RENORMALIZATION FOR INSULATING STATES OF MATTER

BY

SEUNGMIN HONG

DISSERATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Doctoral Committee:

Professor Michael Stone, Chair
Professor Philip W. Phillips, Director of Research
Professor S. Lance Cooper
Professor Rob Leigh
Abstract

In this thesis, we study three cases of insulating states of matter in terms of the renormalization procedure where the conventional renormalization group scheme is not simply applicable.

The first subject is the spectral weight structure of hole-doped Mott insulators. As the mixing between two separate Hubbard bands is dynamically generated, additional charge degrees of freedom is required to give a proper description to the relevant low-energy physics. On this account, we first discuss how the low-energy Hubbard band should be partitioned to account for the extra degrees of freedom. Following the exact integration procedure of the upper Hubbard band, we explicitly demonstrate that the conserved charge cannot be exhausted by counting quasiparticles. In addition, we argue that it is the existence of dynamically generated charge degrees of freedom that gives rise to the coexistence of poles and zeroes in the single-particle Green function. In comparison to the Fermi arc structure, which is intrinsic to cuprate phenomenology, we suggest that the suppression of the spectral weight at the back side of the arc is a consequence of composite excitations, arising from dynamical mixing.

The second topic we study is the nature of the transition between two insulating states of matter in a weakly disordered bosonic system. In particular, we investigate the instabilities of the Mott-insulating phase within a renormalization group analysis of the replica field theory obtained by a strong-coupling expansion around the atomic limit. To this end, we identify a new order parameter and associated correlation length scale that are capable of capturing the transition from a state with zero compressibility, the Mott insulator, to another insulating state with finite compressibility, the Bose glass. The order parameter is the relative variance of the disorder-induced mass distribution. From its distinctive behavior on each phase, we find that the divergence of the relative variance in the Bose glass signals the breakdown of self-averaging. Lastly, we also emphasize that the transition at commensurate filling is governed by a different fixed point in the renormalization group flow.

The last topic discussed is the thermoelectric properties of the correlated Kondo lattice system at low temperature. Specifically, we explore the periodic Anderson impurity model with spin-orbit interactions, using the slave-particle method. As the thermopower is related to the entropy per carrier, particle-hole
asymmetry and a large density of states at the chemical potential make Kondo insulator an ideal platform for thermoelectrics. In this respect, we further explore the possibility to improve the thermoelectric performance by tuning the crystal electric field splitting, the spin-orbit interaction strength, and orbital degeneracy.
To my family.
First and foremost, I have been extremely fortunate to have had Philip Phillips as my advisor and collaborator. Especially in the early years, I struggled with difficult problems, but it was his encouragement and insight that have guided me well and kept me from being lost. Definitely, this dissertation would not have been possible without his guidance and patience over the past few years.

I would also like to thank all of my collaborators: Shiladitya Chakraborty, Frank Küger, Pouyan Ghaemi and Ka-Wai Lo for many illuminating conversations. It was Shiladitya who shared brilliant ideas about the Mott problem that eventually helped me to shape my own research project. I also learned a lot about the Bose glass in a short amount of time from the collaboration with Frank. I also thank Pouyan and who spent many days with me in developing an appropriate model for Kondo-lattice systems. I would also like to thank Ka-Wai who has been patiently working on transport calculations while I have been preparing this thesis. I am also indebted to fellow group members, Ting-Pong Choy for helping me understand the $2e$-boson theory, Jiansheng Wu for introducing me to the Bose glass problem, and Weicheng Lv for teaching me a lot about orbital-ordering in iron-pnictides.

Certainly, my research at Illinois would not have been possible without the support of particular staff members, John Stack, S. Lance Cooper, Wendy Wimmer and Melodee Schweighart. I also would like to thank my thesis committee members, Michael Stone, S. Lance Cooper, and Rob Leigh. My fellow graduate students have been an invaluable source of support and inspiration. I am especially thankful to Ching-Kai Chu, Kiaran Dave, Wade DeGottardi, John Nichol, Tomoki Ozawa, Xianhao Xin, Lei Xing, and Jitong Yu. I also wish to mention my Korean friends, Young Il Joe, Ki Jeong Yim, Woojin Kwon, Kyungsuk Lee, Changmo Yang, Hyeongjoon Kim, Hyeongjin Ma, Minjung Kim, Joonho Jang and Jaseung Ku, who have been putting up with me during my graduate years.

I would also like to acknowledge financial support from the National Science Foundation under Grants No. DMR-0940992 and No. DMR-1104909, the Center for Emergent Superconductivity, a DOE Energy Frontier Research Center under Award No. DE-AC0298CH1088, and the Samsung scholarship.

Last but not least, I would like to show my sincere gratitude to my family, for their unconditional love
and support over the past few years. No words can properly describe how thankful I am. I just wish to dedicate this thesis to them.
# Table of Contents

List of Figures ......................................................... ix

List of Abbreviations .................................................. xii

Chapter 1 Introduction .................................................. 1
  1.1 Renormalization Group ........................................... 1
      1.1.1 Scaling and Universality .................................. 1
      1.1.2 Renormalization group ................................... 3
      1.1.3 Application ............................................. 5
  1.2 Insulating States .................................................. 7
      1.2.1 Mott insulators .......................................... 8
      1.2.2 Bose glass ............................................. 9
      1.2.3 Kondo insulator ....................................... 11
  1.3 Outline .......................................................... 12

Chapter 2 Non-conservation of Fermionic Degrees of Freedom at Low-energy in Doped Mott Insulators .................................................. 14
  2.1 Introduction ..................................................... 14
  2.2 Experimental Motivation ......................................... 14
      2.2.1 Hall coefficient ......................................... 15
      2.2.2 X-ray Absorption and Optical conductivity ............... 16
  2.3 Redefinition of Chemical Potential .............................. 18
  2.4 Confirmation from Exact Low-Energy Theory ....................... 23
  2.5 Concluding Remarks ........................................... 24

Chapter 3 Towards the Standard Model for Fermi Arcs from a Wilsonian Reduction of the Hubbard Model ........................................... 26
  3.1 Introduction ..................................................... 26
  3.2 Charge 2e boson theory ......................................... 28
      3.2.1 Preliminaries ........................................... 28
      3.2.2 Exact Results: The conserved charge and the low energy mode ........................................... 30
      3.2.3 Green function and the approximations ................. 34
      3.2.4 Free-energy Minimum approach to 2e boson ............ 36
  3.3 Final Remarks .................................................. 40

Chapter 4 Two distinct insulating transitions and breakdown of self averaging in the disordered Bose-Hubbard model .................................... 42
  4.1 Introduction ..................................................... 42
  4.2 Breakdown of Self-Averaging .................................... 44
  4.3 Effective Replica Theory ......................................... 46
      4.3.1 Clean system ........................................... 46
      4.3.2 On-site disorder ....................................... 49
### Chapter 4 Renormalization Group Analysis

4.4 Renormalization Group Analysis

4.4.1 Renormalization Group Equations

4.4.2 New Variables

4.5 Results

4.5.1 Numerical Integration

4.5.2 Universality of the MI/BG Transition at Incommensurate Filling

4.5.3 Commensurate Filling

4.6 Final Remarks

### Chapter 5 Tuning thermoelectric power factor by crystal-field and spin-orbit couplings in Kondo lattice materials

5.1 introduction

5.2 Model and methodology

5.3 results

5.3.1 Nearly free electron itinerant bands

5.3.2 Tight binding itinerant electron bands

5.3.3 Effect of multi-orbital degeneracy

5.4 Summary and Remarks

### References
List of Figures

1.1 The energy spectrum of the one-band Hubbard model at half-filling. The higher energy band (upper Hubbard band, UHB for short) and the lower energy band (lower Hubbard band, LHB) are separated by the Coulomb energy scale, $U$, which represents the charge excitation gap. .......................... 9

1.2 The phase diagram of the $D$-dimensional Bose-Hubbard model at zero temperature, in the absence of disorder. In the Mott insulating state, the density of bosons per site, $m$, is fixed to a commensurate number. Either by adjusting the chemical potential, $\mu$, or by pushing the hopping matrix element, $t$, beyond a threshold, one can reach a superfluid phase, with the usual off-diagonal long range order. The superfluid density is one physical observable that becomes finite on the SF side. ................................................................. 10

1.3 Possible phase diagrams of the $D$-dimensional disordered Bose-Hubbard model at zero temperature. Here, $\Delta$ is the upper limit of weak bound disorder, $\mu$ is the chemical potential, and $J/V$ the ratio between the tunneling matrix, $J$ and the Coulomb repulsion, $V$. Whether (a) BG phases always intervene, or (b) there could be a special point of direct transition, remains unanswered. From Ref. [1]. ................................................................. 11

1.4 (a) The itinerant electrons (blue curve) and local moments (straight purple line) are decoupled, at high temperature. (b) At low temperatures, the flat local moment band hybridize with the conduction electrons to form a narrow gap near the Fermi surface. Since the hybridized bands strongly bear the character of the previous local electrons, its mass is highly enhanced near the Fermi energy $\epsilon_F$. ................................................................. 12

2.1 (a) $T$ dependence of $1/n_{\text{eff}}(=eR_H/V_{\text{Cu}})$ for a series of LSCO in underdoped region, $x = 0.01-0.05$, taken from the Ref. [2]. (b) The pseudogap crossover temperature $T^*(x) \simeq -\Delta(x)/\ln x$ obtained from the two fluid model, taken from the Ref. [3]. ................................................................. 15

2.2 The fluorescence yield at the O K-edge of LSCO, with the background subtraction, taken from the Ref. [4]. The absorption features were fit using two Gaussian shapes. ................................................................. 16

2.3 The integrated optical conductivity for a hole-doped (LSCO) and an electron-doped (PCCO) cuprates, plotted as a function of external doping $x$, taken from the Ref. [5]. The dashed line indicates the anticipated carrier density in the case of semiconductor upon doping. ................................................................. 17

2.4 Redistribution of spectral weight in the Hubbard model upon doping the insulating state with $x$ holes. $\alpha$ is the dynamical correction mediated by the doubly occupied sector. To order $t/U$, this correction worked out by Harris and Lange [6]. (a) The traditional approach [7, 8, 9] in which the occupied part of the lower band is fixed to the electron filling $1-x$. (b) New assignment of the spectral weight in terms of dynamically generated charge carriers. In this picture, the weight of the empty part of the LHB per spin is the effective doping level, $x' = x + \alpha$. ................................................................. 20

2.5 Integrated spectral weight in the occupied part of the lower Hubbard band, $\Lambda_{\text{LH}}$, from the effective low-energy theory [10, 11, 12, 13] with $U/t = 8$. Here $x$ is the doping level for the conserved charge, $Q = \sum_{i\sigma} a_i^{\dagger} a_{i\sigma} + 2t^2 \sum_i \varphi_i \varphi_i$. Clearly shown is that the occupied part (red triangles) of the one-particle spectrum has a weight less than $1-x$ (solid blue line). ................................................................. 23
3.1 The disconnected Fermi arcs have been confirmed from ARPES for several cuprates: (a) Ca$_2$-$_x$Na$_x$CuO$_2$Cl$_2$ (Ref. [14]), (b) La$_{2-y}$Ba$_y$CuO$_4$ at $x = 1/8$ (Ref. [15]) (c) ortho-II YBa$_2$Cu$_3$O$_{6.5}$ (Ref. [19]).

3.2 Free energy minimization and probability distribution of charge 2e boson. (a) Plotted here is the $\Delta F = F(\varphi_0^e\mathbf{q}, T) - F(\varphi_0^0)$ as a function of momentum $\mathbf{q}$. $\varphi_0^e$ is the magnitude of the charge 2e boson. As is evident, the free energy is strongly dependent on momentum. As $\varphi_0^e$ increases, the minimum of the free energy shifts to $(\pi, \pi)$. This corresponds to a sign change for the bosonic excitation around a plaquette. (b) Shown here is the probability distribution for $\varphi_0^e$ evaluated from the IR action by the relation, $P(|\varphi|) = 1/Z^\text{IR}_0 \int D\{c, c^\dagger\} e^{-S^\text{IR}}$. The maximum occurs at $\varphi_0^e \approx 1.2$ where the corresponding momentum that minimizes the free energy is $\mathbf{q} = (\pi, \pi)$. 

3.3 The low energy band dispersion along high symmetry directions for (a) a homogeneous configuration, $\varphi_1 = \varphi_0$, (b) a staggered configuration ($\varphi_1 = |\varphi_0|e^{i\pi r}$) evaluated by Eq. (3.15a), and (c) a staggered one evaluated by Eq. (3.15a). Here, we take the broadening factor $\gamma = 0.025t$ for a typical value $U/t = 10$, $t'/t = -0.3$, $t''/t = 0.1$, and the bare hole doping level as $x = 0.12$. In the evaluation, the parameters, $\alpha = x' - x$ are taken from the numerical estimates of the number of double occupancies from Liebsch and Tong. 

3.4 The spectral function of the low energy theory for each hole doping level, (a) $x = 0.05$, (b) $x = 0.08$, (c) $x = 0.12$, and (d) $x = 0.18$. As noted in the text, the “pocket” size is proportional to each doping level. In the present figures, an incoherent background is removed and the intensities of the spectral function are normalized for the first quadrant of the full Brillouin zone. 

4.1 (a) Phase boundary between the MI ($m = 1$) and the BG obtained from numerical integration of the RG equations. The disorder corresponds to site energies increased or decreased by $\delta = \Delta/U = 0.2$ with probability $p = 0.1$. The solid line shows the unphysical mean-field phase boundary ($\bar{r} = 0$) between the MI and SF. (b) Extraction of the correlation length $l_0$ and $G$ corresponding (in the limit of large mass $\bar{r}$) to the inverse mean and the relative variance of the mass distribution, respectively.

5.1 Density of states (DOS) of the parabolic model for different control parameters. The red-dashed lines are for the orbital $l = 1$, the dotted blue lines $l = 2$, and the green lines are for the total DOS. The central gray area displays the thermal window($\sim k_B T$) for each temperature. Figures (a)-(d) compare the DOS for different temperatures. Figures (e)-(h) correspond to different magnitudes of the crystal electric field($\Delta_{\text{CEF}}$) which breaks the degeneracy of the two $f$ orbitals. Figures (i)-(l) correspond to different magnitudes of the spin-orbit interaction($\gamma_{\text{SO}}$) that breaks the degeneracy of the conduction bands with different helicity.
5.5 Transport coefficients for each temperature: thermopower $S$ (left column), conductivity $\sigma$ (middle), and power factor $Z_{PF}$ (right). From the top to the bottom, the temperature is fixed to 25, 15, 10, and 5K, respectively.

5.6 Transport properties are compared for each SOI: $S$ (left column), conductivity $\sigma$ (middle), and power factor $Z_{PF}$ (right). The spin-orbit couplings are chosen to be 0.0, 0.1, 0.2, and 0.5 in units of the hoping amplitude $t_{hop}$, respectively.

5.7 Density of states for bare conduction electrons: quadratic dispersion (top), 3-D tight binding (bottom), helicity $h = -1$ (left, blue curves) and $h = 1$ (right, red lines). For the quadratic dispersion, the SOI is taken to be linear in momentum, as $\gamma_{so}|k|$. The changes in the bandwidth are exactly proportional to $\gamma_{so}$. In figures (c) and (d), the SO term, $\gamma_{so}\Gamma_{k}$ is taken in accordance with the cubic point group symmetry, and the next nearest neighbor hopping parameter $q_{2} = 0.3$. Note that the bandwidths are not drastically affected by the SO, while the shapes become more asymmetric with the increase in $\gamma_{so}$.

5.8 Density of states per orbital for different number of orbitals ($l_{max}$), where the chemical potential at each case is adjusted to the center. The inset displays the DOS without the adjustment.

5.9 Transport properties for $l_{max}$ orbitals: (a) Seebeck coefficient, (b) electrical conductivity, (c) thermal conductivity, (d) thermopower (e) figure of merit and (f) dimensionless figure of merit.
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACL</td>
<td>Algebraic charge liquid.</td>
</tr>
<tr>
<td>ARPES</td>
<td>Angle-resolved photoemission spectroscopy.</td>
</tr>
<tr>
<td>BG</td>
<td>Bose glass.</td>
</tr>
<tr>
<td>DOS</td>
<td>Density of states.</td>
</tr>
<tr>
<td>IR</td>
<td>Infrared.</td>
</tr>
<tr>
<td>KI</td>
<td>Kondo insulator.</td>
</tr>
<tr>
<td>LSCO</td>
<td>$\text{La}_{2-x}\text{Sr}<em>x\text{CuO}</em>{4+\delta}$.</td>
</tr>
<tr>
<td>LHB</td>
<td>Lower Hubbard band.</td>
</tr>
<tr>
<td>MI</td>
<td>Mott insulator.</td>
</tr>
<tr>
<td>RG</td>
<td>Renormalization group.</td>
</tr>
<tr>
<td>SDRG</td>
<td>Strong-disorder renormalization group.</td>
</tr>
<tr>
<td>SF</td>
<td>Superfluid.</td>
</tr>
<tr>
<td>UHB</td>
<td>Upper Hubbard band.</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet.</td>
</tr>
<tr>
<td>XAS</td>
<td>X-ray absorption spectroscopy.</td>
</tr>
<tr>
<td>YBCO</td>
<td>$\text{YBa}_2\text{Cu}<em>3\text{O}</em>{6+x}$.</td>
</tr>
<tr>
<td>YRZ</td>
<td>Yang-Rice-Zhang.</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

In this thesis, we focus on the three insulating states of matter in terms of the renormalization procedure. The first topic is the spectral weight structure of hole doped Mott insulator as a consequence of the dynamical mixing between two separate Hubbard bands \[18, 19\]. The second subject is the nature of transition between two distinct insulating states, namely Mott insulator (MI) to Bose glass (BG) transition in the disordered Bose-Hubbard system \[20\]. The last issue is the thermal transport properties of multi-orbital Kondo insulator with spin-orbit interactions \[21\]. Since the conventional renormalization group (RG) approach \[22, 23, 24\] is not simply applicable to those strongly correlated systems, we consider alternative renormalization methods to explore some aspects of systems with strong repulsive interaction.

1.1 Renormalization Group

Before elaborating on the details of each subject, it is useful to briefly review the conceptual foundations of the renormalization group method and its applicability. Since its formal development \[22\], the method of the renormalization group (RG) has served as one of the most powerful tools to study many-particle interacting systems, particularly when plain perturbative methods are either misleading or uncontrollable. Simply speaking, the crux of the renormalization group method is to reduce a minimal set of recursive relations between the parameters that faithfully define a certain interacting system at each length, or equivalently energy scale. The underlying physical origin that allows such process is certain common scaling relations between the defining parameters at each scale.

1.1.1 Scaling and Universality

In most complex systems that have a large number interacting degrees of freedom, it is not even possible to keep track of all of the constituents, nor practical to do so. Fortunately, it is often the case that one can manage to express such systems with a relatively small number of degrees of freedom and their characteristic parameters, when a specific length scale or energy range is of interest, mostly in the long-wavelength limit.
Even if a few essential physical parameters are successfully isolated, however, it is not apparent whether it is possible to find an exact solution or to apply a perturbative method in a systematic way. Thus, if we wish to uncover the physics of a given system, it is desired to turn to more general methods or notions.

One of the crucial concepts that provides a detour around such difficulties is the correlation function, and its associated length scale, namely the correlation length, $\xi$. The correlation function is formally defined as

$$\begin{align*}
C(x - x') &= \left\langle \left( \mathcal{O}(x) - \langle \mathcal{O}(x) \rangle \right) \left( \mathcal{O}(x') - \langle \mathcal{O}(x') \rangle \right) \right\rangle \\
&= \langle \mathcal{O}(x) \mathcal{O}(x') \rangle - \langle \mathcal{O}(x) \rangle \langle \mathcal{O}(x') \rangle,
\end{align*}$$

(1.1)

where $\mathcal{O}(x)$ represents general observables derived from microscopic degrees of freedom at a given space-time coordinate $x$, and $\langle \cdots \rangle$ denotes the thermodynamic average. Since $\mathcal{O}(x) - \langle \mathcal{O}(x) \rangle$ describes the deviation of the observable $\mathcal{O}$ from its average, these functions give a direct measure to determine how the fluctuations are correlated over the distance $|x - x'|$. Under the normal conditions, or in definite phases of matter, these correlation functions typically display exponential decay over the large distance, $C(r) \sim e^{-|r|/\xi}$. This asymptotic behavior enables us to extract the relevant length scale, so-called correlation length $\xi$. Physically, the correlation length can be seen as the distance over which fluctuations between the microscopic degrees of freedom remain significant. In other words, any sub-system of size comparable to the correlation length can be seen as an independent minimal object that bears exact resemblance to the whole macroscopic system.

When the system undergoes a transition between competing phases in a continuous fashion, the correlation length becomes divergent in general $^1$. This means that the fluctuations are correlated in all length scales. As a result, the entire system should be treated as a single critical block, which is the reason such states of matter are dubbed as being critical. One important consequence of such strong fluctuations close to the criticality is that it becomes effective enough to wash out the microscopic details of the interactions. In effect, what remains crucial is general classifications of criticality.

Typically, near critical points, the singularity of the correlation length is power-law like, $\xi \sim |t|^{-\nu}$. Here, a dimensionless quantity $t$, reduced from the external control parameters, is a measure of the distance away from the critical point and $\nu$ is a positive critical exponent. The occurrence of such singularity in the correlation length also leads to a power-law like divergence in other thermodynamic observables such as the heat capacity, order parameter, susceptibility, compressibility and the like. As the microscopic details are irrelevant near a critical point, so are the other exponents of those macroscopic quantities. Interestingly, in spite of a variation of the system parameters and configurations, there are only a few classes that share the

$^1$At a first order or discontinuous phase transition, the correlation length generally remain finite.
same set of exponents of the thermodynamic observables, namely the *universality class*. A fundamental account for such classifications comes from the renormalization group method.

Indeed, the scaling hypothesis, first suggested by Widom [25], serves as a conceptual bridge between the universality and the formal renormalization group method. The central idea is that, close to criticality, the correlation length $\xi$ is the only relevant length scale. Since $\xi$ diverges exactly at a critical point, the system should look “similar” regardless of the level of the resolution. Hence, if one rescales overall length of the system by any arbitrary factor $b$ as $x \rightarrow bx$, the rest of the characteristic parameters should be accordingly scaled to meet the constraint. More specifically, upon an arbitrary scaling of the length by $b$ ($b > 1$), the singular part of the free energy or similarly any correlation function, that encodes the physical properties of a given system can be expressed in a simple form

$$f_s(t, h) = b^{-d} f_s(t b^{1/\nu}, h b^{\eta h}),$$

(1.2)

where the factor $b^{-d}$ is due to the extensive nature of the free energy in $d$ dimension, and the dimensionless variable $h$ denotes another external parameter. Since singular behavior of many other macroscopic observables are derived from $f_s$, it is also possible to obtain relations between critical exponents.

### 1.1.2 Renormalization group

The general idea of the renormalization group method is to trace the changes of parameters at each length(energy) scale as short-range(high-energy) fluctuations are integrated successively. In carrying out the integration over the short-range disturbances(or high-lying energy sectors), one can use several known mathematical procedures which depend on the configurations of the system at hand. Using any of these iterative trace procedures usually generates new parameters that were not present in the bare configuration; hence the flow should be monitored for all possible coupling parameters, $\{g_a\}$. On the other hand, if the low energy degrees of freedom remain invariant after the partial traces, it can be used to get a formal relation

---

2What decide the actual value of exponents are the dimensionality, the geometry, and the underlying symmetry of the system.

3In some instances, the microscopic cut-off such as lattice spacing between electronic orbitals may come into play an important role in the scaling relation.

4In the real space, a common intuitive way is to perform block spin transformation through decimation. However, it is often more tractable to perform the integration in momentum or frequency space, by gradually changing the cut-off of the momentum(energy) range. Sometimes, in order to handle divergences out of the integrations, one may treat the dimension of the system to be an independent variable, from which a small expansion parameter can be obtained, enabling systematic control of divergences. An alternative RG method, though sounds esoteric, is the operation expansion method. The basic idea is that any product of scaling operators can be put into a linear combination of scaling operators. Once the appropriate coefficient functions are found, from the expansion of the free energy functional, one can easily retreat the relations between coupling coefficients, that is, the RG equations. Whatever the methods are chosen, the ultimate flow of coupling parameters should be identical.

5Whether the low energy degrees of freedom may be changed or not, the partition function of the system should be preserved. It may not be the case, however, for the effective actions or Hamiltonians.
between parameters,
\[ \{ g'_a \} = \mathcal{R}_{ab} (\{ g_b \}). \] (1.3)

The above mapping is nothing but the renowned renormalization group equations that relate the parameters at each energy level of description.

The RG analysis is advantageous in several aspects. It not only enables us to classify the states of the system with a few set of numbers, it also allows us to discover governing interactions at each phase. For example, the characterization and the classification of the state is essentially associated with the flow of parameters under the RG transformation. In the abstract space of complete coupling constants, \( \{ g_a \} \), it is a common practice to search for fixed points where the set of values \( \{ g^*_a \} \) no longer changes under the transformation. Indeed, if the system is invariant under subsequent scale transformations at such points, there are only two possible interpretations. One is that the corresponding state is either at criticality at which scale invariance emerges. Alternatively, the system is in a state with no internal correlations such that any arbitrary subdivisions of the system are essentially identical. In the former case, any small deviation tends to be amplified through the recursive procedure until it is driven to another fixed point. Thus, such points can be viewed as representing the unstable and critical states of the system. In contrast, the other points, corresponding to the later interpretation, are stable as it works in a complementary way. Hence, any region in the parameter space \( \{ g_a \} \) that is attracted to either of these point can be considered essentially in the same phase.

Close to the fixed points, on the other hand, one may further suppose the mapping can be linearized as
\[ g'_a - g^*_a \approx \frac{\partial \mathcal{R}_a}{\partial g_b} \bigg|_{g=g^*} \cdot (\{ g_b - g^*_b \}). \] (1.4)

If we can diagonalize the above linear equations such that \( u'_a = b^{y_a} u_a \) for \( u_a = \sum_b c_b (g_b - g^*_b) \), the nature of fixed points can be easily decided. If \( y_a \) is positive near some point, the new scaling variable \( u_a \) gets more relevant under the successive scale transformation, while it becomes irrelevant for \( y_a < 0 \). In case \( y_a = 0 \), the nature of the scale variable seems to be invariant under the equation flow, at least in the linearized approximation. This is known as marginal. From this seemingly innocent transformation, we can indeed identify the relevant interactions that underpin a particular state of the system.

Finally, the RG machinery also provides a natural account to the homogeneity relation in the singular part of the free energy. Since the partition function is preserved regardless of any mathematical trick, the relation between the free energy at different scales can be derived as,
\[ f_s (\{ g_a \}) = f_n (\{ g'_a \}) + b^{-d} f_s (\{ g'_a \}) \text{ in } d \]
dimension. Aside from its trivial analytical part $f_n$, the singular part of the free energy can be re-expressed in terms of the previously introduced scaling variables $\{u_a\}$ sufficiently close to criticality. Hence, one obtains the advertised scaling relation, $f_s(\{u_a\}) = b^{-d} f_s(\{u'_a\}) = b^{-d} f_s(\{b^{u_a} u_a\})$.

### 1.1.3 Application

Given the general renormalization group steps, it is tempting to apply the strategy to any many-particle interacting system. What we have to keep in mind is that the RG picture is not an omnipotent key that can unlock all the nature of problems at hand. Nonetheless, there have been many successful applications of the RG framework. Here, we give three examples.

**Fermi-liquids**

One important contribution achieved by the RG account is the rigidity of the quasiparticle picture in many weakly interacting electronic systems [26, 27]. In noninteracting spin-$1/2$ fermionic systems, the ground state is obtained by filling all the states below certain energy $\varepsilon_F$, or the Fermi energy. Then, the Fermi surface is defined as $\epsilon(p) = \varepsilon_F$ for some free particle dispersion $\epsilon(p)$. Now if there exists a well-defined surface such that the filled and the empty states are clearly separable, any low energy excitations occur in the vicinity of the Fermi surface either by adding a state outside (particle) or removing a particle inside (hole). Then, as was delineated by Polchinski [26], a scaling transformation should be accomplished with respect to the Fermi surface, since a naive scaling of the momentum can spoil the existence of the Fermi surface itself. The reason can be seen when the momentum is separated into to two components: one along the Fermi surface, and the other orthogonal to it. Then, it is evident that only the orthogonal part should scale under a transformation to preserve the presence of the Fermi surface.

From further dimensional analysis, it was then demonstrated that the existence of a Fermi surface is self-consistent as the mass term in free field theories stands as a relevant coupling. When additional weak scatterings are present such as a repulsive quartic interaction, it was also proven that none of the weak interactions can be relevant [6]. This immediately implies that, any low-energy excitations in a weakly interacting fermionic system can be understood to be essentially identical to excitations in the free system. In other words, the Fermi-liquid picture, or the notion of a quasiparticle is extremely robust regardless of the details of the microscopic interactions and the band structure. Lastly, it should be again stressed that the existence of a well-defined Fermi surface is the key ingredient for all those following scaling analysis. In

---

[6] When scattering between two particles are mediated by large momenta or by time reversal symmetry, such quartic interactions can be marginal. In some occasions, the marginality may bring an instability near the Fermi surface, which typically is a precursor to a phase transition, which includes a Cooper instability.
this thesis, we address the absence of a Fermi surface in Mott insulators and how to obtain a description of low-lying state \[28, 11, 12, 10, 18, 19\].

**Kondo problem**

Another remarkable, also historically important application of the RG framework is the Kondo problem \[29, 22, 30, 31, 27\]. The Kondo problem grew out of experimental observations that the resistivity of conventional metals with dilute magnetic impurities displays a logarithmic upturn below a certain temperature. A general model suggested by Anderson tried to incorporate a strong Coulomb repulsion in the magnetic impurities, while retaining the itinerant nature of electrons in the host metal. More deeply, the hybridization between local and free electrons allows a virtual charge fluctuation which mediates an exchange between the spin of an impurity with that of the itinerant electrons. Hence, within a specific parameter range, the Anderson model can be reduced to one with a spin-exchange interaction between local and itinerant electrons. This model can provide a natural account of the temperature scale where the upturn of the resistivity occurs, that is, at the onset of spin-singlet condensation.

However, the logarithmic divergence already indicates that the perturbative approach is breaking down. In the presence of the itinerant band, high energy fluctuations can be eliminated by gradually narrowing the conduction band, following the common strategy of the RG method. Keeping track of interactions that are allowed by the internal symmetry, \(SU(2)\), a recursive relation between the exchange interactions are obtained. The resultant trajectory of the exchange interaction parameters determines that the system tends toward strong negative interaction as the energy scaled is lowered \[30\]. In the current thesis, we explore the extended problem when more degrees of freedom are present, such as lattices of impurities, multi-orbitals, and spin-orbit interactions.

**Disordered system**

Speaking of impurities, it is not generally realistic to imagine a system to be translationally invariant or isotropic, as presence of randomness\[7\] is inevitable. In general, when disorder is introduced in a pure system that undergoes a phase transition, one may expect the critical temperature to be lowered since it tends to destroy ordering. Depending on the nature of the given system, it may smear the phase transition, or even ravish the ordering itself. In some cases, randomness in systems may bring a new distinct phase near the pure critical point. Insofar as the stability of the criticality in the clean system is concerned, there is a certain standard, namely the *Harris criterion* \[32\], which we will study more in-depth in the disordered

\[\text{7Here, we only consider quenched disorder, which means the positions of impurities are fixed over the relevant time scale.}\]
Bose-Hubbard system.

Methodologically, a pertinent question may arise whether the previously illustrated RG approach in a clean system is still suitable. One way to extend the RG approach is to reshape the problem to be similar to a homogeneous system. A common trick is the so-called replica method [33] which involves writing the average of the free energy, or equivalently $\ln Z$ with $Z$ the partition function, as $\overline{\ln Z} = \lim_{n \to 0} \frac{1}{n}(\overline{Z^n} - 1)$. Here, $\overline{O}$ stands for the quenched disorder average of an observable $O$. Then, the process of cumulant expansion to each replica partition function $Z^n$ reinstates translational invariance. The price of this replica approach is that it will put additional interactions between different replicas, but the advantage is that we can employ the formal perturbative RG method as in a pure system.

Another standard technique is the strong disorder renormalization group (SDRG) [34], which is particularly beneficial if a perturbative replica approach breaks down. The main idea is to find the highest local excitations and to iterate the integration of those local states until one reaches the low-energy region of interest. For instance, in a linear chain with random bond disorder, there is a local region where the bond interaction is the largest. By eliminating the degrees of freedom at this site, one can effectively lower the maximum energy scale, which is in line with the idea of formal RG. In some sense, the philosophy of SDRG is similar to the familiar block spin transformation, though it is enforced for a particular local region. Unlike in standard RG approach, what should be tracked in the SDRG is not the individual parameters, but rather the probability distributions of them. Thus, even if the randomness becomes infinitely large, it is possible to identify the ground state. For example, if only one particular distribution gets infinitely broader while the rests shrink down to a small number, we can say it is in a phase governed by the term with such parameter.

It is commonly believed that SDRG is “the” prescription, capable of attacking the strong-randomness fixed point. Since the normal replica method is perturbative in nature and is only focused on individual parameters, it may fail to make a correct prediction. In this thesis, however, we will develop a modified framework of the replica method so that it can be tailored to the strong-randomness problem.

1.2 Insulating States

With this background about the renormalization theory in general, we proceed to another theme of the current thesis, strong correlation and insulating states of matter. In the following, we will briefly introduce three cases of strongly correlated system: Mott insulator, Bose glass and Kondo insulator.
1.2.1 Mott insulators

The low-energy physics of strongly correlated electronic system is an intriguing subject, since the insight gained from noninteracting particles is not usually applicable. In a Fermi-liquid, the existence of a Fermi surface itself is quintessential, since all renormalizations resulting from repulsive interactions are toward the Fermi surface. Furthermore, as any short-range repulsive interactions remain irrelevant under renormalization, the quasiparticle picture acquires a self-consistent validity. However, the Mott insulator (MI), an insulator with an odd number of electrons per unit cell, poses a serious challenge to the Fermi liquid picture. According to band theory, a lattice system with an odd number of electrons has a partially filled conducting band, which is in contrast to the case of Mott insulators. In this type of insulator, in fact, it is the strong Coulomb repulsion between electrons that gives rise to the formation of the charge gap even at half-filling, by causing the electrons to avoid sitting on top of another [35, 36, 37]. In addition, it has been ubiquitously observed that electronic spectral weights of Mott insulators are strongly mixed across the entire energy domain [7, 38], unlike in a Fermi-liquid. This phenomenon implies that low energy excitations in Mott insulators cannot be sharply defined near a particular energy range. In other words, there is no clear Fermi surface in Mott insulators, which is necessary to perform a perturbative renormalization.

In order to understand Mott physics, it is essential to choose an appropriate model Hamiltonian that retains the key ingredients: kinetics, Coulomb repulsion and relevant orbitals. For the case of the cuprates, it is known that both the Cu $3d_{x^2-y^2}$ and the O $2p_x, 2p_y$ orbitals are participating in the formation of the charge gap [39], which suggests a three-band model for the low-energy electronic states. According to Zhang and Rice, however, a dilute mobile oxygen hole due to an external doping in the low-energy spectrum tends to strongly hybridize with the copper orbitals to form a singlet band, namely the Zhang-Rice singlet band [40]. These hybridized states allow us to ignore the details of the internal low energy states, by substituting them with an effective low-energy band. In other words, a single-orbital model can be thought to be sufficiently effective in describing the states of underdoped cuprates. This is often called the single-band Hubbard model, which was initially studied in the context of NiO, a transition metal oxide, by Mott [35, 36, 37].

In second quantized form, the single-band Hubbard Hamiltonian comprises two terms as

$$H_{\text{Hubb}} = - \sum_{(i,j),\sigma} t_{ij} \left( c_{i\sigma}^\dagger c_{j\sigma} + \text{h.c.} \right) + U \sum_i c_{i\uparrow}^\dagger c_{i\downarrow}^\dagger c_{i\downarrow} c_{i\uparrow},$$

(1.5)

where $c_{i\sigma}^\dagger$ is a creation operator, placing a fermion in the orbital localized at site $i$ with spin $\sigma$. Here, the

---

8 As widely believed, the Mott insulator is the parent state of the copper-oxide superconductors. Numerous experiments on the underdoped samples have produced inconsistencies with Fermi liquid theory [39]. Suppression of single particle density of states, namely the pseudogap, $T$–linear electronic resistivity, and the sign changes of Hall coefficients are a few examples of the failure of the weakly interacting picture.
kinetics is depicted by the transfer of electrons between orbitals that reduce the energy by $t_{ij}$, and the electronic repulsion is described by the *Hubbard* $U$ when two electrons with opposite spins placed on the same orbital. For most practical purposes, it suffices to assume that $t_{ij}$ is nonzero only for (next) nearest neighbors, $i$ and $j$, and that $U$ is point-like in real space. For positive values of $U$, then, the two terms compete one another. While the kinetic terms favor delocalization of electrons, the Coulomb repulsion tends to suppress charge fluctuation. As a result, in the strong-coupling limit $U \gg t$, it develops an insulating phase even at half-filling.

From the studies of the Hubbard model in the strong coupling limit, a couple of fingerprints of the strongly correlated electron physics have been found [41]. First, the retarded Green function, that measures the density of states for adding or removing a particle, revealed a well-separated two-pole structure. These poles allow us to identify the energy spectrum where precisely single-particle excitations occur. As illustrated in Fig. 1.1 there are two distinct bands, the higher energy band (upper Hubbard band, UHB for short) and the lower one (lower Hubbard band, LHB), which is separated roughly by the Coulomb repulsion $U$. In mathematical language, the formation of a charge gap is due to the divergence of the self-energy that not only disturbs the single coherent peak structure, but also causes the suppression of the single-particle density of states at zero energy, consequently forming a charge excitation gap. Further analyses of this model and its relatives have shown that these can also reproduce the rich correlated phenomena such as antiferromagnetism, pseudogap, d-wave pairing of superconductivity, charge and spin density ordering and possibly others [42].

### 1.2.2 Bose glass

The behavior of strongly interacting Bose gas systems subject to disorder is another attractive problem [1] that has drawn considerably more attention lately since it can be realized experimentally\(^9\) in optical lattices\(^4\)Other typical realizations include Josephson array, granular superconductors, flux lattices in type II superconductors, and \(^4\)He in porous.
In the absence of disorder, there are two competing phases: a Mott insulator when on-site repulsion is dominant enough to inhibit particle addition or removal, and a superfluid (SF) obtained when kinetic energy gain is beyond a certain threshold. The phase boundary between two pure phases is determined by the balance of kinetic energy gain and on-site repulsion (See Figure 1.2). Here, it is worth to compare it with the fermionic Hubbard system in the light of strong correlation. Since the off-diagonal long range order in the SF phase is the natural order parameter, it supplies a starting point for the RG study. On the other hand, if one approaches from the MI side, there is no sensible starting point, similar to the Mott insulating states in electronic systems.

In the disordered case, the fate of the Mott-insulating states is at stake as the strength of disorder increases. For example, if there happens to be a large region where disorder is strong enough to compensate the energy to create a particle(hole), then bosons are likely to be delocalized. As a consequence, the low energy excitations are essentially gapless quasiparticles and the system is compressible. Still because of the localization effect, randomness inhibits motion of a particle over long distances, so that such disordered systems should be insulating. This state of matter, an insulator with a finite compressibility, is called as Bose glass (BG).

Regarding the phase diagram, it was first argued by Fisher et al. [1] that this gapless insulating state, or the BG phase is likely to intervene between two pure phases, but it has been the subject of severe debate. Although it was not entirely excluded that commensurate filling could be an exceptional instance where the direct transition from SF to MI takes place (See Fig. 1.3(b)), there has been no general consensus so far. In

\[\text{Figure 1.2: The phase diagram of the } D\text{-dimensional Bose-Hubbard model at zero temperature, in the absence of disorder. In the Mott insulating state, the density of bosons per site, } m, \text{ is fixed to a commensurate number. Either by adjusting the chemical potential, } \mu, \text{ or by pushing the hopping matrix element, } t, \text{ beyond a threshold, one can reach a superfluid phase, with the usual off-diagonal long range order. The superfluid density is one physical observable that becomes finite on the SF side.}\]
a later chapter, we will study the nature of MI-BG transitions.

1.2.3 Kondo insulator

The Kondo lattice is another type of strongly correlated system where a dense array of local moments, typically arising from highly renormalized $f$-electrons, are interacting with the itinerant electrons [31, 49]. At high temperatures, since the local moments are extremely poorly screened, they behave as a normal conductor. On the other hand, at low temperatures, they are like a paramagnetic insulator with a narrow energy gap at the Fermi level, $\epsilon_F$, as shown in Figure 1.4. This insulating state, namely, Kondo insulator (KI) has been reported in several materials; Ce$_3$Bi$_4$Pt$_3$, CeRhAs, CeRhSb, CeRu$_4$Sn$_6$, U$_2$Ru$_2$Sn, and etc. Since the hybridized band is quite flat near the Fermi energy, the effective mass of the dressed fermion is much larger than the original conduction electron, which is the reason it has earned the title, heavy fermions. In addition, the large density of states (DOS) near the chemical potential is an indicator that formation of a many-body resonance is at the very heart of the Kondo insulator, as was in the single impurity Kondo problem.

As in other correlated systems, there is a widely used minimal model that contains the key physics of the Kondo insulator. It is the periodic Anderson model, which is a natural extension of the single impurity Anderson model (For further review see Ref. Hewson [31]). In contrast to the single impurity, the periodic Anderson model is not amenable to any exact techniques, hence only multiple mean-field type methods were used [50, 51, 52] to understand how the hybridization renormalizes the fermionic degrees of freedom. In this thesis, we will employ one of the slave-particle methods that separates spin and charge degrees of freedom from single electron, as it can capture the quasiparticle nature in the low energy.
1.3 Outline

This thesis is organized as follows.

In chapter 2, we discuss the existence of dynamically generated charge degrees of freedom in a doped Mott insulator, along with a few supporting experimental results. As it immediately indicates that the chemical potential in the low energy theory cannot be that of the bare electrons, we will focus on how the spectrum in the low-energy band should be partitioned to account for the extra degrees of freedom. In particular, we will argue that the exact low-energy theory of the one-band Hubbard model is an ideal theoretical tool to understand this spectral rearrangement.

In chapter 3, we consider a unconventional way of renormalization in the context of the hole-doped Hubbard model. In the first part, we will explicitly prove that the number of excitations at low-energies that are coupled to the electromagnetic gauge is less than the conserved charge, which supplements the argument of the previous chapter. Then, we proceed to show two charge $e$ excitations in the lower band, the standard projected electron and a composite entity, give rise to poles and zeros of the single-particle Green function. Then, we demonstrate that the poles generate spectral weight along an arc centered at $(\pi/2, \pi/2)$ while the zeros kill the spectral intensity on the back-side of the arc. The result is the Fermi arc structure intrinsic to cuprate phenomenology.

In chapter 4, we discuss the instabilities of the Mott-insulating phase of weakly disordered Bose-Hubbard model. To this end, we use a renormalization group analysis of the replica field theory obtained by a strong-coupling expansion around the atomic limit. By identifying a new order parameter and associated correlation length scale, we proceed to characterize the MI and BG phases. As will be shown, the divergence of the
relative variance of the disorder distribution indicates the breakdown of self-averaging in the BG phase. We also discuss the issue of the MI-BG transition at commensurate/incommensurate fillings.

In chapter 5, we study thermoelectric transport at low temperatures in correlated Kondo insulators, motivated by the recent observation of a high thermoelectric figure of merit (ZT) in FeSb$_2$. Based on the periodic Anderson model and a slave-particle renormalization prescription, we explore the possibility of improving the thermoelectric properties of correlated Kondo insulators through tuning of crystal field and spin-orbit coupling and present a framework to design more efficient low-temperature thermoelectrics based on our results.
Chapter 2

Non-conservation of Fermionic Degrees of Freedom at Low-energy in Doped Mott Insulators

2.1 Introduction

Based on an exact diagonalization study in their paper, Meinders et al. [7] concluded that because the effective number of low-energy degrees of freedom in a doped Mott insulator is a function of the hybridization and therefore the volume and temperature, “...it is not possible to define a Hamiltonian that describes the low-energy-scale physics unless one accepts an effective nonparticle conservation.” Particle non-conservation as used here refers to the fact that the number of low-energy degrees of freedom is not strictly determined by the electron filling or equivalently the doping level but rather by dynamical degrees of freedom generated from the hybridization and hence the temperature. If this statement is correct, then it must be the case that the chemical potential for the static fermionic low-energy degrees of freedom in any realistic model for a doped Mott insulator is not equivalent to that of the conserved charge, namely the bare electrons. Thus far, an explicit construction demonstrating this has not been advanced.

In this chapter, we directly address the question of how particle conservation breaks down in a low-energy theory of a doped Mott insulator. We first show that experiments on the Hall [53, 54, 2, 3] and optical [5, 55, 56] conductivities and general theoretical considerations support this claim. Finally, we propose a simple partitioning of the spectral weight in the lower Hubbard band (LHB) which isolates the explicit hybridization-dependent degrees of freedom that are responsible for the dynamical generation of charge degrees of freedom and hence effective particle non-conservation as defined above. We show that these degrees of freedom can be understood within the recently [10, 11, 12, 13, 57] derived exact low-energy theory of the Hubbard model.

2.2 Experimental Motivation

Is there any experimental indication in doped Mott systems that the number of charge carriers is dynamically generated? It would suffice to show that either 1) the carrier density is temperature dependent or 2) the
number of charge carriers exceeds the nominal doping level, hereafter referred to as $x$.

### 2.2.1 Hall coefficient

In ordinary metals, the Hall coefficient, $R_H$, can be used to determine the sign and the concentration of the charge carriers. Even though weak electron-scattering processes can affect the relaxation time and the curvature of the Fermi surface, the simple relation $R_H^{-1} = nev_F$ is still valid, where $n$ is the carrier density. Furthermore, it is remarkable that the Hall coefficient is independent of the temperature, since the relaxation rate is rather governed by the electron interactions. Hence, as long as the band structure does not become too complicated, the inverse Hall coefficient remains a useful tool to characterize the Fermi surface topology and the natural charge carriers.

In lightly hole doped compounds such as La$_{2-x}$Sr$_x$CuO$_4$ (LSCO) and YBa$_2$Cu$_3$O$_{7-x}$ (YBCO), prototypical cuprates, careful measurements revealed that $R_H$ is strongly temperature dependent [53, 54, 2, 58]. This behavior has been phenomenologically understood in the two-carrier model, where one of the two components is thermally activated. The empirical two-fluid model, set by Gor’kov and Teitel’baum [3], indeed, turns out to be fit remarkably well.

\[
    n_{\text{Hall}}(x, T) = n_0(x) + n_1(x)e^{-\Delta(x)/T}. \tag{2.1}
\]

In their expression, the second term, $n_1(x)e^{-\Delta(x)/T}$ is the only component that exhibits deviations depending on the temperature, associated with an energy gap $\Delta(x)$. Further analysis of the various LSCO data disclosed the crossover temperature to thermal activation is in good agreement with the line for the pseudogap temperature, $T^*$, below which a depletion of the single-particle density of states occurs without the
superconducting gap. In the context of strong correlations in a doped Mott insulator, this hints that the bound state component is going to be liberated beyond the pseudogap energy scale, to contribute to the effective charge carriers. This further implies that the charge carrier density is not statically determined, but is dynamically generated.

2.2.2 X-ray Absorption and Optical conductivity

Another evidence to the dynamic nature of effective charge carriers is gained from the X-ray oxygen K-edge experiments (XAS) [4, 59, 60, 61]. In these experiments, an incident X-ray beam promotes the core electrons in the 1s oxygen shell to the unoccupied O 2p orbitals. Even if the majority of the excited electron goes to the oxygen band, the Zhang-Rice singlet states [40], formed via an exchange interaction between Cu and O holes, plays a sort of mirror to the LHB. What was revealed through measurements are 1) there are a two distinct peaks, centered around 528.8 and 530.3 eV with the optical gap where the lower peak only emerges upon the introduction of external holes, $x$, and 2) the integrated spectral weight of the low-energy spectrum (Zhang-Rice band) was found to increase faster than $2x$ while the upper one (UHB) decreases faster than $1 - x$. This behavior is not without precedent, as it was also observed in the classic Mott system, Li-doped NiO [62].

In optical absorption experiments, a net transfer of spectral weight from the UHB to the low-energy band was observed [5, 56, 63], which is not expected in a rigid-band insulator or semiconductor. Once the optical conductivity, $\sigma(\omega)$ is obtained from the Kramers-Krönig relation to the reflectivity data, the effective charge carrier density can be calculated by integrating $\sigma(\omega)$ up to the optical gap $\Omega \approx 1.2$ eV,

$$N_{\text{eff}} = \frac{2mV_{\text{cell}}}{\pi e^2} \int_0^\Omega \sigma(\omega)d\omega.$$  (2.2)
Here, $V_{\text{cell}}$ is the unit-cell volume per Cu atoms, and $m$ is the free electron mass. This quantity is helpful in specifying the amount of the spectral weight transfer. As seen from the Fig. 5.4, it is obvious that there are extra contributions to the normalized carrier density, exceeding the nominal doping concentration. In the dilute doping regime, it is even greater than $2x$, which is in agreement with the XAS results.

These observations consistently indicate that the number of effective charge carriers is not simply exhausted by counting the external doping. Hence, the low-energy degrees of freedom in a doped Mott system certainly possess a dynamically generated entity (denoted as $\alpha$, hereafter), in addition to the chemical doping, $x$.

Consequently, experimental probes which couple to the current reveal that the number of charge carriers in the cuprates is 1) temperature dependent and 2) exceeds the nominal doping level, consistent with Meinders et al. [7]. An interesting question is how does one define the chemical potential for such dynamically generated charge degrees of freedom. Clearly it is not equal to that of the bare electrons as the the effective number of charge degrees of freedom exceeds the bare charge count. We argue below that the effective doping level that captures the dynamical generation of the charge degrees of freedom as in Eq. (2.1) is given by

$$x' = x + \alpha$$  \hspace{1cm} (2.3)

where $\alpha$ is a dynamical correction determined by the hybridization. This redefinition of the doping level naturally arises from the exact [10] [11] [12] [13] low-energy theory of the Hubbard model which has been shown to explain [10] [11] [57] both Eq. (2.1) and Eq. (2.2).
2.3 Redefinition of Chemical Potential

The goal in this section is to redefine the chemical potential so that the effective number of fermionic charge carriers is consistent with dynamical generation of charge degrees of freedom discussed in the previous section. In the standard theory of metals, the intensity or spectral weight of a band is completely exhausted by counting the number of electrons it can hold. That is, it is a constant given by one per unit cell and per spin direction. Essential to this view is the robustness of electron quasiparticles even in the presence of interactions. Because of the one-to-one correspondence between electrons and quasiparticles, the chemical potential, $\mu$, can be defined either by counting electrons

$$n = \int_{-\infty}^{\mu} N(\omega) d\omega, \quad (2.4)$$

or by integrating,

$$y = \int_{\mu}^{\Lambda_\rho} N(\omega) d\omega, \quad (2.5)$$

the unoccupied part of the spectrum. Here $N(\omega)$ is the single-particle electron density of states and $\Lambda_\rho$ is a cutoff demarcating the low-energy physics. As a result of the electron-quasiparticle correspondence, $y$ is identical to the number of doped holes, $x$, and the electron filling is given by $n = 2 - x$ (for a single band).

In stark contrast, the empty part of the spectrum per spin at low energies, Eq. (2.5), exceeds the doping level in strongly correlated systems such as doped Mott insulators. The inherent problem with strongly correlated systems is that the energy bands are not the traditional static bands that typify band insulator systems. This can be illustrated simply by considering the Hubbard model

$$H_{\text{Hubb}} = -t \sum_{\langle i,j \rangle, \sigma} c_{i\sigma}^\dagger c_{j\sigma} + U \sum_i c_{i\uparrow}^\dagger c_{i\downarrow}^\dagger c_{i\downarrow} c_{i\uparrow}, \quad (2.6)$$

in which electrons hop among a set of lattice sites, but pay an energy cost $U$ whenever they doubly occupy the same site. Here $i, j$ label lattice sites, $\langle i, j \rangle$ indicates nearest neighbors, $c_{i\sigma}$ annihilates an electron with spin $\sigma$ on site $i$ and $t$ is the nearest-neighbor hopping matrix element. When $t = 0$, the Hamiltonian is diagonal

$$H_U = U \sum_i n_{i\uparrow} n_{i\downarrow} = U \frac{1}{2} \sum_{i\sigma} \eta_{i\sigma}^\dagger \eta_{i\sigma}, \quad (2.7)$$

where $\eta_{i\sigma} = c_{i\sigma} n_{i\bar{\sigma}}$ creates the excitations above the gap in the upper Hubbard band (UHB) on sites occupied
by a single electron. Its complement, $\xi_{i\sigma} = c_{i\sigma}(1 - n_{i\bar{\sigma}})$ creates excitations strictly on empty sites and hence describes particle motion below the gap. Here $\bar{\sigma} = -\sigma$. Consequently, the anticommutator

$$m^0_{LHB} = \frac{1}{N} \sum_{i,\sigma} \langle\{\xi_{i\sigma}, \xi^\dagger_{i\sigma}\}\rangle = 2 - n,$$

(2.8)
determines the spectral weight in the lower Hubbard band (LHB). Since each hole in a half-filled band decreases the single occupancy by one, the weight of the UHB is $1 - x$. Because the total weight of the UHB and LHB must both be 2, we find that $2 - n + 1 - x = 2$ or $n = 1 - x$ and $m^0_{LHB} = 1 + x$ in the atomic limit. The weights $1 + x$ and $1 - x$ also determine the total ways electrons can occupy each of the bands. Thus, in the atomic limit, electrons alone exhaust the total degrees of freedom of each band. Further, since each hole leaves behind an empty site that can be occupied by either a spin up or a spin down electron, the electron addition spectrum in the LHB has weight $y = 2x$ [7, 38, 9]. Hence, the occupied part of the LHB and UHB both have identical weights of $1 - x$ in the atomic limit.

Because the operators $\xi$ and $\eta$ do not diagonalize the hopping term, the total intensity of the LHB

$$m_{LHB} = 1 + x + \frac{2t}{U} \sum_{ij\sigma} g_{ij} \langle\tilde{c}^\dagger_{i\sigma} \tilde{c}_{j\sigma}\rangle + \cdots = 1 + x + \alpha$$

(2.9)
has $t/U$ corrections as shown by Harris and Lange [6]. Here $\tilde{c}_{i\sigma}$ are related to the original bare fermion operators via a canonical transformation that brings the Hubbard model into block diagonal form in which the energy of each block is $nU$. In fact, all orders of perturbation theory [38] increase the intensity of the LHB beyond its atomic limit of $1 + x$. It is these dynamical corrections that $\alpha$ denotes. While the intensity of the LHB increases away from the atomic limit, the total number of ways of assigning electrons to the LHB still remains fixed at $1 + x$. Consequently, the total weight of the LHB exceeds the fermionic phase space and additional degrees of freedom are needed.

Nonetheless, the sum of the spectral weights in the LHB and UHB must be 2 by charge conservation. Consequently, the weight in the UHB, $m_{UHB} = 1 - x - \alpha$, decreases faster than $1 - x$. How should the spectrum in the LHB be partitioned? Harris and Lange [6] did not address this issue possibly because when the total weight of a band exceeds the electron count, the chemical potential for the effective low-energy degrees of freedom is not simply understood within a conventional picture of adding or removing electron quasiparticles. Despite this difficulty, it is common in the strongly correlated community [7, 38, 9, 8] to assign the spectral weight for the bare electrons assuming the doping level is not renormalized by the dynamics. Hence, the weight below the chemical potential, $n$, remains at $1 - x$ and the weight immediately above the chemical potential becomes $y = 2x + \alpha$ as depicted in Fig. 2.4(a). On this account, only the states above
Figure 2.4: Redistribution of spectral weight in the Hubbard model upon doping the insulating state with $x$ holes. $\alpha$ is the dynamical correction mediated by the doubly occupied sector. To order $t/U$, this correction worked out by Harris and Lange [6]. (a) The traditional approach [7, 8, 9] in which the occupied part of the lower band is fixed to the electron filling $1-x$. (b) New assignment of the spectral weight in terms of dynamically generated charge carriers. In this picture, the weight of the empty part of the LHB per spin is the effective doping level, $x' = x + \alpha$.

the chemical potential acquire doubly occupied character dynamically.

This assignment of the chemical potential is valid for the bare electrons alone and does not include the dynamically generated charges as distinct low-energy entities. To address this problem, consider the Lehmann representation,

$$m_{LHB} = \sum_{k,m} \int d\omega \langle \psi_m^{N-1}|c_{k}\psi_g^N\rangle^2 \delta(\omega - E_m^{N-1} + E_g^N)$$

$$+ \sum_{k,m} \int d\omega \langle \psi_m^{N+1}|c_k^\dagger\psi_g^N\rangle^2 \delta(\omega - E_m^{N+1} + E_g^N) \quad (2.10)$$

of the spectral function. In these expressions, $E_m^N$ is the $m$th eigen-energy of the N particle system with ground state $E_g^N$ with associated many-body states $|\psi_m^N\rangle$ and $|\psi_g^N\rangle$, respectively. The filled (first term) and empty (second term) parts of the spectrum do not have traditional definitions in terms of $n$ and $1-n$ respectively if one insists on setting $n = 1-x$. That is, $x$ is not a fundamental property of the spectral function as it is for a rigid-band system such as a band insulator in which it equals the empty part of the spectrum. The disconnect between $x$ and the empty part of the spectrum in the LHB (the second term in Eq. (2.10)) obtains because holes are generated either by doping or mixing with the doubly occupied sector. This is not an option for a band insulator. As a result, the number of holes is not determined strictly by the doping. That is, although $n$ is well-defined, $1-n$ does not have any fundamental meaning in terms of
the integrated spectral weight of the empty part of the LHB. In fact, the empty part of the spectrum has no obvious relation to anything.

We now show how the spectrum can be partitioned so that the chemical potential accounts for a charge number consistent with Eq. (2.1). Note we have some degree of freedom in describing the physics in the LHB since it is not a rigid band. If the dynamical contribution can be removed through a re-definition of the chemical potential, then the empty part of the spectrum per spin will be the effective hole number.

The justification for this picture is as follows. In a hole-doped system, turning on a finite $t/U$ creates pairs of double occupancies and empty sites (doublon-holon pairs). The weight in the UHB corresponds to adding one electron in the high energy sector, in other words creating double occupancy. Doublon-holon pairs clearly deplete this intensity leading to a loss of spectral in the UHB faster than the atomic limit value of $1 - x$. The occupied weight in the LHB corresponds to removing an electron in the low-energy sector. In other words, the occupied part of the spectrum corresponds to removing an electron such that the number of double occupancies remains conserved. Hence the occupied part of the LHB is a measure of single-occupancy whose weight as well must decrease on creation of doublon-holon pairs. In other words, the weights in the occupied part of the LHB and the UHB must be the same, since both provide a measure of the same phase space. Therefore, we propose that the consistent definition of the chemical potential for the low-energy fermionic degrees of freedom can be obtained by demanding that the two weights be equal. Note this says nothing about the nature of the excitations which live in the high-energy scale. Consequently, we arrive at the assignments of the spectral weights in Fig. (2.4b). The occupied part of the LHB has weight $(1 - x - \alpha)$ and the unoccupied part $2(x + \alpha)$. The fermionic degrees of freedom that are associated with this assignment of the chemical potential reflect the dynamical generation of the charge degrees of freedom. As a result of the dynamics, $x' = x + \alpha$ now denotes the effective number of hole degrees of freedom per spin at low energy. Consequently, we propose that it is with respect to $x'$ that a Luttinger theorem exists not $x$, the bare hole number [64].

In the case of electron doping, the chemical potential ($\mu$) lies in the UHB where $2x$ electron removal states are created below $\mu$ and the weight above $\mu$ is given by $1 - x$ in the atomic limit. Turning on a finite $t/U$ creates doublon-holon pairs. In this case, the holes belong to the LHB and represent the high-energy configurations of the system. The weight above $\mu$ represents the amplitude for adding an electron to the UHB, or creating a double occupancy, which is depleted upon creation of doublon-holon pairs since neither holons nor doublons can contribute to the creation of double occupancies upon addition of a single electron. This weight is analogous to that of the occupied part of the LHB in the case of hole doping. For charge-transfer systems, such as the cuprates, the same argument applies because of the equivalence [7] with the
Hubbard model for realistic values of the hybridization between the bands.

To counter the argument that the dynamical corrections might not affect the physics on all energy scales, it suffices to compute the cross correlator between $\xi_{i\sigma} = c_{i\sigma}(1 - n_{i\bar{\sigma}})$ and $\eta_{i\sigma} = c_{i\sigma}n_{i\bar{\sigma}}$. The full electron spectral function,

$$A(k, \omega) = -\frac{1}{\pi} \text{Im} FT(\theta(t - t')\langle\{c_{i\sigma}(t), c_{j\bar{\sigma}}^\dagger(t')\}\rangle) = A_{\eta\eta} + A_{\xi\xi} + 2A_{\eta\xi}, \quad (2.11)$$

contains two diagonal terms $A_{\eta\eta}$ and $A_{\xi\xi}$ and a cross term $A_{\eta\xi}$ which represents the degree to which the high and low energy degrees of freedom are coupled. Here, FT represents the frequency and momentum Fourier transform. It was previously computed that the cross correlator $A_{\eta\xi}$ is non-zero at all frequencies that bracket the turn-on of the spectral weights in the LHB and UHB at half-filling and at finite doping [65]. This is simply a reflection of the fact that at all frequencies, the states in the LHB all have doubly occupied character. The dynamical contribution reduces the spectral weight. Let us call the reduction $q$ and hence the weight is given by $1 - x - q$. The weight in the unoccupied part of the LHB is $2x + \alpha + q$. For the weight of a hole per spin to be equal to that of an electron, we must have that $q = \alpha$. This results in the assignments in Fig. (2.4b).

As a result, the bare electrons and the low-energy dynamically generated fermionic charge carriers in the LHB do not stand in a one-to-one correspondence. The efficient cause of this breakdown is dynamical spectral weight transfer. Insertion of an electron affects the spectrum at all energies while only local changes occur in terms of the low-energy degrees of freedom. Such an orthogonality catastrophe is due entirely to the existence of the UHB [65, 66] and persists as long as the degrees of freedom transferred from the UHB provide a relevant perturbation to those in the LHB. In fact, Fig. (2.4b) provides a possible basis for the Anderson [66] conjecture that the very existence of the UHB (in the form of dynamical spectral weight transfer) leads to a breakdown of Fermi liquid theory.

An experimental prediction of this work is that $\alpha$ should be temperature dependent. Making contact with Eq. (2.1), $\alpha$ should turn on at $T^*$. As a result, the dynamical part of the spectral weight signifies an opening of the pseudogap in the single-particle spectrum as pointed out earlier [10, 57]. This is reasonable for two reasons. First, if $\alpha \neq 0$, the number of ways of adding a particle exceeds the number of ways of adding an electron to the empty part of the spectrum in the LHB. That is, some of the particle addition states in the LHB are orthogonal to the addition of an electron. Second, there is no reason to separate the UHB and LHB’s if there is no gap between them. Consequently, the collapse of the UHB should be coincident with the closing of the pseudogap. Recently, Peets et al. [67] have observed that the UHB collapses once the
Figure 2.5: Integrated spectral weight in the occupied part of the lower Hubbard band, $\Lambda_\mu^-$, from the effective low-energy theory [10, 11, 12, 13] with $U/t = 8$. Here $x$ is the doping level for the conserved charge, $Q = \sum_i a_i^\dagger a_i + 2t^2 \sum_i \psi_i^\dagger \psi_i$. Clearly shown is that the occupied part (red triangles) of the one-particle spectrum has a weight less than $1 - x$ (solid blue line).

pseudogap closes, consistent with our prediction here.

### 2.4 Confirmation from Exact Low-Energy Theory

Since the weight of the band in which the chemical potential resides in a doped Mott insulator exceeds the electron count, new degrees of freedom are required in any consistent low-energy theory. The extra degrees of freedom are generated from mixing with the doubly occupied sector and hence should emerge upon integration of the states far away from the chemical potential. As carried out by Choy et al. [11], a charge $|2e|$ bosonic field emerges upon the integration of the UHB. The boson which is non-propagating has charge $2e$ for hole doping and $-2e$ for electron doping, represents the mixing with double occupancy and double holes respectively. For hole doping, the conserved charge $Q$, which equals the total electron filling $n$ [10, 11, 12, 13, 19], is a sum of two components,

$$Q = \sum_{i\sigma} a_i^\dagger a_i + 2t^2 \sum_i \varphi_i^\dagger \varphi_i,$$  \hspace{1cm} (2.12)

immediately implying that the weight of the fermionic part must be less than the conserved charge. Here $a_i^\dagger$ is the annihilation operator for the fermionic degree of freedom that results when the high-energy scale is integrated out and $\varphi$ is a charge $2e$ boson. That $Q$ is the conserved charge can be verified by inspection as it trivially commutes with the low-energy effective Hamiltonian. In fact, Eq. (2.12) gives a prescription for $\alpha$, namely the bosonic charge, if we interpret $Q$ as $1 - x$ and the fermionic quasiparticle density as $1 - x'$.  

23
In this theory \cite{11}, the quasiparticles are transformed at low energies to

\[ c_{i,\sigma}^\dagger \rightarrow (1 - n_{i,\bar{\sigma}})c_{i,\sigma}^\dagger + V_{\sigma} t U b_{i} c_{i,\bar{\sigma}} + V_{\sigma} t U \varphi_{i,\bar{\sigma}}, \]  

(2.13)
to leading order in \( t/U \) upon the integration of the high energy scale. Here \( b_{i} = \sum_{j} b_{ij} = \sum_{j,\sigma} c_{j,\sigma} V_{\sigma} c_{i,\bar{\sigma}} \)
with \( V_{\uparrow} = -V_{\downarrow} = 1 \) and \( j \) a nearest-neighbour of site \( i \). The first two terms represent the standard electron operator in the lower Hubbard band dressed with spin fluctuations which constitutes the quasiparticles or the effective fermionic degrees of freedom. However, the last term represents the correction due to dynamical spectral weight transfer. Eq. (2.13) lays plain that an electron at low energy contains a propagating part that arises from the charge 2e boson. To illustrate that more than just the fermions are needed to satisfy the \( 1 - x \) sum rule, we computed the pure fermionic part of the spectral function by evaluating the Green function: \[ \int d\varphi^{*} d\varphi FT(\int [Dc_{i}^{*} Dc_{i}(0)] c_{i}(0) \exp(-\int L_{IR} dt)) / Z. \]

\( L_{IR} \) is the low-energy Lagrangian obtained by integrating out the UHB and \( Z \) the partition function. We computed this quantity assuming that the boson is spatially homogeneous, which might be justified \cite{19} in some parameter range. The results in Fig. 2.5 demonstrate that the integrated weight in the occupied part of the spectrum is indeed less than \( 1 - x \).

That this weight is less than \( 1 - x \) is independent of the approximations used to calculate the spectral function. This follows entirely from the fact that the conserved charge, \( Q \) is a sum of a fermionic and a bosonic part. The deficit from \( 1 - x \) is carried in the \( \varphi_{i,\bar{\sigma}} \) term. The difference between the red triangles and the solid line approximates \( \alpha \).

### 2.5 Concluding Remarks

Experimentally, any measurement which probes the fermionic low-energy degrees of freedom should be interpreted in terms of the total number of hole degrees of freedom, \( x + \alpha \) not \( x \). For example, the superfluid density should exceed \( x \) and scale as \( x + \alpha \), already confirmed in YBa\(_{2}\)Cu\(_{3}\)O\(_{6+x}\) (YBCO) \cite{5}. Similarly, Fermi surface volumes, that is the total volume of the hole pockets minus that of the electron pockets, extracted from quantum oscillation experiments \cite{68}, whose origin is still not understood, should be compared with \( x' \) not \( x \) as the experimental probe is the current. This is particularly germane because the Fermi surface volumes extracted experimentally \cite{68,69} for YBCO are not consistent with any integer multiple of the physically doped holes. Interestingly, the first experiments of this type observed oscillations in the Hall coefficient \cite{68}. Hence, it is perfectly reasonable that the effective doping level should be consistent with the physics that leads to Eq. (2.1).

Finally, Fermi liquid theory is recovered when the charge 2e boson decouples from the electronic spectrum.
By decoupling we mean that the UHB collapses and the LHB has a weight of 2. In this limit, there is no true high-energy scale and $\varphi$ should be an irrelevant degree of freedom. To illustrate, using the appropriate scaling such that the kinetic energy remains constant in the limit $d \to \infty$, that is, $t \to O(1/\sqrt{d})$, and averages of the form $\langle c_j^\dagger c_i \rangle \propto 1/\sqrt{d}$ (note as $d \to \infty$ the scaling of $c$ and $\tilde{c}$ (Eq. (2.9)) are not trivially related), we find that the boson-dependent terms in the exact low energy theory, $\sum_i \varphi^\dagger_i c_i \uparrow c_i \downarrow \to O(1/\sqrt{d})$, $t^2/U \sum_i \varphi^\dagger_i \varphi_i \to O(1/d)$ and $t^2/U \sum_{(ij)} \varphi^\dagger_i b_{ij} \to O(d \times (1/\sqrt{d})^3)$, vanish when $d = \infty$. Consequently, no breakdown of Fermi liquid theory obtains as seen numerically for $d = \infty$ and $n \neq 1$. In finite dimensions, the precise value of the coupling constant and doping level at which the bosonic degrees of freedom decouples remains the open problem in Mottness.
Chapter 3

Towards the Standard Model for Fermi Arcs from a Wilsonian Reduction of the Hubbard Model

3.1 Introduction

As revealed by extensive angle-resolved photoemission (ARPES) studies \[71, 72, 73, 14, 74, 15, 75, 16, 76, 77\], lightly doped copper-oxide superconductors (cuprates) in the pseudogap regime possess a band of excitations that only cross the chemical potential once. Such a single crossing generates a set of coherent or pole-like excitations that ultimately form a truncated Fermi region, termed a Fermi arc, as opposed to the traditional Fermi surface generated by a double crossing. (See Figure 3.1.) The coherent excitations, centered around \((0, 0) \rightarrow (\pi, \pi)\), traverse the zone diagonal and terminate in the vicinity of \((\pi, 0)\) or \((0, \pi)\), thereby giving rise to a nodal/anti-nodal dichotomy \[78, 79, 80, 81, 82\], the former being ungapped while the latter is gapped.

While some ARPES experiments \[83\] performed on Bi\(_2\)Sr\(_{2-x}\)La\(_x\)CuO\(_{6+\delta}\) (La-Bi2201) revealed closed hole pockets, and hence consistency with the results from quantum oscillation experiments in high magnetic fields \[84, 85, 69, 86, 68\], this interpretation has been called into question \[76\]. King \textit{et al.} \[76\] observed that the closed pockets seen earlier \[83\] are entirely structural in origin as they originate from overlapping superstructure replicas of the main and shadow bands. Consequently, the preponderance of evidence from ARPES is that the coherent excitations form a disconnected region in momentum space consistent with a single crossing of the chemical potential.

Theoretically, two questions arise. 1) What suppresses the spectral weight on the back-side of the arc? 2) What is the origin of the incoherent excitations or gap at the zone boundary? A natural candidate to explain the former is that two kinds of excitations populate a doped Mott insulator, one of which has no overlap with the electron. Such an excitation will appear in the single-electron Green function as a zero rather than a pole and hence will carry no spectral weight. In this vein, some have proposed neutral composite excitations \[87\] to explain the origin of Fermi arcs. Alternatively, Fermi arcs have been seen in numerics \[88, 89, 90\] on the 2D Hubbard model and have been modeled phenomenologically \[91\] (hereafter YRZ). However, a key assumption of the phenomenological account is that the zero-line is fixed at the diamond-shaped Fermi surface of the non-interacting system. That the diamond-shaped Fermi surface of the non-interacting system
Figure 3.1: The disconnected Fermi arcs have been confirmed from ARPES for several cuprates: (a) \( \text{Ca}_{2-x}\text{Na}_x\text{CuO}_2\text{Cl}_2 \) (Ref. [14]) (b) \( \text{La}_{2-x}\text{Ba}_x\text{CuO}_4 \) at \( x = 1/8 \) (Ref. [15]) (c) ortho-II \( \text{YBa}_2\text{Cu}_3\text{O}_{6.5} \) (Ref. [16]).

constitutes the zero-line of the single-particle Green function is a rigorous mathematical statement [92] only if the underlying Hamiltonian is particle-hole symmetric. In fact, from the precise condition [93, 92, 94] for the vanishing of the real part of the Green function, maintaining that the zero-line is doping independent requires unphysical assumptions regarding the spectral function. Certainly such a conservation of the zero-line is not borne out by numerics on the Hubbard model [95, 96, 97] nor by analytical arguments [92, 64]. In addition, models involving Cooper pairs, fluctuating or otherwise, have been constructed to either yield arcs [75] or hole pockets [98, 99]. Our discussion here, however, will focus entirely on arcs as this seems to what the ARPES experiments are about.

While the physical origin of arcs might not be clear, the mathematics is. Any Green function of the form

\[
G_{\text{toy}}(\omega, \mathbf{k}) = \frac{Z}{\omega - \epsilon_k + |\Delta(\mathbf{k})|^2 / \omega - \epsilon_k} \tag{3.1}
\]

will do. It is important to note that the zero condition is \( \det G = 0 \), and hence two-band systems have an inherent difficulty in obtaining zeros. The Hubbard model we consider here has a single band, and thus if a Green function of the form of \( \text{Eq. (3.1)} \) arises without any symmetry breaking, that would suffice to describe Fermi arc.

The similarity with the BCS Green function is only perfunctory as there is no anomalous component. Aside from having poles, \( \text{Eq. (3.1)} \) has zeros whenever \( \epsilon_k = 0 \) assuming of course the chemical potential corresponds to \( \omega = 0 \). While the two dispersing electronic bands, \( \epsilon_k \) and \( \epsilon_k \) in \( \text{Eq. (3.1)} \) are not hard to come by, the parameter \( \Delta(\mathbf{k}) \) is. It requires some sort of order, fluctuating or otherwise, or a new bosonic degrees of freedom. While YRZ proposed Eq. \( \text{(3.1)} \) phenomenologically, their intuition was based on weak-coupling RPA diagramatics [100] on 2-leg ladder systems. In the context of an algebraic charge liquid (ACL) [101], Qi and Sachdev [98] also obtained Eq. \( \text{(3.1)} \). Thus far, the only system shown to have the properties of the ACL is one with radically different parameters than the cuprates [102].
Our point here is that $\Delta(k)$ arises fundamentally from a new degree of freedom associated with dynamical spectral weight transfer. Ideally, it would be advantageous to derive Eq. (3.1) directly from the strong-coupling parameter space of the basic model for a doped Mott insulator, for example the Hubbard model. Such a derivation has not presented in the literature. Hence, it is this problem that we address. Since zeros arise from a cancellation of the spectral weight in the upper and lower Hubbard bands (hereafter UHB and LHB), an accurate description of the upper band is required in a derivation of Eq. (3.1). Hence, an attempt to obtain Eq. (3.1) from a model that projects out the UHB, for example the $t - J$ model, is initially a non-starter as this model does not have zeros of the type required for Eq. (3.1). Nonetheless, the information regarding the UHB should be correctly encoded into a theory of the lower band if the UHB is integrated out (rather than projected out) exactly *ala* Wilson [22].

In this chapter, we show how the method we have recently developed [28, 11, 12] for carrying out the Wilsonian program for the Hubbard model can be used to derive Eq. (3.1). We show explicitly that two types of excitations emerge, projected electrons (yielding poles in the propagator) and a new bound state that gives rise to zeros. The bound state is not made out of the elemental excitations and hence is orthogonal to an electron (hence the zero). It represents a charge $e$ excitation that originates from the non-rigidity of the Hubbard bands, in other words, the well documented dynamical spectral weight transfer [6, 38, 7], the key fingerprint of the breakdown of the band concept in Mott systems. Since the mathematics of Fermi arcs requires two kinds of excitations, one with poles and the other with zeros, we refer to a physical model that contains both, such as the one presented here, as the standard model. More exotic models relying on some type order would fall outside this framework.

### 3.2 Charge $2e$ boson theory

#### 3.2.1 Preliminaries

Underlying our toy Green function is a two-pole structure of the form,

$$G_{\text{toy}}(\omega, k) = Z \left( \frac{\cos^2 \theta}{\omega - \omega_+} + \frac{\sin^2 \theta}{\omega - \omega_-} \right).$$

(3.2)

Here $\cos^2 \theta = (\omega_+ - \epsilon(k))/(\omega_- - \omega_+)$ and $\omega_{\pm} = \frac{1}{2}(\epsilon(k) + \epsilon(k)) \pm \sqrt{(\epsilon(k) - \epsilon(k))^2 + 4|\Delta(k)|^2}$. Zeros arise from the interference between the poles at $\omega_-$ and $\omega_+$. Any model that admits zeros must have at least this two-pole structure. There are two limits of the Hubbard model in which the zero surface can be calculated.
exactly. In the atomic limit, the zero surface of the exact single-particle Green function

\[ G^R(\omega) = \frac{1 + x}{\omega - \mu + \frac{U}{2}} + \frac{1 - x}{\omega - \mu - \frac{U}{2}}, \]  

(3.3)
is independent of momentum given by \( \omega = \mu \) and \( x = 0 \). When the hopping is non-zero, \( t \neq 0 \), the only limit in which the zero surface can be calculated exactly is at half-filling and particle-hole symmetry. In this limit, the zero surface [92] for a nearest-neighbour band structure is the magnetic Brillouin zone. Since Fermi arcs are absent from both the atomic limit and the half-filled system with hopping, it follows necessarily that Fermi arcs (if they are present at all in the Hubbard model) arise entirely from the dynamical part of the spectral weight.

Dynamical spectral weight transfer represents a concrete example of more being totally different. As is evident from Eq. (3.3), the weight of the lower band in the atomic limit is \( 1 + x \). This spectral weight has a natural interpretation in terms of electron states. There are \( 2x \) electron addition and \( 1 - x \) electron removal states. Hence, in the atomic limit, there is a one-to-one correspondence with the spectral intensity and the number of electron states in the lower band. When the hopping is turned on, the spectral intensity increases in the lower band to \( 1 + x + \alpha \), where

\[ \alpha = \frac{2t}{U} \sum_{ij,\sigma} \langle f^\dagger_{i\sigma} f_{j\sigma} \rangle + O((t/U)^2) \]  

(3.4)

and the \( f_{i\sigma} \)'s are a rotation of the original fermionic operators in the Hubbard model such that it is block diagonal. The energy of each block is \( nU \), \( n \) the number of double occupancies in each block. \( \alpha \) is necessarily positive because any hopping process that creates doubly occupancy decreases the available spectral weight in the upper band. Consequently, counting electrons, fractionalized or otherwise, cannot exhaust the total number of degrees of freedom in the lower band. A new degree of freedom must be present which is distinct and hence orthogonal to electron quasiparticles. This degree of freedom will appear as a zero in the spectral function in the lower band. It is precisely the nature of the states that arise from the mixing with the upper band that we elucidate here.

Several approaches suggest some kind of composite excitation mediates Fermi arcs. Consider the SU(2) gauge theory of the t-J model proposed by [Wen and Lee]103 in which the elemental fields are the appropriate linear combinations of two charge e bosons and two spinons. A mean-field calculation of the coherent spinon-boson Green function reveals that the spectral weight of the occupied part of the spectrum exceeds \( 1 - x \), acquiring a value of \( 1 + x/2 \) instead. [Wen and Lee] alleviated this problem by introducing an interaction which recombined the bosons and fermions back into the elemental fields. The effect of this interaction with
strength $U$ was to enhance the spectral weight in the unoccupied part of the spectrum. The correct spectral weights were obtained simply by adjusting the magnitude of $U$. The unoccupied part of the spectrum shows up as a small hole pocket centered roughly at $(\pi/2, \pi/2)$. The intensity on the back-side of the pocket is greatly suppressed, thereby leading to a structure not too distinct from a Fermi arc. Subsequent work on the $U(1)$ formulation of the $t-J$ model with a phenomenological spinon-holon binding term reached a similar conclusion [104]. However, the most extensive calculation in the gauge theoretic formulations of the Hubbard model reached a rather different conclusion. Working directly with the Hubbard model, Imada and colleagues [87] used a slave-particle construction with the gauge fluctuations treated at the RPA level and concluded it is actually dynamical spectral weight transfer that leads to a fermi arc structure and a Green function of the YRZ form. Aside from suffering from the lack of a systematic way of treating the gauge fluctuations, this formulation generates Fermi arcs from a neutral composite excitation Since neutral entities cannot couple to the current, it is unclear how such entities can influence the spectral function. What we demonstrate here without resorting to a gauge theory is that dynamical spectral weight transfer mediates Fermi arcs.

3.2.2 Exact Results: The conserved charge and the low energy mode

Dynamical spectral weight transfer has two profound consequences. First, we show exactly that the conserved charge $(1-x)$ is not exhausted by counting the degrees of freedom minimally coupled to the electromagnetic gauge. The remainder are carried by an incoherent background. Second, the dynamically transferred degrees of freedom give rise to zeros in the lower band. In fact, the physics we find here is analogous to the finite $U$ charge processes that contribute to the spectral function measured by ARPES in heavy-fermions [105].

We first start with the procedure to integrate out the UHB. Previously, we demonstrated [28, 11, 12] that a theory of the lower Hubbard band (hole doping) can be obtained by introducing a new fermionic field $D_i$ [28, 11] which represents the excitations in the upper band. This field has mass $U$ and hence should be integrated out to construct the exact low-energy theory. As in the derivation of all collective phenomena which dates back to the classic paper of Bohm and Pines [106], a constraint must be introduced so that when solved, the model in the extended space is equivalent to the starting UV-complete theory. The new fermionic degree of freedom enters the action in a quadratic fashion and hence the standard fermionic path integral techniques can be applied to integrate the high-energy scale of the UHB. Since the lower and upper bands are not rigid in the sense that the spectral weights of the two bands are coupled, integrating out the upper band will lead to new degrees of freedom in the lower band. In Euclidean signature, the Hubbard
The action in the extended space is

\[
S_{h}^{\text{UV}} = \int_{0}^{\beta} d\tau \int d^{2}\theta \left\{ \bar{\theta} \sum_{i,\sigma} \Theta (1 - n_{i\sigma}) c_{i\sigma}^{\dagger} \partial_{\tau} c_{i\sigma} + \sum_{i} D_{i}^{*} \bar{\theta}_{i\sigma} D_{i} + U \sum_{j} D_{j}^{*} D_{j} \right. \\
- t \sum_{i,j,\sigma} g_{ij} \left[ \bar{\theta} \Theta (1 - n_{i\sigma})(1 - n_{j\bar{\sigma}}) c_{i\sigma}^{\dagger} c_{j\bar{\sigma}} + D_{i}^{*} c_{j\sigma}^{\dagger} c_{i\sigma} D_{j} + (D_{j}^{*} \Theta c_{i\sigma} V_{\sigma} c_{j\bar{\sigma}} + \text{c.c.}) \right] \\
+ s \bar{\theta} \sum_{j} \varphi^{*}_{j} (D_{j} - \Theta c_{j\uparrow} c_{j\downarrow}) + \text{c.c.} \right\},
\]

(3.5)

where the matrix \(g_{ij}\) selects the relevant neighbors, \(V_{\sigma} = \pm 1(\sigma = \uparrow, \downarrow)\), the constraint is given by \(\delta(D_{i} - \theta c_{i\uparrow} c_{i\downarrow})\), \(\theta\) is a Grassmann, \(s\) is a constant with units of energy so that \(\varphi_{i}\) is dimensionless and \(c_{i\sigma}\) is an electron annihilation operator for site \(i\) with spin \(\sigma\). Because the \(\delta\)-function constraint appears exponentiated in the action, an auxiliary field with charge \(2e\), \(\varphi_{i}\) enters the action as a Lagrange multiplier. As a consequence, the field \(\varphi\) is not made out of the elemental excitations (thereby distinguishing it from other charge \(2e\) scenarios involving pairs of electrons) but rather arises because the UHB and LHB are not rigid bands. In the action, the first two terms represent the dynamics in the lower and upper Hubbard bands, respectively, the third term the mass of the \(D\) field, the fourth term the hopping in the lower band with matrix element \(t\), the next two the dynamical mixing between the upper and lower bands and the last term the constraint. The constant \(s\) has units of energy and is \(O(t)\) \([10]\). It is straightforward to check that solving the constraint by integrating out the auxiliary field, \(\varphi_{i}\), followed by an integration over \(D_{i}\) exactly reduces \(S_{h}^{\text{UV}}\) to the action for the standard Hubbard model. This is the UV limit of our theory. The advantage of the reformulation above is that it cleanly associates the physics of the upper band with a fermionic field \(D_{i}\) which enters the action in a quadratic fashion. To obtain the IR limit, one simply has to perform the Gaussian integration over the massive field, \(D_{i}\). The result is the low-energy or IR action, \(S_{h}^{\text{IR}} = \int d\tau \mathcal{L}_{h}^{\text{IR}}\), with the associated Lagrangian,

\[
\mathcal{L}_{h}^{\text{IR}} = (1 - n_{i\bar{\sigma}}) c_{i\sigma}^{\dagger} \partial_{\tau} c_{i\sigma} - tg_{ij} (1 - n_{i\bar{\sigma}}) c_{i\sigma}^{\dagger} c_{j\bar{\sigma}} (1 - n_{j\bar{\sigma}}) \\
- (s \varphi_{i} - tb_{i})^{*} (M^{-1})_{ij} (s \varphi_{j} - tb_{j}) \\
- (s \varphi^{*}_{i} c_{i\uparrow} c_{i\downarrow} + \text{c.c.}) - \frac{1}{\beta} \text{tr} \ln \mathcal{M},
\]

(3.6)

where a matrix element of \(\mathcal{M}\) is given by \(\mathcal{M}_{ij} = (\partial_{\tau} + U) \delta_{ij} - tg_{ij} c_{j\sigma}^{\dagger} c_{i\sigma}\) and \(b_{i} = \sum_{j} g_{ij} c_{j\sigma} V_{\sigma} c_{i\bar{\sigma}}\). Hereafter, repeated indices are implicitly summed unless otherwise stated. It is important to note that no approximations have been made as of yet.

In both actions, \(S_{h}^{\text{UV}}\) and \(S_{h}^{\text{IR}}\), global \(U(1)\) symmetry guarantees the existence of a conserved charge,
which turns out to be

\[ Q_{h}^{UV} = (1 - n_{i\bar{\sigma}})c_{i\sigma}^{\dagger}c_{i\sigma} + 2D_{i}^{\dagger}D_{i}, \quad (3.7) \]

\[ Q_{h}^{IR} = (1 - n_{i\bar{\sigma}})c_{i\sigma}^{\dagger}c_{i\sigma} + 2\left( s\varphi_{i} - tb_{i} \right)_{i}^{\dagger} \left( M^{-1} \right)_{ik} \left( M^{-1} \right)_{kj} \left( s\varphi_{j} - tb_{j} \right). \quad (3.8) \]

It is a natural consequence that the conserved charge \( \langle Q_{h}^{UV} \rangle \) is consistent with the number of electrons in the original Hubbard model since the operator \( D_{i}^{\dagger}D_{i} \) counts the number of doubly occupied sites. From the Hellman-Feynmann theorem, it is straightforward to check how the number of double occupancies, \( n_{docc}^{UV/IR} \), is related to \( D_{i} \). Since \( n_{docc}^{UV/IR} = \beta^{-1}\partial \ln Z_{h}^{UV/IR} / \partial U \) with \( Z_{h} = \int D[\cdots] e^{-S_{h}} \), one can easily observe the second terms in Eq. (3.7) and (3.8) are identical to \( n_{docc}^{UV} \) and \( n_{docc}^{IR} \), respectively. As a result, the conserved charge \( \langle Q_{h}^{IR} \rangle \) is identified with the number of electrons, \( 1 - x \), with \( x \) the number of holes. This is one of the indications that the low-energy action, \( S_{h}^{IR} \), retains the structure of the Hubbard model even after the integration of the massive modes.

Another advantage of the low energy theory is that the non-Fermi-liquid nature of the low-energy excitations is immediately manifest. To illustrate this, one can add a minimally coupled source term \[ L'_{h}^{UV} = J_{i\sigma}^{*} \left[ \bar{\theta}(1 - n_{i\bar{\sigma}})c_{i\sigma}^{\dagger} + \bar{\theta}c_{i\sigma}^{\dagger}V_{\sigma}D_{i} \right] + c.c. \quad (3.9) \]
so that when the constraint is solved, the bare electron operator is generated \[ L'_{h}^{IR} = J_{i\sigma}^{*}\psi_{i\sigma} + c.c. - J_{i\sigma}^{*}c_{i\bar{\sigma}}^{\dagger} \left( M^{-1} \right)_{ij} c_{j\bar{\sigma}}J_{j\sigma} \quad (3.10) \]
with a new collective field, \( \psi_{i\sigma} \) given by

\[ \psi_{i\sigma}^{*} = (1 - n_{i\bar{\sigma}})c_{i\sigma}^{\dagger} + tb_{i}^{\dagger} \left( M^{-1} \right)_{ji} V_{\sigma}c_{i\bar{\sigma}} - s\varphi_{j}^{*} \left( M^{-1} \right)_{ji} V_{\bar{\sigma}}c_{i\bar{\sigma}}. \quad (3.11) \]

Note \( \psi_{i\sigma} \) is derived not contrived. It is Eq. (9) in the Ref. \[ 11. \] We obtained it by integrating the \( UV \)–complete Lagrangian in the presence of the source term that generates the correct UV current with respect to the massive field \( D_{i} \). \( \psi_{i\sigma} \) is the propagating degree of freedom in the IR. It contains not only an electron-like quasiparticle affected by nearby spin fluctuations, but also a hole (with the opposite spin) that is dressed
with a doubly-charged bosonic mode. Note, we cannot give $\psi_{i\sigma}$ a simple interpretation in terms of bosons or fermions. At best, $\psi_{i\sigma}$ corresponds to the physical field that is minimally coupled to an external gauge field. That is, these are the excitations that couple to light. Hence, it is the field that is probed by an ARPES experiment, for example. While $\psi_{i\sigma}$ was derived earlier, what we did not show explicitly is that it does not stand in a one-to-one correspondence with the bare electrons. This can proven exactly by focusing on the positive-definite quantity,

$$
\psi_{i\sigma}^\dagger \psi_{i\sigma} = (1 - n_{i\sigma}) c_{i\sigma}^\dagger c_{i\sigma} + (tb - s\varphi)_{j} (M^{-1})_{ij} c_{i\sigma}^\dagger c_{i\bar{\sigma}} (M^{-1})_{ij} (tb - s\varphi)_{j}
$$

which is essentially the conserved charge less the number of doubly occupied sites. Since the second term in the last line is positive definite, the number of low-energy collective modes which are minimally coupled to the electromagnetic gauge field is less than $\langle Q_{h} \rangle = 1 - x$. The natural resolution of this conundrum is that the number operator only counts those excitations that have a particle-like interpretation. That is, the number operator only counts the coherent part of the spectrum. All of the stuff mediated by mixing with the upper band is entirely incoherent and hence while it can contribute to the current, it is not enumerated by counting the number of particles. The remainder of the charge count is carried by the last term in Eq. (3.10).

This discrepancy is not a surprise when one considers that the total spectral weight of the lower band exceeds $1 + x$ by a dynamical correction, $\alpha > 0$, that depends on the hopping integral, $t$. Since there are only $1 + x$ electron states in the lower band, and only charge $e$ excitations contribute to the spectral function, there has to be some new charge mode to make up the difference. What $\psi_{i\sigma}$ lays plain is that there are charge $e$ states that contribute to the current that are completely incoherent. It is a composite excitation of $\varphi^\dagger$ and a hole $c_{i\bar{\sigma}}$. In terms of the UV variables, this degree of freedom represents the binding of a doublon and a holon. The new composite excitation, $\varphi^\dagger M^{-1} V_{\sigma} c_{i\bar{\sigma}}$ has internal structure and hence is orthogonal to the projected electron. Since there is no Hilbert space for $\varphi$, interpreting $\varphi^\dagger c_{i\bar{\sigma}}$ in terms of a particle is not possible. It is this additional degree of freedom that creates the Fermi arc structure—that is, the zeros of the Green function. Hence, hidden in $\psi_{i\sigma}$ is an incoherent contribution to the single-particle Green function. What this discussion makes clear is that $\varphi$ should not be considered to be an independent degree of freedom but rather one that is strongly coupled to the fermions.

What we have shown thus far is that there is a dynamical contribution to the charge degrees of freedom that are coupled to the source term that generates the current. Such entities are the physical degrees of freedom that create holes in the lower band. Consequently, when one such excitation is removed from the
lower band, the change in the spectral weight should also depend on $t$. Hence, the doping level should receive a dynamical contribution. To this end, we defined $x' = x + \alpha$ and hence the weight in the UHB is $1 - x'$ and the occupied and empty parts of the lower band are $1 - x'$ and $2x'$, respectively.

### 3.2.3 Green function and the approximations

Thus far, all of our statements are exact. Our calculation of the Green function is not, however. To lend credence to our treatment, we state our assumptions clearly and up front. The complexity arises in treating the $\varphi$ degree of freedom. Our treatment is in the spirit of the results obtained in the previous section, namely that $\varphi$ leads to the creation of a new charge $e$ excitation that is orthogonal to a projected electron on account of its internal structure.

Having determined the generating functional, $\mathcal{L}'_{\chi}\{[J^e_{i\sigma}, J_{i\sigma}]\}$, we proceed to calculate the Green function. In the functional formalism, it is given by

$$G_{\psi}(\mathbf{r}_i - \mathbf{r}_j, \tau) = -\frac{\delta^2}{\delta J_{i\sigma} \delta J^*_{j\sigma}} \ln Z^R_{\mathcal{L}}\{[J^e_{i\sigma}, J_{i\sigma}]\} \bigg|_{J^e_{i\sigma} = J_{i\sigma} = 0} = -\left\langle \mathcal{T}_\tau \psi_i(\tau) \psi_j^\dagger(0) \right\rangle + \left\langle \delta(\tau) c_{i\sigma}^\dagger \mathcal{M}^{-1} c_{j\sigma} \right\rangle,$$

where $\mathcal{T}_\tau$ represents time ordering and $\langle \cdots \rangle$ stands for the average over all possible paths. Since the second term is independent of time, this term contributes to the incoherent part of the Green function. To understand the first term which contains both coherent as well as incoherent responses, it is sufficient to focus on the correlator between the $\psi_{i\sigma}$’s. Since $\psi_{i\sigma}$ contains a composite excitation which has a prefactor of $t/U$, $\langle \mathcal{T}_\tau \psi_{i\sigma}(\tau) \psi_{j\sigma}^\dagger(0) \rangle$ can, in principle, be expanded in power of $t/U$. The presence of the projection operator, $(1 - n_{i\sigma})$, however, does not necessarily guarantee that it gives a dominant contribution compared to the composite entities.

For the purpose of numerical evaluation, we make an approximation to the projection operators, following the idea developed by Zhang, et al. However, the crucial difference is that we make the substitution the bare hole concentration by the effective hole doping level ($x \rightarrow x'$), since the physical entities coupled to the external gauge field are not the bare electrons but are rather dynamically generated. Hence our first approximation is

$$(\mathbf{A-1}) \quad (1 - n_{i\sigma}) c_{i\sigma}^\dagger c_{j\sigma} (1 - n_{j\sigma}) \rightarrow g_t c_{i\sigma}^\dagger c_{j\sigma},$$

where $g_t = 2x'/(1 + x')$. Interestingly, in the strong coupling limit, $(U/t \gg 1)$, a mean-field approach to Kotliar-Ruckenstein’s slave boson construction led to the same renormalization factor for the charged
fermion but with $x'$ replaced with $x$. Likewise,

(A-2) \((1 - n_{i\sigma}) c_{i\sigma}^\dagger (\partial_\tau + \cdots) c_{i\sigma} \to g_p c_{i\sigma}^\dagger (\partial_\tau + \cdots) c_{i\sigma}\),

where $g_p = (1 - x')/(1 - x)$. The multiplicative factors are chosen here for internal consistency with the two assumptions.

Since the action $S_h^{\text{IR}}$ has all relevant degrees of freedom for the low-energy sector, including the spin singlet fluctuations ($h_i$) and mixing between the separate Hubbard bands ($\varphi$), it is reasonable to expand the action in powers of $t/U$. To leading order, the matrix elements $(M^{-1})_{ij}$ is $U^{-1}\delta_{ij}$. From the fact that the collective boson $\varphi_i$ only has dynamics through its coupling to the fermions, we assume the dynamics of the boson to be frozen. Operationally this assumption breaks down at $O(t/U)^2$ where the explicit dynamics of $\varphi$ appears as can be seen from an expansion of the $M$ matrix, $\frac{x^2}{t^2} \varphi^* (\partial_\tau - U + \cdots) \varphi$. In fact, this even at $O(t/U)^2$, the propagator for $\varphi$ lays plain that it has a pole only in the high-energy sector. This justifies the assumption that

(A-3) Bosonic field, $\varphi$, has no dynamics in the LHB.

In other words, it alone is highly massive and is not likely to propagate in the low-energy sector. Finally, although local spin ordering might non-negligible, we will assume it to be at most ancillary to the strong interaction physics arising from the coupled boson-fermion terms. This is a key assumption and certainly not traditional as most treatments of the LHB focus on the spin physics. However, as our emphasis here is on isolating the source of zeros in the LHB, demonstrating that the action possesses such modes in the absence of the spin-spin scattering term would suffice to show that such an interaction is indeed ancillary to the essential charge physics. As will become evident, our treatment does in fact show this to be the case.

Under these considerations, the effective low energy action turns into

$$S_h^{\text{IR}} = \int_0^\beta d\tau \left\{ c_{k\sigma}^\dagger [g_p(\partial_\tau - \mu)\delta_{ij} - g_p \epsilon_k] c_{k\sigma} \right.$$

$$- \frac{1}{U^2} (s\varphi - tb)^* \varphi \right. \left( U + 2\mu \right) (s\varphi - tb) q$$

$$- \left. (s\varphi_q^* c_{q-k\uparrow} c_{k\downarrow} + \text{c.c.}) \right\}, \quad (3.14)$$

where $\mu$ denotes the chemical potential, $k$ and $q$ are the momenta, and $c_{k\sigma} = tg_{ij} e^{i\mathbf{k}(r_j-r_i)}$. This action has a BCS-like coupling and hence will have a Green function of the form of Eq. (3.1). That the Green function must be of the form of Eq. (3.1) is not dependent on the assumptions delineated earlier. It relies solely on the fact that the spectral weight in the lower band exceeds $1 + x$ and hence a new charge $e$ state distinct from the projected electrons must be present. Such an excitation can only be a composite.
Consequently, for a given amplitude of $\varphi_q$, the Fourier transformation of the two point correlator $\mathcal{G} = -\langle T_\tau \psi_i(\tau) \psi_j^\dagger(0) \rangle$ becomes

$$\mathcal{G}(i\omega_n, \mathbf{k}) = \frac{\bar{g}_i}{i\omega_n - \mu - \bar{g}_i \epsilon_k - \Sigma_{\pm}(i\omega_n, \mathbf{k})} + \frac{t}{U} (\cdots), \quad (3.15a)$$

$$\Sigma_{\pm}(i\omega_n, \mathbf{k}) = \frac{s_{k.q}^2 \varphi_q \varphi_q^*}{i\omega_n - \mu \pm \bar{g}_i \epsilon_{q-k}}, \quad (3.15b)$$

where $\omega_n = (2n + 1)\pi/\beta$ for $n \in \mathbb{Z}$, $\bar{g}_i = g_i/g_p$, and $s_{k.q} = 1 - (\epsilon_k + \epsilon_{q-k})/U$. The $\pm$ subscript on the self energy arises from the two choices which are possible to treat the dynamics of the charge $2e$ boson. If $\varphi_i$ is treated as an independent degree of freedom that can condense, then it can be absorbed as a redefinition of the interaction strength, $s \to s\varphi$. This will correspond to a simple condensation of $\varphi$ in a non-zero momentum particle-particle channel, hence the $+$ sign in front of the $\bar{g}_i \epsilon_{q-k}$ factor in the denominator of the self energy. As will be clear, this is not the interpretation of $\varphi$ that is ultimately consistent with the theory outlined here. Alternatively, $\varphi^* c_{i\bar{\sigma}}$ could be viewed as a new composite charge $e$ excitation that results from dynamical spectral weight transfer. With such bound modes, the interaction term, $\varphi^* c^\dagger c_{\downarrow}$, can be interpreted as a particle-hole scattering process. To implement this interpretation in the Green function, we note that since $\varphi^* c^\dagger c_{\downarrow}$ now describes the scattering of an electron $c_{\sigma}$ off a composite particle $\varphi^* V_{\sigma} c_{\bar{\sigma}}/|\varphi|$, the denominator in the one-loop self energy will resemble that of a particle-hole scattering event, thereby leading to a $-\frac{t}{U}$ sign in front of the $\bar{g}_i \epsilon_{q-k}$ term in the denominator of the self energy. In addition, the ellipse in Eq. (3.15a) represents the terms that originate from the mixing between the composite excitations and the projected electron, which are at least suppressed by the factor $t/U$. The number $\bar{g}_i$ results from the rescaling $g_{p,\bar{\sigma}}^{1/2} c_{i\sigma} \to c_{i\sigma}$. Since $\bar{g}_i = g_i/g_p \approx 2x'$, the $t/U$ corrections in Eq. (3.15a) are comparable to the leading term only for $x' < t/2U \sim 0.05$. For example, at half-filling, the Green function only has the $t/U$ term and the spectral weight is governed entirely by the mixing between the projected and composite excitations as shown previously [10]. In the current treatment, we will explore entirely the contribution from the leading term which is of the form of Eq. (3.1).

### 3.2.4 Free-energy Minimum approach to $2e$ boson

Evaluating the Green function is equivalent to a random-matrix problem. In the most general case, the field $\varphi_i$ must be integrated over with a separate value on each site. However, such a multi-variable integration is not tractable in any dimension. From the observation that the collective boson is not canonical, that is, it does not have its own kinetics, it was previously conjectured that the spatially homogeneous configuration was
the most prominent candidate for the ground state \cite{28,10}. Even though such an approach was successful in capturing some experimental \cite{10} findings, it still leaves an open question whether the homogeneous solution minimizes the free energy. To this end, we explore some inhomogeneous solutions for $\varphi_i$ to see where the free-energy is a minimum. In particular, we explore a staggered configuration, $\varphi_i = e^{i\mathbf{q} \cdot \mathbf{r}_i}|\varphi_0|$. It should be noted that a particular choice of the configuration of $\varphi$ does not correspond to spontaneous symmetry breaking, since the bosonic mode is in lack of inherent dynamics.

In Fig. 3.2 we directly compute the free energy difference between a configuration with a spatial texture
and the homogeneous state, $\Delta F = F(\varphi_0 e^{i\mathbf{q}\cdot \mathbf{r}}) - F(\varphi_0)$, where $\mathbf{q}$ determines the spatial dependence of $\varphi_i$. Except for small values of $\varphi_0$ in which the homogeneous solution minimizes the free energy, the distinct minimum occurs at $(\pi, \pi)$ when the magnitude of the bosonic field, $\varphi_0$, approaches unity. This is significant because the probability distribution of $\varphi_0$ computed from $P(|\varphi|) = 1/Z_{IR} \int D[\{c, c^*\}] e^{-S_{IR}}$, has a distinct maximum precisely at the value of $\varphi$ where the $(\pi, \pi)$ solution minimizes the free energy. This state of affairs obtains because a quick inspection of the action reveals that for a staggered configuration of $\varphi$, the $\varphi^\dagger b$ term actually vanishes. This results in a lowering of the energy relative to the homogeneous solution.

That the $(\pi, \pi)$ configuration of $\varphi_i$ minimizes the free energy is highly significant because the evaluation any integral over $\varphi_i$ will be dominated by the staggered solution. What about the single-particle Green function? In our previous treatment of this problem in which we assumed that the mixing with the UHB was mediated by a homogeneous boson, $\varphi_0$ for all sites, we obtained a completely gapped structure at the chemical potential for the spectral function. Given that $\mathbf{q} = (\pi, \pi)$ is the global minimum, our expression for the Green function simplifies to

$$G(i\omega_n, \mathbf{k}) = \int d|\varphi|^2 P(\varphi) G(i\omega_n, \mathbf{k}) |\varphi\rangle = G(i\omega_n, \mathbf{k}) |\varphi\rangle = \delta_{\mathbf{q}, \pi}.$$ (3.16)

The probability distribution function $P(\varphi)$ is shown in Fig. 1b. For completeness, we present in Fig. 3.3 the band dispersion corresponding to the maximum in the spectral function obtained from Eq. (3.15a) for three different cases: 1) homogeneous solution, 2) staggered $(\pi, \pi)$ phase of $\varphi_i$ in the particle-particle channel, $\Sigma_+$ and 3) staggered solution in particle-hole channel, $\Sigma_-$. For the homogeneous phase, we find a hard gap, Fig. 3.3(a), because no momentum states cross the chemical potential. However, as shown earlier, the homogeneous solution does not correspond to a minimum in the free energy. Consider the staggered solutions shown in panels Fig. 3.3(b) and Fig. 3.3(c). Fig. 3.3(b) shows that even a staggered solution in the particle-particle channel, a gap does not occur at the $(\pi, 0)$ region of the Brillouin zone. There is also a crossing along the zone diagonal. This indicates that a simple condensation of $\varphi$ in a non-zero momentum particle-particle channel cannot give rise to the nodal/anti-nodal dichotomy. There is in fact a clear reason why $\varphi$ cannot be treated as an independent degree of freedom that can condense. There is a one-to-one correspondence between Eq. (3.11) and its analogue (Eq. (19) of Ref. [38]) in the standard perturbative treatment of the Hubbard model. In essence, the charge $2e$ boson replaces a string of operators that account for the mixing of double occupancy into the lower band. This is why this approach is simpler. As it would be completely incorrect to replace that string of operators with an average value, it is equally wrong to treat $\varphi$ as a variable that can condense. In fact, it is well known [7] that such mean-field truncations fail to describe
Figure 3.3: The low energy band dispersion along high symmetry directions for (a) a homogeneous configuration, $\varphi_i = \varphi_0$, (b) a staggered configuration ($\varphi_i = |\varphi_0|e^{i\pi r_i}$) evaluated by Eq. (3.15a), and (c) a staggered one evaluated by Eq. (3.15a). Here, we take the broadening factor $\eta = 0.025t$ for a typical value $U/t = 10$, $t'/t = -0.3$, $t''/t = 0.1$, and the bare hole doping level as $x = 0.12$. In the evaluation, the parameters, $\alpha = x' - x$ are taken from the numerical estimates of the number of double occupancies from Liebsch and Tong [17].

Consider the third dispersion, Fig. 3.3 (c) in which $\varphi$ is the mediator of a composite charge $e$ state. This corresponds to a self-energy given by $\Sigma_-$. The break in the dispersion just above the chemical potential is not followed by a re-entrant crossing at a higher momentum. Such a re-entrant crossing would give rise to a closed Fermi surface. It is the presence of the additional propagating degree of freedom which thwarts this re-entrance. In addition, there is no crossing at $(\pi, 0)$, but a broad incoherent feature indicative of the pseudogap. Since the break-up of the bound state results in a band crossing near the $(\pi, 0)$ region, the root cause of the pseudogap is the bound state formed between the bosonic field, $\varphi$ and a hole as we have advocated previously [10]. Consequently, the pseudogap problem is one of confinement. The corresponding Fermi surfaces are shown in Fig. 3.4 The arc-like structure is evident. The line of zeros is given by the divergence of the self-energy and hence it is doublon-holon binding that is responsible for killing the intensity on the back side of the arc. The Fermi surfaces evolve smoothly for the doping levels shown from $x = 0.05$ to 0.39.
Figure 3.4: The spectral function of the low energy theory for each hole doping level, (a) $x = 0.05$, (b) $x = 0.08$, (c) $x = 0.12$, and (d) $x = 0.18$. As noted in the text, the “pocket” size is proportional to each doping level. In the present figures, an incoherent background is removed and the intensities of the spectral function are normalized for the first quadrant of the full Brillouin zone.

$x = 0.18$. Note also the broad feature at the zone boundary. While it is tempting to interpret the broad peak near the antinodal region as an electron pocket, the lack of coherent excitations makes this view untenable.

3.3 Final Remarks

The key point this work demonstrates is that two types of charge carriers go into forming Fermi arcs. The projected electrons are present in any low-energy reduction of the Hubbard model and create the spectral weight on the high-intensity side of the arc. The zeros correspond to composite excitations which are present as a result of dynamical spectral weight transfer and hence are present only if the UHB is retained or treated appropriately. Such composite excitations enter the self-energy through the particle-hole channel, as the relevant scattering process is that of a fermion from the composite excitation. Both of these features leading to an effective two-fluid model [10] [57] are present within a Wilsonian reduction of the high-energy scale in the Hubbard model. The treatment we have derived here should be valid as long as the UHB provides a relevant perturbation to the physics of the LHB. Hence, it cannot describe the crossover to the Fermi liquid regime in which $\varphi$ is unbound. Experimentally, a decoupling of the UHB from the LHB appears to take place around $x \approx 0.25$ [67]. Accompanying the collapse is a transition from a small Fermi surface scaling with $x$ to a large one with effective area $1 - x$. The precise nature of this transition will be the subject of a future
study. However, a prediction of this work is that the in the pseudogap regime, the volume of the Fermi arc region should be given by $x'$ rather than $x$. This follows from the fact that the number of particle-like excitations that minimally couple to the electromagnetic gauge is less than $1 - x' < 1 - x$. Hence, the hole Fermi surface should be given by $x'$. High precision ARPES measurements can be employed to verify this result.

Since our scheme of two type of charge carriers, one giving rise to zeros and the other to poles, seems quite general, it is tempting to rewrite the IR theory in terms of the composite and projected excitations. This would require integrating in an additional field for the composite, $f_{i\sigma}$ degree of freedom. The composite fermion is not canonical, however, and treating it as such would destroy the key feature leading to a suppression of the spectral on the back-side of the arc. Thus far, we have found no consistent way of doing this. Hence, an open problem remains precisely how the new composite excitation should be treated. But that it is present in any standard model of Fermi arcs is not in doubt.
Chapter 4

Two distinct insulating transitions and breakdown of self averaging in the disordered Bose-Hubbard model

4.1 Introduction

Two distinct mechanisms localize interacting bosons moving in a random environment. Either repulsions from local on-site interactions on a $D$-dimensional lattice lock the bosons in place or disorder inhibits tunneling from site to site. The former insulating state is an incompressible Mott insulator (MI), while the second is a compressible Bose glass (BG). In this chapter, we identify an order parameter and an associated length scale that can describe the vanishing of the compressibility at the MI/BG transition. The order parameter is the relative variance of the disorder-induced mass distribution. The key insight leading to this observation is that despite the relevance of the disorder vertex in $D < 4$ spatial dimensions, it is possible to identify a Mott-insulating fixed point where the mass, corresponding to the Mott gap, diverges fast enough such that the relative variance of the mass distribution is renormalized to zero. Sufficiently close to the MI, both the mean and the relative variance diverge, identifying the adjacent phase as an insulating BG, irrespective of the boson filling. We will show that the length scale underlying this order parameter is the effective distance between rare regions.

In their initial treatment of this problem, known as the disordered Bose-Hubbard (BH) model, Fisher and colleagues situated the BG between the MI and superfluid (SF), thereby preventing a direct SF/MI phase transition. They pointed out however, that it is in principle possible but extremely unlikely that for sufficiently weak disorder the BG phase is completely suppressed at commensurate boson fillings. For more than twenty years, this question remained controversial as simulations and analytical arguments both support and negate the possibility of a direct MI/SF transition in the presence of weak disorder. Only recently, strong arguments based on a mathematical theorem have been presented for the existence of the BG upon the destruction of the MI, thereby precluding a direct transition to the SF. Given this result, it is important to establish precisely how the MI to BG transition obtains.

In this chapter, we are concerned with the critical theory for the MI/BG transition for strongly interacting
bosons on a D-dimensional lattice subject to weak disorder. Of particular interest is the universality of the transition, e.g., the identification of a diverging length scale and the determination of the critical exponent describing the divergence of the inverse compressibility. Despite extensive numerical [108, 109, 111, 112, 114, 115, 123] and analytical [110, 118, 119, 120, 121, 122, 113, 124, 116, 117, 125, 126] studies of the disordered BH model, the nature of the MI/BG transition remains elusive.

To be more specific, in the following we consider the disordered BH model in its simplest form,

\[ \hat{H} = -t \sum_{\langle i,j \rangle} (\hat{b}_i^\dagger \hat{b}_j + \text{h.c.}) + \sum_i (\epsilon_i - \mu) \hat{n}_i + \frac{U}{2} \sum_i \hat{n}_i(\hat{n}_i - 1), \]  

(4.1)

describing bosons with creation and annihilation operators \( \hat{b}_i^\dagger, \hat{b}_i \) which are hopping with amplitude \( t \) between neighboring sites \( i, j \) on a D-dimensional hyper-cubic lattice and which are subject to an on-site repulsion \( U \). The chemical potential \( \mu \) controls the boson filling and disorder enters via random uncorrelated on-site potentials \( \epsilon_i \) which must be appropriately bounded for the model to exhibit stable MI phases [1].

Since the formation of the MI is a consequence of strong repulsive interactions, the MI/SF transition in the clean system can only be understood within an effective long-wavelength theory which is dual to the original BH model [4.1] and derived by a strong coupling expansion around the atomic limit [1, 127, 128]. The mass in the theory is given by the Mott gap which vanishes at the transition and turns negative in the SF. In the latter broken-symmetry state the finite bosonic order parameter is proportional to the SF density. Although the underlying strong coupling expansion breaks down in the SF where the renormalization group (RG) flow is towards weak coupling [126], the effective long-wavelength theory can be used to analyze the instability of the MI. Since the theory is controlled in the MI state it should serve as a starting point to also analyze the instability towards the formation of the BG in the presence of weak disorder.

At the outset, it would seem that an effective theory in terms of the bosonic order parameter is completely inapplicable to the MI/BG transition since in both insulating phases the mass gap is finite and the SF correlations are short ranged. Certainly, the relevant length scale at the MI/BG transition has nothing to do with the SF correlation length but instead should be intrinsically related to the disorder. Such a divergent length scale should exist since the transition is expected to be continuous with a divergent inverse compressibility. In this work, we demonstrate that the order parameter for the MI/BG transition can be extracted from the disorder-averaged replica version of the bosonic order parameter theory. The relevant quantity is the relative variance of the disorder-induced mass distribution which renormalizes to zero in the MI and diverges in the BG and hence serves as the order parameter for the MI/BG transition. The
corresponding length scale is the distance between rare regions in the BG which causes the breakdown of self-averaging in the system as indicated by a diverging relative variance. This correlation length is finite in the BG and diverges on approaching the MI.

From the analysis of the RG flow of the mean value and relative variance of the mass distribution, we obtain the key results that (i) the instability of the MI is always towards a BG, in agreement with recent arguments \cite{125}, and that (ii) the MI/BG transitions at incommensurate and commensurate boson fillings are not in the same universality class. In the former case of incommensurate fillings, we calculate the critical exponents at 1-loop order: \( \nu = 1/D \) for the correlation length and \( \gamma = 4/D - 1 \) for the inverse compressibility.

This chapter is organized as follows. In Sec. 4.2, we argue that the relative variance of the disorder-induced mass distribution serves as the order parameter for the MI/BG transition which vanishes in the MI and acquires a finite value in the BG. We further identify the corresponding correlation length as the distance between rare regions in the BG which cause a breakdown of self-averaging in the system. The disorder-averaged replica theory describing the long-wavelength physics at strong coupling is derived in Sec. 4.3. In Sec. 4.4, the RG equations are reviewed and reformulated in a new set of variables which permits us to extract the scale dependence of the relative variance of the induced random-mass distribution. The results are presented in Sec. 4.5. The phase boundary between the MI and the BG is determined by a numerical integration of the RG equations. For incommensurate boson fillings, we obtain analytical results for the correlation length and compressibility exponents. Finally, we show that the MI/BG at commensurate boson fillings is in a different universality class. In Sec. 4.6, our main results are discussed.

### 4.2 Breakdown of Self-Averaging

The MI in the disordered BH model exists in the strong coupling limit. While the disorder must be appropriately bounded for the MI to persist \cite{1}, the key limit that defines strong coupling is the ratio \( U/t \gg 1 \), where \( U \) is the on-site energy and \( t \) the hopping matrix element. As with all strong coupling problems, the natural variables that uncloud the physics are not related straightforwardly to those in the UV-complete Hamiltonian. Further, if a critical theory is to correctly describe the destruction of the MI, it should contain the seeds of the BG phase, which Fisher et al. \cite{11} argued are due to the physics of rare regions. Consequently, the underlying critical theory might have nothing to do with the bare bosonic propagator and the associated superfluid correlation length but rather a new length scale that is intrinsically related to the disorder.

Consider the atomic limit of this problem. A scaling analysis around this regime \cite{129} demonstrates that the correlation length, \( \xi \), defined as the length scale beyond which the system encounters rare-region physics
is finite in the BG and diverges at the transition to the MI with an exponent of $\nu = 1/D$ for generic bounded disorder distributions such as the box distribution $P(\epsilon) = 1/(2\Delta)$ for $|\epsilon| \leq \Delta$ and $P(\epsilon) = 0$ for $|\epsilon| > \Delta$. The violation of the lower bound of $\nu \geq 2/D$ known as the quantum Harris criterium [130] was attributed to a breakdown of self-averaging [129] in the BG phase. That is, the fluctuations are not governed by the central-limit theorem. Several years ago, Aharony and Harris [131] showed that a break-down of self-averaging implies that the relative variance of any thermodynamic quantity must be finite. Taken in tandem, these results imply that characterizing a transition governed by the physics of rare regions requires two quantities to be finite: 1) a finite length scale, $\xi$, over which a rare region is encountered and 2) a finite value of the relative variance of any thermodynamic quantity.

The central idea we advance here is that the nature of the MI/BG transition can be understood using perturbative RG techniques but with a new set of variables and length scales that contain the information about the rare regions. The mass $r$ in the effective long wavelength theory in the clean system corresponds to the Mott gap which renders the correlations of the bosonic order parameter short ranged. Any form of disorder in the microscopic BH model, e.g. the potential disorder $\epsilon_i$ following a distribution $P(\epsilon)$, will induce an effective disorder distribution $\tilde{P}(r)$ of the mass coefficient $r$. Since in any insulating phase the superfluid order parameter must vanish on large scales, in both phases, the MI and the BG, the mean $\bar{r}$ of the distribution must diverge under the RG, $\bar{r}(\ell) \to \infty$. Hence, it cannot be used to distinguish between the MI and BG phases. Given the general considerations above, the natural quantity to distinguish the BG from the MI is the relative variance $R_r$ of the mass distribution,

$$R_r = \frac{(r - \bar{r})^2}{\bar{r}^2}. \quad (4.2)$$

In the MI phase, the system is self averaging and therefore $R_r(\ell) \to 0$ since relative variances of extensive quantities have to vanish in the thermodynamic limit as a consequence of the central-limit theorem. Since the BG is characterized by a breakdown of self averaging, $R_r$ is non-zero in the thermodynamic limit and therefore serves as the order parameter for the MI/BG transition. Here, breakdown of self-averaging refers entirely to the fact that $R_r \neq 0$. Note, $R_r(\ell) \to 0$ in the MI implies that $\xi$ diverges at the transition; that is, there is no finite length scale over which rare regions exist. We point out that a break-down of self averaging, as defined here, has nothing to do with a finite-size effect. Rather, it is a function entirely of whether or not in the thermodynamic limit, $\xi$, the distance over which rare regions exist, is finite. If it is, $R_r$ is necessarily non-zero as we will demonstrate by an explicit calculation.

To work out the RG flow of $R_r$, we will use that the disorder vertex in the disorder-averaged replica theory
is proportional to the variance of the mass distribution. We will proceed by recasting the RG equations in terms of $R_r$ and show that these equations admit a stable MI fixed point at $\bar{r} = \infty$ and $R_r = 0$ as well as a critical MI/BG fixed point at $\bar{r} = \infty$ and finite $R_r$. We point out that such an identification of fixed points is only possible because of the identification of the relevant variables. Whereas the standard equations that are used to study this problem only capture a runaway flow away from the MI/SF fixed point of the clean system \cite{1,126}, in the new variables we find that the unstable MI/SF fixed point is connected with the new critical MI/BG fixed point by a separatrix. Our analysis will further demonstrate that the MI/BG transition at commensurate fillings belongs to a different universality class.

### 4.3 Effective Replica Theory

We analyze the instability of the MI state towards the formation of the BG using the disorder-averaged replica version of the effective long-wavelength theory which is derived by an expansion around the atomic limit of the stable MI state. In this theory, the replica mixing disorder vertex is proportional to the variance of the mass distribution which allows us to extract the RG flow of the relative variance $R_r$ \cite{4.2} which we have identified as the order parameter for the MI/BG transition. In the following subsections, we derive the effective long-wavelength field theory for the weakly disordered BH model in the strong coupling limit.

#### 4.3.1 Clean system

In the clean system at sufficiently low temperature, the MI state takes place in the strong Coulomb repulsion regime, $t/U \ll 1$. In this limit, it is more useful to find a dual theory in which residual interactions are guaranteed to be small by the factor $t/U$ \cite{1,128,126,127}. This can be achieved by factorizing the off-diagonal hoping term near the atomic limit. The first step to set up a dual field theoretical description is to express the partition function, $\mathcal{Z} = \text{tr} e^{-\beta \hat{H}}$ as a coherent state path integral in imaginary time $\tau \in [0, \beta)$ over complex bosonic coherent states, $\hat{b}_i \psi_i = \psi_i \psi_i$, 

\[
\mathcal{Z} = \int \mathcal{D}[\psi, \psi^*] e^{-S^{(0)}_\psi + S^{(1)}_\psi},
\]

\[
S^{(0)}_\psi = \int_0^\beta d\tau \sum_i \left( \psi_i^* \partial_\tau \psi_i - \mu |\psi_i|^2 + \frac{U}{2} |\psi_i|^4 \right),
\]

\[
S^{(1)}_\psi = \int_0^\beta d\tau \sum_{\langle i,j \rangle} t(\psi_i^* \psi_j + \psi_i \psi_j^*).
\]
In order to decouple the nonlocal hoping term, it is convenient to perform the well-known Hubbard-Stratonovich transformation by introducing a complex auxiliary field $\phi_i(\tau)$,

$$Z = \int \mathcal{D}[\psi, \psi^*, \phi, \phi^*] e^{S_{\psi}^{(0)} - S_{\psi \phi} + S_{\phi}^{(0)}},$$  
(4.4a)

$$S_{\phi}^{(0)} = \int_0^\beta d\tau \sum_{ij} \phi_i^*(T^{-1})_{ij} \phi_j,$$  
(4.4b)

$$S_{\psi \phi} = \int_0^\beta d\tau \sum_i (\psi_i^* \phi_i + \psi_i \phi_i^*).$$  
(4.4c)

Here, we have introduced the symmetric matrix $T^{-1}$, which is the inverse of the hoping matrix whose elements equal $t$ if $i$ and $j$ are nearest neighbors and vanish otherwise. The trace over the original fields $\psi_i(\tau)$ enables the partition function to be expressed as

$$Z = Z_0 \int \mathcal{D}[\psi, \psi^*] e^{S_{\psi}^{(0)} - S'_{\phi}},$$  
(4.5)

with an additional term in the action

$$S'_{\phi} = -\ln \left\langle \mathcal{T}_\tau \exp \left[ \int_0^\beta d\tau \sum_i \left( \phi_i(\tau) \hat{b}_i^\dagger(\tau) + \text{h.c.} \right) \right] \right\rangle_0,$$  
(4.6)

where the average $\langle \cdots \rangle_0$ has to be taken with respect to the on-site Hamiltonian $\hat{H}_0 = \sum_i \left[ \frac{U}{2} \hat{n}_i(\hat{n}_i - 1) - \mu \hat{n}_i \right]$ which is diagonal in the local number occupancy basis $|\{n_i\}\rangle$. Here, $\mathcal{T}_\tau$ is the imaginary time ordering operator and $\hat{b}_i(\tau) = e^{\hat{H}_0 \tau} \hat{b}_i e^{-\hat{H}_0 \tau}$. The partition function $Z_0$ is simply the product of the on-site partition functions, $Z_0 = \text{tr} e^{-\beta \hat{H}_0}$, which can be easily computed. It should be stressed that the field $\phi_i$ can serve as an order parameter for superfluidity, since $\langle \phi_i \rangle$ is linearly coupled to $\langle \hat{b}_i^\dagger \rangle$.

In principle, the Eq. (4.6) can be expanded as

$$S'_{\phi} = \sum_{n=1}^{\infty} \frac{(-1)^{2n}}{(2n)!} \int d\tau_1 \cdots d\tau_{2n} G^{(n)}(\tau_1, \cdots, \tau_{2n}) \phi_i(\tau_1)^* \cdots \phi_i(\tau_{n+1})^* \phi_i(\tau_{n+1}) \cdots \phi_i(\tau_{2n}),$$  
(4.7)

where the connected $n$-particle Green function $G^{(n)}(\tau_1, \cdots, \tau_{2n}) = -\langle \mathcal{T}_\tau \hat{b}_i(\tau_1) \cdots \hat{b}_i(\tau_{2n}) \rangle_0$ should be evaluated in the localized limit. Note that $\hat{H}_0$ is the local on-site interacting Hamiltonian, and hence it is not correct to use Wick’s theorem. As a consequence, every connected $n$-particle Green function should be evaluated by inserting a complete set of coherent states, $1 = \sum_n |n\rangle \langle n|$. In the zero temperature limit, the ground state configuration, $|\text{Gnd}\rangle_0$ can be simply determined by finding an integer $m$ that minimizes the on-site energy $\epsilon_m = \frac{U}{2} m(m - 1) - \mu m$, which turns out to be the smallest integer larger than
Since every \( n \)-particle Green function in the local limit involves changes in the local filling \( 2n \)-times, it is not hard to expect its decay rate along the imaginary time direction to increase roughly by the factor \( U \) as the index \( n \) increases. Higher-order vertex corrections decrease sufficiently rapidly, and consequently it is sufficient to expand \( S'_\phi \) up to quartic order to see the effect of residual interactions.

After expanding \( S'_\phi \), we can take the continuum limit and perform a temporal and spatial gradient expansion to obtain the effective action

\[
S_{\text{eff}} = \int_0^\beta d\tau \int d^D r \left( K_1^{(0)} \phi^* \partial_\tau \phi + K_2^{(0)} |\partial_\tau \phi|^2 + K_3^{(0)} |\nabla \phi|^2 + R^{(0)} |\phi|^2 + H^{(0)} |\phi|^4 \right). \tag{4.8}
\]

Here, \( K_3^{(0)} = \frac{1}{2Dt} \) can be obtained from the long-wavelength expansion of \( S_\phi^{(0)} \) in the \( D\)-dimensional hypercubic lattice. The mass coefficient has contributions from both \( S_\phi^{(0)} \) and \( S'_\phi \) and is given by \( R^{(0)} = K_3^{(3)} + G(0) \), where \( G(0) = \int_\beta^0 d\tau G^{(1)}(\tau,0) \). From the insertion, \( 1 = \sum_n |n\rangle \langle n| \), it is straightforward to find

\[
G^{(1)}(\tau,0) = -(m+1)e^{-\epsilon_+\tau} \Theta(\tau) - me^{-\epsilon_-\tau} \Theta(-\tau), \tag{4.9}
\]

where \( \Theta(\tau) \) is Heavyside step function. Here, we defined \( \epsilon_+ = \epsilon_{m+1} - \epsilon_m = mU - \mu \) and \( \epsilon_- = \epsilon_{m-1} - \epsilon_m = (1-m)U + \mu \), which respectively are the energies for particle addition and removal in the atomic limit.

Using this, we obtain the mass coefficient in the zero temperature limit \( \beta U \to \infty \),

\[
R^{(0)} = \frac{1}{2Dt} \left( \frac{m+1}{\epsilon_+} + \frac{m}{\epsilon_-} \right). \tag{4.10}
\]

The temporal coefficients can be also found from the temporal gradient expansions,

\[
K_1^{(0)} = -\frac{\partial R^{(0)}}{\partial \mu}, \quad K_2^{(0)} = -\frac{1}{2} \frac{\partial^2 R^{(0)}}{\partial \mu^2}. \tag{4.11}
\]

Indeed, this result can be quickly obtained from the fact that Eq. (4.4) is invariant under the time-dependent \( U(1) \) gauge transformation:

\[
\psi_i \to \psi_i e^{i\alpha(\tau)}, \quad \phi_i \to \phi_i e^{i\alpha(\tau)}, \quad \mu \to \mu + i\partial_\tau \alpha(\tau), \tag{4.12}
\]

for an arbitrary time-dependent phase \( \alpha(\tau) \). Note that this \( U(1) \) symmetry is inherent from the original BH model.

The quartic interaction vertex can also be computed from the connected two particle Green function,
\(G^{(2)}\) considering all possible time orderings \[128\]. The result is

\[
H^{(0)} = \left(\frac{m+1}{\epsilon_+} + \frac{m}{\epsilon_-}\right) \left(\frac{m+1}{\epsilon_+^2} + \frac{m}{\epsilon_-^2}\right) - \frac{m(m-1)}{\epsilon_+^2 \epsilon_-} - \frac{(m+1)(m+2)}{\epsilon_+^2 \epsilon_-^2} \tag{4.13}\]

with \(\epsilon_{\pm}\) as defined above, \(\epsilon_{+2} = \epsilon_{m+2} - \epsilon_m = (1 + 2m)U - 2\mu\), and \(\epsilon_{-2} = \epsilon_{m-2} - \epsilon_m = (3 - 2m)U + 2\mu\).

From the effective action, Eq. (4.8), we can rescale the length and imaginary time \(\tau\) to dimensionless units, \(x = \Lambda r\) with \(\Lambda\) the momentum cutoff and \(x_0 = U\tau\). In addition, we rescale the fields as \(\phi \rightarrow \phi \sqrt{K_3\Lambda^{2-D}/U}\) to obtain

\[
S = \int_{x, x_0} \left(\gamma_1 \phi^* \partial_0 \phi + \gamma_2 |\partial_0 \phi|^2 + |\nabla \phi|^2 + r|\phi|^2 + h|\phi|^4\right), \tag{4.14}\]

and the mass coefficient which corresponds to the Mott gap, is given by

\[
r = 1 - y \left(\frac{m+1}{m-x} + \frac{m}{1-m+x}\right) \tag{4.15}\]

with \(x = \mu/U\), \(y = 2Dt/U\), and \(m\) the number of bosons per site. The interaction vertex is

\[
h = y^2 \left[\left(\frac{m+1}{m-x} + \frac{m}{1-m+x}\right) \left(\frac{m+1}{1+2m-2x} + \frac{m}{3-2m+2x}\right) - \frac{m(m-1)}{(1-m+x)^2(3-2m+2x)} - \frac{(m+1)(m+2)}{(1+2m-2x)^2(m-x)}\right]. \tag{4.16}\]

Note that \(h\) is proportional to \(y^2\) reflecting the underlying strong-coupling expansion around the atomic limit \((y = 0)\). At the mean-field level, the MI/SF phase boundary is determined by \(r = 0\). The MI states for different fillings \(m\) are characterized by a finite Mott gap, \(r > 0\), and \(\langle \phi \rangle = 0\). The SF obtains for \(r < 0\) where the SF density is proportional to \(\langle \phi \rangle \neq 0\).

The temporal gradient terms are given by \(\gamma_1 = -\partial r/\partial x\) and \(\gamma_2 = -\frac{1}{2} \partial^2 r/(\partial x)^2\) and are therefore related to the slope and the curvature of the mean-field phase boundary. At the tips of the MI lobes at values \(x_m\) of the chemical potential where the boson filling is commensurate with the lattice and the MI states are most stable, \(\gamma_1 = 0\), and the critical theory is characterized by a dynamical exponent \(z = 1\). At incommensurate fillings, \(\gamma_1 \neq 0\), corresponding to \(z = 2\) dynamics.

### 4.3.2 On-site disorder

Consider now the disordered case. The site energies \(\epsilon_i\) will be uncorrelated and chosen from a bounded distribution, \(\epsilon_i \in [-\Delta, \Delta]\). The identical HS transformation can be performed \[126\] yielding a dual theory with coefficients given by the expressions for the clean system but with the chemical potential \(x = \mu/U\)
shifted by the random potential $\epsilon_i/U$ on each site, e.g. $r_i = r(x - \epsilon_i/U)$ for the mass term. This expression is well defined on every site if the disorder distribution $P(\epsilon_i)$ is bounded, and the chemical potential $x$ lies in the intervals $[m - 1 + \delta, m - \delta]$ and $\delta = \Delta/U < 1/2$. Note that this stability condition ensures that the system exhibits stable MI phases in the presence of disorder.

The procedure for constructing the effective action is now standard: restore translational symmetry by performing the disorder average of the free energy using the replica trick [24, 33] and take the continuum limit. In the resultant expression,

$$S_{\text{eff}} = \sum_\alpha S[\phi^*_\alpha, \phi^\alpha] - \frac{\bar{g}}{2} \sum_{\alpha, \beta} \int_{x, x_0, x'_0} |\phi_\alpha(x_0, x)|^2 |\phi_\beta(x'_0, x)|^2,$$

the first term corresponds to $n$ identical copies of the disorder-averaged action with $\alpha$ the replica indices. Note that also for the disorder-averaged coefficients $\bar{\gamma}_1 = -\partial r/\partial x$ and $\bar{\gamma}_2 = -\frac{1}{2} \partial^2 r/(\partial x)^2$. The replica-mixing disorder vertex arises from the quadratic order of the cumulant expansion and is given by the variance of the mass distribution,

$$\bar{g} = (\Delta r)^2 = (r_i - \bar{r})^2.$$

As is evident, it is non-diagonal in imaginary time as a consequence of the perfect correlation of the disorder along the imaginary time direction. Since the disorder also couples to the temporal gradient terms, the replica theory contains disorder vertices with additional time derivatives [118, 124]. These terms turn out to be irrelevant and do not change the universality of the transitions between the MI and the BG and hence are not considered any further.

As a consequence of the underlying strong-coupling expansion, the variance $(\Delta r)^2$ of the mass distribution is not simply given by the variance of the random-site disorder in the BH model [126]. This is easily seen by expanding $r_i = r(x - \epsilon_i/U)$ for small $\epsilon_i/U$ yielding $(\Delta r)^2 = \gamma_1^2 U^2 + \gamma_2^2 (U^2 - \bar{e}^2) + \ldots$ with $\bar{e}$ the $n$-th moment of $P(\epsilon) = P(-\epsilon)$. Hence, the effective disorder $\bar{g}$ is a continuous function of the chemical potential entering via the coefficients $\gamma_{1,2}$ and is strongly suppressed at the tips of the MI lobes where $\gamma_1 = 0$. Therefore, the inclusion of higher moments of the distribution $P(\epsilon)$ is essential not only to account for the boundedness of the distribution but also to capture the coupling to disorder at commensurate fillings. In order to retain the moments to infinite order, without loss of generality, we use the discrete distribution

$$P(\epsilon) = (1 - 2p)\delta(\epsilon) + p\delta(\epsilon - \Delta) + p\delta(\epsilon + \Delta),$$

corresponding to on-site energies which occur with the same probability $p < 1/2$ increased or decreased by
\( \Delta < U/2 \). The resulting mean \( \bar{r} \) and variance \( \bar{g} \) of the induced mass distribution entering the replica theory (4.17) are given by

\[
\bar{r} = \frac{r(x - \epsilon_i/U)}{1 - 2p} r(x) + p[r(x - \delta) + r(x - \delta)], \quad (4.20)
\]

\[
\bar{g} = \frac{(1 - 2p)r^2(x) + p[r^2(x - \delta) + r^2(x + \delta)] - \bar{r}^2}{1 - 2p} \bar{r}^2(x), \quad (4.21)
\]

and contain the moments of \( P(\epsilon) \) to infinite order. The MF phase boundary between the \( m = 1 \) MI and the SF obtained by the condition \( \bar{r} = 0 \) is shown in Fig. 1(a). Whereas at incommensurate fillings the interaction vertex \( \tilde{h} \) is irrelevant for \( D > 2 \), the disorder vertex is always relevant in \( D < 4 \) rendering the MF theory meaningless. In order to determine the stability region of the MI and subsequent phase transitions, we employ the RG.

4.4 Renormalization Group Analysis

We proceed with the RG analysis of the effective replica field theory (4.17). From the RG flow of the mean \( \bar{r} \) and the variance \( \bar{g} \) of the disorder-induced mass distribution, we can subsequently determine the scale dependence of the relative variance \( R_r = \bar{g}/\bar{r}^2 \) to analyze the instability of the MI towards the formation of a BG.

4.4.1 Renormalization Group Equations

After successively eliminating modes of highest energy corresponding to momenta from the infinitesimal shell \( e^{-d\ell} \leq |k| \leq 1 \) and rescaling of momenta \( (k \rightarrow ke^{d\ell}) \), frequencies \( (\omega \rightarrow \omega e^{z\ell}) \), and fields \( (\phi_\alpha \rightarrow \phi_\alpha e^{\lambda\ell}) \), we obtain to one-loop order [126]

\[
\frac{d\bar{r}}{d\ell} = 2\bar{r} + 2I_1\tilde{h} - I_0\tilde{g}, \quad (4.22a)
\]

\[
\frac{d\gamma_1}{d\ell} = (2 - z)\gamma_1 + I_0^2\gamma_1\tilde{g}, \quad (4.22b)
\]

\[
\frac{d\gamma_2}{d\ell} = (2 - 2z)\gamma_2 + I_0^2(I_0\gamma_1^2 + \gamma_2)\tilde{g}, \quad (4.22c)
\]

\[
\frac{d\tilde{h}}{d\ell} = (4 - D - z)\tilde{h} - (J_0 + 4J_1)\tilde{h}^2 + 6I_0^2\tilde{g}\tilde{h}, \quad (4.22d)
\]

\[
\frac{d\tilde{g}}{d\ell} = (4 - D)\tilde{g} + 4I_0^2\tilde{g}^2 - 4J_1\tilde{g}\tilde{h}, \quad (4.22e)
\]
where the scaling dimension $\lambda$ of the fields has been determined such that the coefficient of the spatial gradient term remains constant. Further, we have absorbed a numerical factor arising from the shell integration by redefining $\tilde{h} = 2S_D/(2\pi)^D\tilde{h}$ and $\tilde{g} = S_D/(2\pi)^D\tilde{g}$ with $S_D$ the surface of the $D$-dimensional unit sphere.

The frequency integrals are defined as

\[ I_0 = C(0) = 1/(1 + \bar{r}), \quad (4.23a) \]
\[ I_1 = \int_\omega C(\omega) = [\tilde{\gamma}_1^2 + 4\tilde{\gamma}_2(1 + \bar{r})]^{-1/2}, \quad (4.23b) \]
\[ J_0 = \int_\omega C^2(\omega) = \frac{1}{2}I_0I_1, \quad (4.23c) \]
\[ J_1 = \int_\omega |C(\omega)|^2 = 2\tilde{\gamma}_2I_1^3, \quad (4.23d) \]

with $\int_\omega = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi}$ and $C(\omega) = 1/(1 + \bar{r} - i\tilde{\gamma}_1\omega + \tilde{\gamma}_2\omega^2)$ the on-shell propagator.

### 4.4.2 New Variables

From the RG equations, it is evident that for $D < 4$, the variance $\tilde{g}$ increases with the scale which suggests that the RG flow is always towards strong disorder. This interpretation is inconsistent with perturbative arguments which show that the disordered BH model exhibits stable MI phases in the regime of weak disorder [1], and does not take into account that in the insulating phases the mean $\bar{r}$ also diverges under the RG.

Therefore the question of whether disorder is strong or not can only be answered by analyzing the relative variance $R_r \sim \tilde{g}/\bar{r}^2$. In the stable MI phase, disorder is irrelevant and we expect that $R_r(\ell) \to 0$.

Realizing that $\bar{r}$ diverges for both insulating phases, we implement the approach of Aharony and Harris [131] by focusing on the relative variance $R_r$ of the disorder-induced mass distribution. As we argued in Sec. 4.2, $R_r$ serves as the order parameter for the MI/BG transition characterized by a breakdown of self-averaging.

Putting this into practice in the RG equations requires a new set of variables. Instead of the mass $\bar{r}$, we use $I_0 = 1/(1 + \bar{r})$ which corresponds to the on-shell propagator in the static limit, $C(0)$, and asymptotically becomes the inverse mass in the insulating phases where $\bar{r}(\ell) \to \infty$. In addition, we work with $\Gamma_i = I_0\gamma_i$ ($i = 1, 2$) for the temporal gradient terms and $H = I^2_0\tilde{h}$, $G = I^2_0\tilde{g}$ for the interaction and disorder vertices, respectively. In the following, we identify the MI phase by the conditions that $\bar{r}(\ell) \to \infty$ but $R_r(\ell) \to 0$.

In the BG, $R_r(\ell) \neq 0$. In the large mass regime, $G$ corresponds precisely to the relative variance $R_r$ of the mass distribution.
4.5 Results

4.5.1 Numerical Integration

We start with a numerical integration of the full set of RG equations \(\text{(4.22)}\) in terms of the variables introduced above to see if we can indeed identify the MI/BG transition. Since \(\tilde{\gamma}_1 \neq 0\) away from the tips of the Mott lobes, we use a dynamical exponent \(z = 2\). Note that for small \(\tilde{r}\) close to the MI-SF transition in the clean system, the spatial gradient term \(\gamma_x = 1\) serves as a cut-off in \(C(\omega)\) and the corresponding frequency integrals.

Indeed we find a regime where the system flows towards a stable MI fixed point \(P_{\text{MI}}\) with \(I_0 = 0\) and \(G = 0\). Therefore, in the MI phase the mean \(\bar{r}\) of the mass distribution diverges fast enough that the relative variance \(R_r\) of the distribution renormalizes to zero. As expected, \(\Gamma_{1,2}\) and \(H\) renormalize to zero in the MI.

Using the analytic expressions for the bare coefficients as a function of the microscopic parameters, we obtain the MI phases in the presence of disorder. Fig. 1(a) shows the resulting \(m = 1\) MI lobe in \(D = 3\) as a function of \(x = \mu/U\) and \(y = 2Dt/U\) for \(p = 0.1\) and \(\delta = 0.2\). Sufficiently close to the MI state, \(G\) diverges at a certain scale \(\ell^*\) whereas \(I_0(\ell^*)\) is practically zero. This clearly identifies the adjacent phase as an insulating BG! Note that the mass distribution spreads faster than it shifts to infinity. Hence it is the tail of the distribution which destabilizes the MI, indicative of rare-region physics.

In the following, we determine the phase boundary \(y_c(x)\) by bisection and follow the divergence of the correlation length \(\xi = e^{\ell^*}\) as \(y \to y_c^+\). Note that the correlation length defined here is determined by the scale \(\ell^*\) where \(R_r\) diverges and self-averaging breaks down. The correlation length \(\xi\) therefore corresponds to the distance of rare regions in the BG which cause the breakdown of self-averaging (see Sec. 4.2). Sufficiently close to the MI, we expect a power-law divergence, \(\xi \sim (y - y_c)^{-\nu}\), defining the correlation-length exponent \(\nu\). To extract \(\nu\), in Fig. 1(b) we plot \(\partial(\ln \xi)/\partial(\ln(y - y_c))\) as a function of \(\ln(y - y_c)\) for different values of \(x\). For \(y \to y_c^+\), the functions indeed converge to a constant value corresponding to \(\nu = 1/3\).

Interestingly, on approaching commensurate boson fillings, the length scale upon which universal behavior is observed increases and possibly even diverges. Whereas an increase might be simply a quantitative effect originating from the suppression of the bare disorder vertex by commensuration, a divergence would indicate a different universality class at the tips of the MI lobes.

4.5.2 Universality of the MI/BG Transition at Incommensurate Filling

To better understand the nature of the transitions we proceed with an analytical investigation of the incommensurate case \((z = 2)\). Since for \(D > 2\) the interaction vertex \(\tilde{h}\) is irrelevant at the MI-SF transition of the
Figure 4.1: (a) Phase boundary between the MI \((m = 1)\) and the BG obtained from numerical integration of the RG equations. The disorder corresponds to site energies increased or decreased by \(\delta = \Delta / U = 0.2\) with probability \(p = 0.1\). The solid line shows the unphysical mean-field phase boundary \((\bar{r} = 0)\) between the MI and SF. (b) Extraction of the correlation length exponent for different values \(x\) of the chemical potential as indicated in (a).

In the clean system, in the vicinity of the fixed point \(P_{\text{MI-SF}} (G = 0, I_0 = 1)\), we can simplify the RG equations by setting \(\bar{h} = 0\). The resulting RG equations for \(I_0\) and \(G\),

\[
\frac{dI_0}{d\ell} = (G - 2) I_0 + 2 I_0^2, \quad (4.24a)
\]

\[
\frac{dG}{d\ell} = (4 I_0 - D) G + 6 G^2, \quad (4.24b)
\]

are decoupled from \(\Gamma_{1,2}\). Our analysis of these equations is strictly in the perturbative regime where \(G(\ell) \leq 1\). The divergence of \(I_0\) which obtains for \(G(\ell) > 2\) is spurious and outside the range of validity of the perturbative RG equations. In addition to \(P_{\text{MI-SF}}\) and the stable MI fixed point \(P_{\text{MI}} (I_0 = 0, G = 0)\), the RG equations \((4.24)\) exhibit an additional fixed point \(P_{\text{MI-BG}} (0, D/6)\), which will turn out to be critical fixed point determining the universality of the MI-BG transition at incommensurate fillings. From linearization around \(P_{\text{MI-BG}}\) we find \(dH/d\ell = (D/3 - 2) H\) indicating that the interaction vertex remains irrelevant.

For \(G = 0\) we obtain a trivial flow equation for \(I_0\) capturing the MI-SF of the clean system. For \(I_0 < 1\), the flow is towards \(P_{\text{MI}}\), while for \(I_0 > 1 \ (\bar{r} < 0)\), \(I_0(\ell) \to \infty\) signaling the instability toward the formation of a SF.
Figure 4.2: RG flow for (a) $D = 3$ and (b) $D = 5$ for incommensurate boson fillings as a function of $I_0$ and $G$ corresponding (in the limit of large mass $\bar{r}$) to the inverse mean and the relative variance of the mass distribution, respectively.

For $D < 4$, $P_{\text{MI-SF}}$ is unstable for small disorder, $G > 0$. However, instead of a runaway flow, we find a separatrix $G_s(I_0)$ which connects $P_{\text{MI-SF}}$ and $P_{\text{MI-BG}}$ as shown in the case $D = 3$ in Fig. 2(a). Linearizing around the fixed-points, we find that asymptotically $G_s(I_0) \simeq (D-2)(1-I_0)$ in the vicinity of $P_{\text{MI-SF}}$ and $G_s(I_0) \simeq D/6 - 4I_0/(12-D)$ close to $P_{\text{MI-BG}}$, respectively.

For initial values $I_0(0) < 1$ and $G(0) < G_s(I_0(0))$, the system flows towards the MI fixed point, whereas for $G(0) > G_s(I_0(0))$, $G(\ell)$ will eventually diverge. For initial values very close to the MI phase, the trajectory $(I_0(\ell), G(\ell))$ will track the separatrix and consequently, sufficiently close to the transition, the divergence of $G(\ell)$ is controlled by the fixed point $P_{\text{MI-BG}}$. Linearizing around this point, we find $G(\ell) = (\Delta G)e^{D\ell}$ with $\Delta G = G_0 - G_s$ the infinitesimal distance to the separatrix. From the condition $G(\ell^*) \simeq 1$ we obtain the correlation length $\xi \sim (\Delta G)^{-1/D}$, implying a correlation length exponent of $\nu = 1/D$, in perfect agreement with the value extracted from numerical integration of the full set of RG equations for $D = 3$ (see Fig. 1(b)) and a previous real-space RG and scaling analysis \[113\] \[129\].

Since $G \neq 0$ at the fixed point and in the BG, the fluctuations are no longer \[131\] governed by the central-limit theorem. This implies that self-averaging breaks down. Consequently, the bound $\nu \geq 2/D$ \[130\] no longer applies. That the BG is mediated by rare localized regions is the efficient cause of this effect. In quantum systems, such rare events play a more pronounced \[132\] role than in classical systems because of the perfect correlation of the disorder along the imaginary time direction.
It is also the divergence of the relative variance that is responsible for a finite compressibility \( \kappa = (\beta U)^{-1} \partial^2 \ln \bar{Z} / \partial \bar{r}^2 \) in the BG. Using the asymptotic form of \( S_{\text{eff}} \) in which only the \( \bar{r} \) and \( \bar{g} \) terms are retained, we obtain to leading order \( \kappa \approx 1 / \bar{r}^2 (1 + 4 \bar{g} / \bar{r}^2) \) which clearly vanishes in the MI. However, in the BG sufficiently close to the transition, \( \kappa \equiv \kappa(l^*) \sim \xi^{D-4} \sim (y - y_c)^\gamma \) with \( \gamma = 4 / (D - 1) \).

For completeness, we briefly discuss the case \( D > 4 \) in which a small amount of disorder is irrelevant at the MI-SF transition. The inversion of the RG flow along the \( G \)-direction is a consequence of the presence of an additional unstable fixed point \( Q(1 - \epsilon/8, \epsilon/4) \) in \( D = 4 + \epsilon \) located on the separatrix (see Fig. 2(b)). Therefore, in \( D > 4 \) the disorder has to exceed a critical value in order to induce a BG phase in contrast to recent claims \[125\].

### 4.5.3 Commensurate Filling

We now focus on \textit{commensurate} boson fillings to understand the nature of the transition at the tips of the MI lobes where \( \bar{g}_1 = 0 \) and \( z = 1 \). The smaller dynamical exponent makes the interaction vertex \( H = I_0^2 \bar{h} \) more relevant and in fact marginal at the MI-SF transition of the clean system in \( D = 3 \). However, for \( D = 4 - \epsilon \), \( H \) remains irrelevant at \( P_{\text{MI-SF}} \) suggesting that the critical behavior is controlled by the same separatrix as for incommensurate fillings. Interestingly, this is not the case. A numerical integration of the RG equations shows that in the disordered phase close to the MI, \( H(l) \) diverges simultaneously with \( G(l) \). This behavior is easily understood upon linearizing around \( P_{\text{MI-BG}} \) yielding \( dH / dl = (D/3 - 1)H \) which demonstrates that \( H \) becomes relevant for \( D > 3 \). Therefore, for any finite value \( H(0) > 0 \), the separatrix obtained by projection into the \( I_0 - G \) plane does not terminate in \( P_{\text{MI-BG}} \) but diverges as \( G_s \sim I_0^{-(D/(12-D))} \). Consequently, the MI-BG transition at commensurate fillings is not controlled by \( P_{\text{MI-BG}} \) but by a different fixed point which is not accessible in the present calculation. However, from the divergence along the separatrix, \( G_s \sim e^{2l} \) we estimate \( \nu = 6/D \) which is significantly larger than the incommensurate value and satisfies the bound \( \nu \geq 2/D \) \[130\].

### 4.6 Final Remarks

To summarize, we have analyzed the instabilities of the MI state of the weakly disordered BH model within an RG analysis of the disorder-averaged replica theory valid in the strong-coupling regime. Our analysis shows that the MI always becomes unstable towards a BG regardless of the boson filling, in agreement with recent work \[125\] which, based on a mathematical theorem, excludes the possibility of a direct MI/SF transition in the presence of disorder.
While in both insulating phases the mass coefficient diverges on large scales, rendering SF correlations short ranged, the relative variance of the induced random mass distribution renormalizes to zero in the MI and diverges in the BG. This allowed us to demonstrate that the relative variance serves as the order parameter for the MI/BG transition. We further extracted the diverging correlation length of the MI/BG transition from the scale where the relative variance diverges under the RG. Since the finite relative variance of an extensive quantity signals a breakdown of the central-limit theorem \[131\], the correlation length corresponds to the separation between rare regions in the BG, causing a breakdown of self-averaging in the system. We point out that the correlation length defined here has nothing to do with the SF correlation length which remains finite over the MI/BG transition.

The reformulation of the RG equations in terms of new variables, in particular the relative variance, allows us to identify fixed points and to subsequently analyze the universality of the MI/BG transition. That new variables, not apparent in the UV-complete theory, are necessary to access the fixed point is typical of strong-coupling problems. At incommensurate fillings, we find a separatrix which connects the MI/SF fixed point in the clean system with the critical MI/BG fixed point in the weakly disordered system. The correlation-length and compressibility exponents characterizing the MI/BG transition are determined as \(\nu = 1/D\) and \(\gamma = 4/D - 1\), respectively. In principle, the power-law divergence of the inverse compressibility could be verified either experimentally or numerically provided that the critical region can be accessed sufficiently. The violation of the strict lower bound \(\nu \geq 2/D\) known as the quantum Harris criterium \[130\] has been argued some time ago \[131, 129\] to be a key signature for the breakdown of self averaging. Interestingly, our correlation-length exponent \(\nu = 1/D\) is identical to the exponent estimated by a scaling analysis in the atomic limit for a box distribution of site energies \[129\].

Our analysis shows that the MI/BG transition at commensurate fillings is controlled by a different fixed point which is not accessible within the present RG approach \[133\]. The two distinct transitions found here are, in retrospect, not out of the ordinary, since in the clean system, the MI/SF transitions are in the universality class of the \((D+1)\)-dimensional XY model \[134, 135\] and of mean-field type \[1\] for commensurate and incommensurate fillings, respectively. This dichotomy persists in the disordered system as well.

What this analysis indicates is that within a replica theory, instabilities driven by local rare regions, indicative of Griffiths singularities, are completely accessible if one focuses on the scaling dependence of the relative variance of the mass distribution. Since the relative variance of any extensive quantity acquires a non-zero value \[131\] once self-averaging breaks down, the method presented here can in principle be applied to any transition driven by local rare-region physics.
Chapter 5

Tuning thermoelectric power factor by crystal-field and spin-orbit couplings in Kondo lattice materials

5.1 introduction

Thermoelectrics support a voltage drop in response to a modest temperature gradient. Since a temperature gradient affects the electrons and the lattice degrees of freedom, optimizing thermoelectrics involves not only the thermopower or Seebeck coefficient ($S$), but also the electrical ($\sigma$) and thermal ($\kappa$) conductivities. The holy grail of thermoelectrics is to achieve a figure of merit

$$ZT = \left( \frac{S^2\sigma}{\kappa} \right) T$$  \hspace{1cm} (5.1)

that exceeds unity at room temperature. This tall order remains a grand challenge problem [136, 137, 138, 139]. Two promising recent directions have focused on either decreasing the thermal conductivity as in the case of nanocrystalline arrays of $\text{Bi}_x\text{Sb}_{2-x}\text{Te}_3$ in which a $ZT$ of 1.4 was achieved [137] at $T = 373K$ or maximizing the numerator of $Z$ through strong electron correlations. An example of the latter is the report [140] that FeSb$_2$ achieves a colossal thermopower of 45000$\mu$V/K at 10K resulting in the largest power factor, $S^2\sigma$ witnessed to date. In this paper, we follow-up on the role strong correlations play in maximizing the power factor by focusing on Kondo insulators. We show explicitly that multi-orbital physics in Kondo insulators lies at the heart of the problem of maximizing the power factor.

Because the thermopower is related to the entropy per carrier, particle-hole asymmetry and large density of states at the chemical potential are central to the optimization of $Z$. In this regard, the Anderson model of a single impurity in a metal [111], which is among the few solvable strongly correlated systems solvable exactly, presents a density of states with demanding features for efficient thermoelectric transport. For a single $SU(2)$ spin on a localized impurity, the density of states appears as a single infinite symmetric peak at the chemical potential leading to a divergent density of states but vanishing Seebeck coefficient by virtue of the particle-hole symmetry. Increasing the degeneracy of the localized orbital and the metallic band to $SU(N)$ ($N > 2$) softens the peak in the density of states and at the same time moves above the chemical
potential leading to an asymmetric density of states and as a result a larger Seebeck coefficient [31].

It makes sense then to consider systems in which such physics is naturally present, for example Kondo insulators in which a regular lattice of Anderson impurities is hybridized with multiple bands of itinerant electrons. The electrons in the local orbitals are poorly screened and the strong Coulomb repulsion prohibits them from being multiply occupied. Contrary to the single impurity, the periodic Anderson model is not exactly solvable but multiple mean-field type methods were used [50, 51, 52] to understand many of their features. Motivated by the single impurity model, we examine the effect of degeneracy of the local impurities and the conduction band on the thermoelectric properties of Kondo insulators. In addition to directly studying the degeneracy of the local and conduction bands, we study the effect lifting the degeneracy by a crystal-field (which mainly effects the local orbitals) and spin-orbit coupling (which mainly effects the conduction band) have on thermoelectric efficiency. In this way, we can continuously lift the level of degeneracy. Interestingly, we observe that there is an optimum value of the crystal field and spin orbit coupling. As was shown in a previous study [142], the presence of multiple orbitals close to the chemical potential is a common feature of Kondo insulators. Our results indeed present a possible route for using strong correlations to enhance the thermoelectric performance through controlling the orbital degeneracy of local and itinerant bands.

5.2 Model and methodology

Heavy fermion materials typically contain rare earth or actinide ions forming a lattice of localized magnetic moments [49]. The strong Coulomb repulsion of electrons localized in f or d orbitals leads to formation of these local moments [141] which then hybridize with the itinerant electron bands and form the heavy electron bands. If the chemical potential is in the heavy electron bands, a heavy fermion metal is formed. The volume of the Fermi surface in this correlated state corresponds to a sum of the number of itinerant and localized electrons. If the chemical potential is in the hybridization gap, the heavy electron band will be fully occupied and a Kondo insulator obtains [143, 144]. Notice that such an insulating state is fundamentally different from a non-interacting insulator. For example, in order to reach a filled valance band, we need to add the number of localized and itinerant electrons which is solely developed as a result of strong interactions.
The underlying microscopic model of this correlated system is:

\[
H = \sum_{kls\sigma'} \varepsilon_{\sigma\sigma'}(k)c_{kl\sigma}^\dagger c_{kl\sigma'} + \sum_{kls} \varepsilon_{fi} d_{kl\sigma}^\dagger d_{kl\sigma} + \sum_{ikl\sigma} \left( V_{kl\sigma} e^{ikr} c_{kl\sigma}^\dagger d_{i\sigma} + \text{h.c.} \right) + \frac{U}{2} \left( \sum_{i\sigma} n_{i\sigma}^d n_{i\sigma'}^d + \sum_{i\neq i',\sigma\sigma'} n_{i\sigma}^d n_{i'\sigma'}^d \right)
\]

where \( c_{kl\sigma}^\dagger \) (\( d_{kl\sigma}^\dagger \)) is the creation of a conduction(local) electron with momentum \( k \), orbital \( l \), and spin \( \sigma = (\uparrow, \downarrow) \), and \( n_{i\sigma}^d = d_{i\sigma}^\dagger d_{i\sigma} \) is the number operator of a local \( d \) orbital at site \( r_i \). The dispersion of the \( c \)-electron \( \varepsilon_{\sigma\sigma'}(k) = \epsilon_k \delta_{\sigma\sigma'} + \Gamma_k \cdot \sigma_{\sigma\sigma'} \) includes spin coupling. The non-dispersive energy of local states \( (\epsilon_{fi}) \) depends on the orbital index \( l \). The pseudovector \( \Gamma_k \) represents the amplitude of the spin-orbit (SO) coupling [145, 146] and its form depends on the crystal symmetry of the underlying lattice. Typically, the hybridization matrix element, \( V_{kl\sigma} \) encodes the complex orbital structures of local states which can have novel effects on the properties of the strongly correlated heavy fermion phase [147], but as in other studies, we consider \( V_{kl\sigma} \) to be independent of \((k, \sigma)\) to make the calculation more tractable.

**U(1) slave-particle renormalization**

Using the model Hamiltonian, Eq. (5.2), we can capture the effect of the degeneracy of both localized and itinerant bands, as well as the effect of crystal field and spin orbit coupling in breaking the degeneracy of these bands. As a result of weak screening of electrons in \( f \) and \( d \) orbitals, the associated on-site repulsive potential \( U \) is much larger than the hopping energies of the itinerant electrons. To treat the large on-site repulsion term, we use the \( U(1) \) slave-boson mean-field theory [51, 148]. In this treatment, the creation operator of a local electron \( d_{i\sigma}^\dagger = f_{i\sigma}^\dagger b_i \) is partitioned into a neutral fermion \( f_{i\sigma}^\dagger \), and a charged boson \( b_i \) that accounts for annihilation of an empty state. Since the local Hilbert space is restricted to either an empty or a singly occupied state, the additional local constraint,

\[
\tilde{Q}_i = b_i^\dagger b_i + \sum_{i\sigma} f_{i\sigma}^\dagger f_{i\sigma} = 1
\]

should be enforced at every site \( r_i \). The Hamiltonian in terms of these slave particles then becomes

\[
H = \sum_{kl\sigma\sigma'} \varepsilon_{\sigma\sigma'}(k)c_{kl\sigma}^\dagger c_{kl\sigma'} + \sum_{k\sigma} \varepsilon_{fi} f_{k\sigma}^\dagger f_{k\sigma} + \sum_{i\sigma} \left( V_i^* e^{-i k r} f_{i\sigma}^\dagger b_i c_{kl\sigma} + \text{h.c.} \right) + \sum_{i} \lambda_i (\tilde{Q}_i - 1)
\]

where \( \lambda_i \) is a Lagrange multiplier to maintain the local constraint. In the above Hamiltonian, the effect of the crystal field is to break the degeneracy of the local orbital states \( \epsilon_{fi} \) whereas the spin-orbit coupling breaks
the spin degeneracy of the conduction band. As a result, by tuning the crystal field and spin-orbit coupling, we can change the degeneracy of the local and conduction orbitals in a continuous manner. Consequently, we have a tunable knob to gain the optimum thermoelectric performance.

The mean field approximation to the model Hamiltonian can be obtained by taking the coherent expectation \( b = \langle b_i \rangle = \langle b_i^\dagger \rangle \) and \( \lambda = \langle \lambda_i \rangle \). This replacement effectively renormalizes the mixing matrix element \( V_l \rightarrow bV_l^* \), and the local energy \( \epsilon_{f_l} \rightarrow \epsilon_{f_l} + \lambda \) and leads to the quadratic Hamiltonian

\[
H_{MF} = \sum_{k,l,h} (\epsilon_k + h|\Gamma_k|) c_{k,l,h}^\dagger c_{k,l,h} + \sum_{k,l,\sigma} (\epsilon_{f_l} + \lambda) f_{k,l,h}^\dagger f_{k,l,h} \\
+ \sum_{k,l,h} (bV_l^* f_{k,l,h}^\dagger c_{k,l,h} + \text{h.c.}) + \lambda \sum_i (b^2 - 1).
\] (5.5)

Instead of working in the spin basis, we will use a helical basis that diagonalizes the single-electron dispersion \( \epsilon_{\sigma\sigma'}(k) \rightarrow [U_k^\dagger \epsilon(k) U_k]_{hh'} = (\epsilon_k + h|\Gamma_k|) \delta_{hh'} \) with \( h, h' = \pm 1 \). Then \( c_{k,lh}(f_{k,lh}) \) is accordingly rotated by the unitary matrix \( U_k \) from the spin basis, \( c_{k,l\sigma}(f_{k,l\sigma}) \). By performing the Bogoliubov transformation,

\[
a_{k,lh}^\pm = \alpha_{k,lh} c_{k,lh} + \beta_{k,lh} f_{k,lh} \quad (5.6)
\]
\[
a_{k,lh}^- = -\beta_{k,lh} c_{k,lh} + \alpha_{k,lh} f_{k,lh}, \quad (5.7)
\]

we obtain the diagonal mean-field Hamiltonian,

\[
H_{MF} = \sum_{k,l,h,\pm} E_{k,lh}^\pm a_{k,lh,\pm}^\dagger a_{k,lh,\pm} + \lambda \sum_i (b^2 - 1), \quad (5.8)
\]

where the dispersion is given by

\[
E_{k,lh}^\pm = \frac{1}{2} (\epsilon_k + h|\Gamma_k| + \epsilon_{f_l} + \lambda \pm W_{k,lh}), \quad (5.9)
\]
\[
W_{k,lh} = \sqrt{(\epsilon_k + h|\Gamma_k| - \epsilon_{f_l} - \lambda)^2 + 4b^2 V_l^2}. \quad (5.10)
\]

The Bogoliubov parameters are

\[
\begin{pmatrix}
\alpha_{k,lh}^2 \\
\beta_{k,lh}^2
\end{pmatrix} = \frac{1}{2} \left[ 1 \pm \frac{(\epsilon_k + h|\Gamma_k|) - (\epsilon_{f_l} + \lambda)}{W_{k,lh}} \right]. \quad (5.11)
\]

Minimization of the free energy with respect to the mean field parameters \( b \) and \( \lambda \), and the total chemical
potential $\mu$ leads to two coupled equations

\begin{align}
1 &= b^2 + \sum_{k_{lh}} \alpha_{k_{lh}}^2 n_F(E_{k_{lh}+}) + \beta_{k_{lh}}^2 n_F(E_{k_{lh}-}), \quad (5.12) \\
\lambda &= \sum_{k_{lh}} \frac{V^2}{W_{k_{lh}}} \left[ n_F(E_{k_{lh}-}) - n_F(E_{k_{lh}+}) \right], \quad (5.13) \\
n_{tot} &= \sum_{k_{lh}} \left[ n_F(E_{k_{lh}-}) + n_F(E_{k_{lh}+}) \right], \quad (5.14)
\end{align}

where, the total density of electrons is fixed to be $n_{tot} = 2l_{max}$ for $l = 1, 2, \cdots, l_{max}$. The transport of this non-interacting mean-field Hamiltonian is now tractable.

**Relaxation time approximation**

To compute the transport properties, we use the relaxation-time approximation to the Boltzmann equation \[149\]. Under this scheme, the electrical resistivity, $\rho = \sigma^{-1}$, and the thermopower tensors, $S$, are explicitly given by

\begin{align}
\rho &= L^{-1}_0, \quad S = -\frac{k_B}{|e|} L^{-1}_0 L_1, \quad (5.15)
\end{align}

where the tensors $L_m$ are

\begin{align}
(L_m)_{ab} &= -\frac{e^2}{V_{\text{volume}}} \sum_{k_{lh}} \frac{\partial n_F(E_{k_{lh}+})}{\partial E_{k_{lh}+}} \tau_{k_{lh}}(v_{k_{lh}+})_a (v_{k_{lh}+})_b \left( \frac{E_{k_{lh}+} - \mu}{k_B T} \right)^m. \quad (5.16)
\end{align}

Here we set $v_{k_{lh}+} = \frac{1}{\hbar} \nabla E_{k_{lh}+}$. Considering that the electrons are scattered by $N_{\text{imp}}$ impurities with an interaction strength of $V_{\text{imp}}$ (i.e., $H_{\text{scatt}} \sim V_{\text{imp}} c^\dagger_{k_l \sigma} c_{k_l \sigma}$), the relaxation time $\tau_{k_{lh} \pm}$ for each state is given by

\begin{align}
\frac{1}{\tau_{k_{lh} \pm}} &= \frac{2\pi N_{\text{imp}}}{h} |V_{\text{imp}}|^2 \left[ \frac{\partial E_{k_{lh} \pm}}{\partial (\epsilon_k + \hbar |\Gamma_k|)} \right]^2 \rho_{lh}(E_{k_{lh} \pm}), \quad (5.17)
\end{align}

with $\rho_{lh}(E_{k_{lh} \pm})$ the density of the states of the Bogoliubov quasiparticles.

**5.3 results**

We now present our results on the dependence of the transport properties on the orbital degeneracies of both localized and itinerant electron bands which form correlated Kondo insulators. In the first two subsections, we consider double degeneracy of the conduction and localized bands. This model is indeed consistent with the models previously proposed for Kondo insulators \[150\]. The crystal field will then split the degeneracy...
of the two $f$ levels in $5.4$ and spin-orbit coupling breaks the degeneracy of the conduction band states with differing helicity [146]. In these two sections, we change the size of the degeneracy-breaking gap continuously. Using the relaxation-time approximation, we can then calculate the transport properties of a Kondo insulator. For most of the materials, the dominant contribution to the thermal conductivity comes from lattice vibrations; as a consequence, the electronic contribution to the thermoelectric performance is measured through the power factor $Z_{\text{PF}} = \sigma S^2$ where $\sigma$ is the electrical conductivity and $S$ is the Seebeck coefficient. In order to confirm that the enhancement of the thermoelectric efficiency can properly be attributed to strong correlations, we consider two different band structures of itinerant electrons in $5.3.1$ and $5.3.2$ and we see that similar features emerge.

Finally, in $5.3.3$ we present the effect of multiple orbital degeneracy. Contrary to the treatment in $5.3.1$ and $5.3.2$ where the double degeneracy is continuously lifted by crystal field and spin-orbit couplings, in section $5.3.3$ we discretely change the number of degenerate conduction and localized bands. We show that indeed there is also an optimum degeneracy associated with the maximum power factor.

5.3.1 Nearly free electron itinerant bands

We first focus on the effect of crystal field and spin-orbit coupling on the power factor within the context of a parabolic band for the itinerant electrons $\epsilon_k = \epsilon_0 + W(k/k_{BZ})^2$ with $W = 2\text{eV}$, taken from ref. [151]. In principle, the SO coupling should be expressed as a periodic function under the crystal environment, but we model it to be isotropic as well, $\Gamma_k = \gamma_{\text{so}}(k/k_{BZ})$ ($\gamma_{\text{so}} \leq 0.2\text{eV}$). In the following, we carry out the numerical calculation based on this isotropic band dispersion with $l_{\text{max}} = 2$. Here, we choose $\epsilon_{f_1} = 1.0606\text{eV}$, $V_1 = 0.2236\text{eV}$ and $V_2 = 1.05V_1 = 0.2348\text{eV}$. The parameter $\frac{N_{\text{imp}}}{N_{\text{site}}}|V_{\text{imp}}|^2 = 0.045\text{eV}^2$ [151]. The other control parameters are temperature ($T \leq 100K$), crystal electric field (CEF) splitting ($\Delta_{\text{CEF}} = \epsilon_{f_2} - \epsilon_{f_1} \leq 15\text{meV}$), and the SO coupling ($\gamma_{\text{so}} \leq 0.15\text{eV}$). Although we do not include the supporting data here, we found that our conclusions are insensitive to the strength of $V_2$ as long as $0.5 \lesssim V_2/V_1 \lesssim 2.0$.

Figure 5.1 shows the density of states (DOS) for different control parameters. From Figs. 5.1(a) to (d), we notice that the temperature only controls the number of thermally activated charge carriers, while it does not significantly change the DOS compared to the other parameters. When the degeneracy of the two local orbitals is broken by the crystal electric field, one of the hybridized bands moves closer to the chemical potential. Consequently, the system is driven from an insulator to a conductor [Figs. 5.1(e)-(h)], at which point the power factor is significantly enhanced (See Fig. 5.2). Likewise, the spin-orbit interaction breaks the degeneracy of the two helical modes, which turns an insulator into a metallic state [Figs. 5.1(i)-(l)]. We point out that the metallic state is characterized either by a local orbital, $l = 2$ [Fig. 5.1(h)] or by a helicity,
Figure 5.1: Density of states (DOS) of the parabolic model for different control parameters. The red-dashed lines are for the orbital $l = 1$, the dotted blue lines $l = 2$, and the green lines are for the total DOS. The central gray area displays the thermal window ($\sim k_B T$) for each temperature. Figures (a)-(d) compare the DOS for different temperatures. Figures (e)-(h) correspond to different magnitudes of the crystal electric field ($\Delta_{CEF}$) which breaks the degeneracy of the two $f$ orbitals. Figures (i)-(l) correspond to different magnitudes of the spin-orbit interaction ($\gamma_{SO}$) that breaks the degeneracy of the conduction bands with different helicity.

$h = +$ [Fig. 5.1(l)], since only the bands with corresponding quantum numbers are conducting.

In order to further examine the effect of the control parameters, we first calculate the transport coefficients as a function of the CEF and SO. In Fig. 5.2 we show the results of a calculation of the thermopower ($S$), the electrical conductivity ($\sigma$), and the power factor ($Z_{PF}$). As can be seen from the right column, the power factor is enhanced either by finding the optimal CEF or by adjusting the SOI. Since both CEF and SOI shift some of the lower energy bands toward the chemical potential [Figs. 5.1(e)-(l)], the number of lower energy bands relevant for thermal transport is controlled by CEF and SOI simultaneously. For the temperature range $T \lesssim 20K$, where the thermal windows are sufficiently narrow, CEF and SOI compete; hence there are two distinctive optimal regimes. For a sufficiently wide thermal window, attainable at intermediate temperatures, $T \sim 30K$, CEF and SOI are working cooperatively to form a single optimal region. For $T > 30K$, the enhancement in $Z_{PF}$ is not as drastic as at low temperature. $Z_{PF}$ is maximized in the vicinity of the insulator-metal transition (see the conductivity $\sigma$ at $T = 5 - 20K$), resulting from a competition between $S$ and $\sigma$. For instance, at $T = 5K$, the thermopower $S$ decreases with SOI and CEF, while the
Figure 5.2: Transport coefficients for different temperatures as a function of crystal field splitting and spin-orbit interaction: thermopower $S$ (left column), conductivity $\sigma$ (middle), and power factor $Z_{PF}$ (right). From the top to the bottom, the temperature varies from 50 to 5K, and the solid lines are equally spaced constant contours.

system acquires a finite conductivity. Note that the metallic state here has one dominant helical state over the other.

In Fig. 5.3, we repeat the calculation of the transport coefficients for a fixed SOI as a function of CEF and temperature. Consistent with Fig. 5.2 is that the optimal point for the power factor is located in the vicinity where the insulator-metal transition occurs. For instance, when $\gamma_{SO} \lesssim 0.1eV$, there is the optimal CEF and temperature for the power factor, at which point the electric conductivity acquires a noticeable finite value. From the left column, one finds that the thermopower generally decreases with increasing temperature as a widened thermal widow implies the reduction of the asymmetry in the DOS within the thermal region [Figs. 5.1(a)-(d)]. This obtains because as the temperature increases, more of the bands (lower and upper) are involved in the thermal transport. In other words, the asymmetry of the DOS within the thermally active
Figure 5.3: Transport coefficients for each SOI as a function of CEF and temperature: thermopower $S$ (left column), conductivity $\sigma$ (middle), and power factor $Z_{PF}$ (right). The range of SOI is 0.02 – 0.15 eV from the top to bottom panels.

region is relieved. Beyond a certain of the threshold SOI, $\gamma_{so} \geq 0.13$ eV, there is no phase transition (at mean field level); hence optimization cannot be realized.

### 5.3.2 Tight binding itinerant electron bands

Next, we consider the 3-dimensional tight binding case. We see that as in the case of a quadratic band, tuning the crystal-field and spin-orbit coupling can optimize the thermoelectric performance. This result indicates that the effect of orbital degeneracy in controlling the thermoelectric performance is not that sensitive to the details of the band structure. Here, we choose the hopping parameter $t_{hop} = 0.2167$ eV (band width $W = 2.6$ eV), and we located the local energy $\epsilon_{f_1} = -0.8 t_{hop}$. The hybridization strength $V_1 = t_{hop}$ and $V_2 = 1.01 t_{hop}$. In the SOI, we take the next-nearest neighbor hopping parameter $g_2 = 0.3$. 
As in the simplified parabolic model, the role of CEF and SOI is not different; both efficiently control the system to drive it from an insulator to a conductor as seen from Fig. 5.4. Compared to the corresponding panels in Figs. 5.1, however, Figs. 5.4(e)-(h) show that the CEF also pushes one of the upper bands toward the chemical potential, hence reducing the gap size significantly. In fact, the parabolic model is rather exceptional since the bottom of the upper bands corresponds to the point \( k = 0 \), which is not usual for typical 3D tight-binding models. Figs. 5.4(i)-(l) display the evolution of DOS with the increase of the SOI. Even though the degeneracy of the two helical modes are broken with a finite SOI, it cannot be seen clearly as was in the linearized SOI case [Figs. 5.1(i)-(l)]. The reason is that \( |\Gamma_k| \) decreases as \( k \) approaches the boundary of the Brillouin zone due to the periodic form of the SOI, while it does not for the linearized SOI. Unlike the CEF which affected drastically only one of the orbitals, the effect of the SOI is quite different. Up to \( \gamma_{so} = 0.2t_{hop} \simeq 40\text{meV} \), the changes in the DOS are not significant. For \( \gamma_{so} \gtrsim 0.2t_{hop} \), the system undergoes a phase transition to a (helically polarized) metal, beyond which point \( Z_{PF} \) is reduced (See Fig. 5.5).

From Fig. 5.5, we observe consistency with the parabolic model: the power factor can be enhanced

![Figure 5.4: Density of states(DOS) of the 3D tight-binding model for different control parameters. Both \( \Delta_{CEF} \) and \( \gamma_{so} \) are in units of the hoping amplitude, \( t_{hop} = 0.216\text{eV} \). The red-dashed lines are for the orbital \( l = 1 \), the dotted blue lines are for \( l = 2 \), and the green lines are for the total DOS. The central gray area indicates the thermally active region for each temperature. Figs. (a)-(d) compare the DOS for different temperature, (e)-(h) for the crystal electric field(\( \Delta_{CEF} \)), and (i)-(l) for the spin-orbit interaction(\( \gamma_{SO} \)).](image-url)
by adjusting the CEF, while the SOI slightly lowers the optimal value of the CEF. For $T > 15K$, the enhancement in $Z_{PF}$ is not as drastic as was in the low temperature case. As in the parabolic model, this trend occurs because the thermally active region is too wide to encompass only one band [see Fig. 5.4(c) and (d)]. Comparison with the other columns reveals that $Z_{PF}$ is also maximized near an insulator to a (helical) metal transition (see the conductivity $\sigma$ at $T = 5K$), which is the consequence of the competition between $S$ and $\sigma$. Here, one can observe that $S$ becomes maximal at $\Delta_{CEF} \simeq 0.01t_{hop}$, which is a consequence of the choice $V_2/V_1 = 1.01$. With $V_2/V_1 = 1$, $S$ only decreases with CEF (not shown). Fig. 5.6 similarly confirms the consistency with the parabolic model. The only difference is that the effect of SOI is not as remarkable, though it works to shift the optimal value of the CEF. The reason mainly lies in the changes of the DOS depending on SOI: linearized SOI changes the bandwidth significantly, while 3D tight-binding SOI does not due to its periodic structure. (See Fig. 5.7)

Figure 5.5: Transport coefficients for each temperature: thermopower $S$ (left column), conductivity $\sigma$ (middle), and power factor $Z_{PF}$ (right). From the top to the bottom, the temperature is fixed to 25, 15, 10, and 5K, respectively.
Figure 5.6: Transport properties are compared for each SOI: $S$ (left column), conductivity $\sigma$ (middle), and power factor $Z_{PF}$ (right). The spin-orbit couplings are chosen to be 0.0, 0.1, 0.2, and 0.5 in units of the hoping amplitude $t_{hop}$, respectively.

### 5.3.3 Effect of multi-orbital degeneracy

In addition to the continuous control of orbital degeneracy through crystal-field and spin-orbit coupling, we can specifically study the effect of increasing the number of degenerate orbitals ($l_{\text{max}} = 1, 2, \cdots, 5$). To minimize the number of free parameters, we will set the orbital degeneracy of the two bands (the allowed values of $l$) to be equal. Although continuous control is not possible in this case, one can then consider changing the material content to achieve a better thermoelectric. Here, the bare conduction electron dispersion is taken to be that of the 3D tight binding model.

First, we compare the DOS depending on the number of orbitals involved [Fig. 5.8]. As $l_{\text{max}}$ increases, the asymmetry between the upper and the lower bands becomes more pronounced. At the same time, the insulating gap increases with $l_{\text{max}}$ for $l_{\text{max}} \geq 2$. Note that this feature is quite similar to the single impurity problem with $N$ flavors. The inset of Fig. 5.8 displays the DOS without adjusting the chemical potential. Since the slave-boson method renormalizes the local energy by $\epsilon_f \to \epsilon_f + \lambda$, the relative location of the Kondo resonance for each case (near each gap) indicates that the amount of renormalization $\lambda$ increases with the
number of available orbitals. Since each band below and above the insulating gap should accommodate one electron, the deformation of the lower bands becomes less significant as $l_{\text{max}}$ increases (see the lower bands for different $l_{\text{max}}$). In other words, since the area below and above the gap should be equal, the asymmetry of the DOS becomes more significant as $\lambda$, or equivalently $l_{\text{max}}$, increases.

Given the band structures at the mean-field level, we proceed to calculate the transport properties as shown in Fig. 5.9. As $l_{\text{max}}$ increases, maxima of $Z_{\text{PF}}$ also increases as the temperature is elevated. The thermopower is also enhanced with the number of available orbitals, which is caused by pronounced asymmetry in the DOS (See Fig. 5.8). Since the gap size increases, the conductivity generally decreases with the number of orbitals for $l_{\text{max}} \geq 2$. Though not shown here, the maximal power factor per orbital, $Z_{\text{PF}}/l_{\text{max}}$, also increases with $l_{\text{max}}$ until $l_{\text{max}} < 7$. In Fig. 5.9(c), the thermal conductivity due to electronic structure is evaluated, excluding any contribution from lattice vibrations. Typically, phonons are dominant contributors to the thermal conductivities, but it may not be so prevalent at the low temperature range considered here.
presumably $T \lesssim 10 \text{K}$. The resultant (dimensionless) figure of merit, assuming $\kappa = \kappa_{\text{electron}} + \kappa_{\text{phonon}} \simeq \kappa = \kappa_{\text{electron}}$, is strongly enhanced with the number of orbitals at least by an order of 10. This is one of the key results of this paper.

5.4 Summary and Remarks

We have shown that orbital degeneracy opens a new window into increasing the power factor for strongly interacting thermoelectrics. Our key findings are that the power factor is maximized in strongly correlated systems by tuning 1) the gap between nearly degenerate local f-orbitals through the crystal field effect, 2) ) the gap between nearly degenerate itinerant electron bands through the spin-orbit coupling, and 3) the the number of degenerate local and itinerant orbitals.

This approach provides a new parameter space for the design of strongly correlated thermoelectric materials. Our result was derived using the slave-particle mean-field theory, which is not expected to be quantitatively reliable but should capture general trends. The effect of degeneracy in enhancing thermoelectric performance of strongly correlated systems could also be investigated using other methods such dynamical mean-field theory \cite{152} and finite frequency methods \cite{153}. Another direction for future work is to consider more profound effects of spin-orbit coupling combined with correlations, as in the proposed
“topological Kondo insulators” \cite{154}, whose surface states are currently being sought experimentally; such surface states have the potential to increase thermoelectric performance at low temperatures \cite{136}.

Finally, a direction that is difficult theoretically but may be important for actual materials is to find ways of interpolating between the effectively itinerant calculation here (i.e., there are plane-wave states of the slave particles) with the “atomic limit” \cite{155}, where the effects of multiple orbitals have also been considered \cite{156}. The atomic limit, which is valid when the hopping is the smallest energy scale in the problem, has been argued to be relevant to experiments on sodium cobaltates near room temperature \cite{157}. The results of this paper should motivate continued investigation of correlated materials for thermoelectricity and suggest that a guided search with controlled crystal-field splitting may lead to further improvements in thermoelectric figure of merit, especially in the low-temperature regime.

Figure 5.9: Transport properties for $l_{\text{max}}$ orbitals: (a) Seebeck coefficient, (b) electrical conductivity, (c) thermal conductivity, (d) thermopower (e) figure of merit and (f) dimensionless figure of merit.
References


