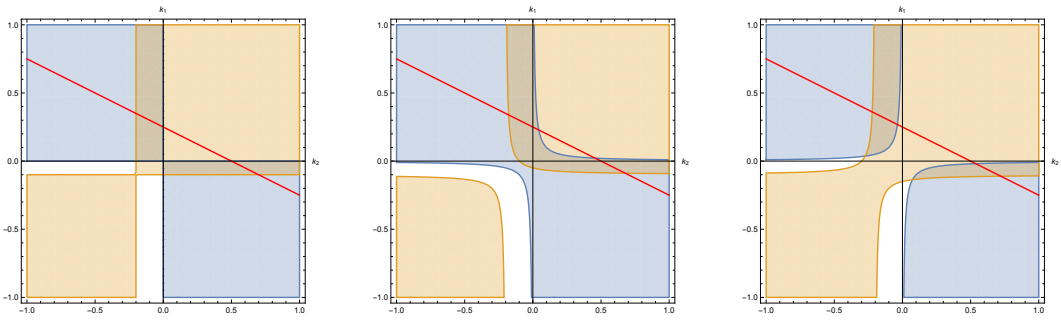


Figure C.1: Regions to calculate the imaginary part of the susceptibility with zero, positive, and negative chemical potential and a Fermi velocity cutoff.

long as $q_- \neq 0$. If q_- does equal zero, we should do the integral over k_- instead; the result will be the same as with $q_+ = 0$ but with the indices exchanged. Fig. C.1 has three plots of the regions and lines. All three have $q_+ = 0.1$, $q_- = 0.2$, and $\omega = 0.07$, while their chemical potentials differ, with $\mu = 0$, 0.01, and -0.01.

When $\mu = 0$, we see that $D_1 \cap D_2$ consists of two skinny rectangles with small dimension $|q_+|$ and $|q_-|$. Increasing ω increases the k_+-intercept of the line, while varying q_+ and q_- modify both the slope and the k_+-intercept.

If we hold q_+ and q_- fixed, the cutoff will begin to modify Im $\tilde{\chi}$ once the k_+-intersect is large enough that the line intersects the corner of the skinny regions. Take $\mu = 0$ for simplicity. For small μ, the results will be qualitatively similar: we replace the straight boundaries of the rectangles with sections of hyperbolas, but the intersection must occur far out along the Fermi surface if the cutoff is large compared with $|q|$, so the location of the intersection will not change much. If $q_- > q_+ > 0$, the line will first intersect the bottom right corner of the lower rectangle as we increase ω. The corner of this rectangle is at $k_+ = -q_+$ and $k_- + q_- = \Upsilon$, so it occurs when

$$-q_+q_- + (\Upsilon - q_-)q_+ = \omega - \varepsilon(q) \tag{C.16}$$

or

$$\omega = q_+\Upsilon - \varepsilon(q). \tag{C.17}$$

If q_+ and q_- have nearly the same value, the line will intersect the top left corner of the vertical rectangle before it reaches the upper right corner of the horizontal rectangle. Otherwise, Im $\tilde{\chi}$ will then decrease linearly with increasing ω until the line intersects the top right corner of the horizontal rectangle, where $k_+ = 0$ and $k_- + q_- = \Upsilon$, so

$$\omega = q_+\Upsilon. \tag{C.18}$$

At this point Im $\tilde{\chi}$ will be half of its value without the cutoff. Im $\tilde{\chi}$ will then remain constant until the line intersects the top left corner of the vertical rectangle. Notice that exchanging the 1 and 2 labels reverses the roles of the two rectangles, so we can deduce that this second drop occurs when

$$\omega = q_{-}\Upsilon - \varepsilon(q) \tag{C.19}$$

and ends when

$$\omega = q_{-}\Upsilon, \tag{C.20}$$

at which point Im $\tilde{\chi}$ becomes zero.

C.3 The integration region for nonzero chemical potential

We can find the general formula for values of ω at which Im $\tilde{\chi}$ drops or stops dropping by identifying when the line intersects with both the hyperbolas and the cutoffs. Work in the first quadrant of q-space; we can then generalize using Eq. (C.14). Define q_{s} to be the component of q with smaller magnitude and q_{b} to be the component with larger magnitude. The first drop in Im $\tilde{\chi}$ relative to its value without a cutoff will occur when the line first intersects the corner of one of the skinny regions. This occurs when $k_{\mathrm{b}} + q_{\mathrm{b}} = \Upsilon$ on the hyperbola

$$(k_{\mathrm{s}} + q_{\mathrm{s}})(k_{\mathrm{b}} + q_{\mathrm{b}}) = \mu, \tag{C.21}$$

which implies

$$k_{\mathrm{s}} = \frac{\mu}{\Upsilon} - q_{\mathrm{s}}, \tag{C.22}$$

so the value of ω for which the first drop begins is given by

$$\left(\frac{\mu}{\Upsilon} - q_{\mathrm{s}}\right) q_{\mathrm{b}} + (\Upsilon - q_{\mathrm{b}}) q_{\mathrm{s}} = \omega - \varepsilon(q) \tag{C.23}$$

or

$$\omega = q_{\mathrm{s}}\Upsilon + \frac{q_{\mathrm{b}}}{\Upsilon}\mu - \varepsilon(q). \tag{C.24}$$

The first drop ends once

$$\omega = q_{\mathrm{s}}\Upsilon + \frac{q_{\mathrm{b}}}{\Upsilon - q_{\mathrm{b}}}\mu, \tag{C.25}$$

and exchanging k_{s} and k_{b} tells us the second drop begins once

$$\omega = q_{\mathrm{b}}\Upsilon + \frac{q_{\mathrm{s}}}{\Upsilon}\mu - \varepsilon(q) \tag{C.26}$$

and ends once

$$\omega = q_{\mathrm{b}}\Upsilon + \frac{q_{\mathrm{s}}}{\Upsilon - q_{\mathrm{s}}}\mu. \tag{C.27}$$

Therefore the points at which Im $\tilde{\chi}$ changes behavior only differ from the $\mu = 0$ case by terms $O(q_i\mu/\Upsilon)$. Finally, we can use Eq. (C.14) to generalize the location of these drops:

$$\omega_1 = |q_s|\Upsilon + \frac{|q_b|}{\Upsilon}\mu \text{sign}\, q_s \text{sign}\, q_b - |\varepsilon(q)|, \tag{C.28}$$

$$\omega_2 = |q_s|\Upsilon + \frac{|q_b|}{\Upsilon - |q_b|}\mu \text{sign}\, q_s \text{sign}\, q_b, \tag{C.29}$$

$$\omega_3 = |q_b|\Upsilon + \frac{|q_s|}{\Upsilon}\mu \text{sign}\, q_s \text{sign}\, q_b - |\varepsilon(q)|, \tag{C.30}$$

$$\omega_4 = |q_b|\Upsilon + \frac{|q_s|}{\Upsilon - |q_s|}\mu \text{sign}\, q_s \text{sign}\, q_b. \tag{C.31}$$

In the case with no cutoff, if we take either component of q to be zero the line will only intersect one of the skinny rectangles. In other words, Im $\tilde{\chi}$ as a function of q_+ has point discontinuities at $q_{+\pm} = 0$. The effect of the cutoff is to smooth out these point discontinuities, so that when one of the components of q drops below ω/Υ, Im $\tilde{\chi}$ already begins to decrease. This indicates that our results will converge nonuniformly to the results without a cutoff as we take Υ to infinity. The nonuniform convergence is not a problem since the we will always smear over a range of momentum.

C.4 The imaginary part of the susceptibility

Take $q_- > q_+ > 0$ and $\omega > 0$. As long as we are in the regime where the cutoff has no effect, the line intersects the $-q$ displaced curve on the right when

$$(k_+ + q_+)(k_- + q_-) = \mu \tag{C.32}$$

and

$$k_+ q_- + k_- q_+ = \omega - \varepsilon(q), \tag{C.33}$$

which implies

$$k_- = \frac{\omega - \varepsilon(q) + \sqrt{[\omega + \varepsilon(q)]^2 - 4\varepsilon(q)\mu}}{2q_+}. \tag{C.34}$$

We know that we should take the positive root because we are looking for the intercept where $k_- > k_+$. Now let us find the right intercept of the line with $k_+k_- = \mu$ when ω is big enough that there are in fact two such intercepts but not so big that Υ comes into play. This happens when

$$k_+ = \frac{\mu}{k_-}, \tag{C.35}$$

so

$$k_- = \frac{\omega - \varepsilon(q) + \sqrt{[\omega - \varepsilon(q)]^2 - 4\varepsilon(q)\mu}}{2q_+}. \tag{C.36}$$

Once again we take the positive (right) root. Clearly the difference of Eq. (C.34) and Eq. (C.36) gives the extent of the interval in k_--space. Once we divide by q_- to deal with the factor from the delta function, these expressions will be symmetric in the labels 1 and 2. This tells us that the width of the interval on the other skinny section must be the same, since we can interchange the two labels to interchange the legs.

We have now nearly recovered the final expression from [63], save for the issue when ω is small enough that the discriminant in Eq. (C.36) becomes negative, at which point ω is small enough that the line no longer intersects $k_+k_- = \mu$. To deal with this, we take the result from [63], which we can write as

$$\text{Im}\ \tilde{\chi} = \frac{1}{2}\text{Re}\left[\frac{1}{q_-}\left(\frac{\alpha_+}{2q_+} - \frac{\alpha_-}{2q_+}\right) + \frac{1}{q_+}\left(\frac{\alpha_+}{2q_-} - \frac{\alpha_-}{2q_-}\right)\right], \tag{C.37}$$

where

$$\alpha_\pm \equiv \omega - \varepsilon(q) + \sqrt{[\omega \pm \varepsilon(q)]^2 - 4\varepsilon(q)\mu}, \tag{C.38}$$

and note that the effect of the cutoff is to replace the roots from Eq. (C.34) and Eq. (C.36) with $\Upsilon - q_\text{b}$ or $\Upsilon - q_\text{s}$ once ω reaches the values in Eqs. (C.28)–(C.31). Therefore the full expression for Im $\tilde{\chi}$ is

$$\begin{aligned}
\text{Im}\ \tilde{\chi} = \frac{1}{2}\text{sign}\,\omega \\
\times \text{Re}\left\{\frac{1}{|q_\text{b}|}\left[\frac{\tilde{\alpha}_+}{2|q_\text{s}|}\theta(\omega_1 - |\omega|) + (\Upsilon - q_\text{b})\theta(|\omega| - \omega_1)\right.\right. \\
\left. -\frac{\tilde{\alpha}_-}{2|q_\text{s}|}\theta(\omega_2 - |\omega|) - (\Upsilon - q_\text{b})\theta(|\omega| - \omega_2)\right] \\
+ \frac{1}{|q_\text{s}|}\left[\frac{\tilde{\alpha}_+}{2|q_\text{b}|}\theta(\omega_3 - |\omega|) + (\Upsilon - q_\text{s})\theta(|\omega| - \omega_3)\right. \\
\left.\left. -\frac{\tilde{\alpha}_-}{2|q_\text{b}|}\theta(\omega_4 - |\omega|) + (\Upsilon - q_\text{s})\theta(|\omega| - \omega_4)\right]\right\},
\end{aligned} \tag{C.39}$$

where

$$\tilde{\alpha}_\pm \equiv |\omega| - |\varepsilon(q)| + \sqrt{[|\omega| \pm |\varepsilon(q)|]^2 - 4\varepsilon(q)\mu}. \tag{C.40}$$

Fig. C.2 contains a plot comparing the behavior of Im $\tilde{\chi}$ with and without the cutoff.

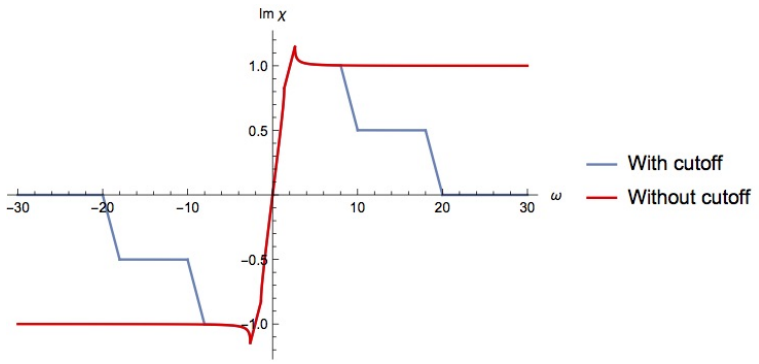


Figure C.2: A comparison of the imaginary part of the susceptibility with and without a cutoff. $q_+ = 1$, $q_- = 2$, $\mu = 0.05$, and $\Upsilon = 10$.

C.5 Integration region for the imaginary part of the self-energy

Consider now the imaginary part of the self-energy Im Σ, Eq. (17) in [63]. The energy argument of Im $\tilde{\chi}$ in this equation is

$$\varepsilon(k) - \varepsilon(k+q) = -(k_+q_- + k_-q_+ + q_+q_-) = -(k_+q_- + k_-q_+) - \varepsilon(q). \tag{C.41}$$

The step functions constrain the integration region such that

$$\mu < \varepsilon(k+q) < \varepsilon(k). \tag{C.42}$$

This ensures that the integrand is always positive in the region where we are integrating. Fig. C.3 contains a plot showing the modification of the integrand by the presence of a cutoff on the van Hove region.

C.6 Boundary of the modified region

We can also from figure C.3 that the integrand is modified in the region satisfying (C.42) in the vicinity of the negative q_+0 and q_- axes and far out along the legs of the allowed hyperbolic region. We will now find the equations for the curves indicating the onset of the first drop.

(C.28) tells us that the first drop within the integrated region occurs when

$$-(k_+q_- + k_-q_+) - \varepsilon(q) = |q_\mathrm{s}|\Upsilon + \frac{|q_\mathrm{b}|}{\Upsilon}\mu \operatorname{sign} q_\mathrm{s} \operatorname{sign} q_\mathrm{b} - |\varepsilon(q)|. \tag{C.43}$$

If $|q_+| < |q_-|$, we have

$$q_- = -\frac{(k_- + \frac{\mu}{\Upsilon}\operatorname{sign} q_+)q_+ + (1 - \operatorname{sign} q_+ \operatorname{sign} q_-)\varepsilon(q)}{\Upsilon \operatorname{sign} q_- + k_+}. \tag{C.44}$$

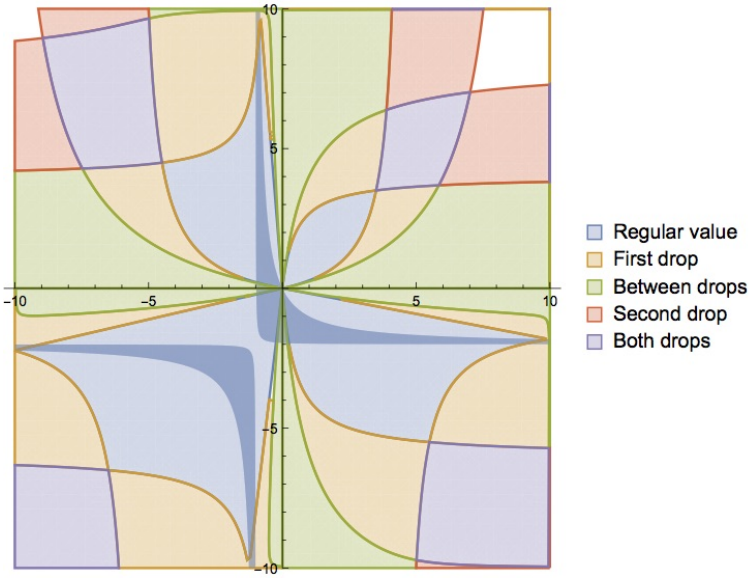


Figure C.3: The integrand for the imaginary part of the self-energy is modified by a Fermi velocity cutoff differently for different values of the loop momentum. The integral is only performed over the shaded region. $k_+ = 1$, $k_- = 2$, $\mu = 0.05$, and $\Upsilon = 10$.

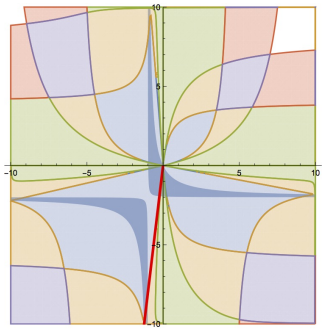
If $|q_+| > |q_-|$,

$$q_- = -\frac{(\operatorname{sign} q_+ \Upsilon + k_-)q_+ + (1 - \operatorname{sign} q_+ \operatorname{sign} q_-)\varepsilon(q)}{\frac{\mu}{\Upsilon}\operatorname{sign} q_- + k_+}. \tag{C.45}$$

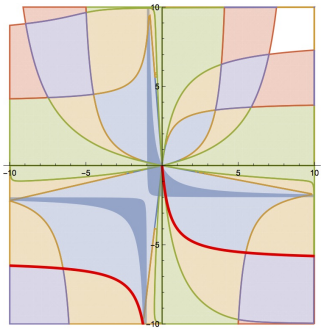
Inspection of Fig. C.3 indicates that there are in fact eight curves emanating from the origin that bound the region of momentum space with an unmodified value for Im $\tilde{\chi}$. These eight curves correspond to the two possibilities for the relative sizes of $|q_+|$ and $|q_-|$ and the four combinations for the signs of q_+ and q_-. Fig. C.4 contains plots of the eight curves corresponding to the parameters given for the momentum space in Fig. C.3.

C.7 Cutoff dependence of the imaginary part of the self-energy

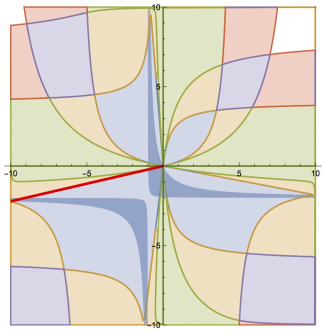
As stated above, these plots indicate that for small enough μ and $\varepsilon(k)$, there are six regions between the hyperbolas $\varepsilon(k+q) = \mu$ and $\varepsilon(k+q) = \varepsilon(k)$ where the integrand of Im Σ is modified by the presence of a cutoff: the two regions along the negative q_+ and q_- axes near the origin and the four regions far out along the legs of space between the hyperbolas. Fig. C.5 demonstrates that as we increase Υ, the two regions near the origin shrink in size while the regions along the legs are pushed out to larger and larger values of q. The value of Im $\tilde{\chi}$ in the region near



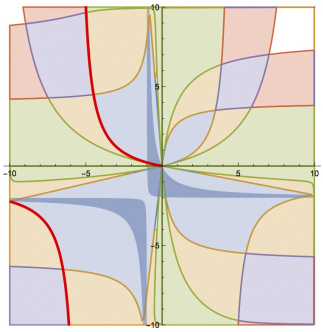
(a) $-\left(q_+ = \frac{k_+ - \Upsilon}{k_- - \mu/\Upsilon}\right) q_-$



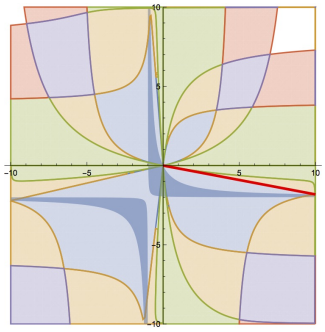
(b) $q_+ = -\left(\frac{k_+ + \Upsilon}{k_- + \mu/\Upsilon + 2q_-}\right) q_-$



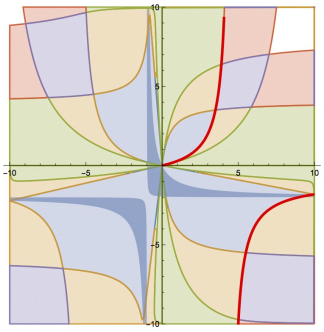
(c) $q_+ = -\left(\frac{k_+ - \mu/\Upsilon}{k_- - \mu}\right) q_-$



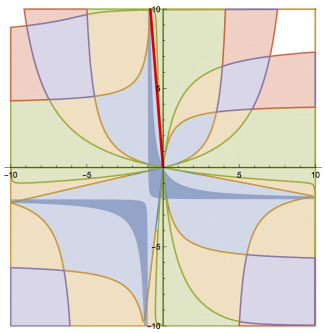
(d) $q_+ = -\left(\frac{k_+ + \mu/\Upsilon}{k_- + \mu + 2q_-}\right) q_-$



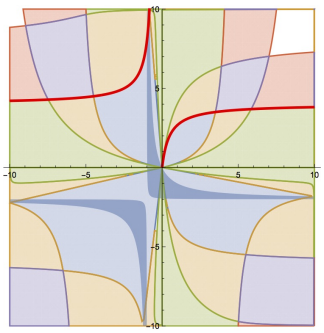
(e) $q_+ = -\left(\frac{k_+ + \mu/\Upsilon}{k_- + \mu}\right) q_-$



(f) $q_+ = -\left(\frac{k_+ - \mu/\Upsilon}{k_- - \mu + 2q_-}\right) q_-$

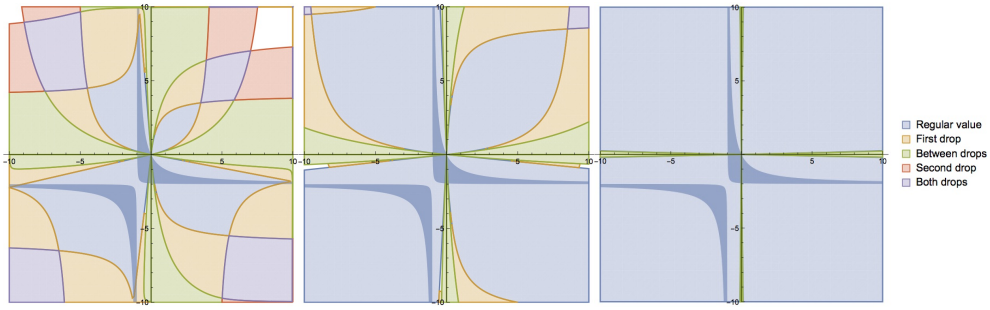


(g) $q_+ = -\left(\frac{k_+ + \mu}{k_- + \mu/\Upsilon}\right) q_-$



(h) $q_+ = -\left(\frac{k_+ - \mu}{k_- - \mu/\Upsilon + 2q_-}\right) q_-$

Figure C.4: The eight curves denoting the onset of the effects of the rapidity cutoff, with their equations given below. Curves are in red. $k_+ = 1$, $k_- = 2$, $\mu = 0.05$, and $\Upsilon = 10$.



(a) $\Upsilon = 10$ (b) $\Upsilon = 20$ (c) $\Upsilon = 100$

Figure C.5: The effect of changing the Fermi velocity cutoff on the region where the integrand for the loop integral is modified. $k_+ = 1$, $k_- = 2$, and $\mu = 0.05$ for each of the plots, while Υ takes on different values.

the origin is approximately constant for small values of μ, so we expect that the difference caused by the cutoff in this region will become smaller and smaller as the regions shrink. For the regions along the legs, Im $\tilde{\chi}$ drops of like $1/|q_b|$ along the legs of the hyperbolic region, and the width of the hyperbolic region also shrinks like $1/|q_b|$. Therefore the contribution from these regions should also shrink as we push them out to larger values of q.

Numerical analysis (Table C.1) indicates that that the contribution to Im Σ from these regions is bounded and shrinks with increasing Υ, and hence that Im Σ converges to the results from [63] in the limit $\Upsilon \to \infty$. Im Σ_Υ indicates the value of Im Σ with a cutoff, while Im Σ is the value without a cutoff. Fig. C.6 shows that Im Σ_Υ is still linear in $\varepsilon(k)$ even when one of the components of k is near the cutoff. The presence of a finite cutoff merely alters the slope of the line. Finally, Fig. C.7 shows how this dependence on the Fermi velocity cutoff modifies the expression for the imaginary part of the self-energy for large values of the momentum along the collinear direction.

k_+	k_-	μ	Υ	Im Σ_Υ	Im Σ
0.5	1	0.05	10	1.344	1.438
0.5	1	0.05	100	1.431	1.438
0.5	1	0.05	1000	1.437	1.438
0.5	1	0.05	10000	1.438	1.438
0.5	1	0.05	100000	1.438	1.438
0.1	5	0.05	10	1.030	1.438
0.1	5	0.05	100	1.407	1.438
0.1	5	0.05	1000	1.436	1.438
0.1	5	0.05	10000	1.438	1.438
0.1	5	0.05	100000	1.438	1.438
0.02	5	0.01	10	0.206	0.287
0.02	5	0.01	100	0.281	0.287
0.02	5	0.01	1000	0.287	0.287
0.02	5	0.01	10000	0.287	0.287
0.02	5	0.01	100000	0.287	0.287

Table C.1: Numerical analysis of the effects of a cutoff on Im Σ. Im Σ_Υ converges to Im Σ as we increase Υ.

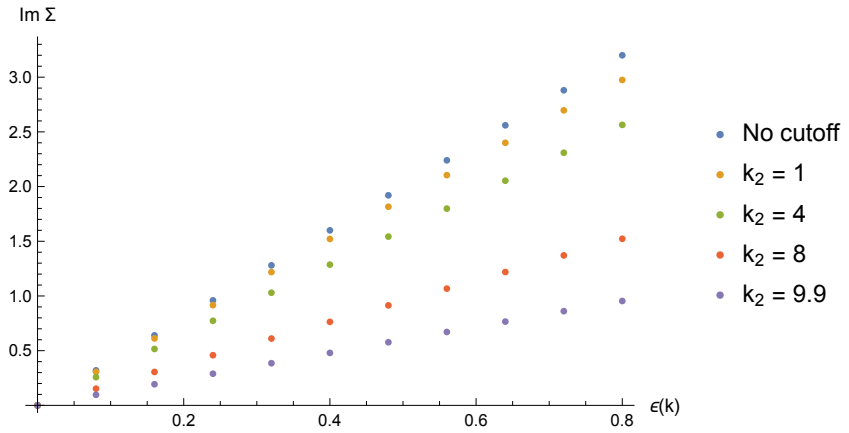


Figure C.6: Plot demonstrating that Im Σ_Υ is linear in $\varepsilon(k)$ for various values of the large component of k. $\mu = 0$, $\Upsilon = 10$.

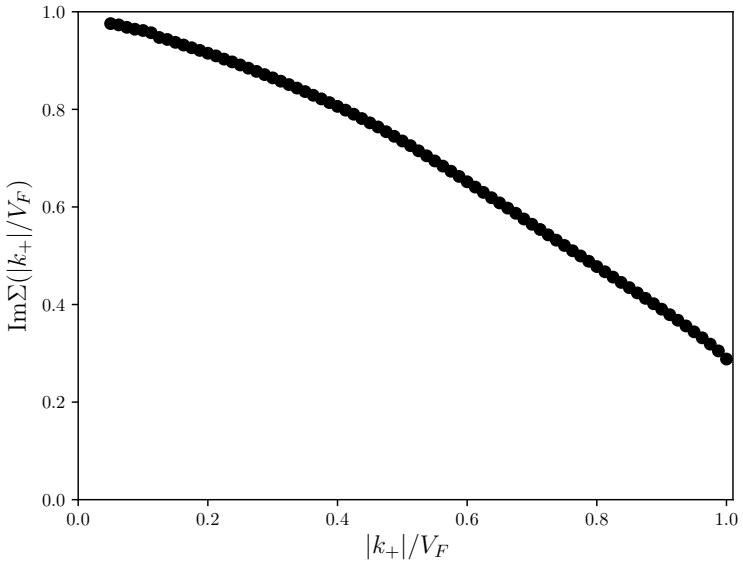


Figure C.7: Numerical results for the dependence of the imaginary part of the self-energy on $|k_+|/V_F$ in units of g^2 assuming a constant coupling. Im Σ is normalized to one for $V_F \to \infty$.