Stability of Ferromagnetism in Hubbard Models with Nearly Flat Bands

Hal Tasaki[1]

Whether spin-independent Coulomb interaction in an electron system can be the origin of ferromagnetism has been an open problem for a long time. Recently, a "constructive" approach to this problem has been developed, and the existence of ferromagnetism in the ground states of certain Hubbard models was established rigorously. A special feature of these Hubbard models is that their lowest bands (in the corresponding single-electron problems) are completely flat. Here we study models obtained by adding small but arbitrary translation-invariant perturbation to the hopping Hamiltonian of these flat-band models. The resulting models have nearly flat lowest bands. We prove that the ferromagnetic state is stable against a single-spin flip provided that Coulomb interaction U is sufficiently large. (It is easily found that the same state is unstable against a single-spin flip if U is small enough.) We also prove upper and lower bounds for the dispersion relation of the lowest energy eigenstate with a single flipped spin, which bounds establish that the model has "healthy" spin-wave excitation. It is notable that the (local) stability of ferromagnetism is proved in nonsingular Hubbard models, in which we must overcome competition between the kinetic energy and the Coulomb interaction. We also note that this is one of the very few rigorous and robust results which deal with truly nonperturbative phenomena in many-electron systems. The local stability strongly suggests that the Hubbard models with nearly flat bands have ferromagnetic ground states. We believe that the present models can be studied as paradigm models for (insulating) ferromagnetism in itinerant electron systems.

KEY WORDS: Hubbard models; ferromagnetism; local stability; spin-wave excitation.

[1] Department of Physics, Gakushuin University, Mejiro, Toshima-ku, Tokyo 171, Japan; e-mail: hal.tasaki@gakushuin.ac.jp.

CONTENTS

1. INTRODUCTION

1.1. Background

The origin of the strong ferromagnetic ordering observed in some materials has long been a mystery.[29] Since noninteracting electron systems universally exhibit paramagnetism, the origin of ferromagnetism should be sought in electron–electron interaction. In most solids, however, the dominant part of the interaction between electrons is the Coulomb interaction, which is perfectly spin independent (e.g., see ref. 3, Chapter 32, p. 674). Therefore we are faced with a very interesting and fundamental problem in theoretical physics, to determine *whether spin-independent interaction in an itinerant electron system can be the origin of ferromagnetic ordering*. This problem is important not only because ferromagnetism is a very common (and useful) phenomenon, but because it focuses on the fundamental role of nonlinear interactions in many-body quantum mechanical systems.

The present work is a continuation of our work in refs. 47 and 34, where we dealt with the above fundamental problem from the standpoint of "constructive condensed matter physics." Our goal is to provide concrete models in which the existence of ferromagnetic ordering can be established rigorously. Such models should shed light on mechanisms by which the Coulomb interaction generates ferromagnetic ordering.

It was Heisenberg[14] who first realized that ferromagnetism is an intrinsically quantum mechanical phenomenon. In Heisenberg's approach to ferromagnetism, one starts from the picture that each electron (relevant to magnetism) is almost localized at an atomic orbit. By treating the effect of Coulomb interaction and overlap between nearby atomic orbits in a perturbative manner, Heisenberg concluded that "exchange interaction" appears between nearby electronic spins, which determines the magnetic properties of the system.

The validity of the Heisenberg picture has been challenged from both theoretical and·physical points of views (see, for example, ref. 15). It has been realized that in most situations the exchange interaction is antiferromagnetic rather than ferromagnetic. Moreover, conditions which would justify the basic assumption that electrons can be treated as localized at atomic sites are not well understood.[2]

---

[2] This issue is closely related to the problem of Mott–Hubbard insulators.

In a different approach to the problem of ferromagnetism originated by Bloch[5] one starts from the quantum mechanical free electron gas, in which electrons are in plane-wave-like states. One then treats the effect of Coulomb interaction perturbatively and tries to find instability against certain magnetic ordering. When combined with the Hartree–Fock approximation (or a mean-field theory), this approach leads to the picture that there is an instability against ferromagnetism when the density of states at the Fermi energy and the Coulomb interaction are sufficiently large.

It has been realized, however, that the Hartree–Fock approximation drastically overestimates the tendency toward ferromagnetism, thus predicting the existence of ferromagnetism in many situations where it does not take place. From a theoretical point of view, the approximation is unsatisfactory since it artificially replaces the fundamental $SU(2)$ symmetry (i.e., rotation symmetry in the spin space) of the electron systems with a discrete Z_2 symmetry. Although there have been many improvements of the simple Hartree–Fock theory, this approach does not provide a conclusive answer to the fundamental problem about the origin of ferromagnetism that we raised in the beginning of the present subsection. See, for example, ref. 16 for a review.

1.2. Ferromagnetism in the Hubbard Model

A modern version of the problem of the origin of ferromagnetism was formulated by Kanamori,[18] Gutzwiller,[12] and Hubbard[17] in 1960s. (A similar formulation was given earlier, for example, in ref. 41.) They studied simple tight-binding models of electrons with on-site Coulomb interaction[3] of strength U. The model is usually called the Hubbard model. When there is no electron–electron interaction (i.e., $U = 0$), the model exhibits paramagnetism as an inevitable consequence of the Pauli exclusion principle. Among other things, Kanamori, Gutzwiller, and Hubbard asked *whether the paramagnetism found for $U = 0$ can be converted into ferromagnetism when there is a sufficiently large Coulomb interaction U.* This is a concrete formulation of the fundamental problem that we raised in the opening of the previous subsection.

It is worth noting that the on-site Coulomb interaction itself is completely independent of electronic spins, and does not favor any magnetic

---

[3] It is sometimes argued that the originally long-ranged Coulomb interaction becomes short ranged by the screening effect from electrons in the bands (or orbits) which are *not* taken into account in the Hubbard model. But it is still true that the assumption that there is only on-site interaction is highly artificial.

ordering. Therefore one does not find any terms in the Hubbard Hamiltonian which explicitly favor ferromagnetism (or any other ordering). Our theoretical goal will be to show that magnetic ordering arises as a consequence of a subtle interplay between the kinetic motion of electrons and the short-ranged Coulomb interaction. It is interesting to compare the situation with that in spin systems, where one is often given a Hamiltonian which favors some kind of magnetic ordering and the major task is to investigate if such ordering really takes place. We can say that the Hubbard model formulation goes deeper into fundamental mechanisms of magnetism than that of spin systems. It offers a challenging problem to theoretical physicists to derive magnetic interaction from models which do not explicitly contain such interactions. Perhaps the best justification of the Hubbard model as a standard model of itinerant electron systems comes from such a theoretical consideration, rather than its accuracy in modeling narrow-band electron systems. See also the introductions in refs. 34, 35, and 43 for discussions about ferromagnetism in the Hubbard model.

We stress that ferromagnetism is *not* a universal property of the Hubbard model. The Hubbard model is believed to exhibit various phenomena including paramagnetism, antiferromagnetism, ferrimagnetism, ferromagnetism, or superconductivity, depending on various conditions. Such drastic "nonuniversality" of the model motivated us to take the present "constructive" approach rather than to prove theorems which apply to general Hubbard models.

The problem of ferromagnetism in the Hubbard model was extensively studied by using various heuristic methods. The Hartree–Fock approximation discussed above leads one to the so-called Stoner criterion. It says that the Hubbard model exhibits ferromagnetism if one has $UD_F > 1$, where D_F is the density of states of the corresponding single-electron problem measured at the Fermi level (of the corresponding noninteracting system). Although the criterion cannot be trusted literally, it guides us to look for ferromagnetism in models with not too small U and/or large density of states.

The first rigorous result about ferromagnetism in the Hubbard model was provided by Nagaoka[36] and independently by Thouless.[53] It was proved that certain Hubbard models have ground states with saturated ferromagnetism when there is exactly one hole and the Coulomb repulsion U is infinite. See refs. 26 and 46 for shorter proofs. Whether the Nagaoka–Thouless ferromagnetism survives in the models with finite density of holes and/or finite Coulomb repulsion is a very interesting but totally unsolved problem.[8,39,44,45,54,13,37,25,22] See also the introduction of ref. 35 for a compact review of this subject.

Very recently, Müller-Hartmann[35] argued that the Hubbard model with $U = \infty$ on a one-dimensional zigzag chain exhibits ferromagnetism.[4] Interestingly, the geometry of the chain is similar to that of one-dimensional models studied in the present paper.

Remark. It should be noted that the Hubbard model is by no means the unique formulation for studying strong correlation effects in narrow-band itinerant-electron systems. If one recalls how a tight-binding model is derived (or supposed to be derived) from a continuum model, there is a good reason to study models with more complicated interactions than mere on-site Coulomb repulsion. One can even include interactions which explicitly favor ferromagnetism, and still formulate interesting problems. See ref. 42 for an approach to ferromagnetism in such extended Hubbard models.

1.3. Flat-Band Ferromagnetism

In 1989, Lieb proved an important general theorem for the Hubbard model at half-filling on a bipartite lattice.[27] As a corollary of the theorem, Lieb showed that a rather general class of Hubbard models exhibits ferrimagnetism.[5] See also ref. 40.

In 1991, Mielke[30,31] came up with a new class of rigorous examples of ferromagnetism in the Hubbard model. He showed that the Hubbard models on a general class of line graphs have ferromagnetic ground states. A special feature of Mielke's model is that the corresponding single-electron Schrödinger equation[6] has highly degenerate ground states. In other words, Mielke's models have flat (or dispersionless) bands. The original results of Mielke's were for the electron number which corresponds to the half-filling of the lowest flat band, but later they were extended to different electron densities in two-dimensional models.[32]

A similar but different class of examples of ferromagnetism in the Hubbard models was proposed in refs. 47 and 34. These models were defined on a class of decorated lattices, and were also characterized by flat bands at the bottom of the single-electron spectrum. In a class of models

---

[4] Although Müller-Hartmann's argument is quite interesting, it does not form a mathematically rigorous proof (as far as we can read off from ref. 35). The argument involves an uncontrolled continuum limit of a strongly interacting system. To make it into a rigorous proof seems to be a nontrivial task.

[5] Ferrimagnetism is a kind of antiferromagnetism on a bipartite lattice such that the numbers of sites in two sublattices are different.

[6] Here (and throughout the present paper) we are talking about the genuine one-particle problem, not an effective (and usually ill-defined) one-particle problem in an interacting system which is often discussed in heuristic works.

in two and higher dimensions, it was proved that the ferromagnetism is stable against fluctuation of electron numbers.[47,34]

The examples of ferromagnetism in refs. 30–32, 47, and 34 are common in that they treat special models with flat lowest bands.[7] The ferromagnetism established for these models is now called flat-band ferromagnetism.[22] There is a general theorem due to Mielke[33] which states a necessary and sufficient condition for a Hubbard model with a flat lowest band to exhibit ferromagnetism when the flat-band is half-filled. Although flat-band ferromagnetism sheds light on very important aspects of the role of strong interaction in itinerant electron systems, it relies on the rather singular assumption that the models have completely flat bands. As we discuss in Section 2.1, we do not have true "competition" between the kinetic energy and the Coulomb interaction.

If one adds a small perturbation to the hopping Hamiltonian of a flat-band model, one generically gets a model with slightly dispersive lowest band. It was conjectured[47,34] that such models with nearly flat bands exhibit ferromagnetism, provided that the Coulomb interaction U is large enough. Kusakabe and Aoki[24,23] presented a detailed study of this problem by numerical experiments and careful variational calculations. Their results provide strong support for the conjecture that the flat-band ferromagnetism is stable against small perturbations to the band structure.

We stress that this is a very delicate conjecture for the following reasons.

- When the ground states are ferromagnetically ordered, there inevitably exist spin-wave (magnon, or Nambu–Goldstone) modes whose excitation energies are of order L^{-2}, where L denotes the linear size of the lattice. The total energy of the perturbation, on the other hand, is always proportional to the system volume L^d. This means that the total perturbation always exceeds the energy gap when the system size becomes large. Such a situation can never be dealt with naive perturbation theories.

- When the lowest band is nonflat, the model with $U = 0$ exhibits Pauli paramagnetism. It is strongly believed that, for sufficiently small U, the ground states of the models (in finite volumes) are spin singlet. Therefore one must have sufficiently large U to get ferromagnetism. This means that the problem is a truly *nonperturbative* one.

In other words, one must directly face the notoriously difficult problem of "competition" between the kinetic energy and the Coulomb interaction.

---

[7] Lieb's examples also have flat bands in the middle of the single-electron spectra.

Technically speaking, this nature of the problem inhibits one from making use of the common strategy of constructing exact ground states by minimizing local Hamiltonians.[8] This strategy has been used to derive exact ground states of various (extended) Hubbard models[6,42,48,50,7] as well as in our early work[47,34] on the flat-band Hubbard models. While there are rigorous perturbation theories for various many-body problems, including classical and quantum spin systems and quantum field theories, there is, as far as we know, no general theory which enables one to control generic perturbation in models which exhibit continuous symmetry breaking.

1.4. About the Present Paper

In the present paper, we report the first important step toward the solution of the above problem about the stability of flat-band ferromagnetism. We treat models with nearly flat bands obtained by adding almost arbitrary perturbations to the hopping matrices of the flat-band models. For sufficiently large U, we prove that the ferromagnetic state is locally stable. More precisely, we show that the lowest energy among ferromagnetic states is strictly less than the lowest energy among states with a single flipped spin. The local stability, along with the global stability established for the flat-band models, strongly suggests that the ferromagnetic states are the true ground states of the present models for sufficiently large U. (See the remark below.) We also prove that, in a certain range of the parameter space, the spin-wave dispersion relations of the present models behave exactly like those in the Heisenberg ferromagnet. This confirms the conjecture of Kusakabe and Aoki.[23] These results were first announced in ref. 49.

As far as we know, this is the first time that the (local) stability of ferromagnetism is proved in truly nonsingular Hubbard models, overcoming the competition between the kinetic energy and the Coulomb interaction. We also note that this is one of the very few rigorous and robust works in which nonperturbative aspects of many-electron problems are treated. Recently there have been remarkable progress in rigorous treatment of interacting many-fermion systems based on renormalization group techniques. However, these treatments deal only with weak coupling phenomena such as the Tomonaga–Luttinger liquid[4] and superconductivity.[10]

The present paper is organized as follows. In Section 2, we restrict ourselves to the simplest one-dimensional models and discuss our main results

---

[8] It turned out that there are exceptions to this statement.[51] See the remark at the end of Section 1.4. However, we still believe that the above comment is true for generic models.

and the ideas behind the proof. We have tried to make this section accessible to a wide range of readers. In Section 3, we introduce a general class of models in arbitrary dimensions and state our rigorous results precisely. Sections 4–10 are devoted to the proof of our theorems. We have carefully organized the lengthy proof so as to make it as readable as possible. One can read off the organization of these sections by taking a look at the table of contents. In general, earlier sections contain physically interesting ideas, and later sections contain technical materials. A browse through Sections 4–6 should give the reader a clear idea about the detailed structure of our proof.

Remark. After the completion of the present paper, we finally succeeded in proving the global stability of ferromagnetism in Hubbard models obtained by adding a *special* perturbation to the flat-band models.[51] We stress that this new result does not diminish the importance of the present work. Even for the models treated in the ref. 51 the only way (that we know of) to prove meaningful lower bounds for spin-wave excitation energy is via the machinery developed here. The robustness of the present results (in the sense that we allow arbitrary weak translation-invariant perturbation) is also important.

2. STABILITY OF FERROMAGNETISM IN ONE-DIMENSIONAL MODELS

In the present section, we discuss our main results and the basic ideas of their proof in the context of simplest one-dimensional Hubbard models. The advantage of restricting ourselves to one-dimensional models is that we can discuss the essence of our theory without being bothered by many technical details. In particular the analysis of the band structure (Section 2.3) and the construction of localized bases (Section 2.4) can be carried out in an explicit and elementary manner thanks to special features of the simple models. These explicit calculation will be a good introduction to more elaborate analyses in the general class of models. Fortunately, the ideas developed in one-dimensional models can be used in the study of the general models in higher dimensions with only technical modifications.

We have. tried to make the present section self-contained and accessible to a wide range of readers. We urge the reader to take a look at this section, whether or not he or she is planning to study the later sections.

We note that what we present in this section is far from a complete mathematical proof. We often neglect "small" contributions without any justification, and some of the formulas are not perfectly correct (in view of the rigorous analysis presented in the later sections). Nevertheless we

believe that the material presented here will give a clear idea about the philosophy and the structure of our proof.

2.1. Models and Main Results

We define the simplest two-band models in one dimension and describe what we can prove about the stability of ferromagnetism and the spin-wave dispersion relations. We stress that the restriction to one dimension is by no means essential. All the results here extend to corresponding models in higher dimensions (i.e., two, three, or even higher). The reader who is *not* planning to study the later sections is invited to take a brief look at Section 3, especially at Figs. 7–9 of two-dimensional lattices and band structures.

Let L be a fixed odd integer, and denote by

$$\Lambda_o = \left\{ -\frac{L-1}{2}, ..., -1, 0, 1, ..., \frac{L-1}{2} \right\} \subset \mathbf{Z} \tag{2.1}$$

the chain of length L (identified with a set of integers). We also define

$$\Lambda' = \Lambda_o + \frac{1}{2} = \left\{ -\frac{L}{2} + 1, ..., -\frac{1}{2}, \frac{1}{2}, \frac{3}{2}, ..., \frac{L}{2} \right\} \tag{2.2}$$

which is the chain obtained by shifting Λ_o by 1/2. Our lattice Λ is obtained by "decorating" the chain Λ_o by the sites from Λ' as $\Lambda = \Lambda_o \cup \Lambda'$. See Fig. 1. One may regard our lattice structure as mimicking that of an oxide, where sites in Λ_o correspond to metallic atoms and sites in Λ' correspond to oxygen atoms. We have no intention of building models which are realistic from the viewpoint of condensed matter physics. But this analogy
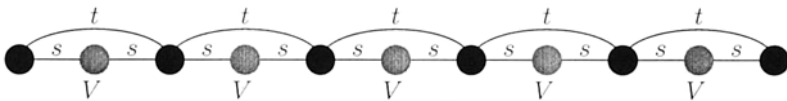
Fig. 1. The one-dimensional lattice studied in Section 2. We identify the leftmost site with the rightmost site to get a closed chain. The black dots represent sites in Λ_o (metallic atoms) and the gray dots represent sites in Λ' (oxygen atoms). There are two types of hopping t and s and on-site (one-body) potential V. In addition we have on-site Coulomb repulsion $U > 0$ at each site. There are $2L$ sites in the lattice, and we put L electrons in the system. (Here $L = 5$.) For the flat-band models characterized by $s = \lambda t$, $V = (\lambda^2 - 2) t$ with $\lambda > 0$, $t > 0$, the ground states of the models are proved to be ferromagnetic. Here we prove the local stability of ferromagnetism for models obtained by adding small perturbations to the flat-band models.

proves to be helpful in understanding various aspects of our work, including the basic mechanism of ferromagnetism.

We shall study the Hubbard model on Λ with the Hamiltonian

$$
H = t \sum_{\substack{x \in \Lambda_o \\ \sigma = \uparrow, \downarrow}} (c_{x,\sigma}^\dagger c_{x+1,\sigma} + \text{h.c.}) + s \sum_{\substack{x \in \Lambda \\ \sigma = \uparrow, \downarrow}} (c_{x,\sigma}^\dagger c_{x+(1/2),\sigma} + \text{h.c.})
$$

$$
+ V \sum_{\substack{x \in \Lambda' \\ \sigma = \uparrow, \downarrow}} n_{x,\sigma} + U \sum_{x \in \Lambda} n_{x,\uparrow} n_{x,\downarrow} \tag{2.3}
$$

where we use periodic boundary conditions to identify x with $x - L$ if necessary. Here $c_{x,\sigma}^\dagger$ and $c_{x,\sigma}$ are the creation and the annihilation operators, respectively, of an electron at site[9] $x \in \Lambda$ with spin $\sigma = \uparrow, \downarrow$. They satisfy the standard fermionic anticommutation relations. [See (3.11), (3.12) for details.] The corresponding number operator is $n_{x,\sigma} = c_{x,\sigma}^\dagger c_{x,\sigma}$. Finally, "h.c." in (2.3) stands for the Hermitian conjugate.

The real parameters t and s represent the amplitudes that an electron hops between neighboring sites in Λ_o (separated by a distance 1) and between neighboring sites in Λ (separated by a distance 1/2), respectively. The real parameter V is the on-site potential energy for the sites in Λ' (see Fig. 1). The first three terms in (2.3) determine single-electron properties of the model. The fourth term is the on-site Coulomb interaction characteristic in the Hubbard model with the interaction energy $U > 0$.

We consider many-electron states with the total electron number fixed to L. (See the end of Section 3.2 for an explicit construction of the Hilbert space.) Since there are $2L$ sites in the lattice Λ, the present electron number corresponds to the quarter-filling of the whole bands (or the half-filling of the lower band). This electron number is natural if one imagines that each site in Λ_o (which corresponds to a metallic atom) emits one electron to the band.[10]

---

[9] More precisely, these operators correspond to an orbital state around the atom at x. We have here assumed that each orbit is nondegenerate. Usually models based on such an assumption are referred to as single-band Hubbard models. We find this terminology confusing since our model indeed has multiple bands in its single-electron spectrum. We think a better terminology is "single-state" Hubbard model. Then the models we consider are classified as "single-state multiband Hubbard models."

[10] Of course one gets the same electron number if each site in Λ' emits one electron. But we want to insist on the present picture since it gives the desired electron number for the general class of models studied later. Moreover, the picture to identify Λ_o sites as metallic atoms is consistent with the nature of the "ferromagnetic ground states."

The first result about ferromagnetism deals with the so-called flat-band Hubbard model. To define the model, we introduce a parameter $\lambda > 0$, and set[11]

$$s = \lambda t, \qquad V = (\lambda^2 - 2) t \qquad (2.4)$$

The following strong result was proved in refs. 47 and 34.

Theorem 2.1 (Flat-band ferromagnetism). Let $t > 0$ and $\lambda > 0$ be arbitrary, and let s and V be determined by (2.4). Then, for any $U > 0$, the ground states of the Hamiltonian (2.3) exhibit saturated ferromagnetism, and are nondegenerate apart from the trivial spin degeneracy.

More precisely, a state is said to "exhibit saturated ferromagnetism" if the total spin S_{tot} of the state takes the maximum possible value $S_{max} = L/2$. See the end of Section 3.2 for a precise definition of S_{tot}. See also Theorem 3.1 for the general theorem and Section 5.4 for a proof. The flat-band ferromagnetism has been established for a general class of models including those in higher dimensions.[33] In a class of models in two and higher dimensions, the existence of ferromagnetism for lower electron densities as well as the existence of a paramagnetism–ferromagnetism transition (as the electron density is changed) are established rigorously.[32,47,34]

A model determined by the conditions (2.4) with $t > 0$ and $\lambda > 0$ has a very special feature that the ground states of the corresponding single-electron Schrödinger equation are L-fold degenerate. In other words, the lower band (in its single-electron spectrum) is dispersionless (or flat). We shall see this explicitly in Section 2.3. See Fig. 2a in Section 2.3. As a consequence, the many-electron ground states in the noninteracting model with $U = 0$ are highly degenerate. The total spin can take any of the allowed values $S_{tot} = 1/2$, $3/2,..., L/2$. This is a kind of parramagnetism, but is certainly different from the Pauli paramagnetism which allows only unique (or twofold degenerate) ground state(s) with the minimum possible S_{tot} (which is 0 or 1/2).

The role of the Coulomb interaction U in flat-band ferromagnetism is to lift the above-mentioned degeneracy and to "select" only the ferromagnetic states as ground states. This is why even infinitesimaly small U is sufficient for stabilizing ferromagnetism. Although the flat-band ferromagnetism focuses on a nontrivial and important effect caused by electron interactions, it avoids dealing with the truly difficult problem about "competition" between the kinetic energy and the electron interactions.

---

[11] The model studied here is obtained by setting $d = \nu = 1$ in the general class of models introduced in Section 3.2 and studied in the later sections. In the Hamiltonian of the later sections, the energy is shifted by a constant so that the lowest band in the flat-band models has vanishing energy.

Let us now turn to the models with nearly flat bands obtained by perturbing the above models. In order to simplify the discussion, we consider the simplest possible perturbation.[12] Instead of (2.4), let us set

$$s = \lambda t, \qquad V = (\lambda^2 - 2 + p) t \qquad (2.5)$$

where the parameter p measures the strength of the perturbation. As we see soon in Section 2.3, the lower band is no longer flat for $p \neq 0$.

Let $E_{\min}(S_{\text{tot}})$ denote the lowest energy among the L-electron states with a given total spin S_{tot}. The Pauli exclusion principle implies that, for a model with $p \neq 0$ and $U = 0$, these energies satisfy the monotonicity inequality

$$E_{\min}(1/2) < E_{\min}(3/2) < \cdots < E_{\min}(S_{\max} - 1) < E_{\min}(S_{\max}) \qquad (2.6)$$

with $S_{\max} = L/2$. This is nothing but the Pauli paramagnetism.

We want to examine if these strict inequalities can be reversed as a consequence of on-site Coulomb interaction. We stress that this is a truly nonperturbative problem in which one must directly face the "competition" between the kinetic energy and the interaction. In fact it is quite easy to see that we must have a sufficiently large U to stabilize ferromagnetism.

Theorem 2.2 (Instability of "ferromagnetic ground states" for small U). Let $t > 0$ and $\lambda > 0$ be arbitrary, and let s and V be determined by (2.5). We assume $p \neq 0$. Let

$$\bar{\varepsilon}(p) = (t/2) \, |4 - \{(\lambda^2 + 4)^2 + 2p(\lambda^2 - 4)\}^{1/2} + (\lambda^2 + 2p\lambda^2)^{1/2}|$$

$$= \{4/(\lambda^2 + 4)\} \, t \, |p| + O(p^2)$$

denote the band width of the lower band. Then for U satisfying $0 \leqslant U < \bar{\varepsilon}(p)$, we have

$$E_{\min}(S_{\max} - 1) < E_{\min}(S_{\max}) \qquad (2.7)$$

This is the one-dimensional version of Theorem 3.3.

We call the states with $S_{\text{tot}} = S_{\max}$ which have the energy $E_{\min}(S_{\max})$ the "ferromagnetic ground states."[13] It is easily found that the "ferromagnetic ground states" are nondegenerate apart from the trivial $(2S_{\max} + 1) = (L + 1)$-fold degeneracy. (See Lemma 3.2.)

---

[12] In the general treatment described in the later sections, we allow completely general perturbations with translation invariance and certain summability. See Section 3.2.

[13] This is a slight abuse of the word, since the states are not necessarily the true ground states.

Theorem 2.2 states that the "ferromagnetic ground states" are unstable against a single-spin flip. Although the inequality (2.7) does not tell us what the ground state of the model is, it does establish that the "ferromagnetic ground states" are *not* the true ground states.

Of course results like Theorem 2.2 can be proved rather easily by the standard variational argument. What is really interesting (and difficult) is to get a reversed inequality for models with larger values of U. The following is the most important result of the present paper.

Theorem 2.3 (Local stability of "ferromagnetic ground states"). Let $t > 0$, and let s and V be determined by (2.5). We further assume that $\lambda \geqslant \lambda_2$, $|p| \leqslant p_1$, $\lambda |p| \leqslant p_1$, and

$$U \geqslant K_1 \lambda^2 t |p| \qquad (2.8)$$

where λ_2, p_1, p_1, and K_1 are positive constants.[14] Then we have

$$E_{min}(S_{max} - 1) > E_{min}(S_{max}) \qquad (2.9)$$

This is the one-dimensional version of Theorem 3.4.

The bound (2.9) states that the "ferromagnetic ground states" are stable under a single-spin flip. Clearly the most important condition for the above local stability theorem is (2.8), which says we must have sufficiently large Coulomb interaction (compared with the bandwidth $\propto |p| t$). This is natural since the opposite inequality (2.7) holds if U is small. We can say that the above local stability theorem establishes a truly nonperturbative result in which the "competition" between the kinetic energy and the electron interaction is successfully dealt with.

Recall that both the energies $E_{min}(S_{max} - 1)$ and $E_{min}(S_{max})$ grow proportionally to the lattice size L^d, while their difference should be proportional to L^{-2}. In such a situation, there seems to be little hope in proving the desired inequality (2.9) for large L by combining a suitable lower bound for the left-hand side and an upper bound for the right-hand side. However, there are some nice features that save our task from being impossible. In the subspace with $S_{tot} = S_{max}$, the on-site Coulomb repulsion is completely irrelevant because of the Pauli principle. Therefore the energy $E_{min}(S_{max})$ in the right-hand side of (2.9) is nothing but the ground-state energy of the corresponding noninteracting spinless fermion, which energy is known exactly (at least formally). In the subspace with $S_{tot} = S_{max} - 1$, the on-site repulsion does play a highly nontrivial role, but one can still

---

[14] We use the same symbols for the constants as in the later sections. In general models, the constants depend on the basic model parameters d, v, and R, but here they are simply constants.

imagine that its effect is (at most) of order 1 rather than of order L^d. This is because (in a suitable representation) there is only one electron with down spin, and this single electron interacts with the rest of the electrons with up spin. This intuitive observation is indeed the basic starting point of our proof.

We are also able to establish rather strong results about the excitation energy above the "ferromagnetic ground states." Let $\mathscr{K} = \{k = 2\pi n/(L-1) \mid n \in \mathbf{Z} \text{ s.t. } |n| \leq (L-1)/2\}$ be the set of wave numbers allowed in the present model. For $k \in \mathscr{K}$, we denote by $\mathscr{H}_k$ the Hilbert space of the states which have a definite crystal momentum k and which contain $(L-1)$ up-spin electrons and one down-spin electron. [See (3.34) for a precise definition.] We let $E_{\mathrm{SW}}(k)$ be the lowest energy among the states in $\mathscr{H}_k$. Note that $E_{\mathrm{SW}}(k)$ can be interpreted as the energy of an elementary spin-wave excitation. The following theorem essentially determines the behavior of $E_{\mathrm{SW}}(k)$.

Theorem 2.4 (Bounds on the spin-wave energy).   Let $t > 0$, and let s and V be determined by (2.5). Assume that $\lambda \geq \lambda_3$, $|\rho| \leq \rho_0$, and $K_2 \lambda t \geq U \geq A_3 \lambda^2 t \, |\rho|$, where λ_3, ρ_0, K_2, and A_3 are positive constants. Then we have

$$F_2 \frac{4U}{\lambda^4} \left(\sin \frac{k}{2} \right)^2 \leq E_{\mathrm{SW}}(k) - E_{\min}(S_{\max}) \leq F_1 \frac{4U}{\lambda^4} \left(\sin \frac{k}{2} \right)^2 \qquad (2.10)$$

with

$$F_1 = 1 + \frac{A_4}{\lambda} + A_5 \lambda \, |\rho| + \frac{A_6 \lambda^2 t \, |\rho|^2}{U} \qquad (2.11)$$

and

$$F_2 = 1 - A_1 \, |\rho| - \frac{A_2}{\lambda} - \frac{A_3 \lambda^2 t \, |\rho|}{U} \qquad (2.12)$$

where A_i ($i = 1, ..., 6$) are positive constants.

This is the one-dimensional version of Theorems 3.5 and 3.6.

It is remarkable that we have $F_1 \simeq F_2 \simeq 1$ if $\lambda \gg 1$, $\lambda \, |\rho| \ll 1$, and $U \gg \lambda^2 t \, |\rho|$. In this case the bounds (2.10) imply

$$E_{\mathrm{SW}}(k) - E_{\min}(S_{\max}) \simeq \frac{4U}{\lambda^4} \left(\sin \frac{k}{2} \right)^2 = 2 J_{\mathrm{eff}} \left(\sin \frac{k}{2} \right)^2 \qquad (2.13)$$

which is nothing but the spin-wave dispersion relation for the ferromagnetic Heisenberg spin system. (See Section 2.6.) This result is very important since it guarantees that our Hubbard model develops low-lying excited states with the precise structure expected in a "healthy" ferromagnetic system.

Theorem 2.4 is also meaningful when applied to the flat-band model with $\rho = 0$. The theorem guarantees that the exchange interaction $J_{\text{eff}} \simeq 2U/\lambda^4$ [which appears in (2.13)] remains finite even for the flat-band models, thus confirming the conjecture of Kusakabe and Aoki.[23] We can conclude that the ferromagnetism in the flat-band models is not at all pathological[15] in spite of the rather artificial condition imposed on the models.

The reader may notice that Theorem 2.4, unlike Theorem 2.3, requires an upper bound for the Coulomb interaction U. There indeed is a *physical* reason for this limitation. Our proof of Theorem 2.4 is based on an explicit construction of the state which approximates the elementary spin-wave excitation.

Our approximate excited state, however, takes into account the effect of interaction U in a rather crude way. This inhibits us from getting precise estimate in the models with larger values of U. That our analysis is not efficient for large U can be easily seen from our formula for the effective exchange interaction $J_{\text{eff}} = 2U\lambda^{-4}$, which is proportional to U. For larger values of U, we expect J_{eff} to be "renormalized" to a less increasing function of U. In particular, Kusakabe and Aoki[23] pointed out that J_{eff} remains finite even in the limit $U \uparrow \infty$. A proof of this fascinating conjecture might be possible if one extends the present work by devising a more efficient approximate excited state which takes into account the large-U "renormalization" in a proper manner.

2.2. Discussions and Open Problems

The inequality (2.9) stated in our main Theorem 2.3 only establishes the local stability of the "ferromagnetic ground states," not the desired global stability. However, the strong result for the flat-band models summarized in Theorem 2.1 suggests that the local stability (2.9) implies that the "ferromagnetic ground states" are the true ground states. In the course of constructing our proof of the local stability theorem, we have developed a heuristic picture about the mechanism underlying ferromagnetism in our

---

[15] We recall that Nagaoka's example of ferromagnetism is known to have a pathological spin-wave dispersion relation.[36,22] As for the other rigorous examples, no results about spin-wave excitations are known.

model. The picture, which is briefly described in Section 2.6, also indicates that ferromagnetic states are the true ground states. As we have noted in the remark at the end of Section 1.4, this conjecture has been verified for a special class of perturbations.

It is interesting to look at our rigorous results in the light of traditional approaches to ferromagnetism discussed in Sections 1.1 and 1.2. In order to guarantee the existence of ferromagnetism in our model, we assumed that ρ is small enough so that the band is nearly flat and the Coulomb interaction U is large enough. Since a nearly flat band has large density of states, our requirement shares something in common with the Stoner criterion. Of course there is no hope that the criterion $UD_F > 1$ gives reliable conditions for the range of parameters where ferromagnetism takes place. The improved criterion for ferromagnetism due to Kanamori[18] and the accompanying formula for effective U do not seem to coincide with our results.

If one looks into the proof of the theorems, however, it becomes clear that there is a picture quite similar to that developed by Heisenberg. We use a basis in which each electron is treated as almost localized at each lattice site in Λ_o. The basic mechanism for stabilizing ferromagnetism comes from the "exchange" part of the interaction Hamiltonian, which is in principle the same as what Heisenberg treated. See also Section 2.6.

It is amusing that the ferromagnetism in our model may be understood in terms of the above two heuristic pictures. Usually the band electron picture and the Heisenberg's localized electron picture of ferromagnetism are regarded as incompatible with each other.

All the rigorous results summarized in the previous subsection strongly suggest that our Hubbard model exhibits nonpathological ferromagnetism in the vicinity of the flat-band models characterized by (2.4). However we are far from understanding precise (necessary and sufficient) condition for ferromagnetism. We believe that the one-dimensional Hubbard model with the Hamiltonian (2.3) at quarter-filling can be studied as a paradigm model for itinerant electron ferromagnetism (in insulators). To determine the region (in the three-dimensional parameter space spanned by s, V, and U, as well as the sign of t) where ferromagnetism takes place is a challenging and very illuminating problem that can be studied by various methods, including numerical ones.

One might regard the models with only nearest neighbor hoppings (obtained by setting $t = 0$) as "standard." However, the Lieb–Mattis theorem[28] ensures[16] that such models do not exhibit ferromagnetism for any values of V and U. This shows that the appearance of ferromagnetism

---

[16] Rigorously speaking, this is true only for the models with open boundary conditions.

is a rather delicate phenomenon which cannot be determined by simple criteria like the Stoner criterion.

There is a perturbative argument[52] (similar to that in Section 2.6) which suggests that the Hubbard model with Hamiltonian (2.3) exhibits ferromagnetism in a finite but not very large region including the flat-band models. Perhaps this observation is consistent with the empirical fact that most of the known insulators appear to be antiferromagnets.[17]

The electron number we have chosen corresponds to the half-filling of the lower (nearly flat) band. This is also the case for the general class of models studied in the later sections. From the standard band-theoretic point of view, an electron system with such a filling becomes metallic. When the Coulomb interaction U is sufficiently large in our models, however, the strong correlation makes the lower band (effectively) fully filled. Since the lower band is separated by an energy gap from the upper band, the system is expected to become an insulator. In this sense, our models provide examples of Mott–Hubbard insulators. This is also true for the general models in higher dimensions.

We expect to get ferromagnetic metals by lowering the electron density in the present models. In the flat-band case[34] we found that the model must be at least two dimensional in order for ferromagnetism to be stable against the change of electron density. We argued that the one-dimensional flat-band model exhibits ferromagnetism only when the lower band is exactly half-filled, and exhibits paramagnetism for any lower electron densities.[18] We believe that this dimensional dependence is a special feature of the flat-band models in which electrons "cannot move" (in some sense).

We believe that our Hubbard models with nearly flat band in any dimensions with lower electron density are among the best candidates of itinerant electron systems which exhibit metallic ferromagnetism. Unfortunately we have no rigorous results in this direction.

Finally we recall that, in dimensions one or two, ferromagnetism in any short-ranged model with a rotation symmetry is inevitably destroyed by infinitesimally small thermal fluctuation.[11,21] In order to have ferromagnetism stable at finite temperatures, we must treat models in (at least) three dimensions. We expect that ferromagnetism in the three-dimensional versions of our models survives at finite temperatures, but have no rigorous results.[19]

---

[17] Recently there have appeared a few organic compounds which are insulating ferromagnets.

[18] We did not give a proof of the latter statement in ref. 34, but we believe there is no essential difficulty in proving it rigorously.

[19] We recall that the existence of a ferromagnetic order in the ferromagnetic quantum Heisenberg model at low enough temperatures is not yet proved.[9] It is very likely that the corresponding problem in the Hubbard model is much harder.

When one recalls the fact that we are so familiar in our daily lives with metallic ferromagnetism stable at room temperatures, to prove the existence of metallic ferromagnetism (say, in our models with lower electron densities) at low enough temperatures may appear as a modest goal. From the theoretical and mathematical points of views, however, the problem looks formidably difficult. It seems that not only mathematical techniques, but a fundamental understanding of the "physics" of itinerant electron ferromagnetism are still lacking.

2.3. Band Structure in the Single-Electron Problem

Before going into the full many-body problem, it is useful to investigate the corresponding single-electron problem. The single-electron Schrödinger equation corresponding to the Hubbard model (2.3) with the parametrization (2.5) is written as[20]

$$
\varepsilon \varphi_x = \begin{cases} t(\varphi_{x-1} + \varphi_{x+1}) + \lambda t(\varphi_{x-(1/2)} + \varphi_{x+(1/2)}) & \text{if } x \in \Lambda_o \\ (\lambda^2 - 2 + \rho) \, t\varphi_x + \lambda t(\varphi_{x-(1/2)} + \varphi_{x+(1/2)}) & \text{if } x \in \Lambda' \end{cases} \quad (2.14)
$$

where ε is the energy eigenvalue. By using the translation invariance of Eq. (2.14), we can write an eigenstate $(\varphi_x)_{x \in \Lambda}$ in the form of the Bloch state as $\varphi_x = e^{ikx} v_x(k)$ with $k \in \mathscr{K}$ and $v_x(k)$ such that $v_{x+1}(k) = v_x(k)$ for any $x \in \Lambda$. The Schrödinger equation in k-space which determines ε and $v_x(k)$ is

$$
\varepsilon \begin{pmatrix} v_0(k) \\ v_{1/2}(k) \end{pmatrix} = \begin{pmatrix} 2t \cos k & 2\lambda t \cos(k/2) \\ 2\lambda t \cos(k/2) & (\lambda^2 - 2 + \rho) \, t \end{pmatrix} \begin{pmatrix} v_0(k) \\ v_{1/2}(k) \end{pmatrix} \quad (2.15)
$$

The eigenvalue problem (2.15) can be solved easily, and for each $k \in \mathscr{K}$, we find two energy eigenvalues

$$
\varepsilon_{1,2}(k) = \frac{t}{2} \left(\lambda^2 - 2 + \rho + 2 \cos k \right.
$$

$$
\left. \pm \left\{ [\lambda^2 - 2(1 + \cos k) + \rho]^2 + 4 \left(2\lambda \cos \frac{k}{2} \right)^2 \right\}^{1/2} \right) \quad (2.16)
$$

where 1, 2 are the band indexes with 1 (respectively 2) corresponding to the $-$ (respectively $+$) sign. The energy $\varepsilon_j(k)$, as a function of k, is usually called the dispersion relation of the jth band. When $\rho = 0$, (2.16) becomes

---

[20] See Section 4.1 if it is not clear how the single-electron Schrödinger equation is derived.

$\varepsilon_1(k) = -2t$ and $\varepsilon_2(k) = \lambda^2 t + 2t \cos k$. Note that the lower band is completely flat (dispersionless), and there is an energy gap $\lambda^2 t$ between the two bands as in Fig. 2a. When the perturbation to the flat-band model is sufficiently small (i.e., $|\rho| t \ll \lambda^2 t$), the lower band is nearly flat, and there remains a gap close to $\lambda^2 t$ as in Fig. 2b. See Lemma 4.1.

We choose an eigenvector $\mathbf{v}^{(0)}(k) = (v_0^{(0)}(k), v_{1/2}^{(0)}(k))$ corresponding to the eigenvalue $\varepsilon_1(k)$ as

$$\mathbf{v}^{(0)}(k) = \begin{pmatrix} v_0^{(0)}(k) \\ v_{1/2}^{(0)}(k) \end{pmatrix} = \begin{pmatrix} \frac{1}{2}(F(k) + \{F(k)^2 + 4[2\lambda^{-1} \cos(k/2)]^2\}^{1/2}) \\ -2\lambda^{-1} \cos(k/2) \end{pmatrix} \quad (2.17)$$

where $F(k) = 1 - 2(1 + \cos k)/\lambda^2 + \rho/\lambda^2$. Note that we did not normalize the vector $\mathbf{v}^{(0)}(k)$. The eigenvector $\mathbf{v}^{(1/2)}(k)$ which corresponds to the eigenvalue $\varepsilon_2(k)$ can be written in terms of $\mathbf{v}^{(0)}(k)$ as

$$\mathbf{v}^{(1/2)}(k) = \begin{pmatrix} v_0^{(1/2)}(k) \\ v_{1/2}^{(1/2)}(k) \end{pmatrix} = \begin{pmatrix} -v_{1/2}^{(0)}(k) \\ v_0^{(0)}(k) \end{pmatrix} \quad (2.18)$$

2.4. Localized Bases for Single-Electron States

The band structure discussed above plays a fundamental role in the corresponding many-body problem as well. But the k-space picture, which was very useful in analyzing the band structure, turns out to be not quite effective in treating strong short-range interactions. This dilemma (which originates from the wave–particle dualism in quantum mechanics) suggests the need for a new description of electronic states which takes into account the band structure and at the same time treats electrons as "particles" rather than "waves."
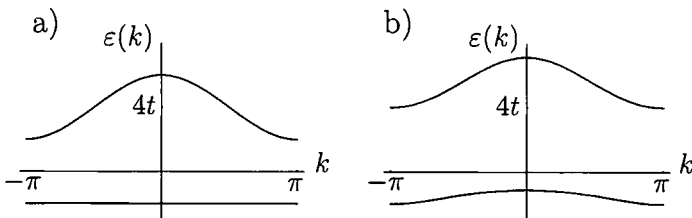

Fig. 2. The dispersion relations $\varepsilon_1(k)$, $\varepsilon_2(k)$ in the one-dimensional models. (a) The flat-band model with $\lambda = 2$, $\rho = 0$. (b) The perturbed model with $\lambda = 2$, $\rho = 0.2$, which has a nearly flat lower band and an energy gap between the two bands.

Let $\mu(x) = 0$ for $x \in \Lambda_o$ and $\mu(x) = 1/2$ for $x \in \Lambda'$. For $x, y \in \Lambda$, we define

$$\varphi_x^{(y)} = (2\pi)^{-1} \int dk \; e^{ik(x-y)} v_{\mu(x)}^{(\mu(y))}(k) \tag{2.19}$$

where $\int dk \; (\cdots)$ is a shorthand for the sum $(2\pi/L) \sum_{k \in \mathscr{K}} (\cdots)$. Suppose that $y \in \Lambda_o$ is fixed. Then one can regard[21] $\varphi^{(y)} = (\varphi_x^{(y)})_{x \in \Lambda}$ as (the wave function of) a single-electron state which is a superposition of the Bloch states $e^{ikx} v_{\mu(x)}^{(0)}(k)$ with various k. This means that, for any $y \in \Lambda_o$, the state $\varphi^{(y)}$ belongs to the Hilbert space of the lower band. By examining the definition (2.19), it follows that the collection $\{\varphi^{(y)}\}_{y \in \Lambda_o}$ forms a (non-orthonormal) basis of the Hilbert space corresponding to the lower band. Similarly the collection $\{\varphi^{(y)}\}_{y \in \Lambda'}$ forms a basis of the Hilbert space of the upper band.

Moreover, the states $\varphi^{(y)} = (\varphi_x^{(y)})_{x \in \Lambda}$ have rather nice localization properties. When $\rho = 0$, an explicit calculation shows that, for $y \in \Lambda_o$, $\varphi_y^{(y)} = 1$, and $\varphi_x^{(y)} = -1/\lambda$ if $|x - y| = 1/2$ and $\varphi_x^{(y)} = 0$ otherwise. (See Section 4.2, where $\varphi_x^{(y)}$ with $\rho = 0$ is denoted as $\psi_x^{(y)}$.) These are the strictly localized basis states constructed and used in refs. 47 and 34.

For $\rho \neq 0$, the basis states $\varphi^{(y)}$ are no longer strictly localized. Expanding the term $[F(k)^2 + 4\{2\lambda^{-1} \cos(k/2)\}^2]^{1/2}$ in (2.17) into a power series in λ^{-2} and (ρ/λ^2), we can still prove that the state $\varphi^{(y)} = (\varphi_x^{(y)})_{x \in \Lambda}$ is almost localized at the site y. More precisely, we have

$$\varphi_x^{(y)} \simeq \begin{cases} 1 & \text{if} \quad x = y \\ \pm 1/\lambda & \text{if} \quad |x - y| = 1/2 \\ O(|\rho|/\lambda^2) & \text{if} \quad |x - y| = 1 \\ \text{smaller, decays exponentially} & \text{for} \quad |x - y| > 1 \end{cases} \tag{2.20}$$

when $\lambda \gg 1$ and $|\rho|/\lambda^2 \ll 1$. (We take the $+$ sign if $y \in \Lambda'$ and the $-$ sign if $y \in \Lambda_o$). This sharp localization property of the states $\varphi^{(y)}$ plays a fundamental role throughout our proof.

Since the states $\varphi^{(y)}$ with different reference sites y are not necessarily orthogonal to each other, it is useful to introduce the dual of the basis $\{\varphi^{(y)}\}$. We shall construct the dual basis states $\tilde{\varphi}^{(y)} = (\tilde{\varphi}_x^{(y)})_{x \in \Lambda}$ so that $\sum_{x \in \Lambda} (\tilde{\varphi}_x^{(y)})^* \varphi_x^{(y')} = \delta_{y, y'}$ holds. [It also holds that $\sum_{y \in \Lambda} (\tilde{\varphi}_x^{(y)})^* \varphi_{x'}^{(y)} = \delta_{x, x'}$.] Then $\{\tilde{\varphi}^{(y)}\}_{y \in \Lambda_o}$ and $\{\tilde{\varphi}^{(y)}\}_{y \in \Lambda'}$ automatically form bases of the Hilbert spaces for the upper and lower bands, respectively. See Sections 10.1 and

---

[21] In a symbol like $\varphi_x^{(y)}$, the upper index y is the "name" given to the state, while the lower index x is the argument in the standard wave function representation. When we refer to the state itself, we write $\varphi^{(y)}$. Such a notation is used throughout the present paper.

10.2 for concrete procedures for constructing $\tilde{\varphi}^{(y)}$ from the vectors $\mathbf{v}^{(0)}(k)$ and $\mathbf{v}^{(1/2)}(k)$.

For $\lambda \gg 1$ and $|\rho|/\lambda^2 \ll 1$, the dual basis state $\tilde{\varphi}^{(y)}$ is localized at the site y as

$$\tilde{\varphi}_x^{(y)} \simeq \begin{cases} 1 & \text{if} \quad x = y \\ \pm 1/\lambda & \text{if} \quad |x - y| = 1/2 \\ -1/\lambda^2 & \text{if} \quad |x - y| = 1 \\ \text{smaller, decays exponentially} & \text{for} \quad |x - y| > 1 \end{cases} \quad (2.21)$$

where the $\pm$ sign is chosen as in (2.20). It should be noted that the states $\tilde{\varphi}^{(y)}$ are only moderately localized as compared with the sharp localization of $\varphi^{(y)}$. Even for $\rho = 0$, $\tilde{\varphi}_x^{(y)}$ has nonvanishing exponentially decaying tail.

Remark. It is interesting to compare the states $\varphi^{(y)}$ and $\tilde{\varphi}^{(y)}$ with Wannier functions.[20] Wannier functions are the standard machinery in condensed matter physics for providing the particlelike picture of electronic states by also taking into account band structures.

The Wannier functions $\omega^{(y)} = (\omega_x^{(y)})_{x \in \Lambda}$ are constructed as in (2.19), but with the vectors $\mathbf{v}^{(u)}(k)$ (with $u = 0, 1/2$) replaced by their normalized versions $\mathbf{v}^{(u)}(k)/|\mathbf{v}^{(u)}(k)|$. As a consequence $\{\omega^{(y)}\}_{y \in \Lambda_o}$ and $\{\omega^{(y)}\}_{y \in \Lambda'}$ form *orthonormal* bases of the Hilbert spaces for the upper and lower bands, respectively. As for the localization property, we have

$$\omega_x^{(y)} \simeq \begin{cases} 1 & \text{if} \quad x = y \\ \pm 1/\lambda & \text{if} \quad |x - y| = 1/2 \\ -1/(2\lambda^2) & \text{if} \quad |x - y| = 1 \\ \text{smaller, decays exponentially} & \text{for} \quad |x - y| > 1 \end{cases} \quad (2.22)$$

which is, roughly speaking, intermediate between those of $\varphi^{(y)}$ and $\tilde{\varphi}^{(y)}$.

Although the orthonormality of the Wannier basis is a clear advantage of this machinery, the poor localization property (2.22) is not optimal for our analysis of the Hubbard model. The sharp localization (2.20) is so important for us that we can give up the orthonormality of bases.

It is interesting that, in the context of a band calculation, Anderson[2] suggested using nonorthonormal basis states which are more sharply localized than the Wannier states. One can regard our $\varphi^{(y)}$ as a concrete (and typical) example of Anderson's ultralocalized functions, used in mathematical proofs of ferromagnetism rather than in band calculations.

In ref. 49, where the main results of the present paper were first announced, we claimed that the proof of local stability of ferromagnetism is impossible if we use the Wannier states instead of the sharply localized

states $\varphi^{(y)}$. We now feel, however, that a similar proof based on the Wannier functions may be constructed if one is careful enough in estimating various matrix elements.

2.5. Representation of the Hamiltonian in Terms of the Localized Basis

In order to analyze many-body problems by using the particlelike picture developed above, we introduce the fermion operators

$$a^{\dagger}_{x,\sigma} = \sum_{y \in \Lambda} (\varphi^{(x)}_{y})^{*} c^{\dagger}_{y,\sigma}, \qquad b_{x,\sigma} = \sum_{y \in \Lambda} \tilde{\varphi}^{(x)}_{y} c_{y,\sigma} \qquad (2.23)$$

for $x \in \Lambda$ and $\sigma = \uparrow, \downarrow$. It turns out that these operators obey the standard anticommutation relations such as $\{a^{\dagger}_{x,\sigma}, b_{y,\tau}\} = \delta_{x,y}\delta_{\sigma,\tau}$. This means that the "right" annihilation operator to be used with $a^{\dagger}_{x,\sigma}$ is $b_{x,\sigma}$, not $a_{x,\sigma} = (a^{\dagger}_{x,\sigma})^{\dagger}$.

As we show in Section 5.2, we can rewrite the Hamiltonian (2.3) in terms of these new operator as

$$H = \sum_{\substack{x, y \in \Lambda_{o} \\ \sigma = \uparrow,\downarrow}} \tau_{x,y} a^{\dagger}_{x,\sigma} b_{y,\sigma} + \sum_{\substack{x, y \in \Lambda' \\ \sigma = \uparrow,\downarrow}} \tau_{x,y} a^{\dagger}_{x,\sigma} b_{y,\sigma}$$

$$+ \sum_{y,v,w,z \in \Lambda} \tilde{U}_{y,v;w,z} a^{\dagger}_{y,\uparrow} a^{\dagger}_{v,\downarrow} b_{w,\downarrow} b_{z,\uparrow} \qquad (2.24)$$

Note that there is no hopping between Λ_{o} and Λ' in the hopping parts of H. This is because the operators $a^{\dagger}_{x,\sigma}$ and $b_{y,\sigma}$ "know" about the band structure. As for the properties of the effective hopping $\tau_{x,y}$, we only need to note that $\tau_{x,x+1} = O(|\rho| \, t)$ for $x \in \Lambda_{o}$, and $\tau_{x,x} \simeq \lambda^{2} t$ for $x \in \Lambda'$.

Note that the interaction term in (2.24) is no longer on-site. This fact is essential for the appearance of ferromagnetism in these models. The effective (four Fermi) coupling $\tilde{U}_{y,v;w,z}$ in (2.24) is given by

$$\tilde{U}_{y,v;w,z} = U \sum_{x \in \Lambda} \tilde{\varphi}^{(y)}_{x} \tilde{\varphi}^{(v)}_{x} (\varphi^{(w)}_{x} \varphi^{(z)}_{x})^{*} \qquad (2.25)$$

This expression means that the coupling function $\tilde{U}_{y,v;w,z}$ is determined by the overlap between the four states $\tilde{\varphi}^{(y)}$, $\tilde{\varphi}^{(v)}$, $\varphi^{(w)}$, and $\varphi^{(z)}$, where the former two states are created and the latter two states are annihilated. Since each state $\varphi^{(y)}$ or $\tilde{\varphi}^{(y)}$ is localized at the reference site y, we find that $\tilde{U}_{y,v;w,z}$ is also short ranged. We can say that our representation success-fully took into account the particlelike nature of electrons. We also note

that the coupling function satisfies the translation invariance $\tilde{U}_{y,v;w,z} = \tilde{U}_{y+p,v+p;w+p,z+p}$ for any $p \in \mathbf{Z}$.

Let us assume $\lambda \gg 1$ and $|\rho|/\lambda^2 \ll 1$. Then we can substitute the properties (2.20) and (2.21) of the basis states into (2.25) and evaluate $\tilde{U}_{y,v;w,z}$ explicitly as

$$\tilde{U}_{0,0;0,0} \simeq U\tilde{\varphi}_0^{(0)}\tilde{\varphi}_0^{(0)}(\varphi_0^{(0)}\varphi_0^{(0)})^* \simeq U \tag{2.26}$$

$$\tilde{U}_{0,1;0,1} \simeq U\tilde{\varphi}_{1/2}^{(0)}\tilde{\varphi}_{1/2}^{(1)}(\varphi_{1/2}^{(0)}\varphi_{1/2}^{(1)})^* \simeq \frac{U}{\lambda^4} \tag{2.27}$$

$$\tilde{U}_{0,1;0,0} \simeq U\tilde{\varphi}_0^{(0)}\tilde{\varphi}_0^{(1)}(\varphi_0^{(0)}\varphi_0^{(0)})^* \simeq -\frac{U}{\lambda^2} \tag{2.28}$$

$$\tilde{U}_{0,0;0,1} \simeq U\{\tilde{\varphi}_0^{(0)}\tilde{\varphi}_0^{(0)}(\varphi_0^{(0)}\varphi_0^{(1)})^* + \tilde{\varphi}_{1/2}^{(0)}\tilde{\varphi}_{1/2}^{(0)}(\varphi_{1/2}^{(0)}\varphi_{1/2}^{(1)})^*\}$$

$$\simeq O\left(\frac{U|\rho|}{\lambda^2}\right) + \frac{U}{\lambda^4} \tag{2.29}$$

and

$$\tilde{U}_{0,1;1/2,1/2} \simeq \tilde{U}_{1/2,1/2;0,1} \simeq \frac{U}{\lambda^2} \tag{2.30}$$

These are the components of $\tilde{U}$ which play important roles when we investigate low-lying excited states of our Hubbard model. Note that $\tilde{U}_{0,0;0,1}$ and $\tilde{U}_{0,1;0,0}$ are drastically different. This asymmetry, which originates from the difference in the localization properties (2.20), (2.21) of the states $\varphi^{(y)}$ and $\tilde{\varphi}(y)$, is important for our proof.

2.6. Perturbative Analysis and Effective Spin Hamiltonian

At this stage we shall develop a heuristic theory which reveals why our Hubbard model exhibits a stable ferromagnetism. This subsection is different from all the others in that it is devoted to arguments which are not yet made rigorous. This, however, allows us to go beyond our technical limitation, and discuss the stability of ferromagnetism beyond a single-spin flip.

Here we focus on the region of parameters characterized by $|\rho|\, t \ll U \ll \lambda^2 t$. Recall that $|\rho|\, t$, U, and $\lambda^2 t$ roughly represent the bandwidth of the lower band, local Coulomb interaction, and the band gap, respectively. By

examining the representation (2.24) of the Hamiltonian, we extract the most dominant part as the "unperturbed" Hamiltonian

$$H_0 = \sum_{\substack{x, y \in \Lambda' \\ \sigma = \uparrow, \downarrow}} \tau_{x, y} a^\dagger_{x, \sigma} b_{y, \sigma} + \sum_{x \in \Lambda} \tilde{U}_{x, x; x, x} \tilde{n}_{x, \uparrow} \tilde{n}_{x, \downarrow} \qquad (2.31)$$

Here we introduced the pseudo number operator $\tilde{n}_{x, \sigma} = a^\dagger_{x, \sigma} b_{x, \sigma}$. Although $\tilde{n}_{x, \sigma}$ is not Hermitian, it works in exactly the same way as the standard number operator as long as one uses $a^\dagger_{x, \sigma}$ and $b_{x, \sigma}$ as creation and annihilation operators, respectively. By recalling $\tau_{x, x} \simeq \lambda^2 t$ for $x \in \Lambda'$ and $\tilde{U}_{x, x; x, x} \simeq U$, we find that the conditions for minimizing H_0 are (i) there are only electrons from the lower band, i.e., those created by $a^\dagger_{x, \sigma}$ with $x \in \Lambda_o$, and (ii) there are no doubly occupied sites in the language of $\tilde{n}_{x, \sigma}$. Since the number of electrons L is the same as the number of the sites in Λ_o, such states can be written as linear combinations of the states

$$\Phi_\sigma = \left(\prod_{x \in \Lambda_o} a^\dagger_{x, \sigma(x)} \right) \Phi_{\text{vac}} \qquad (2.32)$$

Here the multi-index $\sigma = (\sigma(x))_{x \in \Lambda_o}$ with $\sigma(x) = \uparrow, \downarrow$ represents spin configurations. Clearly we have $H_0 \Phi_\sigma = 0$ for any σ. The unperturbed Hamiltonian H_0 has 2^L-fold degenerate ground states.

Let us examine how the degeneracy is lifted when we consider the remainder of the Hamiltonian, which is

$$H_{\text{pert}} = \sum_{\substack{x, y \in \Lambda_o \\ \sigma = \uparrow, \downarrow}} \tau_{x, y} a^\dagger_{x, \sigma} b_{y, \sigma} + \sum_{\substack{y, v, w, z \in \Lambda \\ (\text{except } y = v = w = z)}} \tilde{U}_{y, v; w, z} a^\dagger_{y, \uparrow} a^\dagger_{v, \downarrow} b_{w, \downarrow} b_{z, \uparrow} \qquad (2.33)$$

We wish to develop a standard first-order perturbation theory, but using the nonorthonormal basis consisting of the states $(\prod_{x \in A} a^\dagger_{x, \uparrow})(\prod_{x \in B} a^\dagger_{x, \downarrow}) \Phi_{\text{vac}}$, where A, B are arbitrary subsets of Λ. Let P_0 be the projection operator[22] (defined with respect to the present basis) onto the 2^L-dimensional ground-state space spanned by Φ_σ. The basic object in the first-order degenerate perturbation theory is then the effective Hamiltonian $H_{\text{eff}} = P_0 H_{\text{pert}} P_0$. Note that H_{eff} is not a self-adjoint operator. This is inevitable since we are developing a perturbation theory based on a nonorthonormal basis. Since the standard perturbation theory can be applied to non-Hermitian matrices as well, the situation is by no means pathological. There is a similar perturbation theory that uses an orthonormal basis constructed from the Wannier states.[52]

---

[22] The procedure for defining P_0 is as follows. Given a many-electron state Φ, one (uniquely) expands it in terms of the basis states. Then one throws away all the basis states which are not of the form Φ_σ, (2.32). The resulting state is $P_0 \Phi$. Note that P_0 is not an orthogonal projection.

Obviously a term contributing to H_{eff} should not affect the locations of the electrons. As a consequence, contributions come from the so-called "exchange" terms (and the diagonal elements of $\tau_{x,y}$) as

$$H_{\text{eff}} = \left\{ E_0 + \sum_{\substack{x,y \in \Lambda_o \\ (x \neq y)}} \tilde{U}_{x,y;x,y}(a^\dagger_{x,\uparrow} a^\dagger_{y,\downarrow} b_{y,\downarrow} b_{x,\uparrow} + a^\dagger_{y,\uparrow} a^\dagger_{x,\downarrow} b_{y,\downarrow} b_{x,\uparrow}) \right\} P_{\text{p}} \quad (2.34)$$

where $E_0 = \sum_{x \in \Lambda_o} \tau_{x,x}$ turns out to be the energy of the "ferromagnetic ground states." (Figure 4 in Section 2.7 illustrates how the "exchange" terms act on a state.)

It turns out that the "exchange" term is the ultimate origin of ferromagnetism in our Hubbard model. In the present model, the "exchange" takes place between the spins of two electrons in neighboring Λ_o sites (metallic atoms). By recalling that there is a Λ' site (oxygen atom) in between them, one might prefer to call the present process "super-exchange."[1] While this terminology is also possible, we stress that we never get ferromagnetism if the direct hopping between Λ_o sites [represented by t in the Hamiltonian (2.3)] is absent, as we discussed in Section 2.2. We think there is a much more delicate mechanism going on here than what one would naively expect from a "superexchange" process.

Let us define the pseudo spin operators by $\tilde{S}^{(j)}_x = \sum_{\sigma,\tau=\uparrow,\downarrow} a^\dagger_{x,\sigma} p^{(j)}_{\sigma,\tau} b_{x,\tau}/2$ for $j = 1$, 2, and 3, where $p^{(j)}_{\sigma,\tau}$ are the Pauli matrices (3.26). Again these operators are not Hermitian, but work in exactly the same way as the standard spin operators. Then the effective Hamiltonian is rewritten as

$$H_{\text{eff}} = \left[E_0 - \sum_{\substack{x,y \in \Lambda_o \\ (x \neq y)}} \tilde{U}_{x,y;x,y} \left\{ \left(\sum_{j=1}^{3} \tilde{S}^{(j)}_x \tilde{S}^{(j)}_y \right) - \frac{3}{4} \right\} \right] P_0$$

$$\simeq \left[E_0 - \frac{2U}{\lambda^4} \sum_{x \in \Lambda_o} \left\{ \left(\sum_{j=1}^{3} \tilde{S}^{(j)}_x \tilde{S}^{(j)}_{x+1} \right) - \frac{3}{4} \right\} \right] P_0 \quad (2.35)$$

where we used the estimate (2.27) for $\tilde{U}$ to get the final line. The right-hand side of (2.35) is nothing but the Hamiltonian of the nearest neighbor Heisenberg chain with the ferromagnetic interaction $J_{\text{eff}} \simeq 2U/\lambda^4$. We have successfully derived a ferromagnetic spin system starting from the Hubbard model for itinerant electrons.

If we believe in this first-order perturbation theory, then we can conclude from the "spin Hamiltonian" (2.35) that the ground states of the present Hubbard model are the ferromagnetic states given by

$$\Phi_\uparrow = \left(\prod_{x \in \Lambda_o} a^\dagger_{x,\uparrow} \right) \Phi_{\text{vac}} \quad (2.36)$$

and its $SU(2)$ rotations. Moreover, low-energy excitations of the Hubbard model should coincide with those of the ferromagnetic Heisenberg model (2.35). The elementary spin-wave excitation should then have the dispersion relation

$$E_{SW}(k) - E_0 = 2J_{eff}\left(\sin\frac{k}{2}\right)^2 \simeq \frac{4U}{\lambda^4}\left(\sin\frac{k}{2}\right)^2 \qquad (2.37)$$

Note that this heuristic estimate exactly coincides with our rigorous result (2.13)!

It should be stressed, however, that the above naive perturbation theory remains to be justified in many aspects. We have been neglecting many contributions without giving any estimates. The most important contribution that has been neglected comes from the second-order perturbation from the hopping terms or the effective hopping terms (as illustrated in Fig. 5 in Section 2.7). Since such a perturbation lowers the energy of electron pairs in a spin singlet, it weakens the tendency toward ferromagnetism. Fortunately, a rough estimate shows that this effect is small provided that $|\rho| \, t \ll U \ll \lambda^2 t$.

We do not argue here that the validity of the present perturbation theory can be established. By comparing it with our rigorous results about the local stability of ferromagnetism and the spin-wave excitation, however, it seems rather likely that this treatment gives sensible conclusions about low-energy properties of our Hubbard model. In ref. 52 we further discuss the derivation of low-energy effective spin Hamiltonians in the Hubbard models.

2.7. Sketch of the Proof

We will now illustrate how the theorems discussed in Section 2.1 are proved. The heart of the proof is to construct rather accurate trial states for the spin-wave excitations and carefully examine the action of the Hamiltonian on them.

To begin with, we note that one of the "ferromagnetic ground states" (defined in Section 2.1 as the lowest energy states within the sector with $S_{tot} = S_{max}$) can be written as in (2.36). Note that (2.36) is nothing but the state obtained by "completely filling" the (single-electron) states in the lower band with up-spin electrons. The energy of $\Phi_\uparrow$ is given by $E_0 = \sum_{x \in A_o} t_{x,x}$.

As a candidate for the spin-wave excitation, we shall consider the state in which a single down-spin propagates in $\Phi_\uparrow$ with a momentum k as

$$\Omega(k) = \alpha(k)^{-1} \sum_{x \in \Lambda_o} e^{ikx} \Gamma_x \tag{2.38}$$

where we introduced

$$\Gamma_x = a_{x,\downarrow}^\dagger b_{x,\uparrow} \Phi_\uparrow \tag{2.39}$$

The normalization $\alpha(k)$ will be determined later.

Since the annihilation operator $b_{x,\uparrow}$ properly cancels out with the creation operator $a_{x,\uparrow}^\dagger$, the state Γ_x is such that (one of) $a_{x,\uparrow}^\dagger$ in (2.36) is replaced with $a_{x,\downarrow}^\dagger$. Recall that, as can be seen from the definition (2.23), the operator $a_{x,\sigma}^\dagger$ creates an electron in the sharply localized state $\varphi^{(x)}$. As the localization property (2.20) of $\varphi^{(x)}$ shows, two neighboring states $\varphi^{(x)}$ and $\varphi^{(x+1)}$ have very small overlap (of order $1/\lambda^2$). This means that the down-spin electron inserted in (2.38) costs very small energy due to the Coulomb repulsion $U \sum_{x \in \Lambda} n_{x,\uparrow} n_{x,\downarrow}$. At the same time it costs small kinetic energy since it only contains (single-electron) states from the lower band. These observations suggest that the $\Omega(k)$ of (2.38) are good trial states for low-lying excitations in which both the kinetic energy and the Coulomb interaction are properly taken into account. See Fig. 3.

To prove the upper bound for the spin-wave dispersion relation in (2.10), we employ the standard variational inequality [see (9.1)] and calculate the expectation value of H in the state $\Omega(k)$. We then find that the main contribution comes from the "exchange" Hamiltonian (2.35), which leads us to the desired upper bound. See Section 9 for details.

To further investigate the accuracy of the trial state and to get the lower band in (2.10), we apply the Hamiltonian to $\Omega(k)$. Although an

x
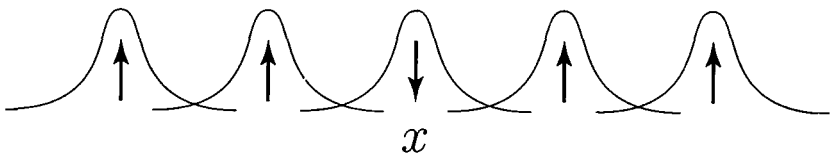
Fig. 3. A schematic picture of the state Γ_x, (2.39), which appears in the definition of $\Omega(k)$, (2.38). Since Γ_x is constructed by using sharply localized basis states for the lower band, it costs small Coulomb repulsion energy and small kinetic energy. In the state $\Omega(k)$, the down spin propagates with momentum k and further reduces the total energy to $\simeq E_0 + (4U/\lambda^4) \{\sin(k/2)\}^2$.

enormous number of terms can appear, the major contributions[23] come from three basic short-range processes that we now describe. The three processes are represented by the following three operators (2.40)–(2.42), which are extracted from the Hamiltonian in the form (2.24). The first process is the nearest neighbor "exchange" discussed in Section 2.6, which is represented by

$$H_1 = \sum_{\substack{x \in \Lambda_o \\ \sigma = \uparrow, \downarrow}} \tilde{U}_{x,x+1;x,x+1} (a^\dagger_{x,\sigma} a^\dagger_{x+1,-\sigma} b_{x+1,-\sigma} b_{x,\sigma}$$
$$+ a^\dagger_{x+1,\sigma} a^\dagger_{x,-\sigma} b_{x+1,-\sigma} b_{x,\sigma}) \tag{2.40}$$

The second process is the nearest neighbor hopping[24] represented by

$$H_2 = \sum_{\substack{x \in \Lambda_o \\ r = \pm 1 \\ \sigma = \uparrow, \downarrow}} \{ \tilde{U}_{x+r,x+r;x+r,x} a^\dagger_{x+r,\sigma} a^\dagger_{x+r,-\sigma} b_{x+r,-\sigma} b_{x,\sigma} + \tau_{x+r,x} a^\dagger_{x+r,\sigma} b_{x,\sigma} \} \tag{2.41}$$

The third process is represented by

$$H_3 = \sum_{\substack{x \in \Lambda_o \\ \sigma = \uparrow, \downarrow}} \tilde{U}_{x+(1/2),x+(1/2);x,x+1} a^\dagger_{x+(1/2),\sigma} a^\dagger_{x+(1/2),-\sigma} b_{x,-\sigma} b_{x+1,\sigma} \tag{2.42}$$

which lets two electrons in neighboring Λ_o sites $x, x+1$ hop simultaneously to the site between them. Note that $a^\dagger_{x+(1/2),\sigma}$ creates an electron in the upper band.

Let us investigate the action of these partial Hamiltonians (2.40)–(2.42) on the state $\Omega(k)$. It is useful to first consider the action on the state Γ_y defined in (2.39) which contains a down-spin electron at site y. By operating the "exchange" Hamiltonian (2.40) on Γ_y, we find

$$H_1 \Gamma_y = \tilde{U}_{y-1,y;y-1,y} (\Gamma_y - \Gamma_{y-1}) + \tilde{U}_{y,y+1;y,y+1} (\Gamma_y - \Gamma_{y+1})$$
$$= \tilde{U}_{0,1;0,1} (2\Gamma_y - \Gamma_{y-1} - \Gamma_{y+1}) \tag{2.43}$$

where the minus signs come from fermion ordering for the "exchanged" configurations. Figure 4 illustrates how these four terms arise. We also used

---

[23] In the later sections, we of course control all the possible contributions.

[24] The first operator annihilates an electron at $x+r$ with spin $-\sigma$, and then creates the same thing. Therefore its action is the same as the second operator $a^\dagger_{x+r,\sigma} b_{x,\sigma}$ provided that there is a spin $(-\sigma)$ electron at $x+r$.
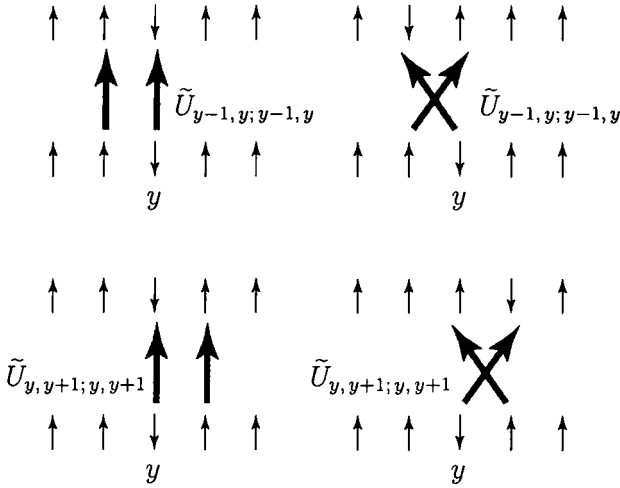
Fig. 4.   When the "exchange" Hamiltonian H_1 acts on the state Γ_y, four terms are generated. Two of them are the same as Γ_y, while the electronic spins are exchanged in the other two states. The process illustrated here can be regarded as the ultimate origin of ferromagnetism in the present model.

the translation invariance of $\tilde{U}$, which is indeed essential for the present proof. Recalling (2.38), we get the expected result

$$H_1 \Omega(k) = \left\{ E_0 + 4\tilde{U}_{0,1;0,1} \left(\sin \frac{k}{2} \right)^2 \right\} \Omega(k) \tag{2.44}$$

To see the role of H_2, (2.41), let us operate with it on Γ_y to get

$$H_2 \Gamma_y = \sum_{r=\pm 1} \left\{ (\tilde{U}_{y+r,y+r;y+r,y} + \tau_{y+r,y}) \tilde{\Gamma}_{y+r}^{(r)} - (\tilde{U}_{y,y;y,y-r} + \tau_{y,y-r}) \tilde{\Gamma}_y^{(r)} \right\}$$

$$= (\tilde{U}_{r,r;r,0} + \tau_{r,0})(\tilde{\Gamma}_{y+r}^{(r)} - \tilde{\Gamma}_y^{(r)}) \tag{2.45}$$

where $\tilde{\Gamma}_y^{(r)} = a_{y,\downarrow}^\dagger b_{y-r,\uparrow} \Phi_\uparrow$ is the state with an empty site $y - r$ and a doubly occupied site y. We also used the translation invariance of $\tilde{U}$ and τ. In Fig. 5, we illustrate the action of a part of H_2 on Γ_y. Again from (2.38), we find that

$$H_2 \Omega(k) = \sum_{r=\pm 1} \frac{e^{-ikr} - 1}{\alpha(k)} (\tilde{U}_{r,r;r,0} + \tau_{r,0}) \, \Xi_r(k)$$

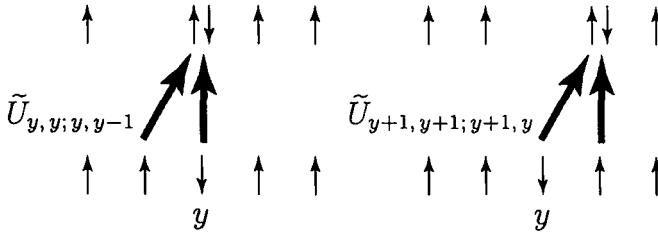$$= \sum_{r=\pm 1} \frac{e^{-ikr} - 1}{\alpha(k)} (\tilde{U}_{1,1;1,0} + \tau_{1,0}) \, \Xi_r(k) \tag{2.46}$$

Fig. 5. When the (effective hopping) operator $\sum_{x \in \Lambda, \sigma = \uparrow, \downarrow} \tilde{U}_{x+1,x+1;x+1,x} a^\dagger_{x+1,\sigma} a^\dagger_{x+1,-\sigma} b_{x+1,-\sigma} b_{x,\sigma}$ acts on Γ_y, two states with a bound pair of a hole and a doubly occupied site are generated. Note that the two resulting states are related through the translation by a distance 1. This process is the major source of instability against antiferromagnetism.

where we used the reflection symmetry[25] to get the final line. Here

$$\Xi_r(k) = \sum_{x \in \Lambda_o} e^{ikx} \tilde{\Gamma}^{(r)}_x = \sum_{x \in \Lambda_o} e^{ikx} a^\dagger_{x,\downarrow} b_{x-r,\uparrow} \Phi_\uparrow \qquad (2.47)$$

is the state in which a bound pair of an empty site $x - r$ and a doubly occupied site x is propagating with momentum k. We shall abbreviate $\Xi_{\pm 1}(k)$ as $\Xi_\pm(k)$.

Similarly we obtain

$$H_3 \Omega(k) = \frac{e^{ik} - 1}{\alpha(k)} \tilde{U}_{1/2,1/2;0,1} \Theta(k) \qquad (2.48)$$

where

$$\Theta(k) = \sum_{x \in \Lambda_o} e^{ikx} a^\dagger_{x+(1/2),\downarrow} a^\dagger_{x+(1/2),\uparrow} b_{x,\uparrow} b_{x+1,\uparrow} \Phi_\uparrow \qquad (2.49)$$

is the state in which two adjacent empty sites in Λ_o and a doubly occupied site between them form a bound state and propagating with momentum k. See Fig. 6 for schematic pictures of the states[26] $\Omega(k)$, $\Xi_+(k)$, and $\Theta(k)$. We also note here that these states all belong to the Hilbert space $\mathcal{H}_k$, which we introduced just above Theorem 2.4.

The relations (2.46) and (2.48) clearly show that our trial state $\Omega(k)$ cannot be the exact eigenstate of the Hamiltonian. To investigate the low-lying spectrum of H, we have to consider (at least) the subspace spanned

[25] Such a symmetry exists in the present model. In the general class of models studied later, we do not assume reflection or rotation symmetries.

[26] In the general notation used in the later sections, the states $\Xi_r(k)$ and $\Theta(k)$ are denoted as $\Phi_{0,r}(k)$ and $\Phi_{1/2,1/2,0,1}(k)$, respectively. See Section 6.2.

by the states $\Omega(k)$, $\varXi_{\pm}(k)$, and $\Theta(k)$. As before, we calculate the action of H on the latter states to find

$$H\varXi_{\pm}(k) \simeq (E_0 + \tilde{U}_{0,0;0,0})\, \varXi_{\pm}(k) + \alpha(k)(\tilde{U}_{0,1;1,1} + \tau_{0,1})(e^{\mp ik} - 1)\, \Omega(k)$$
$$+ (\text{other states}) \tag{2.50}$$

and

$$H\Theta(k) \simeq (E_0 - 2\tau_{0,0} + 2\tau_{1/2,1/2} + \tilde{U}_{1/2,1/2;1/2,1/2})\, \Theta(k)$$
$$+ \alpha(k)(e^{-ik} - 1)\, \tilde{U}_{0,1;1/2,1/2}\Omega(k) + (\text{other states}) \tag{2.51}$$

Although it might not be clear at this stage, it turns out that the "other states" in (2.50) and (2.51) do not play essential roles. We leave such estimates (as well as the precise definition of the "other states") to the latter sections, and simply neglect them here.

Equations (2.44), (2.46), (2.48), (2.50), and (2.51) provide, for each $k \in \mathscr{K}$, the representation of the Hamiltonian in the four-dimensional subspace of $\mathscr{H}_k$ spanned by the states $\Omega(k)$, $\varXi_{\pm}(k)$, and $\Theta(k)$. We now read off the matrix elements[27] from these equations and then use the estimates (2.26)–(2.30) of $\tilde{U}$ to evaluate them as

$$h[\Omega(k), \Omega(k)] \simeq E_0 + 4\tilde{U}_{0,1;0,1}\left(\sin\frac{k}{2}\right)^2$$
$$\simeq E_0 + 4\frac{U}{\lambda^4}\left(\sin\frac{k}{2}\right)^2 \tag{2.52}$$

$$h[\varXi_{\pm}(k), \varXi_{\pm}(k)] \simeq E_0 + \tilde{U}_{0,0;0,0} \simeq E_0 + U \tag{2.53}$$

$$h[\Theta(k), \Theta(k)] \simeq E_0 - 2\tau_{0,0} + 2\tau_{1/2,1/2} + \tilde{U}_{1/2,1/2;1/2,1/2}$$
$$\simeq E_0 + 2\lambda^2 t + U \tag{2.54}$$

$$h[\varXi_{\pm}(k), \Omega(k)] \simeq \alpha(k)^{-1}(\tilde{U}_{1,1;1,0} + \tau_{1,0})(e^{\pm ik} - 1)$$
$$\simeq \alpha(k)^{-1}\left(\frac{cU\rho}{\lambda^2} + \frac{U}{\lambda^4} + c'\rho t\right)(e^{\pm ik} - 1) \tag{2.55}$$

$$h[\Omega(k), \varXi_{\pm}(k)] \simeq \alpha(k)(\tilde{U}_{0,1;1,1} + \tau_{0,1})(e^{\mp ik} - 1)$$
$$\simeq \alpha(k)\left(-\frac{U}{\lambda^2} + c'\rho t\right)(e^{\mp ik} - 1) \tag{2.56}$$

---

[27] As usual, matrix elements $h[\varPsi, \varPhi]$ are defined by the unique expansion $H\varPhi = \sum_{\varPsi} h[\varPsi, \varPhi]\, \varPsi$. See also (6.2).

$$h[\Theta(k), \Omega(k)] \simeq \alpha(k)^{-1} \, \tilde{U}_{1/2, 1/2; 0, 1}(e^{ik} - 1)$$

$$\simeq \alpha(k)^{-1} \frac{U}{\lambda^2} (e^{ik} - 1) \tag{2.57}$$

and

$$h[\Theta(k), \Omega(k)] \simeq \alpha(k) \, \tilde{U}_{0, 1; 1/2, 1/2}(e^{-ik} - 1)$$

$$\simeq \alpha(k) \frac{U}{\lambda^2} (e^{-ik} - 1) \tag{2.58}$$

where the approximate values are obtained for $\lambda \gg 1$ and $|\rho|/\lambda^2 \ll 1$, and c, c' are constants. Reflecting the use of the nonorthonormal basis, these matrix elements are not symmetric. In particular, the drastic difference between the elements $h[\Xi_{\pm}(k), \Omega(k)]$ of (2.55) and $h[\Omega(k), \Xi_{\pm}(k)]$ of (2.56) plays a fundamental role in our proof. Figure 6 shows these matrix elements.
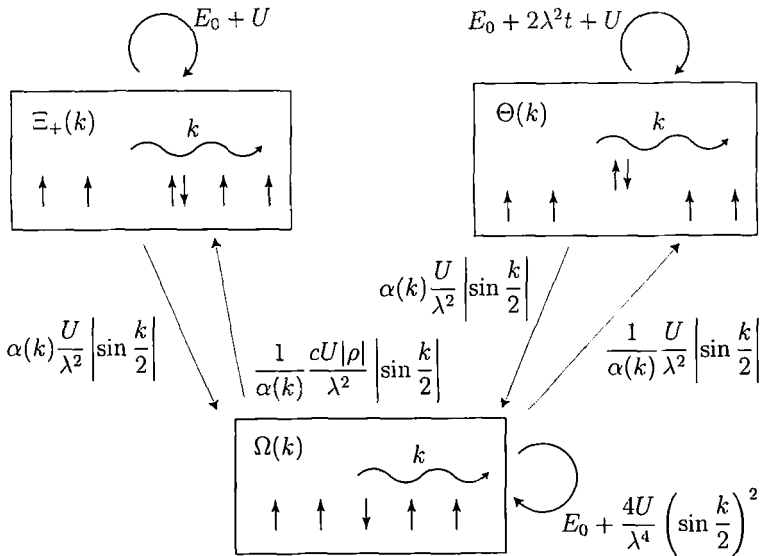
Fig. 6. Schematic pictures of the states $\Omega(k)$, $\Xi_{+}(k)$, and $\Theta(k)$ and the matrix elements between them. As for the off-diagonal matrix elements, we only present the main part of their absolute values. Note that the matrix elements between $\Omega(k)$ and $\Xi_{+}(k)$ are highly asymmetric [apart from the artificial asymmetry factor $\alpha(k)$]. The small outgoing matrix elements from $\Omega(k)$ indicate that the state $\Omega(k)$ is a good trial state for the spin-wave excitation.

In order to bound the excitation energy from below, we use the following well-known fact about the minimum eigenvalue of a matrix. Let $(h_{i,j})_{i,j=1,\ldots,N}$ be an $N \times N$ matrix with real eigenvalues. Then the lowest eigenvalue h_0 of the matrix satisfies $h_0 \geqslant \min_{i=1,\ldots,N} D_i$ with $D_i = \mathrm{Re}[h_{i,i}] - \sum_{j \neq i} |h_{i,j}|$. This is almost trivial, but see Lemma 6.1 for a proof. We stress that this simple-minded inequality is expected to yield physically meaningful results only when one uses a basis which "almost diagonalizes" the low-energy part of the Hamiltonian.

We now apply this inequality to the 4×4 matrix representation of H in each sector with a fixed $k \in \mathcal{K}$. The quantities corresponding to D_i are evaluated for each state as

$$D[\Omega(k)] = h[\Omega(k), \Omega(k)] - \sum_{\pm} |h[\Omega(k), \Xi_{\pm}(k)]| - |h[\Omega(k), \Theta(k)]|$$

$$\geqslant E_0 + \frac{4U}{\lambda^4} \left(\sin \frac{k}{2} \right)^2 - 4\alpha(k) \left(\frac{U}{\lambda^2} + c' |\rho| t \right) \left| \sin \frac{k}{2} \right|$$

$$- 2\alpha(k) \frac{U}{\lambda^2} \left| \sin \frac{k}{2} \right| \tag{2.59}$$

$$D[\Xi_{\pm}(k)] = h[\Xi_{\pm}(k), \Xi_{\pm}(k)] - |h[\Xi_{\pm}(k), \Omega(k)]|$$

$$\geqslant E_0 + U - 2\alpha(k)^{-1} \left(\frac{cU |\rho|}{\lambda^2} + \frac{U}{\lambda^4} + c' |\rho| t \right) \left| \sin \frac{k}{2} \right| \tag{2.60}$$

and

$$D[\Theta(k)] = h[\Theta(k), \Theta(k)] - |h[\Theta(k), \Omega(k)]|$$

$$\geqslant E_0 + 2\lambda^2 t + U - 2\alpha(k)^{-1} \frac{U}{\lambda^2} \left| \sin \frac{k}{2} \right| \tag{2.61}$$

where we used the estimates (2.52)–(2.58) to get the lower bounds.

At this stage, we choose[28] the constant $\alpha(k)$ as

$$\alpha(k) = 4 \left(\frac{c |\rho|}{\lambda^2} + \frac{1}{\lambda^4} + \frac{c' |\rho| t}{U} \right) \left| \sin \frac{k}{2} \right| \tag{2.62}$$

---

[28] The choice of $\alpha(k)$ here is different from that in the full proof in Section 6. [See (6.31).] This is because the actual estimate of $D[\Xi_{\pm}(k)]$ in the later sections takes into account various small terms which are simply neglected here.

This choice makes the bound (2.60) into

$$D[\Xi_\pm(k)] \geqslant E_0 + \frac{U}{2} \tag{2.63}$$

and the bound (2.61) into

$$
\begin{aligned}
D[\Theta(k)] &\geqslant E_0 + 2\lambda^2 t + U - \frac{U}{2\lambda^2}\left(\frac{c\,|\rho|}{\lambda^2} + \frac{1}{\lambda^4} + \frac{c'\,|\rho|\,t}{U}\right)^{-1} \\
&\geqslant E_0 + 2\lambda^2 t + U - \frac{\lambda^2 U}{2} \\
&\geqslant E_0 + U \tag{2.64}
\end{aligned}
$$

where we have further assumed[29] $U \leqslant 4t$. Finally we substitute (2.62) into the bound (2.59) to get

$$
\begin{aligned}
D[\Omega(k)] &\geqslant E_0 + \frac{4U}{\lambda^4}\left(\sin\frac{k}{2}\right)^2 \\
&\quad - 4\left(\frac{c\,|\rho|}{\lambda^2} + \frac{1}{\lambda^4} + \frac{c'\,|\rho|\,t}{U}\right)\left(\frac{6U}{\lambda^2} + 4c'\,|\rho|\,t\right)\left(\sin\frac{k}{2}\right)^2 \\
&= E_0 + \frac{4U}{\lambda^4}\left\{1 - \left(c\,|\rho| + \frac{1}{\lambda^2} + \frac{c'\,|\rho|\,\lambda^2 t}{U}\right)\right. \\
&\quad \times \left.\left(6 + \frac{4c'\,|\rho|\,\lambda^2 t}{U}\right)\right\}\left(\sin\frac{k}{2}\right)^2 \\
&\geqslant E_0 + \frac{4U}{\lambda^4}\left\{1 - A_1\,|\rho| - \frac{A_2}{\lambda} - \frac{A_3\lambda^2 t\,|\rho|}{U}\right\}\left(\sin\frac{k}{2}\right)^2 \tag{2.65}
\end{aligned}
$$

with constants[30] A_1, A_2, and A_3. Since the lower bounds (2.60) and (2.61) for $D[\Xi_\pm(k)]$ and $D[\Theta(k)]$ are strictly larger than that for $D[\Omega(k)]$, we find that the right-hand side of (2.65) gives the desired lower bound for the lowest eigenvalue of the Hamiltonian in the space $\mathscr{H}_k$. Therefore the lower bound for the spin-wave excitation in (2.10) (which is the main statement of Theorem 2.4) has been derived.

[29] The upper bound required for U depends sensitively on the choice of $\alpha(k)$. The requirement $U \leqslant K_2 \lambda t$ made in Theorem 2.4 (and which appears in the full proof) is somewhat different from the present one.

[30] Here the term A_2/λ in the right-hand side of (2.65) can be replaced with A_2/λ^2 if we simply equate the above expression. Since the actual matrix elements have many "small" terms that are neglected here, what we can prove (in the later sections with perfect rigor) is the bound in terms of the quantity in the right-hand side of (2.65).

In the remainder, we sketch how we get Theorem 2.3 about the local stability of ferromagnetism from the above lower bounds. The lower bounds in (2.10) gives strict bounds $E_{SW}(k) > E_0$ for all $k \in \mathcal{K}$ except for $k = 0$. This means that the desired local stability inequality (2.9) has been proved except in the translation-invariant sector with $k = 0$. To deal with the $k = 0$ sector is easy once we realize that $\Omega(0)$ is nothing but an $SU(2)$ rotation of the "ferromagnetic ground state" $\Phi_\uparrow$. By simply repeating the above arguments for the *three*-dimensional subspace spanned by $\Xi_\pm(0)$ and $\Theta(0)$, we easily find that the desired bound (2.9) also holds in the sector with $k = 0$. It only remains to extend the parameter region in which the statement is valid. This is easily done by a general consideration about the monotonicity of energies as a function of U. See Section 6.4.

3. DEFINITIONS AND MAIN RESULTS

In the present section, we define the general class of models treated in the present paper and precisely state our main theorems.

3.1. Lattice

We describe the lattice on which our Hubbard model is defined. The lattice is characterized by the dimension of the lattice $d = 1, 2, 3,...,$ the dimension of "cells" $v = 1, 2,..., d$, and the linear size L, which is taken to be an odd integer. Throughout the present paper we assume that the three parameters d, v, and L are fixed to allowed values. All the bounds proved in the present paper are independent of the system size L.

Let Λ_o be the d-dimensional $L \times \cdots \times L$ hypercubic lattice with periodic boundary conditions,

$$\Lambda_o = \{x = (x_1,..., x_d) \mid x_i \in \mathbf{Z}, |x_i| \leqslant (L-1)/2 \text{ for } i = 1,..., d\} \quad (3.1)$$

We "decorate" the lattice Λ_o by adding sites taken at the center of each v-dimensional cell. Let $\mathcal{U}'$ be the set of vectors defined as

$$\mathcal{U}' = \left\{ u = (u_1,..., u_d) \mid u_i = 0 \text{ or } 1/2, \text{ and } 2 \sum_{i=1}^{d} u_i = v \right\} \quad (3.2)$$

Note that each $u \in \mathcal{U}'$ has the length $|u| = \sqrt{v}/2$. For each $u \in \mathcal{U}'$, we let

$$\Lambda_u = \{x + u \mid x \in \Lambda_o\} \quad (3.3)$$

By introducing the unit cell $\mathscr{U}$ of the lattice by

$$\mathscr{U} = \{o\} \cup \mathscr{U}' \tag{3.4}$$

where $o = (0,...,0)$, our decorated hypercubic lattice is defined as

$$\Lambda = \bigcup_{u \in \mathscr{U}} \Lambda_u \tag{3.5}$$

We often decompose Λ as $\Lambda = \Lambda_o \cup \Lambda'$, where

$$\Lambda' = \bigcup_{u \in \mathscr{U}'} \Lambda_u \tag{3.6}$$

As is discussed in Section 2.1, we can imagine that sites in Λ_o represent metallic atoms and sites in Λ' represent oxygen atoms. The numbers of sites (vectors) in the unit cell[31]

$$b = |\mathscr{U}| = \binom{d}{v} + 1 \tag{3.7}$$

is important, since it gives the number of the bands in the corresponding single-electron problem.

For $d = 1$, the only possible choice of v is $v = 1$, and we get the chain with two kinds of atoms discussed in Section 2 (see Fig. 1). For $d = 2$, we can either set $v = 1$ to get the lattice in Fig. 7a with the band number $b = 3$, or set $v = 2$ to get the lattice in Fig. 7b with $b = 2$. For $d = 3$, there are three choices for v. The lattices with $v = 2$ and $v = 3$ have the structures similar to the fcc and the bcc lattices, respectively.

We introduce some sets of lattice vectors which will become useful. We define

$$\mathscr{F}_o = \left\{ f = (f_1,..., f_d) \mid f_i = 0 \text{ or } \pm 1/2, \text{ and } 2 \sum_{i=1}^{d} |f_i| = v \right\} \tag{3.8}$$

which is the collection of the sites in Λ' adjacent to the origin o. We have $|\mathscr{F}_o| = 2^v \binom{d}{v}$. For $f \in \mathscr{F}_o$, we define

$$\mathscr{F}_f = \{ g \in \mathscr{F}_o \mid |g_i| = |f_i| \text{ for } i = 1,..., d \} \tag{3.9}$$

Note that for a fixed $f \in \mathscr{F}_o$, $\{ f + g \mid g \in \mathscr{F}_f \}$ is the set of sites in Λ_o which are adjacent to f. We have $|\mathscr{F}_f| = 2^v$. We also note that for $g \in \mathscr{F}_f$, we have

$$\mathscr{F}_g = \mathscr{F}_f \tag{3.10}$$

---

[31] Throughout the present paper, $|S|$ denotes the number of elements in a set S.
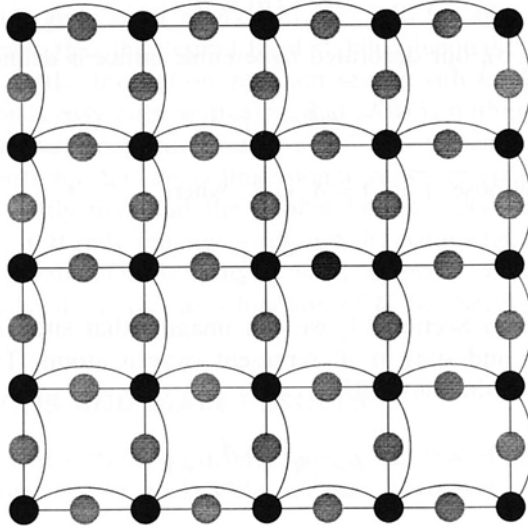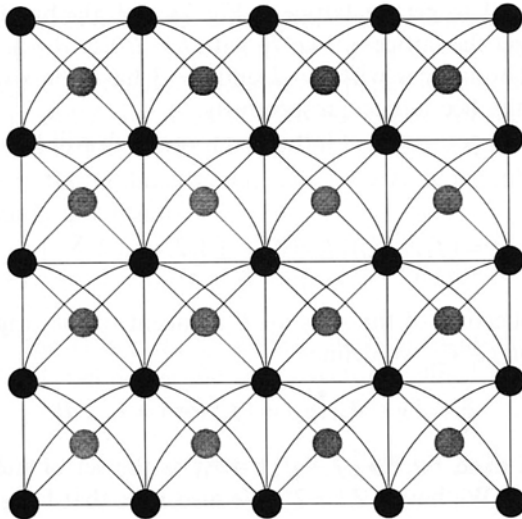
Fig. 7.   The lattice Λ in two dimensions ($d = 2$) with (a) $\nu = 1$ and (b) $\nu = 2$. The black dots are sites in Λ_o and the gray dots are sites in Λ'. One may interpret black sites as metallic atoms and gray sites as oxygen atoms.

3.2. Hubbard Model

We define the Hubbard model on the decorated hypercubic lattice Λ. As usual, we denote by $c_{x,\sigma}^{\dagger}$ and $c_{x,\sigma}$ the creation and the annihilation operators, respectively, of an electron at site $x \in \Lambda$ with spin $\sigma = \uparrow, \downarrow$. These operators satisfy the standard anticommutation relations

$$\{c_{x,\sigma}, c_{y,\tau}\} = \{c_{x,\sigma}^{\dagger}, c_{y,\tau}^{\dagger}\} = 0 \tag{3.11}$$

and

$$\{c_{x,\sigma}, c_{y,\tau}^{\dagger}\} = \delta_{x,y}\delta_{\sigma,\tau} \tag{3.12}$$

for any $x, y \in \Lambda$ and $\sigma, \tau = \uparrow, \downarrow$, where $\{A, B\} = AB + BA$. The number operator for an electron at site x with spin σ is defined as

$$n_{x,\sigma} = c_{x,\sigma}^{\dagger} c_{x,\sigma} \tag{3.13}$$

We consider the standard Hubbard Hamiltonian

$$H = H_{\text{hop}} + H_{\text{int}} \tag{3.14}$$

The interaction Hamiltonian is

$$H_{\text{int}} = U \sum_{x \in \Lambda} n_{x,\uparrow} n_{x,\downarrow} \tag{3.15}$$

where $U > 0$ is the on-site Coulomb repulsion energy. The hopping Hamiltonian is further decomposed as

$$H_{\text{hop}} = H_{\text{hop}}^{(0)} + \rho H_{\text{hop}}' \tag{3.16}$$

where $H_{\text{hop}}^{(0)}$ is the hopping Hamiltonian of the flat-band model defined as

$$H_{\text{hop}}^{(0)} = t \sum_{\sigma = \uparrow,\downarrow} \sum_{x \in \Lambda'} \left(\lambda c_{x,\sigma}^{\dagger} + \sum_{\substack{y \in \Lambda_0 \\ |x-y| = \sqrt{v}/2}} c_{y,\sigma}^{\dagger} \right) \left(\lambda c_{x,\sigma} + \sum_{\substack{y \in \Lambda_0 \\ |x-y| = \sqrt{v}/2}} c_{y,\sigma} \right) \tag{3.17}$$

where $t > 0$ and $\lambda > 0$ are parameters. It is, of course, possible to represent the Hamiltonian in the "standard" form as

$$H_{\text{hop}}^{(0)} = \sum_{\sigma = \uparrow,\downarrow} \sum_{x,y \in \Lambda} t_{x,y}^{(0)} c_{x,\sigma}^{\dagger} c_{y,\sigma} \tag{3.18}$$

where the hopping matrix elements are given by

$$
t^{(0)}_{x,y} = t^{(0)}_{y,x} = \begin{cases} \lambda^2 t & \text{if } x = y \in \Lambda' \\ \lambda t & \text{if } x \in \Lambda_o, y \in \Lambda' \text{ with } |x - y| = \sqrt{\nu}/2 \\ 2^{(\nu-\mu)} \binom{d-\mu}{\nu-\mu} t & \text{if } x, y \in \Lambda_o \text{ with } |x - y| = \sqrt{\mu} \quad (3.19) \\ & \text{where } \mu = 0, 1, ..., \nu \\ 0 & \text{otherwise} \end{cases}
$$

The representation (3.17) shows that the hopping Hamiltonian $H^{(0)}_{\text{hop}}$ is characterized by mean-field-like hoppings within each ν-dimensional cell which consists of $x \in \Lambda'$ and the sites $y \in \Lambda_o$ adjacent to it. This rather artificial choice of the hopping produces the single-electron spectrum with a completely flat band. See Section 4.1.

The perturbation Hamiltonian, on the other hand, is rather arbitrary. The magnitude of the perturbation is controlled by the real parameter p. The Hamiltonian H'_{hop} has the standard form

$$
H'_{\text{hop}} = \sum_{\sigma = \uparrow, \downarrow} \sum_{x, y \in \Lambda} t'_{x,y} c^{\dagger}_{x,\sigma} c_{y,\sigma} \tag{3.20}
$$

The hopping matrix elements $t'_{x,y} = t'_{y,x} \in \mathbf{R}$ are arbitrary except for the following conditions. We require the translation invariance

$$
t'_{x,y} = t'_{x+z, y+z} \tag{3.21}
$$

for any $z \in \mathbf{Z}^d$ and any $x, y \in \Lambda$, and the summability

$$
\sum_{y \in \Lambda} |t'_{x,y}| \leqslant t \tag{3.22}
$$

$$
\sum_{y \in \Lambda} |x - y| \cdot |t'_{x,y}| \leqslant tR \tag{3.23}
$$

for any $x \in \Lambda$. Here t is the same as before, and R is a constant which measures the range of the hopping $\{t'_{x,y}\}$. When R chosen to optimize (3.23) is less than $\sqrt{\nu}/2$, we redefine it[32] as $R = \sqrt{\nu}/2$.

The Hilbert space of the model is spanned by the basis states

$$
\Phi_{A,B} = \left(\prod_{y \in A} c^{\dagger}_{y,\uparrow} \right) \left(\prod_{z \in B} c^{\dagger}_{z,\downarrow} \right) \Phi_{\text{vac}} \tag{3.24}
$$

---

[32] This is done for a purely technical reason to make some formulas simple. See (10.90).

where A, B are subsets of Λ, and Φ_{vac} is the unique vacuum state charac-terized by $c_{x,\sigma}\Phi_{\text{vac}} = 0$ for any $x \in \Lambda$ and $\sigma = \uparrow, \downarrow$. In the present work, we only consider the Hilbert space $\mathcal{H}$ with the electron number fixed to $L^d = |\Lambda_o|$, which is spanned by the basis states (3.24) with $|A| + |B| = L^d$.

3.3. Local Stability Theorem

The total spin operator of the Hubbard model is defined as usual by

$$S_{\text{tot}}^{(\alpha)} = \frac{1}{2} \sum_{x \in \Lambda} \sum_{\sigma, \tau = \uparrow, \downarrow} c_{x,\sigma}^{\dagger} (p^{(\alpha)})_{\sigma, \tau} c_{x, \tau} \qquad (3.25)$$

for $\alpha = 1, 2, 3$, where $p^{(\alpha)}$ are the Pauli matrices

$$p^{(1)} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad p^{(2)} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad p^{(3)} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad (3.26)$$

An explicit calculation shows that the spin operators $(\mathbf{S}_{\text{tot}})^2 = \sum_{\alpha = 1, 2, 3} (S_{\text{tot}}^{(\alpha)})^2$, $S_{\text{tot}}^{(3)}$, and the Hamiltonian H commute with each other. This means that we can find simultaneous eigenstates of these operators. The eigenvalue of $(\mathbf{S}_{\text{tot}})^2$ is denoted as $S_{\text{tot}}(S_{\text{tot}} + 1)$, where S_{tot} can take values $1/2, 3/2, ..., S_{\text{max}}$ with $S_{\text{max}} = L^d/2$.

We are now able to state the theorem due to Tasaki[47] and Mielke and Tasaki.[34]

Theorem 3.1 (Flat-band ferromagnetism). Consider the Hubbard model with the Hamiltonian (3.14). Assume $t > 0$, $\lambda > 0$, and $\rho = 0$ to get a model with a flat band. Then, for any $U > 0$, the ground states of the Hamiltonian H have $S_{\text{tot}} = S_{\text{max}}$ and are nondegenerate apart from the trivial $(2S_{\text{max}} + 1)$-fold spin degeneracy.

This theorem is desirable in the sense that it completely determines the ground states of the model. But the result is not robust, since it applies only to the models with a completely flat band. Since refs. 47 and 34 only discuss the models with $v = 1$, we will prove the theorem in Section 5.4.

We now describe the new *robust* results for the models with a nearly flat band. For $S_{\text{tot}} = 1/2, 3/2, ..., S_{\text{max}}$, we denote by $E_{\text{min}}(S_{\text{tot}})$ the lowest eigenvalue of the Hamiltonian (3.14) in the sector which consists of the states Φ such that $(\mathbf{S}_{\text{tot}})^2 \Phi = S_{\text{tot}}(S_{\text{tot}} + 1) \Phi$. Then we have the following simple lemma for the sector with $S_{\text{tot}} = S_{\text{max}}$.

Lemma 3.2 ("Ferromagnetic ground states"). Assume that $t > 0$, $\lambda \geqslant \lambda_1$, and $|\rho| \lambda^{-2} \leqslant r_1$, where λ_1 and r_1 are finite constants which depend only on the dimensions d and v. (See Lemma 4.1 for explicit formulas of λ_1

and r_1.) Then for arbitrary U, the states Φ such that $(S_{tot})^2\,\Phi = S_{max}(S_{max}+1)\,\Phi$ and $H\Phi = E_{min}(S_{max})\,\Phi$ are nondegenerate apart from the trivial $(2S_{max}+1)$-fold spin degeneracy.

This lemma is almost trivial, but will be proved in Section 5.3. For convenience, we call the state Φ characterized by the above lemma the "ferromagnetic ground states." These states are the energy minimizers in the sector with $S_{tot} = S_{max}$, and are not necessarily the true ground states. We shall remind the reader of this abuse of terminology by always putting the "ferromagnetic ground states" within quotation marks.

The first theorem establishes the *instability* of the "ferromagnetic ground states" against a single-spin flip. Let $\bar{\varepsilon}$ denote the bandwidth of the lowest band. (See Section 4.1.) For $\rho \neq 0$ and a generic choice of $\{t'_{x,y}\}$, the bandwidth $\bar{\varepsilon}$ is strictly positive.

Theorem 3.3 (Instability of the "ferromagnetic ground states"). Assume the conditions for Lemma 3.2. We also assume that $\bar{\varepsilon} > 0$ (which is generically true if $\rho \neq 0$) and

$$0 \leqslant U < \bar{\varepsilon} \tag{3.27}$$

Then the "ferromagnetic ground states" are unstable under a single-spin flip in the sense that

$$E_{min}(S_{max}-1) < E_{min}(S_{max}) \tag{3.28}$$

The theorem will be proved in Section 5.3.

Theorem 3.3 shows that one can lower the energy by flipping a single spin in the "ferromagnetic ground states." It only shows that the "ferromagnetic ground states" are not the true ground states. To identify the true ground states (for $U \neq 0$) in this situation is a highly nontrivial and interesting question.

To show the instability of a certain state (as in the above theorem) is not a hard task since one can rely on the standard variational argument. A really important (and difficult) part of the present work is to show the following theorem, which states the *stability*.

Theorem 3.4 (Local stability of the "ferromagnetic ground states"). Consider the Hubbard model with the Hamiltonian (3.14). Assume that the parameters satisfy

$$\lambda \geqslant \lambda_2, \qquad |\rho| \leqslant \rho_1, \qquad \lambda\,|\rho| \leqslant \rho_1 \tag{3.29}$$

and

$$U \geqslant K_1 \lambda^2 t \, |\rho| \tag{3.30}$$

where λ_2, ρ_1, p_1, and K_1 are positive constants which depend only on the basic parameters d, ν, and R. Then the "ferromagnetic ground states" are stable under a single-spin flip in the sense that

$$E_{\min}(S_{\max} - 1) > E_{\min}(S_{\max}) \tag{3.31}$$

We stress that the problem of stability against a single-spin flip is already a highly nontrivial many-body problem. The restriction to the sector with $S_{\mathrm{tot}} = S_{\max} - 1$ does *not* reduce the problem to that of a single-particle (such as a magnon), since there are plenty of spaces for the electrons to move around. Moreover there is no way of expressing the eigenstates as Slater determinant states since there are both up-spin and down-spin electrons interacting via local Coulomb repulsion. See also the discussion after Theorem 2.3.

3.4. Bounds for the Spin-Wave Excitation Energy

Finally we describe our results about the elementary spin-wave excitation. The lower bound for the spin-wave energy in Theorem 3.6 is closely related to the above local stability theorem.

For $x \in \mathbf{Z}^d$, we let T_x denote the translation operator acting on the Hilbert space $\mathscr{H}$ as

$$T_x \left[\left(\prod_{y \in A} c_{y,\uparrow}^\dagger \right) \left(\prod_{z \in B} c_{z,\downarrow}^\dagger \right) \Phi_{\mathrm{vac}} \right] = \left(\prod_{y \in A} c_{y+x,\uparrow}^\dagger \right) \left(\prod_{z \in B} c_{z+x,\downarrow}^\dagger \right) \Phi_{\mathrm{vac}} \tag{3.32}$$

where A and B are arbitrary (ordered) subsets of Λ as in (3.24). Let us define the space of wavenumber vectors by

$$\mathscr{K} = \{ k = (k_1, ..., k_d) \, | \, k_i = 2\pi n_i / L \text{ with } n_i \in \mathbf{Z}$$

$$\text{such that } |n_i| \leqslant (L-1)/2 \} \tag{3.33}$$

For each $k \in \mathscr{K}$, we denote by $\mathscr{H}_k$ the Hilbert space of the states with the crystal momentum k and with $L^d - 1$ up-spin electrons and one down-spin electron. More precisely, we set

$$\mathscr{H}_k = \{ \Phi \in \mathscr{H} \, | \, T_x[\Phi] = e^{-ik \cdot x} \Phi \text{ for any } x \in \mathbf{Z}^d$$

$$\text{and } S_{\mathrm{tot}}^{(3)} \Phi = (S_{\max} - 1) \, \Phi \} \tag{3.34}$$

We can now define the energy $E_{SW}(k)$ of the elementary spin-wave excitation with the wave number $k \in \mathcal{K}$ as the lowest energy among the states in $\mathcal{H}_k$. Then we have the following two theorems.

Theorem 3.5 (Upper bound for the spin-wave energy). Assume that $\lambda \geqslant \lambda_0$ and $|\rho| \leqslant \rho_0$, where λ_0 and ρ_0 are positive constants which depend only on d, v, and R. Then we have

$$E_{SW}(k) - E_{\min}(S_{\max}) \leqslant F_1 \frac{U}{\lambda^4} G(k) \qquad (3.35)$$

where

$$G(k) = 2 \sum_{f \in \mathcal{F}_0} \sum_{g \in \mathcal{F}_f} \left(\sin \frac{k \cdot (f+g)}{2} \right)^2 \qquad (3.36)$$

The prefactor F_1 can be written as

$$F_1 = 1 + \frac{A_4}{\lambda} + A_5 \lambda |\rho| + \frac{A_6 \lambda^2 t |\rho|^2}{U} \qquad (3.37)$$

with the constants A_i $(i = 4, 5, 6)$, which depend only on d, v, and R.

Like Theorem 3.3 about the instability of the "ferromagnetic ground states," the above theorem is proved by the standard variational argument. See Section 9.

The major achievement in the present paper is the lower bound which corresponds to the above (3.35).

Theorem 3.6 (Lower bound for the spin-wave energy). Assume that $\lambda \geqslant \lambda_3$, $|\rho| \leqslant \rho_0$, and $K_2 \lambda t \geqslant U \geqslant A_3 \lambda^2 t |\rho|$, where λ_3, ρ_0, K_2, and A_3 are positive constants which depend only on d, v, and R. Then we have

$$E_{SW}(k) - E_{\min}(S_{\max}) \geqslant F_2 \frac{U}{\lambda^4} G(k) \qquad (3.38)$$

with $G(k)$ defined in (3.36). The prefactor F_2 can be written as

$$F_2 = 1 - A_1 |\rho| - \frac{A_2}{\lambda} - \frac{A_3 \lambda^2 t |\rho|}{U} \qquad (3.39)$$

with the constants A_1, A_2, and A_3, which depend only on d, v, and R.

Note that (3.37) and (3.39) imply that $F_1 \simeq F_2 \simeq 1$ when λ is large and $|\rho|$ is (very) small. In this case the dispersion relation $E_{SW}(k)$ of the elementary spin-wave excitation is given by

$$E_{SW}(k) - E_{\min}(S_{\max}) \simeq \frac{U}{\lambda^4} G(k) \tag{3.40}$$

This dispersion relation is exactly what one expects in the ferromagnetic Heisenberg quantum spin system defined on the hypercubic lattice Λ_o with the exchange interaction $J_{eff} = 2U\lambda^{-4}$.

As stressed in Section 2.1, Theorem 3.6 requires an upper bound for the Coulomb interaction U. By noting that $E_{SW}(k)$ is increasing in U, however, it is easy to prove nontrivial lower bounds for $E_{SW}(k)$ for larger values of U.

Corollary 3.7. Assume that $\lambda \geqslant \lambda_3$, $|\rho| \leqslant \rho_0$, $A_3 \lambda |\rho|/K_2 \leqslant 1$, and $U \geqslant K_2 \lambda t$. Then we have

$$E_{SW}(k) - E_{\min}(S_{\max}) \geqslant F_3 G(k) \tag{3.41}$$

with $G(k)$ defined in (3.36). The prefactor F_3 can be written as

$$F_3 = \left(1 - A_1 |\rho| - \frac{A_2}{\lambda} - \frac{A_3 \lambda |\rho|}{K_2}\right) \frac{K_2 t}{\lambda^3} \tag{3.42}$$

Proof. The first three conditions assumed here guarantee that we can use Theorem 3.6 when $U = K_2 \lambda t$. We claim that, for each $k \in \mathcal{K}$, $E_{SW}(k)$ is an increasing function of U. This is because both H and H_{int} commute with T_x and $S_{tot}^{(3)}$, and $E_{SW}(k)$ is defined to be the lowest energy in the sector with the fixed momentum k and the fixed eigenvalue of $S_{tot}^{(3)}$. Then it is trivial that $E_{SW}(k)$ for $U \geqslant K_2 \lambda t$ satisfies the desired bound (3.41), where the right-hand side of (3.41) is obtained by substituting $U = K_2 \lambda t$ into (3.38). ∎

4. SINGLE-ELECTRON PROBLEM

We shall investigate the properties of the single-electron system corresponding to our Hubbard model. A careful study of single-electron properties is indispensable when we work with interacting many-electron systems.

4.1. Band Structure of the Model

If there is only a single electron with, say, up spin in the whole system, a general state can be written as

$$\Phi(\varphi) = \sum_{x \in \Lambda} \varphi_x c^\dagger_{x,\uparrow} \Phi_{\text{vac}} \tag{4.1}$$

with $\varphi_x \in \mathbf{C}$. As in the standard quantum mechanics, we regard the collection $\varphi = (\varphi_x)_{x \in \Lambda}$ as a vector in a $|\Lambda|$-dimensional complex linear space $\mathcal{H}_{\text{single}} \cong \mathbf{C}^{|\Lambda|}$, which we call the single-electron Hilbert space.

Since it obviously holds that $H_{\text{int}} \Phi(\varphi) = 0$, the Schrödinger equation $H\Phi(\varphi) = \varepsilon \Phi(\varphi)$ reduces to

$$\sum_{y \in \Lambda} t_{x,y} \varphi_y = \varepsilon \varphi_x \tag{4.2}$$

where $t_{x,y} = t^{(0)}_{x,y} + \rho t'_{x,y}$, and we denote the (single-electron) energy eigenvalue as ε.

By rewriting the expression (3.17) for H_{hop} as

$$H_{\text{hop}} = t \sum_{\sigma = \uparrow, \downarrow} \sum_{u \in \mathscr{U}'} \sum_{x \in \Lambda_u} \left(\lambda c^\dagger_{x,\sigma} + \sum_{f \in \mathscr{F}_u} c^\dagger_{x+f,\sigma} \right) \left(\lambda c_{x,\sigma} + \sum_{f \in \mathscr{F}_u} c_{x+f,\sigma} \right) \tag{4.3}$$

we can write down (4.2) in a concrete form as

$$\varepsilon \varphi_x = t \sum_{f \in \mathscr{F}_o} \left(\lambda \varphi_{x+f} + \sum_{g \in \mathscr{F}_f} \varphi_{x+f+g} \right) + \rho \sum_{y \in \Lambda} t'_{x,y} \varphi_y \tag{4.4}$$

and

$$\varepsilon \varphi_{x+u} = \lambda^2 t \varphi_{x+u} + \lambda t \sum_{f \in \mathscr{F}_u} \varphi_{x+u+f} + \rho \sum_{y \in \Lambda} t'_{x+u,y} \varphi_y \tag{4.5}$$

where $x \in \Lambda_o$ and $u \in \mathscr{U}'$. We recall that $\mathscr{U}$ is the unit cell of the lattice, and[33] $\mathscr{U}' = \mathscr{U} \setminus \{o\}$.

Since the hopping matrix elements $t'_{x,y}$ are invariant under the translation by any integer vector $z \in \mathbf{Z}^d$, we can use the Bloch theorem to write an eigenstate of (4.2) as

$$\varphi_x = e^{ik \cdot x} v_x(k) \tag{4.6}$$

---

[33] For any sets A and B, $A \setminus B$ denotes the set $\{ x \in A \mid x \notin B \}$.

with $k \in \mathcal{K}$ [see (3.33)], and $v_x(k)$ satisfying

$$v_x(k) = v_{x+y}(k) \tag{4.7}$$

for any $y \in \mathbf{Z}^d$. With the translation invariance (4.7) in mind, we can identify, for each fixed $k \in \mathcal{K}$, the function $v_x(k)$ (of x) with a b-dimensional vector[34]

$$\mathbf{v}(k) = (v_u(k))_{u \in \mathcal{U}} \in \mathbf{C}^b \tag{4.8}$$

where $b = |\mathcal{U}| = \binom{d}{v} + 1$ will turn out to be the number of the bands in the Schrödinger equation (4.2).

By substituting the representation (4.6) into the Schrödinger equation (4.4), (4.5), we find the equation (the Schrödinger equation in k-space)

$$\varepsilon \mathbf{v}(k) = (\lambda^2 t \, \mathsf{M}(k) + \rho t \, \mathsf{Q}(k)) \, \mathbf{v}(k) \tag{4.9}$$

which determines, for each $k \in \mathcal{K}$, the eigenvalue ε of the original Schrödinger equation (4.2). Here $\mathsf{M}(k) = (M_{u,u'}(k))_{u,u' \in \mathcal{U}}$ and $\mathsf{Q}(k) = (Q_{u,u'}(k))_{u,u' \in \mathcal{U}}$ are $b \times b$ matrices.[35] They are defined by

$$M_{u,u'}(k) = M_{u',u}(k) = \begin{cases} A(k)/\lambda^2 & \text{if} \quad u = u' = o \\ C_u(k)/\lambda & \text{if} \quad u \in \mathcal{U}' \text{ and } u' = o \\ 0 & \text{if} \quad u, u' \in \mathcal{U}' \text{ and } u \neq u' \\ 1 & \text{if} \quad u = u' \in \mathcal{U} \end{cases} \tag{4.10}$$

and

$$Q_{u,u'}(k) = \frac{1}{t} \sum_{y \in A_{u'}} t'_{u,y} e^{ik \cdot (y-u)} \tag{4.11}$$

Here we have introduced

$$C_f(k) = \sum_{g \in \mathcal{F}_f} e^{ik \cdot g} \tag{4.12}$$

for $f \in \mathcal{F}_o$, and

$$A(k) = \sum_{f \in \mathcal{F}_o} \sum_{g \in \mathcal{F}_f} e^{ik \cdot (f+g)} = \sum_{u \in \mathcal{U}'} (C_u(k))^2 \tag{4.13}$$

---

[34] In the present paper, boldface symbols are reserved to indicate elements of the b-dimensional vector space introduced here.

[35] Sans serif symbols denote matrices in the b-dimensional vector space.

Since the matrix $(\lambda^2 t M(k) + pt Q(k))$ is Hermitian, it generically has b eigenvalues and eigenstates for each $k \in \mathcal{K}$. We denote these eigenvalues as $\varepsilon_j(k)$, where the band index $j = 1, 2, ..., b$ is assigned so that $\varepsilon_j(k) \leqslant \varepsilon_{j+1}(k)$. When viewed as a function of k, the eigenvalues $\varepsilon_j(k)$ are usually called the dispersion relations of the jth band.

When $p = 0$, the eigenvalue problem (4.9) can be solved easily, and we obtain the dispersion relations

$$\varepsilon_j(k) = \begin{cases} 0 & \text{for} \quad j = 1 \\ \lambda^2 t & \text{for} \quad j = 2, ..., b-1 \\ \lambda^2 t + t A(k) & \text{for} \quad j = b \end{cases} \tag{4.14}$$

Note that the model has a rather singular band structure where most of the bands have constant energies (i.e., are flat), and all the bands with $j = 2, ..., b-1$ are completely degenerate. Another important feature of (4.14) is that the lowest band ($j = 1$) is separated from the higher bands by an energy gap $\lambda^2 t$. See Fig. 2a for the dispersion relation in $d = 1$. We have also drawn the dispersion relation of the flat-band model with $d = 2$ and $v = 1$ in Fig. 8.
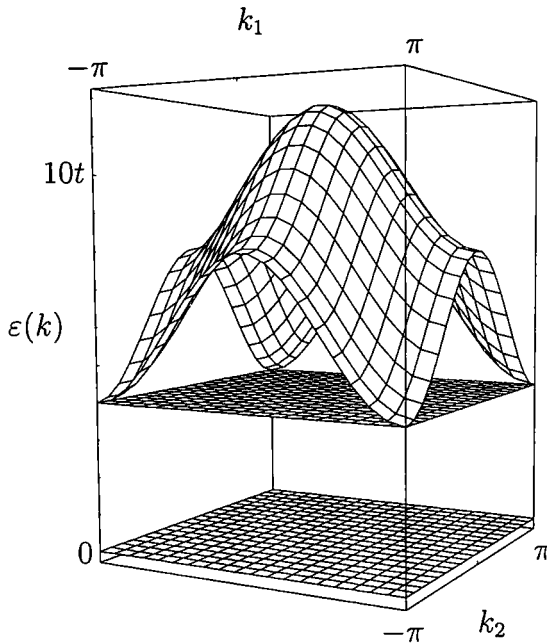

Fig. 8.   The dispersion relation for the three-band model with $d = 2$, $v = 1$. We have set $t > 0$, $\lambda = 2$, and $p = 0$ to get a flat-band model. There are two flat bands and one cosine band.

For $\rho \neq 0$, with a generic choice of H'_{hop}, the dispersion relations $\varepsilon_j(k)$ become k dependent and the bands are no longer flat. See Fig. 2b for the dispersion relation with slightly perturbed model in $d = 1$, and Fig. 9 for that in $d = 2$. The degeneracy between the bands with $j = 2,..., b - 1$ is also likely to be lifted (unless the perturbation has a certain symmetry). Actual band structures depend delicately on the choice of the perturbation, and are not easy to calculate. It generically holds, however, that the lowest band is still separated from the rest of the bands by an energy gap, provided that $|\rho|$ is not too large. We present the following crude estimate, which is sufficient for our purpose.

Lemma 4.1. Assume that $\lambda \geqslant \lambda_1 = 2^\nu (b-1)^{1/2}$ and $|\rho|\, \lambda^{-2} \leqslant r_1 = 9 \times 10^{-3}/b$. Then we have

$$\varepsilon_1(k) \leqslant \lambda^2 t/4 \qquad (4.15)$$
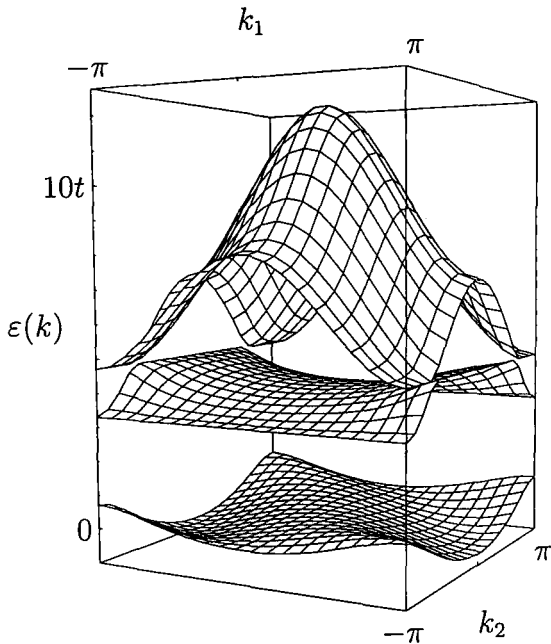
Fig. 9. The dispersion relation for the three-band model with $d = 2$, $\nu = 1$. The perturbation is given by $t'_{x,x} = t$ if $x \in \mathscr{F}_{(0,1/2)}$, $t'_{x,x} = -t$ if $x \in \mathscr{F}_{(1/2,0)}$, $t'_{x,y} = t'_{y,x} = t$ if $x \in \mathscr{F}_o$ and $y = x + (1, 1)$, and $t'_{x,y} = 0$ otherwise. We have set $t > 0$, $\lambda = 2$, and $\rho = 0.7$. Note that the two lower bands become dispersive, and a gap appears between the second and the third bands.

and

$$\varepsilon_j(k) \geqslant 3\lambda^2 t/4 \tag{4.16}$$

for $j = 2, 3,..., b$.

Proof. The statement is almost trivial, but we give a proof for completeness. Since the eigenvalues of $\lambda^2 t \, \mathsf{M}(k)$ are either $= 0$ or $\geqslant \lambda^2 t$, we have

$$\left(\lambda^2 t \, \mathsf{M}(k) - \frac{\lambda^2 t}{2} \right)^2 \geqslant \left(\frac{\lambda^2 t}{2} \right)^2 \tag{4.17}$$

Consider the similar quantity for the perturbed matrix, and note that

$$\left[\{ \lambda^2 t \, \mathsf{M}(k) + \rho t \, \mathsf{Q}(k) \} - \frac{\lambda^2 t}{2} \right]^2$$

$$= \left(\lambda^2 t \, \mathsf{M}(k) - \frac{\lambda^2 t}{2} \right)^2 + \rho^2 t^2 (\mathsf{Q}(k))^2$$

$$+ \left(\lambda^2 t \, \mathsf{M}(k) - \frac{\lambda^2 t}{2} \right) \rho t \, \mathsf{Q}(k) + \rho t \, \mathsf{Q}(k) \left(\lambda^2 t \, \mathsf{M}(k) - \frac{\lambda^2 t}{2} \right)$$

$$\geqslant \left(\frac{\lambda^2 t}{2} \right)^2 - \rho^2 t^2 \, \| \mathsf{Q}(k) \|^2 - 2 \, |\rho| \, t \, \| \mathsf{Q}(k) \| \left(\lambda^2 t \, \| \mathsf{M}(k) \| + \frac{\lambda^2 t}{2} \right) \tag{4.18}$$

By substituting the assumed bound for $|\rho|$ and the bounds

$$\| \mathsf{Q}(k) \| \leqslant b, \qquad \| \mathsf{M}(k) \| \leqslant 1 + |A(k)| \, \lambda^{-2} \leqslant 1 + (b-1) \, 4^\nu \lambda^{-2} \leqslant 2$$

we observe that the right-hand side of (4.18) is not less than $(\lambda^2 t/4)^2$. This means that the Schrödinger equation (4.9) cannot have eigenvalues in the range $\lambda^2 t/4 \leqslant \varepsilon \leqslant (3/4) \, \lambda^2 t$. Since the eigenvalue $\varepsilon_j(k)$ with j and k fixed is continuous in ρ, the statement of the lemma follows. ∎

From now on, we assume that the condition for Lemma 4.1 is satisfied. The existence of a gap allows us to treat the lowest band in a special manner. Let us decompose the single-electron Hilbert space as

$$\mathcal{H}_{\text{single}} = \mathcal{H}^{(1)}_{\text{single}} \oplus \mathcal{H}'_{\text{single}} \tag{4.19}$$

where $\mathcal{H}^{(1)}_{\text{single}}$ is the Hilbert space corresponding to the lowest band (with the band index $j = 1$). It is spanned by the eigenstates of (4.2) with the

eigenvalue $\varepsilon_1(k)$ for $k \in \mathcal{K}$. The space $\mathcal{H}'_{\text{single}}$ is the orthogonal complement of $\mathcal{H}^{(1)}_{\text{single}}$. The dimensions of the spaces $\mathcal{H}^{(1)}_{\text{single}}$ and $\mathcal{H}'_{\text{single}}$ are $L^d = |\Lambda_o|$ and $(b-1)L^d = |\Lambda'|$, respectively.

Let $P^{(1)}_{\text{single}}$ be the orthogonal projection onto the space $\mathcal{H}^{(1)}_{\text{single}}$, and denote by T the hopping operator on $\mathcal{H}_{\text{single}}$, whose matrix representation is given by $(t_{x,y})_{x,y \in \Lambda}$. We define the modified hopping operator $\tilde{T}$ by

$$\tilde{T} = TP^{(1)}_{\text{single}} + \tfrac{3}{4}\lambda^2 t(1 - P^{(1)}_{\text{single}}) \tag{4.20}$$

and denote by $(\tilde{t}_{x,y})_{x,y \in \Lambda}$ the matrix representation of $\tilde{T}$. Note that the bound (4.16) implies the operator inequality $T \geqslant \tilde{T}$. Define the modified hopping Hamiltonian by

$$\tilde{H}_{\text{hop}} = \sum_{\sigma = \uparrow, \downarrow} \sum_{x,y \in \Lambda} \tilde{t}_{x,y} c^{\dagger}_{x,\sigma} c_{y,\sigma} \tag{4.21}$$

which also satisfies

$$H_{\text{hop}} \geqslant \tilde{H}_{\text{hop}} \tag{4.22}$$

Although the introduction of $\tilde{H}_{\text{hop}}$ is not essential for our proof, it considerably simplifies the required estimates.

4.2. Localized Bases

We introduce bases for the single-electron spaces $\mathcal{H}^{(1)}_{\text{single}}$ and $\mathcal{H}'_{\text{single}}$, in which each basis state is localized at a lattice site. The use of such localized bases enables us to treat electrons as "particles," but with taking into account the band structure of the model. The actual construction of the bases will be presented in Section 10.

We start from the easy case with $\rho = 0$, i.e., the flat-band models. For $x \in \Lambda_o$, we define the state $\psi^{(x)} = (\psi^{(x)}_y)_{y \in \Lambda} \in \mathcal{H}_{\text{single}}$ by

$$\psi^{(x)}_y = \begin{cases} 1 & \text{if } x = y \\ -1/\lambda & \text{if } y \in \Lambda' \text{ and } |x - y| = \sqrt{v}/2 \\ 0 & \text{otherwise} \end{cases} \tag{4.23}$$

An explicit calculation shows that $\sum_{y \in \Lambda} t^{(0)}_{x,y} \psi^{(x)}_y = 0$ for $x \in \Lambda_o$. This can be done by using (3.19), but it is easier to use (3.17). This means that $\psi^{(x)} \in \mathcal{H}^{(1)}_{\text{single}}$, since the lowest band has a constant energy $\varepsilon = 0$ when $\rho = 0$ as in (4.14). Since the states $\psi^{(x)}$ with $x \in \Lambda_o$ are linearly independent and $|\Lambda_o|$ is equal to the dimension of $\mathcal{H}^{(1)}_{\text{single}}$, we find that the collection of the states $\{\psi^{(x)}\}_{x \in \Lambda_o}$ forms a basis of $\mathcal{H}^{(1)}_{\text{single}}$.

For $x \in \Lambda'$, we similarly define $\psi^{(x)} \in \mathcal{H}_{\text{single}}$ as

$$\psi_y^{(x)} = \begin{cases} 1 & \text{if } x = y \\ 1/\lambda & \text{if } y \in \Lambda_o \text{ and } |x - y| = \sqrt{\nu}/2 \\ 0 & \text{otherwise} \end{cases} \tag{4.24}$$

It is evident that $\psi^{(x)}$ with $x \in \Lambda'$ and $\psi^{(x')}$ with $x' \in \Lambda_o$ are orthogonal to each other. This means that $\psi^{(x)}$ with any $x \in \Lambda'$ is orthogonal to the space $\mathcal{H}_{\text{single}}^{(1)}$. By counting the dimension, it then follows that $\{\psi^{(x)}\}_{x \in \Lambda'}$ form a basis of $\mathcal{H}'_{\text{single}}$.

Both the bases $\{\psi^{(x)}\}_{x \in \Lambda_o}$ and $\{\psi^{(x)}\}_{x \in \Lambda'}$ are not orthonormal, but the states in the bases are sharply localized at lattice sites. The introduction and use of the localized basis for $\mathcal{H}_{\text{single}}^{(1)}$ was essential in the study of the flat-band Hubbard models in refs. 47 and 34.

To deal with non-flat-band models, we shall construct similar bases for the models with $\rho \neq 0$. Since this is a problem of perturbation theory in a one-body quantum mechanics, there is no essential difficulty when the strength of the perturbation $|\rho| t$ is sufficiently smaller than the energy gap $\lambda^2 t$. In Section 10, we prove the following.

Lemma 4.2. Suppose that $\lambda \geqslant \lambda_0$ and $|\rho| \lambda^{-2} \leqslant r_0$, where λ_0 and r_0 are positive constants which depend only on the dimensions d, ν. Then we can take for each $x \in \Lambda$ a state $\varphi^{(x)} = (\varphi_y^{(x)})_{y \in \Lambda} \in \mathcal{H}_{\text{single}}$ such that $\varphi_y^{(x)} = \varphi_{y+z}^{(x+z)}$ holds for any $z \in \mathbf{Z}^d$. The collections of the states $\{\varphi^{(x)}\}_{x \in \Lambda_o}$ and $\{\varphi^{(x)}\}_{x \in \Lambda'}$ form bases of $\mathcal{H}_{\text{single}}^{(1)}$ and $\mathcal{H}'_{\text{single}}$, respectively. These basis states are summable as

$$\sum_{y \in \Lambda} |\varphi_y^{(x)} - \psi_y^{(x)}| \leqslant B_1 \frac{|\rho|}{\lambda^2} \tag{4.25}$$

$$\sum_{y \in \Lambda} |x - y| \cdot |\varphi_y^{(x)} - \psi_y^{(x)}| \leqslant B_1 R \frac{|\rho|}{\lambda^2} \tag{4.26}$$

$$\sum_{x \in \Lambda} |\varphi_y^{(x)} - \psi_y^{(x)}| \leqslant B_1 \frac{|\rho|}{\lambda^2} \tag{4.27}$$

and

$$\sum_{x \in \Lambda} |x - y| \cdot |\varphi_y^{(x)} - \psi_y^{(x)}| \leqslant B_1 R \frac{|\rho|}{\lambda^2} \tag{4.28}$$

where B_1 is a positive constant which depends only on d and ν.

The bounds (4.25)–(4.28) imply that each state $\varphi^{(x)}$ is sharply localized at the reference site x. The bounds also show that the states $\varphi^{(x)}$ are chosen so that they become identical to $\psi^{(x)}$ when $\rho = 0$.

Let us investigate how the modified hopping operator $\tilde{T} = (\tilde{t}_{x,y})_{x,y \in \Lambda}$ introduced in (4.20) acts on these basis states. From the definition (4.20), it is obvious that

$$\tilde{T}\varphi^{(x)} = \tfrac{3}{4}\lambda^2 t \varphi^{(x)} \tag{4.29}$$

if $x \in \Lambda'$. For $x \in \Lambda_o$, the basis state is transformed as

$$\tilde{T}\varphi^{(x)} = \sum_{y \in \Lambda_o} \tau_{y,x} \varphi^{(y)} \tag{4.30}$$

where the effective hopping matrix elements $\tau_{y,x}$ are given by

$$\tau_{y,x} = (2\pi)^{-d} \int dk\, e^{ik \cdot (y-x)} \varepsilon_1(k) \tag{4.31}$$

where $\varepsilon_1(k)$ is the dispersion relation of the lowest band discussed in Section 4.1, and $\int dk\,(\cdots)$ is a shorthand for the sum $(2\pi/L)^d \sum_{k \in \mathcal{K}}(\cdots)$. Note that only y in Λ_o appears in the right-hand side of (4.30), reflecting the band structure.

The precise form of $\tau_{y,x}$ depends on specific perturbation. But the following general bound is sufficient for our purpose.

Lemma 4.3. When $\lambda \geqslant \lambda_0$ and $|\rho|\,\lambda^{-2} \leqslant r_0$, we have

$$\sum_{x \in \Lambda_o} |\tau_{y,x}| = \sum_{y \in \Lambda_o} |\tau_{y,x}| \leqslant B_1 |\rho|\, t \tag{4.32}$$

and

$$\sum_{x \in \Lambda_o} |x-y| \cdot |\tau_{y,x}| = \sum_{y \in \Lambda_o} |x-y| \cdot |\tau_{y,x}| \leqslant B_1 R\, |\rho|\, t \tag{4.33}$$

Since the bases $\{\varphi^{(x)}\}_{x \in \Lambda_o}$ and $\{\varphi^{(x)}\}_{x \in \Lambda'}$ are not orthonormal, it is convenient to introduce the bases which are dual to them. The dual bases are constructed uniquely by a standard procedure (in Section 10), and we can prove the following.

Lemma 4.4. Suppose that $\lambda \geqslant \lambda_0$ and $|\rho|\,\lambda^{-2} \leqslant r_0$. Then we can take for each $x \in \Lambda$ a state $\tilde{\varphi}^{(x)} = (\tilde{\varphi}_y^{(x)})_{y \in \Lambda} \in \mathscr{H}_{\text{single}}$ such that $\tilde{\varphi}_y^{(x)} = \tilde{\varphi}_{y+z}^{(x+z)}$ holds for any $z \in \mathbf{Z}^d$. The collections of the states $\{\tilde{\varphi}^{(x)}\}_{x \in \Lambda_o}$ and $\{\tilde{\varphi}^{(x)}\}_{x \in \Lambda'}$

form bases of $\mathscr{H}^{(1)}_{single}$ and $\mathscr{H}'_{single}$, respectively. They are dual of the bases $\{\varphi^{(x)}\}_{x \in \Lambda_o}$ and $\{\varphi^{(x)}\}_{x \in \Lambda'}$ in the sense that we have

$$\sum_{y \in \Lambda} (\tilde{\varphi}_y^{(x)})^* \varphi_y^{(x')} = \delta_{x,x'} \tag{4.34}$$

for any $x, x' \in \Lambda$, and

$$\sum_{x \in \Lambda} (\tilde{\varphi}_y^{(x)})^* \varphi_{y'}^{(x)} = \delta_{y,y'} \tag{4.35}$$

for any $y, y' \in \Lambda$. These basis states are summable as

$$\sum_{y \in \Lambda} |\tilde{\varphi}_y^{(x)} - \psi_y^{(x)}| \leqslant B_1 \frac{|\rho|}{\lambda^2} + \frac{B_2}{\lambda^2} \tag{4.36}$$

$$\sum_{y \in \Lambda} |x - y| \, |\tilde{\varphi}_y^{(x)} - \psi_y^{(x)}| \leqslant B_1 R \frac{|\rho|}{\lambda^2} + \frac{B_2}{\lambda^2} \tag{4.37}$$

$$\sum_{x \in \Lambda} |\tilde{\varphi}_y^{(x)} - \psi_y^{(x)}| \leqslant B_1 \frac{|\rho|}{\lambda^2} + \frac{B_2}{\lambda^2} \tag{4.38}$$

and

$$\sum_{x \in \Lambda} |x - y| \cdot |\tilde{\varphi}_y^{(x)} - \psi_y^{(x)}| \leqslant B_1 R \frac{|\rho|}{\lambda^2} + \frac{B_2}{\lambda^2} \tag{4.39}$$

where B_2 is a positive constant which depends only on d and v.

   Note that the right-hand sides of (4.36)–(4.39) do not vanish when $\rho = 0$. This is because the dual basis state $\tilde{\varphi}_y^{(x)}$ has nonvanishing exponentially decaying tail even in the flat-band model. This remarkable asymmetry between the state $\varphi^{(x)}$ and its dual $\tilde{\varphi}^{(x)}$ plays a fundamental role in our work.

5. LOCALIZED BASIS FOR THE HUBBARD MODEL

   In the present section, we discuss the framework for describing many-electron systems by using the localized basis introduced in Section 4.2. Elementary statements about the "ferromagnetic ground states" and the theorem for flat-band ferromagnetism are also proved.

5.1. Fermion Operators for the Localized Bases

We rewrite the Hubbard Hamiltonian (3.14) by using the new fermion operators. The new representation turns out to be suitable for our purpose for taking into account both the particlelike nature of electrons and the band structure of the model.

We first define the creation operator corresponding to the basis state $\varphi^{(x)}$ as

$$a_{x,\sigma}^\dagger = \sum_{y \in \Lambda} (\varphi_y^{(x)})^* \, c_{y,\sigma}^\dagger \tag{5.1}$$

for $x \in \Lambda$ and $\sigma = \uparrow, \downarrow$. Similarly we define the annihilation operator corresponding to the dual basis state $\tilde{\varphi}^{(x)}$ as

$$b_{x,\sigma} = \sum_{y \in \Lambda} \tilde{\varphi}_y^{(x)} c_{y,\sigma} \tag{5.2}$$

for $x \in \Lambda$ and $\sigma = \uparrow, \downarrow$.

By using the basic anticommutation relations (3.11) and (3.12), the definitions (5.1) and (5.2), and the duality relation (4.34), we find that these operators satisfy the anticommutation relations

$$\{a_{x,\sigma}^\dagger, a_{y,\tau}^\dagger\} = \{b_{x,\sigma}, b_{y,\tau}\} = 0 \tag{5.3}$$

and

$$\{a_{x,\sigma}^\dagger, b_{y,\tau}\} = \delta_{x,y} \delta_{\sigma,\tau} \tag{5.4}$$

for any $x, y \in \Lambda$ and $\sigma, \tau = \uparrow, \downarrow$. Note that (5.3) and (5.4) have exactly the same forms as the canonical anticommutation relations.

By using the other duality relation (4.35), we can invert (5.1) and (5.2) to get

$$c_{x,\sigma}^\dagger = \sum_{y \in \Lambda} \tilde{\varphi}_x^{(y)} a_{y,\sigma}^\dagger \tag{5.5}$$

and

$$c_{x,\sigma} = \sum_{y \in \Lambda} (\varphi_x^{(y)})^* \, b_{y,\sigma} \tag{5.6}$$

5.2. Representation of the Hamiltonian

We shall rewrite the Hamiltonian using the operators $a_{x,\sigma}^\dagger$ and $b_{x,\sigma}$. As for the hopping part, we treat the modified Hamiltonian $\tilde{H}_{\text{hop}}$ defined in (4.21), rather than the original H_{hop}. By substituting (5.5) and (5.6) into (4.21), we find that

$$
\begin{aligned}
\tilde{H}_{\text{hop}} &= \sum_{\sigma=\uparrow,\downarrow} \sum_{x,y,v,w \in \Lambda} \tilde{\varphi}_x^{(v)} \tilde{t}_{x,y} (\varphi_y^{(w)})^* \, a_{v,\sigma}^\dagger b_{w,\sigma} \\
&= \sum_{\sigma=\uparrow,\downarrow} \left\{ \sum_{x,y \in \Lambda_o} \tau_{x,y} a_{x,\sigma}^\dagger b_{y,\sigma} + \frac{3\lambda^2 t}{4} \sum_{x \in \Lambda'} a_{x,\sigma}^\dagger b_{x,\sigma} \right\}
\end{aligned}
\tag{5.7}
$$

where we have used (4.29) and (4.30), which determine the action of $\tilde{T}$, and the duality relation (4.34). The representation (5.7) makes the band structure manifest.

Similarly we can rewrite the interaction Hamiltonian (3.15) as

$$
\begin{aligned}
H_{\text{int}} &= U \sum_{x,y,v,w,z \in \Lambda} (\tilde{\varphi}_x^{(y)} a_{y,\uparrow}^\dagger)((\varphi_x^{(z)})^* b_{z,\uparrow})(\tilde{\varphi}_x^{(v)} a_{v,\downarrow}^\dagger)((\varphi_x^{(w)})^* b_{w,\downarrow}) \\
&= \sum_{y,v,w,z \in \Lambda} \tilde{U}_{y,v;w,z} \, a_{y,\uparrow}^\dagger a_{v,\downarrow}^\dagger b_{w,\downarrow} b_{z,\uparrow}
\end{aligned}
\tag{5.8}
$$

where the effective interaction is given by

$$
\tilde{U}_{y,v;w,z} = U \sum_{x \in \Lambda} \tilde{\varphi}_x^{(y)} \tilde{\varphi}_x^{(v)} (\varphi_x^{(w)} \varphi_x^{(z)})^*
\tag{5.9}
$$

Note that the interaction Hamiltonian H_{int} in the new representation (5.8) is no longer local.

Remark. It is also possible to write down a representation similar to (5.7) for the original hopping Hamiltonian

$$
H_{\text{hop}} = \sum_{\sigma=\uparrow,\downarrow} \left\{ \sum_{x,y \in \Lambda_o} \tau_{x,y} a_{x,\sigma}^\dagger b_{y,\sigma} + \sum_{x,y \in \Lambda'} \tau_{x,y} a_{x,\sigma}^\dagger b_{y,\sigma} \right\}
\tag{5.10}
$$

with properly defined $\tau_{x,y}$ for $x, y \in \Lambda'$.

5.3. Elementary Facts About the "Ferromagnetic Ground States"

We can now prove the basic statement about the "ferromagnetic ground states."

Proof of Lemma 3.2. Since we are interested in states with $S_{tot} = S_{max}$, we can concentrate on the sector with $S_{tot}^{(3)} = S_{max}$. States (with $S_{tot} = S_{max}$) in other sectors can be obtained by suitably applying the total spin lowering operator. Clearly H_{int} annihilates a state with $S_{tot}^{(3)} = S_{max}$ as it contains only up-spin electrons.

Because the conditions for Lemma 4.1 are satisfied, there is a finite energy gap between the lowest band and the remaining bands. In order to make the eigenvalue of H_{hop} small, we need to use as many states from the lowest band. Since the electron number L^d is identical to the dimension of the Hilbert space $\mathscr{H}_{single}^{(1)}$ for the loest band, this can be done in a unique way, and we find

$$\Phi_\uparrow = \left(\prod_{x \in \Lambda_o} a_{x,\uparrow}^\dagger \right) \Phi_{vac} \tag{5.11}$$

is the desired "ferromagnetic ground state." By operating with H_{hop} in the representation (5.10), we find that $H\Phi_\uparrow = E_0 \Phi_\uparrow$ with

$$E_0 = \sum_{x \in \Lambda_o} \tau_{x,x} = L^d \tau_{o,o} \quad \blacksquare \tag{5.12}$$

We also prove the theorem about the instability of the "ferromagnetic ground states." The proof is based on the standard variational argument.

Proof of Theorem 3.3. Let $d_{k,\sigma}^\dagger$ be the creation operator for the Bloch state (4.6) in the lowest band with the wavenumber vector $k \in \mathscr{K}$, and let $\varepsilon_1(k)$ be the corresponding energy eigenvalue. Let k_{min} and k_{max} be such that

$$\varepsilon_1(k_{min}) \leqslant \varepsilon_1(k) \leqslant \varepsilon_1(k_{max}) \tag{5.13}$$

holds for any $k \in \mathscr{K}$. The bandwidth is given by $\bar{\varepsilon} = \varepsilon_1(k_{max}) - \varepsilon_1(k_{min})$. Take a variational state

$$\Phi_{var} = d_{k_{min},\downarrow}^\dagger d_{k_{max},\uparrow} \Phi_\uparrow \tag{5.14}$$

The energy expectation value of the state Φ_{var} is easily shown to satisfy

$$\frac{(\Phi_{var}, H\Phi_{var})}{(\Phi_{var}, \Phi_{var})} \leqslant E_0 - \bar{\varepsilon} + U = E_{min}(S_{max}) - \bar{\varepsilon} + U \tag{5.15}$$

The claimed instability follows when $\bar{\varepsilon} > U$. $\quad \blacksquare$

5.4. Flat-Band Ferromagnetism

In refs. 47 and 34, Theorem 3.1, which establishes flat-band ferro-magnetism, was proved for the models with $v = 1$. Although the extension to the general case is not hard, we present it here for completeness.

Proof of Theorem 3.1. The flat-band model is characterized by $\tau_{x,y} = 0$ for any $x, y \in \Lambda_o$. Then it is easily verified [from, say, (5.10)] that $H_{\text{hop}} \geqslant 0$. We also know that $H_{\text{int}} \geqslant 0$, and hence $H \geqslant 0$. From (5.12), on the other hand, one finds that the "ferromagnetic ground state" $\Phi_\uparrow$ of (5.11) has vanishing energy, and hence is a ground state of H. The remaining task is to determine all the other ground states.

Let Φ be an arbitrary ground state with L^d electrons. We obviously have

$$H_{\text{hop}} \Phi = 0 \tag{5.16}$$

and

$$H_{\text{int}} \Phi = 0 \tag{5.17}$$

which means that Φ is at the same time a ground state of H_{hop} and of H_{int}. As we discussed in Section 2.1, this is a special feature of flat-band models.

Since H_{int} of (3.15) is a sum of nonnegative terms, (5.17) implies $n_{x,\uparrow} n_{x,\downarrow} \Phi = 0$ for each $x \in \Lambda$. Since $n_{x,\uparrow} n_{x,\downarrow} = (c_{x,\uparrow} c_{x,\downarrow})^\dagger (c_{x,\uparrow} c_{x,\downarrow})$, this further implies $c_{x,\uparrow} c_{x,\downarrow} \Phi = 0$ for each $x \in \Lambda$. By using the inversion formula (5.6) and noting that $(\varphi_x^{(y)})^* = \psi_x^{(y)}$ for the flat-band models [see Section 4.2, especially (4.23)], this reduces to the following useful condition:

$$\sum_{y,z \in \Lambda} \psi_x^{(y)} \psi_x^{(z)} b_{y,\uparrow} b_{z,\downarrow} \Phi = 0 \tag{5.18}$$

The relation (5.16) implies that the state Φ consists only of the single-electron states from the lowest (flat) band. Therefore we expand it as

$$\Phi = \sum_{A, B \in \Lambda_o} f(A, B) \left(\prod_{x \in A} a_{x,\uparrow}^\dagger \right) \left(\prod_{x \in B} a_{x,\downarrow}^\dagger \right) \Phi_{\text{vac}} \tag{5.19}$$

where the sum is taken over all subsets $A, B \subset \Lambda_o$ such that $|A| + |B| = L^d$, and $f(A, B)$ are coefficients.

For $x \in \Lambda_o$ and Φ of the form (5.19), the condition (5.18) becomes

$$b_{x,\uparrow} b_{x,\downarrow} \Phi = 0 \tag{5.20}$$

because of the definition (4.23) of the ψ states. By using the anticommutation relation (5.4), (5.20) implies that $f(A, B) = 0$ whenever $A \cap B \neq \varnothing$. Thus the expansion (5.19) can be reorganized as

$$\Phi = \sum_{\sigma} g(\sigma) \left(\prod_{x \in \Lambda_o} a^{\dagger}_{x, \sigma(x)} \right) \Phi_{\text{vac}} \tag{5.21}$$

where the sum is now taken over all the possible "spin configurations" $\sigma = (\sigma(x))_{x \in \Lambda_o}$ with $\sigma(x) = \uparrow, \downarrow$.

For $x \in \Lambda'$ and Φ of the form (5.21), the condition (5.18) becomes

$$\sum_{\substack{y, z \in \Lambda_o(x) \\ y > z}} (b_{y, \uparrow} b_{z, \downarrow} - b_{z, \uparrow} b_{y, \downarrow}) \Phi = 0 \tag{5.22}$$

where $\Lambda_o(x) = \{ y \in \Lambda_o \mid |y - x| = \sqrt{v}/2 \}$, and we ordered this set in an arbitrary manner. Since any site $x \in \Lambda_o$ is "occupied" in the representation (5.21), the condition (5.22) is satisfied only when we have[36]

$$(b_{y, \uparrow} b_{z, \downarrow} - b_{z, \uparrow} b_{y, \downarrow}) \Phi = 0 \tag{5.23}$$

for any $y, z \in \Lambda_o(x)$ with $y \neq z$ for some $x \in \Lambda'$.

By substituting the expansion (5.21) into the condition (5.23), we find that the coefficients satisfy

$$g(\sigma) = g(\sigma_{y, z}) \tag{5.24}$$

where $\sigma_{y, z}$ is the spin configuration obtained by switching $\sigma(y)$ and $\sigma(z)$ in the original σ. The relation (5.24) along with the expansion (5.21) implies that Φ can be written as

$$\Phi = \sum_{M = 0}^{L^d} \alpha_M (S^{-}_{\text{tot}})^M \Phi_{\uparrow} \tag{5.25}$$

with suitable coefficients α_M. Here $S^{-}_{\text{tot}} = S^{(1)}_{\text{tot}} - iS^{(2)}_{\text{tot}}$ is the spin-lowering operator. This proves the desired theorem. ∎

---

[36] This is only true when the electron number is $L^d = |\Lambda_o|$. We treated only the special models with $v = 1$ in refs. 47 and 34, where this step can be extended to other electron numbers.

5.5. Basis for the Many-Electron System

We shall introduce a basis for describing many-electron problems. Let $s \in \Lambda$, and let $A \subset \Lambda$ be a subset with $|A| = L^d - 1$. We define

$$\Psi_{s,A}(k) = \sum_{x \in \Lambda_o} e^{ik \cdot x} T_x \left[a_{s,\downarrow}^\dagger \left(\prod_{t \in A} a_{t,\uparrow}^\dagger \right) \Phi_{\text{vac}} \right] \tag{5.26}$$

where T_x is the translation operator (3.32). The state $\Psi_{s,A}(k)$ is an element of the Hilbert space $\mathscr{H}_k$, (3.34), of the states with momentum k and a single down-spin electron. Clearly $\Psi_{s,A}(k)$ with different (s, A) can define the same state. For $s \in \Lambda_u$ (with a $u \in \mathscr{U}$), one can take a unique $y \in \Lambda_o$ such that $s = u - y$. Then we have

$$\Psi_{s,A}(k) = e^{i\theta} \Psi_{u,A+y}(k) \tag{5.27}$$

where $A + y = \{ x + y \mid x \in A \}$, and $\theta \in \mathbf{R}$.

Let $\bar{\Lambda}_o = \Lambda_o \backslash \{ o \}$. We define

$$\Omega(k) = \frac{1}{\alpha(k)} \Psi_{o,\bar{\Lambda}_o}(k) \tag{5.28}$$

where $\alpha(k) > 0$ is a real function of k which will be determined later in the proof. We note that $\Omega(k)$ is our approximate spin-wave excitation, which plays the central role in our proof.

Finally, we define our basis $\mathscr{B}_k$ for the space $\mathscr{H}_k$ as

$$\mathscr{B}_k = \left\{ \Omega(k) \right\} \cup \left\{ \Psi_{u,A}(k) \mid u \in \mathscr{U}, A \subset \Lambda \text{ with } |A| = L^d - 1 \text{ and} \right.$$
$$\left. (u, A) \neq (o, \bar{\Lambda}_o) \right\} \tag{5.29}$$

6. PROOF OF THE MAIN THEOREMS

In the present section, we shall describe the proof of our main theorems on the stability of ferromagnetism and the lower bound for the spin-wave dispersion relation. We make use of various estimates which will be proved in later sections.

6.1. Basic Lemma

Let us define

$$\tilde{H} = \tilde{H}_{\text{hop}} + H_{\text{int}} \tag{6.1}$$

where $\tilde{H}_{\text{hop}}$ is the modified hopping Hamiltonian (4.21) and H_{int} is the standard interaction Hamiltonian (3.15). For basis states Φ, $\Psi \in \mathcal{B}_k$, we define the matrix elements $h[\Psi, \Phi] \in \mathbb{C}$ of the Hamiltonian $\tilde{H}$ above by the unique expansion

$$\tilde{H}\Phi = \sum_{\Psi \in \mathcal{B}_k} h[\Psi, \Phi] \Psi \tag{6.2}$$

Note that only states from $\mathcal{B}_k$ with a fixed k appear in the right-hand side of (6.2) since $\tilde{H}$ is translation invariant and the momentum k is conserved.
   For $\Phi \in \mathcal{B}_k$, we define

$$D[\Phi] = \text{Re}[h[\Phi, \Phi]] - \sum_{\Psi \in \mathcal{B}_k \setminus \{\Phi\}} |h[\Phi, \Psi]| \tag{6.3}$$

Then we have the following lemma. The basic statement is elementary and well known (in standard linear algebra), but it serves as a basis of our proof.

   Lemma 6.1. Let $E_{\text{SW}}(k)$ be the energy of the spin-wave excitation defined in Section 3.4. Then for each $k \in \mathcal{K}$, we have

$$E_{\text{SW}}(k) \geqslant \min_{\Phi \in \mathcal{B}_k} D[\Phi] \tag{6.4}$$

   Proof. Let $\tilde{E}(k)$ be the lowest eigenvalue of $\tilde{H}$ in the Hilbert space $\mathcal{H}_k$, (3.34). We first claim that $E_{\text{SW}}(k) \geqslant \tilde{E}(k)$. This is a straightforward consequence of the operator inequality $H \geqslant \tilde{H}$ [which follows from (6.1) and (4.22)] and the fact that both H and $\tilde{H}$ commute with T_x ($x \in \mathbb{Z}^d$) and $S_{\text{tot}}^{(3)}$.
   Thus the desired bound (6.4) follows from the inequality

$$\tilde{E}(k) \geqslant \min_{\Phi \in \mathcal{B}_k} D[\Phi] \tag{6.5}$$

which is indeed a straightfoward consequence of a well-known relation in elementary linear algebra. To show (6.5), let E be an eigenvalue of $\tilde{H}$, and $\Phi_0 \in \mathcal{H}_k$ be the corresponding eigenstate. We expand Φ_0 as $\Phi_0 = \sum_{\Psi \in \mathcal{B}_k} C(\Psi) \Psi$, where $C(\Psi)$ are coefficients. From (6.2) and the eigenvalue equation $\tilde{H}\Phi_0 = E\Phi_0$, we find that $C(\Psi)$ satisfy

$$EC(\Phi) = \sum_{\Psi \in \mathcal{B}_k} h[\Phi, \Psi] C(\Psi) \tag{6.6}$$

for any $\Phi \in \mathcal{B}_k$. Let $\Phi' \in \mathcal{B}_k$ be the state such that $|C(\Psi)/C(\Phi')| \leqslant 1$ holds for any $\Psi \in \mathcal{B}_k$. Then we have

$$
\begin{aligned}
E &= \sum_{\Psi \in \mathcal{B}_k} h[\Phi', \Psi] \frac{C(\Psi)}{C(\Phi')} \\
&\geqslant \operatorname{Re}[h[\Phi', \Phi']] - \sum_{\Psi \in \mathcal{B}_k \setminus \{\Phi'\}} |h[\Phi', \Psi]| \\
&= D[\Phi']
\end{aligned}
\tag{6.7}
$$

Since $\tilde{E}(k)$ is the smallest eigenvalue, the desired inequality (6.5) follows. ∎

Being a very crude bound, we cannot expect (6.4) to yield meaningful results unless we use a basis which "almost diagonalizes" the low-energy part of the Hamiltonian. As we shall see below, it turns out that the basis we constructed in Section 5.5 indeed has such properties.

6.2. Estimates of the Matrix Elements

We shall summarize the results of Sections 7 and 8, where we estimate various matrix elements.

Before stating the results, it is convenient to introduce a new labeling of the special states $\Psi \in \mathcal{B}_k$ which have nonvanishing matrix elements $h[\Psi, \Omega(k)]$. For any $u \in \mathcal{U}$ and $r \in \Lambda_o$, we define

$$
\Phi_{u,r}(k) = \sum_{x \in \Lambda_o} e^{ik \cdot x} a^\dagger_{x+u,\downarrow} b_{x+r,\uparrow} \Phi_\uparrow
\tag{6.8}
$$

where $\Phi_\uparrow = \left(\prod_{y \in \Lambda_o} a^\dagger_{y,\uparrow} \right) \Phi_{\mathrm{vac}}$ is the "ferromagnetic ground state." By noting that $T_x[\prod_{y \in \Lambda_o} a^\dagger_{y,\uparrow}] = \prod_{y \in \Lambda_o} a^\dagger_{y,\uparrow}$, we can relate the state (6.8) to the general state $\Psi_{s,A}(k)$ of (5.26) as

$$
\begin{aligned}
\Phi_{u,r}(k) &= \sum_{x \in \Lambda_o} e^{ik \cdot x} T_x \left[a^\dagger_{u,\downarrow} b_{r,\uparrow} \left(\prod_{y \in \Lambda_o} a^\dagger_{y,\uparrow} \right) \Phi_{\mathrm{vac}} \right] \\
&= \operatorname{sgn}[r] \sum_{x \in \Lambda_o} e^{ik \cdot x} T_x \left[a^\dagger_{u,\downarrow} \left(\prod_{y \in \Lambda_o \setminus \{r\}} a^\dagger_{y,\uparrow} \right) \Phi_{\mathrm{vac}} \right] \\
&= \operatorname{sgn}[r] \, \Psi_{u, \Lambda_o \setminus \{r\}}(k)
\end{aligned}
\tag{6.9}
$$

where $\mathrm{sgn}[r] = \pm 1$. By using (6.9), we can rewrite (5.28) as

$$\Omega(k) = \frac{1}{\alpha(k)} \Phi_{o,o}(k) \qquad (6.10)$$

Let $\Lambda_o^+ \subset \Lambda_o$ be a special subset with the property that for any $s, t \in \Lambda_o$ such that $s \neq t$, we have either $s - t \in \Lambda_o^+$ or $t - s \in \Lambda_o^+$ (and not both). An example is

$$\Lambda_o^+ = \left\{ x \in \Lambda_o \,\bigg|\, 0 < x_1 \leqslant \frac{L-1}{2} \right\} \cup \left\{ x \in \Lambda_o \,\bigg|\, x_1 = 0, \, 0 < x_2 \leqslant \frac{L-1}{2} \right\}$$

$$\cup \left\{ x \in \Lambda_o \,\bigg|\, x_1 = x_2 = 0, \, 0 < x_3 \leqslant \frac{L-1}{2} \right\} \cup \cdots$$

$$\cup \left\{ x \in \Lambda_o \,\bigg|\, x_1 = x_2 = \cdots = x_{d-1} = 0, \, 0 < x_d \leqslant \frac{L-1}{2} \right\} \qquad (6.11)$$

For $u \in \mathcal{U}$, $r \in \Lambda'$, and $s, t \in \Lambda_o$ such that $s - t \in \Lambda_o^+$, we define

$$\Phi_{u,r,t,s}(k) = \sum_{x \in \Lambda_o} e^{ik \cdot x} a^\dagger_{x+u,\downarrow} a^\dagger_{x+r,\uparrow} b_{x+t,\uparrow} b_{x+s,\uparrow} \Phi_\uparrow \qquad (6.12)$$

It can be shown that the only states $\Psi \in \mathcal{B}_k$ such that $h[\Psi, \Omega(k)] \neq 0$ can be written in the form $\Phi_{u,r}(k)$ or $\Phi_{u,r,t,s}(k)$ with suitable u, r, t, and s. See Section 7.

By using the representations (5.7) and (5.8) for the Hamiltonians, we can express the matrix elements $h[\Psi, \Phi]$ explicitly in terms of the effective hopping $\tau_{x,y}$ and the effective interaction $\tilde{U}_{y,v;w,z}$. We leave the derivation to Section 7, and summarize the results as the following lemma.

Lemma 6.2. For any u, r, t, and s as in (6.8) or (6.12), we have

$$h[\Omega(k), \Omega(k)] = E_0 + 2 \sum_{s \in \Lambda_o} \left(\sin \frac{k \cdot s}{2} \right)^2 \tilde{U}_{s,o;s,o} \qquad (6.13)$$

$$h[\Omega(k), \Phi_{u,r}(k)] = \delta_{u,o} \alpha(k)(e^{-ik \cdot r} - 1) \, \tau_{r,o}$$

$$+ \alpha(k) \sum_{s \in \Lambda_o} (e^{-ik \cdot r} - e^{-ik \cdot s}) \, \tilde{U}_{s,r;u,s} \qquad (6.14)$$

$$h[\Omega(k), \Phi_{u,r,t,s}(k)] = \alpha(k)(e^{-ik\cdot s} - e^{-ik\cdot t})\, \tilde{U}_{s,t;u,r} \tag{6.15}$$

$$h[\Phi_{u,r}(k), \Omega(k)] = \delta_{u,o}\frac{1}{\alpha(k)}(e^{ik\cdot r} - 1)\, \tau_{o,r}$$

$$+ \frac{1}{\alpha(k)}\sum_{s\in A_o}(e^{ik\cdot r} - e^{ik\cdot s})\, \tilde{U}_{u,s;s,r} \tag{6.16}$$

and

$$h[\Phi_{u,r;t,s}(k), \Omega(k)] = \frac{1}{\alpha(k)}(e^{ik\cdot s} - e^{ik\cdot t})\, \tilde{U}_{u,r;s,t} \tag{6.17}$$

It should be noted that these matrix elements are not symmetric, reflecting that the basis $\mathscr{B}_k$ is not orthonormal.

By combining the expressions (6.13)–(6.17), the bounds (4.32) and (4.33) for $\tau_{x,y}$, the representation (5.9) for $\tilde{U}_{y,v;w,z}$ in terms of the basis states $\varphi_y^{(x)}$ and $\tilde{\varphi}_y^{(x)}$, and the bounds (4.25)–(4.28) and (4.36)–(4.39) for these states, we can derive explicit bounds for the matrix elements and their sums. Again we leave all the derivations to Section 8, and summarize the results as the following lemma.

Lemma 6.3. Under the assumptions that $\lambda \geqslant \lambda_0$ and $|\rho| \leqslant \rho_0$, we have

$$\mathrm{Re}[h[\Omega(k)\,\Omega(k)]] \geqslant E_0 + \frac{U}{\lambda^4}\left(1 - C_1|\rho| - \frac{C_2}{\lambda}\right)G(k) \tag{6.18}$$

with $G(k)$ defined in (3.36),

$$\sum_{\Psi\in\mathscr{B}_k\setminus\{\Omega(k)\}}|h[\Omega(k), \Psi]| \leqslant \alpha(k)\left(B_1 R\,|\rho|\,t + \frac{C_3 U}{\lambda^2}\right)|k| \tag{6.19}$$

$$|h[\Phi_{u,r}(k), \Omega(k)]| \leqslant \frac{1}{\alpha(k)}\left(B_1 R\,|\rho|\,t + \frac{C_4 U\,|\rho|}{\lambda^2} + \frac{C_5 U}{\lambda^3}\right)|k| \tag{6.20}$$

and

$$|h[\Phi_{u,r,t,s}(k), \Omega(k)]| \leqslant \frac{1}{\alpha(k)}\frac{C_6 U}{\lambda^2}|k| \tag{6.21}$$

Here C_i $(i = 1, 2, 3, 4, 5, 6)$ are positive constants which depend only on d, v, and R.

We can perform a similar analysis for the matrix elements which do not involve the state $\Omega(k)$. For $\Phi \in \mathscr{B}_k \backslash \{\Omega(k)\}$, we define

$$\tilde{D}[\Phi] = \text{Re}[h[\Phi, \Phi]] - \sum_{\Psi \in \mathscr{B}_k \backslash \{\Phi, \Omega(k)\}} |h[\Phi, \Psi]| \qquad (6.22)$$

Then we prove the following in Section 8.4.

Lemma 6.4. Assume that $\lambda \geqslant \lambda_4$, $|\rho| \leqslant \rho_0$, and $K_3 t |\rho| \leqslant U \leqslant K_4 \lambda^3 t$, where λ_4, ρ_0, K_3, and K_4 are constants which depend only on d, v, and R. Then we have

$$\tilde{D}[\Psi_{u,A}(k)] \geqslant E_0 + \frac{\lambda^2 t}{2} \qquad (6.23)$$

for any $u \in \mathscr{U}$ and $A \subset \Lambda$ such that $|A| = L^d - 1$ and $A \cap \Lambda' \neq \varnothing$,

$$\tilde{D}[\Phi_{u,r}(k)] \geqslant E_0 + \frac{\lambda^2 t}{2} \qquad (6.24)$$

for $u \neq o$, and

$$\tilde{D}[\Phi_{o,r}(k)] \geqslant E_0 + \frac{U}{2} \qquad (6.25)$$

for $r \neq o$.

6.3. Proof of Theorem 3.6

We will now prove Theorem 3.6 for the lower bound of the spin-wave excitation energy, which is one of most important results. In the proof, we make use of Lemmas 6.3 and 6.4. We will later confirm that the conditions for these lemmas are satisfied.

From Lemma 6.1, we find that the desired lower bound (3.38) follows if we show

$$D[\Phi] \geqslant E_0 + F_2 \frac{U}{\lambda^4} G(k) \qquad (6.26)$$

for any $\Phi \in \mathscr{B}_k$. [Recall that $E_0 = E_{\min}(S_{\max})$.]

We shall first verify (6.26) for $\Phi = \Psi_{u,A}(k) \in \mathscr{B}_k$ such that

$$h[\Psi_{u,A}(k), \Omega(k)] = 0 \qquad (6.27)$$

Then, comparing the definitions (6.3) and (6.22), we find $D[\Psi_{u,A}(k)] = \tilde{D}[\Psi_{u,A}(k)]$ for such $\Psi_{u,A}(k)$. We also claim that the condition (6.27) inevitably implies $A \cap A' \neq \varnothing$. To see this, we note that the converse $A \cap A' = \varnothing$ means $A = A_o \backslash \{r\}$ for some $r \in A_o$, and hence $\Psi_{u,A}(k)$ is equal to $\pm \Phi_{u,r}(k)$. [See (6.9).] Therefore we can use the lower bound (6.23) to find

$$D[\Psi_{u,A}(k)] = \tilde{D}[\Psi_{u,A}(k)] \geqslant E_0 + \frac{\lambda^2 t}{2} \geqslant E_0 + F_2 \frac{U}{\lambda^4} G(k) \qquad (6.28)$$

where the final bound is derived by noting that $G(k) \leqslant 2^{2\nu+1}\binom{d}{\nu}$ and $F_2 \leqslant 1$, and assuming that

$$\frac{\lambda^6 t}{U} \geqslant 2^{\nu+2}\binom{d}{\nu} \qquad (6.29)$$

Therefore the desired inequality (6.26) is verified for $\Phi = \Psi_{u,A}(k)$ such that (6.27) holds.

Next we examine the inequality (6.26) for the states which do not satisfy the condition (6.27). They are the states $\Omega(k)$, $\Phi_{u,r}(k)$, and $\Phi_{u,r,s}(k)$ defined in (5.28) [see also (6.10)], (6.8), and (6.12), respectively.

As for the state $\Phi_{u,r}(k)$, we use the definitions (6.3), (6.22), and the bounds (6.24) and (6.20) to get

$$D[\Phi_{u,r}(k)] = \tilde{D}[\Phi_{u,r}(k)] - |h[\Phi_{u,r}(k), \Omega(k)]|$$

$$\geqslant E_0 + \frac{U}{2} - \frac{1}{\alpha(k)}\left(B_1 R |\rho| t + \frac{C_4 U |\rho|}{\lambda^2} + \frac{C_5 U}{\lambda^3}\right) |k| \qquad (6.30)$$

Let us choose $\alpha(k)$ as

$$\alpha(k) = \frac{4 |k|}{U}\left(B_1 R |\rho| t + \frac{C_4 U |\rho|}{\lambda^2} + \frac{C_5 U}{\lambda^3}\right) \qquad (6.31)$$

Then (6.30) becomes

$$D[\Phi_{u,r}(k)] \geqslant E_0 + \frac{U}{4} \geqslant E_0 + F_2 \frac{U}{\lambda^4} G(k) \qquad (6.32)$$

To get the final bound, we have made a further assumption that

$$\lambda^4 \geqslant 2^{2\nu+3} \binom{d}{\nu} \tag{6.33}$$

We have shown the desired bound (6.26) for $\Phi = \Phi_{u,r}(k)$.

The state $\Phi_{u,r,t,s}(k)$ (where we require $r \in \Lambda'$) satisfies the condition for the bound (6.23). By combining (6.23) with the bound (6.21), and using (6.31), we have

$$D[\Phi_{u,r,t,s}(k)] = \tilde{D}[\Phi_{u,r,t,s}(k)] - |h[\Phi_{u,r,t,s}(k), \Omega(k)]|$$

$$\geqslant E_0 + \frac{\lambda^2 t}{2} - \frac{1}{\alpha(k)} \frac{C_6 U}{\lambda^2} |k|$$

$$= E_0 + \frac{\lambda^2 t}{2} - \frac{C_6 U}{4\lambda^2 (B_1 R |\rho| \, t + C_4 U |\rho|/\lambda^2 + C_5 U/\lambda^3)}$$

$$\geqslant E_0 + \frac{\lambda^2 t}{2} - \frac{C_6 \lambda}{4 C_5} U$$

$$\geqslant E_0 + \frac{\lambda^2 t}{4}$$

$$\geqslant E_0 + F_2 \frac{U}{\lambda^4} G(k) \tag{6.34}$$

where, to get the final bound, we required

$$0 \leqslant U \leqslant K_2 \lambda t \tag{6.35}$$

with $K_2 = C_5/C_6$, and

$$\frac{\lambda^6 t}{U} \geqslant 2^{2\nu+3} \binom{d}{\nu} \tag{6.36}$$

We have shown the bound (6.26) for $\Phi = \Phi_{u,r,t,s}(k)$.

Finally we examine the state $\Omega(k)$, which is our trial state for the elementary spin-wave excitation. By using the bounds (6.18) and (6.19) and the choice (6.31) of $\alpha(k)$, we get

$$D[\Omega(k)] \geqslant E_0 + \frac{U}{\lambda^4} \left(1 - C_1 |\rho| - \frac{C_2}{\lambda} \right) G(k)$$

$$- \alpha(k) \left(B_1 R |\rho| \, t + \frac{C_3 U}{\lambda^2} \right) |k|$$

$$= E_0 + \frac{U}{\lambda^4} \left(1 - C_1 |\rho| - \frac{C_2}{\lambda} \right) G(k)$$

$$- 4 \frac{U}{\lambda^4} \left(C_3 + B_1 R \frac{\lambda^2 t |\rho|}{U} \right) \left(C_4 |\rho| + \frac{C_5}{\lambda} + B_1 R \frac{\lambda^2 t |\rho|}{U} \right) |k|^2$$

$$\geqslant E_0 + \frac{U}{\lambda^4} \left(1 - A_1 |\rho| - \frac{A_2}{\lambda} - A_3 \frac{\lambda^2 t |\rho|}{U} \right) G(k)$$

$$= E_0 + F_2 \frac{U}{\lambda^4} G(k) \tag{6.37}$$

with suitable positive constants A_1, A_2, and A_3. Here we used the bound

$$|k|^2 \leqslant \pi^2 \sum_{i=1}^{d} \left(\sin \frac{k_i}{2} \right)^2 \leqslant \frac{\pi^2}{4} G(k) \tag{6.38}$$

which follows from $|k_i| \leqslant \pi$, and further assumed that

$$A_3 \frac{\lambda^2 t |\rho|}{U} \leqslant 1 \tag{6.39}$$

We have thus confirmed the desired bound (6.26) for all $\Phi \in \mathscr{B}_k$. This means that the desired lower bound (3.38) for the spin-wave excitation energy has been proved.

It remains to examine the conditions for the model parameters assumed in the proof. The assumptions made during the proof are (6.29), (6.33), (6.35), (6.36), and (6.39). Among them (6.35) and (6.39) are explicitly assumed in the statement of the theorem.

Since we shall choose λ_3 so that $\lambda_3 \geqslant \lambda_4 \geqslant \lambda_0$, the conditions about λ and ρ stated in Lemmas 6.3 and 6.4 are satisfied.

Let us set

$$\lambda_3 = \max \left\{ \lambda_0, \lambda_4, \left(K_2 2^{\nu+3} \binom{d}{\nu} \right)^{1/5}, \left(2^{2\nu+3} \binom{d}{\nu} \right)^{1/4}, (K_2/K_4)^{1/2}, (K_3/A_3)^{1/2} \right\} \tag{6.40}$$

From the assumption $\lambda \geqslant \lambda_3$ (with the above λ_3) and the assumed (6.35) and (6.39), we can verify that the conditions (6.29), (6.33), (6.36), and $K_3 |\rho| t \leqslant U \leqslant K_4 \lambda^3 t$ (which is required in Lemma 6.4) are satisfied. Finally the conditions $\lambda \geqslant \lambda_0$ and $\lambda \geqslant \lambda_4$ required in Lemmas 6.3 and 6.4, respectively, are satisfied since $\lambda \geqslant \lambda_4 \geqslant \lambda_0$. This completes the proof of the theorem.

6.4. Proof of Theorem 3.4

We now prove our main theorem, which states the local stability of the "ferromagnetic ground states."

Theorem 3.4 follows from the following statement, which has more general (but more compliated) conditions.

Lemma 6.5. The local stability inequality (3.31) is valid if either of the following is satisfied:

(i) $\lambda \geq \lambda_3$, $|\rho| \leq \rho_0$, $A_1 |\rho| + A_2 \lambda^{-1} + A_3 \lambda^2 t |\rho| U^{-1} < 1$, and $0 < U \leq K_2 \lambda t$.

(ii) $\lambda \geq \lambda_3$, $|\rho| \leq \rho_0$, $A_1 |\rho| + A_2 \lambda^{-1} + A_3 \lambda |\rho| (K_2)^{-1} < 1$, and $U \geq K_2 \lambda t$.

The constants A_1, A_2, and A_3 are the same as those in Theorem 3.6.

The lemma actually is the most natural way of expressing our stability theorem. The conditions (3.29) and (3.30) in Theorem 3.4 were introduced to give an easily accessible sufficient condition for condition (i) or (ii) in Lemma 6.5.

Proof of Theorem 3.4, Given Lemma 6.5. We set $\lambda_2 = \max\{\lambda_3, A_2/4\}$, $\rho_1 = \min\{\rho_0, (4A_1)^{-1}\}$, $p_1 = K_2(4A_3)^{-1}$, and $K_1 = 4A_3$. Suppose that the conditions in Theorem 3.4 are satisfied.

We first assume $0 \leq U \leq K_2 \lambda t$. Then we have $A_1 |\rho| \leq 1/4$, $A_2/\lambda \leq 1/4$, and $A_3 \lambda^2 t |\rho|/U \leq 1/4$. It is obvious that all the conditions in (i) are satisfied.

Next we assume $U \geq K_2 \lambda t$. Again we have $A_1 |\rho| \leq 1/4$, $A_2/\lambda \leq 1/4$, and $A_3 \lambda |\rho|/K_2 \leq 1/4$. The conditions in (ii) are satisfied. $\blacksquare$

In what follows we prove Lemma 6.5.

For each state Φ which is an eigenstate of $(\mathbf{S}_{\text{tot}})^2$ with $S_{\text{tot}} = S_{\max} - 1$, we can take its $SU(2)$ rotation $\tilde{\Phi}$ which satisfies $S_{\text{tot}}^{(3)} \tilde{\Phi} = (S_{\max} - 1) \tilde{\Phi}$. Since Φ and $\tilde{\Phi}$ have the same energy, it suffices to concentrate on the space

$$\mathcal{H}_{S_{\max} - 1} = \{ \Phi \mid S_{\text{tot}}^{(3)} \Psi = (S_{\max} - 1) \Phi \} \tag{6.41}$$

and prove the stability theorem. By using $\mathcal{H}_k$ defined in (3.34), the above space is decomposed

$$\mathcal{H}_{S_{\max} - 1} = \bigoplus_{k \in \mathcal{K}} \mathcal{H}_k \tag{6.42}$$

We first assume that the condition (i) in Lemma 6.5 is satisfied. Then the assumptions of Theorem 3.6 are automatically satisfied, and we also have $F_2 > 0$. Thus for any $k \in \mathcal{K}$ such that $k \neq o = (0,..., 0)$ the lowest energy $E_{SW}(k)$ in the sector $\mathcal{H}_k$ satisfies

$$E_{SW}(k) > E_0 = E_{min}(S_{max}) \tag{6.43}$$

Recalling the decomposition (6.41), onre finds that this proves the desired bound $E_{min}(S_{max} - 1) > E_{min}(S_{max})$ except in the sector $\mathcal{H}_o$.

To deal with the sector $\mathcal{H}_o$ is not hard. We note that the state $\Omega(o)$ is written as

$$\Omega(o) = \frac{1}{\alpha(k)} S_{tot}^{-} \Phi_\uparrow \tag{6.44}$$

where $S_{tot}^{-} = S_{tot}^{(1)} - i S_{tot}^{(1)}$ is the spin-lowering operator. This means that $\Omega(o)$ is nothing but one of the "ferromagnetic ground states" and has the total spin $S_{tot} = S_{max}$. Let $E_{min}(S_{max} - 1, o)$ be the lowest energy in the sector $\mathcal{H}_o$ with $S_{tot} = S_{max} - 1$. Then, by repeating the argument in the proof of Lemma 6.1, we find that

$$E_{min}(S_{max} - 1, o) \geq \min_{\Phi \in \mathcal{B}_o \setminus \{\Omega(o)\}} D[\Phi] \tag{6.45}$$

The right-hand side can be bounded from below by using the inequalities (6.28), (6.32), and (6.34). We get

$$E_{min}(S_{max} - 1, o) \geq E_0 + \min \left\{ \frac{\lambda^2 t}{2}, \frac{U}{4}, \frac{\lambda^2 t}{4} \right\} > E_0 \tag{6.46}$$

which completes the proof of the desired local stability inequality (3.31).

The only remaining task is to prove the inequality (3.31) when the condition (ii) in Lemma 6.5 is satisfied.[37] Note that U is not bounded from above in this case.

The key ingredient in the extension is to realize that $E_{min}(S_{max})$ does not depend on U, while $E_{min}(S_{max} - 1)$ is increasing in U. The latter fact follows by noting that H_{int} is increasing in U (as an operator), both H_{hop} and H_{int} commute with the total spin operator, and $E_{min}(S_{max} - 1)$ is the lowest energy in the sector with the fixed S_{tot}.

Suppose that the condition (ii) in the Remark after Theorem 3.4 is satisfied. Then by setting $U = K_2 \lambda t$, the condition (i) in Lemma 6.5 is satisfied, and we have $E_{min}(S_{max} - 1) > E_{min}(S_{max})$. Because of the increasing

---

[37] The following argument was suggested to the author by Andreas Mielke.

property of $E_{\min}(S_{\max}-1)$, this inequality remains valid if we increase U with other parameters kept fixed. This proves the local stability inequality (3.31).

7. REPRESENTATION OF THE MATRIX ELEMENTS

Here we will prove Lemma 6.2 about the representation of the matrix elements involving the states $\Omega(k)$, $\Phi_{u,r}(k)$, and $\Phi_{u,r,t,s}(k)$.

7.1. Treatment of the Hopping Hamiltonian

By operating $\tilde{H}_{\text{hop}}$ with in the form (5.7) on the state $\Phi_{o,r}(k)$ [see (6.8)] and using the anticommutation relations (5.3), (5.4), we get

$$\tilde{H}_{\text{hop}}\Phi_{o,r}(k) = \left(\sum_{\substack{y,v\in\Lambda_o \\ \sigma=\uparrow,\downarrow}}\tau_{y,v}a^\dagger_{y,\sigma}a^\dagger_{v,\sigma}\right)\sum_{p\in\Lambda_o}e^{ik\cdot p}a^\dagger_{p,\downarrow}b_{p+r,\uparrow}\Phi_\uparrow$$

$$= -\sum_{p,v\in\Lambda_o}\tau_{p+r,v}e^{ik\cdot p}a^\dagger_{p,\downarrow}b_{v,\uparrow}\Phi_\uparrow$$

$$+\sum_{p,y\in\Lambda_o}\tau_{y,y}e^{ik\cdot p}a^\dagger_{p,\downarrow}b_{p+r,\uparrow}\Phi_\uparrow$$

$$+\sum_{p,y\in\Lambda_o}\tau_{y,p}e^{ik\cdot p}a^\dagger_{y,\downarrow}b_{p+r,\uparrow}\Phi_\uparrow \tag{7.1}$$

We shall make the change of variables $p=x$, $v=x+s$ (with x, $s\in\Lambda_o$) in the first term, and the change of variables $y=x$, $p=x+s-r$ (with x, $s\in\Lambda_o$) in the second term. By also using (5.12), we have

$$\tilde{H}_{\text{hop}}\Phi_{o,r}(k) = E_0\Phi_{o,r}(k) - \sum_{x,s\in\Lambda_o}\tau_{r,s}e^{ik\cdot x}a^\dagger_{x,\downarrow}b_{x+s,\uparrow}\Phi_\uparrow$$

$$+\sum_{x,s\in\Lambda_o}\tau_{o,s-r}e^{ik\cdot(s-r)}e^{ik\cdot x}a^\dagger_{x,\downarrow}b_{x+s,\uparrow}\Phi_\uparrow$$

$$= E_0\Phi_{o,r}(k) + \sum_{s\in\Lambda_o}\tau_{r,s}(e^{ik\cdot(s-r)}-1)\Phi_{o,s}(k) \tag{7.2}$$

where we made use of the translation invariance of $\tau_{x,y}$. Following the definition (6.2) of matrix elements, we define the matrix elements $h_{\text{hop}}[\Psi,\Phi]$ by the unique expansion

$$\tilde{H}_{\text{hop}}\Phi = \sum_{\Psi\in\mathscr{B}_k}h_{\text{hop}}[\Psi,\Phi]\Psi \tag{7.3}$$

By comparing (7.2) with this definition, we find

$$h_{\text{hop}}[\Phi_{o,s}(k), \Phi_{o,r}(k)] = \delta_{r,s}E_0 + \tau_{r,s}(e^{ik\cdot(s-r)} - 1) \tag{7.4}$$

By recalling $\Omega(k) = \alpha(k)^{-1}\Phi_{o,o}(k)$, we find that (7.4) yields

$$h_{\text{hop}}[\Omega(k), \Omega(k)] = E_0 \tag{7.5}$$

$$h_{\text{hop}}[\Omega(k), \Phi_{o,r}(k)] = \alpha(k)(e^{-ik\cdot\tau} - 1)\tau_{r,o} \tag{7.6}$$

and

$$h_{\text{hop}}[\Phi_{o,r}(k), \Omega(k)] = \frac{1}{\alpha(k)}(e^{ik\cdot r} - 1)\tau_{o,r} \tag{7.7}$$

7.2. Treatment of the Interaction Hamiltonian

Before calculating the matrix elements of the interaction Hamiltonian, we recall the representation (5.8), and decompose it as $H_{\text{int}} = H_{\text{int}}^{(1)} + H_{\text{int}}^{(2)}$ with

$$H_{\text{int}}^{(1)} = \sum_{y\in\Lambda_0}\sum_{v,w,z\in\Lambda} \tilde{U}_{y,v;w,z}a_{y,\uparrow}^\dagger a_{v,\downarrow}^\dagger b_{w,\downarrow}b_{z,\uparrow} \tag{7.8}$$

and

$$H_{\text{int}}^{(2)} = \sum_{y\in\Lambda'}\sum_{v,w,z\in\Lambda} \tilde{U}_{y,v;w,z}a_{y,\uparrow}^\dagger a_{v,\downarrow}^\dagger b_{w,\downarrow}b_{z,\uparrow} \tag{7.9}$$

Note that $H_{\text{int}}^{(1)}$ and $H_{\text{int}}^{(2)}$ are not Hermitian.

We apply $H_{\text{int}}^{(1)}$ to $\Phi_{u,r}(k)$ and simplify the expression by using the anticommutation relations (5.3), (5.4) to get

$$
\begin{aligned}
H_{\text{int}}^{(1)}\Phi_{u,r}(k) &= \sum_{\substack{y\in\Lambda_o \\ v,w,z\in\Lambda}} \tilde{U}_{y,v;w,z}a_{y,\uparrow}^\dagger a_{v,\downarrow}^\dagger b_{w,\downarrow}b_{z,\uparrow}\sum_{p\in\Lambda_o}e^{ik\cdot p}a_{p+u,\downarrow}^\dagger b_{p+r,\uparrow}\Phi_\uparrow \\
&= \sum_{\substack{v\in\Lambda \\ y,p\in\Lambda_o}} \tilde{U}_{y,v;u+p,y}e^{ik\cdot p}a_{v,\downarrow}^\dagger b_{p+r,\uparrow}\Phi_\uparrow \\
&\quad - \sum_{\substack{v\in\Lambda \\ z,p\in\Lambda_o}} \tilde{U}_{p+r,v;p+u,z}e^{ik\cdot p}a_{v,\downarrow}^\dagger b_{z,\uparrow}\Phi_\uparrow
\end{aligned}
\tag{7.10}
$$

We note that $v \in \Lambda$ can be uniquely decomposed as $v = x + u'$ with $x \in \Lambda_o$ and $u' \in \mathcal{U}$. We further make the change of variables $p = x + w - r$,

$y = x + s - r$ (with $w, s \in \Lambda_o$) in the first term [in the right-hand side of (7.10)], and the change of variables $z = x + w$, $p = x - s$ (with $w, s \in \Lambda_o$) in the second term. Then we get

$$H_{int}^{(1)} \Phi_{u,r}(k)$$

$$= \sum_{\substack{u' \in \mathscr{U} \\ x,s,w \in \Lambda_o}} \tilde{U}_{x+s-r,x+u';x+u+w-r,x+s-r} e^{ik \cdot (x+w-r)} a_{x+u',\downarrow}^{\dagger} b_{x+w,\uparrow} \Phi_{\uparrow}$$

$$- \sum_{\substack{u' \in \mathscr{U} \\ x,s,w \in \Lambda_o}} \tilde{U}_{x+r-s,x+u';x+u-s,x+w} e^{ik \cdot (x-s)} a_{x+u',\downarrow}^{\dagger} b_{x+w,\uparrow} \Phi_{\uparrow}$$

$$= \sum_{\substack{u' \in \mathscr{U} \\ s,w \in \Lambda_o}} (\tilde{U}_{s,u'+r;u+w,s} e^{ik \cdot (w-r)} - \tilde{U}_{r,u'+s;u,w+s} e^{-ik \cdot s}) \Phi_{u',w}(k) \quad (7.11)$$

where we used the translation invariance of $\tilde{U}_{y,v;w,z}$.

We again define the matrix elements $h_{int}[\Psi, \Phi]$ by the unique expansion

$$H_{int} \Phi = \sum_{\Psi \in \mathscr{B}_k} h_{int}[\Psi, \Phi] \Psi \quad (7.12)$$

Then we can read off from (7.11) that

$$h_{int}[\Phi_{u',w}(k), \Phi_{u,r}(k)]$$

$$= \sum_{s \in \Lambda_o} (\tilde{U}_{s,u'+r;u+w,s} e^{ik \cdot (w-r)} - \tilde{U}_{r,u'+s;u,w+s} e^{-ik \cdot s}) \quad (7.13)$$

By setting $u' = w = o$ in (7.13), we get

$$h_{int}[\Omega(k), \Phi_{u,r}(k)] = \alpha(k) \sum_{s \in \Lambda_o} (e^{-ik \cdot r} - e^{-ik \cdot s}) \tilde{U}_{s,r;u,s} \quad (7.14)$$

Next, we set $u = r = o$ in (7.3) to get

$$h_{int}[\Phi_{u',w}(k), \Omega(k)] = \frac{1}{\alpha(k)} \sum_{s \in \Lambda_o} (\tilde{U}_{s,u';w,s} e^{ik \cdot w} - \tilde{U}_{o,u'+s;o,w+s} e^{-ik \cdot s})$$

$$= \frac{1}{\alpha(k)} \sum_{s \in \Lambda_o} (e^{ik \cdot w} - e^{ik \cdot s}) \tilde{U}_{u',s;s,w} \quad (7.15)$$

where in the second term we used the translation invariance and the symmetry as

$$\tilde{U}_{o,u'+s;o,w+s} = \tilde{U}_{-s,u';-s,w} = \tilde{U}_{u',-s;-s,w}$$

and then replaced $s \to -s$. Finally we set $u = u' = r = w = o$ in (7.13) to get

$$
h_{\text{int}}[\Omega(k), \Omega(k)] = \sum_{s \in \Lambda_o} (1 - e^{-ik \cdot s})\, \tilde{U}_{s,o;s,o}
$$

$$
= \tfrac{1}{2} \sum_{s \in \Lambda_o} (2 - e^{-ik \cdot s} - e^{ik \cdot s})\, \tilde{U}_{s,o;s,o}
$$

$$
= 2 \sum_{s \in \Lambda_o} \left(\sin \frac{k \cdot s}{2} \right)^2 \tilde{U}_{s,o;s,o}
\tag{7.16}
$$

where we used $\tilde{U}_{s,o;s,o} = \tilde{U}_{o,-s;o,-s} = \tilde{U}_{-s,o;-s,o}$, which follows from the translation invariance and the symmetry of $\tilde{U}_{y,v;w,z}$. Note that we do not assume any reflection invariance.

We are now ready to prove some of the expressions in Lemma 6.2. The expression (6.13) follows by summing (7.5) and (7.16), the expression (6.14) follows by summing (7.6) and (7.14), and the expression (6.16) follows by summing (7.7) and (7.15).

We next calculate the action of (7.9) as

$$
H^{(2)}_{\text{int}} \Phi_{o,o}(k) = \sum_{\substack{y \in \Lambda' \\ v,w,z \in \Lambda}} \tilde{U}_{y,v;w,z} a^\dagger_{y,\uparrow} a^\dagger_{v,\downarrow} b_{w,\downarrow} b_{z,\uparrow} \sum_{p \in \Lambda_o} e^{ik \cdot p} a^\dagger_{p,\downarrow} b_{p,\uparrow} \Phi_\uparrow
$$

$$
= - \sum_{\substack{y \in \Lambda' \\ v \in \Lambda \\ p,z \in \Lambda_o}} \tilde{U}_{u,v;p,z} e^{ik \cdot p} a^\dagger_{v,\downarrow} a^\dagger_{y,\uparrow} b_{z,\uparrow} b_{p,\uparrow} \Phi_\uparrow
\tag{7.17}
$$

In the final expression, we note that the summand is vanishing for $p = z$, and decompose the sum over p, z as

$$
\sum_{\substack{p,z \in \Lambda_o \\ p \ne z}} (\cdots) = \sum_{\substack{p,z \in \Lambda_o \\ p-z \in \Lambda_o^+}} (\cdots) + \sum_{\substack{p,z \in \Lambda_o \\ z-p \in \Lambda_o^+}} (\cdots)
\tag{7.18}
$$

where Λ_o^+ is defined in (6.11). We then switch the variables z and p in the second sum to get

$$
H^{(2)}_{\text{int}} \Phi_{o,o}(k) = \sum_{\substack{p,z \in \Lambda_o \\ p-z \in \Lambda_o^+}} (e^{ik \cdot p} - e^{ik \cdot z}) \sum_{\substack{v \in \Lambda \\ y \in \Lambda'}} \tilde{U}_{y,v;p,z} a^\dagger_{v,\downarrow} a^\dagger_{y,\uparrow} b_{z,\uparrow} b_{p,\uparrow} \Phi_\uparrow
\tag{7.19}
$$

We write $v = x + u$ with $x \in \Lambda_o$ and $u \in \mathcal{U}$, and make the change of variables $y = x + r$, $p = x + s$, and $z = x + t$ (with $r \in \Lambda'$, s, $t \in \Lambda_o$ such that $s - t \in \Lambda_o^+$) to get

$$H^{(2)}_{int} \Phi_{o,o}(k)$$

$$= \sum_{\substack{u \in \mathscr{U} \\ r \in \Lambda' \\ s,t \in \Lambda_o \\ s-t \in \Lambda_o^+}} (e^{iks \cdot s} - e^{ik \cdot t}) \, \tilde{U}_{r,u;s,t} \sum_{x \in \Lambda_o} e^{ik \cdot x} a^\dagger_{x+u,\downarrow} a^\dagger_{x+r,\uparrow} b_{x+t,\uparrow} b_{x+s,\uparrow} \Phi_\uparrow$$

$$= \sum_{\substack{u \in \mathscr{U} \\ r \in \Lambda' \\ s,t \in \Lambda_o \\ s-t \in \Lambda_o^+}} (e^{ik \cdot s} - e^{ik \cdot t}) \, \tilde{U}_{r,u;s,t} \Phi_{u,r,t,s}(k) \tag{7.20}$$

This leads us to

$$h_{int}[\Phi_{u,r,t,s}(k), \Omega(k)] = \frac{1}{\alpha(k)} (e^{ik \cdot s} - e^{ik \cdot t}) \, \tilde{U}_{r,u;s,t} \tag{7.21}$$

which gives the desired expression (6.17) since there are no corresponding contributions from $\tilde{H}_{hop}$ or $H^{(1)}_{int}$.

To prove the only remaining expression (6.15), we calculate

$$H^{(1)}_{int} \Phi_{u,r,t,s}(k) = \sum_{\substack{v \in \Lambda \\ x \in \Lambda_o}} \tilde{U}_{x+t,v;x+u,x+r} e^{ik \cdot x} a^\dagger_{v,\downarrow} b_{x+s,\uparrow} \Phi_\uparrow$$

$$- \sum_{\substack{v \in \Lambda \\ x \in \Lambda_o}} \tilde{U}_{x+s,v;x+u,x+r} e^{ik \cdot x} a^\dagger_{v,\downarrow} b_{x+t,\uparrow} \Phi_\uparrow$$

$$+ (\text{other terms}) \tag{7.22}$$

where (other terms) do not contain any contributions to $\Phi_{o,o}(k)$. Since we are interested in calculating the matrix elements $h_{int}[\Phi_{o,o}(k), \Phi_{u,r,t,s}(k)]$, we shall pick up only those terms which have some contributions to $\Phi_{o,o}(k)$. This allows us to sum only over $v \in \Lambda_o$ instead of $v \in \Lambda$. We can also consider only x such that $x + s = v$ in the first term, and $x + t = v$ in the second term. Then we get

$$H^{(1)}_{int} \Phi_{u,r,t,s}(k) = \sum_{v \in \Lambda_o} \tilde{U}_{v+t-s,v;v+u-s,v+r-s} e^{ik \cdot (v-s)} a^\dagger_{v,\downarrow} b_{v,\uparrow} \Phi_\uparrow$$

$$- \sum_{v \in \Lambda_o} \tilde{U}_{v+s-t,v;v+u-t,v+r-t} e^{ik \cdot (v-t)} a^\dagger_{v,\downarrow} b_{v,\uparrow} \Phi_\uparrow$$

$$+ (\text{other terms})$$

$$= \tilde{U}_{s,t;u,r}(e^{-ik \cdot s} - e^{-ik \cdot t}) \, \Phi_{o,o}(k)$$

$$+ (\text{other terms}) \tag{7.23}$$

which implies

$$h_{\mathrm{int}}[\Omega(k), \Phi_{u,r,t,s}(k)] = \alpha(k)(e^{-ik\cdot s} - e^{-ik\cdot t})\, \tilde{U}_{s,t;u,r} \qquad (7.24)$$

Since there are no corresponding contributions from $\tilde{H}_{\mathrm{hop}}$ or $H_{\mathrm{int}}^{(1)}$, this gives the desired expression (6.15). This completes the proof of Lemma 6.2.

8. BOUNDS ON THE MATRIX ELEMENTS

Here we prove Lemmas 6.3 and 6.4, which state various bounds for the matrix elements and their sums. In the proofs we shall use the properties of the localized bases summarized in Lemmas 4.2–4.4, which will be proved in Section 10. In order to make use of these lemmas, we have to assume that $\lambda \geqslant \lambda_0$ and $|\rho|\,\lambda^{-2} \leqslant r_0$. The bound for λ is assumed in the statement of Lemma 6.3. In Lemma 6.4, we assumed the stronger condition $\lambda \geqslant \lambda_4$. (We will choose λ_4 so that $\lambda_4 \geqslant \lambda_0$.) The bound $|\rho|\,\lambda^{-2} \leqslant r_0$ follows from the assumption $|\rho| \leqslant \rho_0$ in Lemmas 6.3 and 6.4, since we shall now set $\rho_0 = (\lambda_0)^2 r_0$.

8.1. Bound for $h[\Omega(k), \Omega(k)]$

We first prove the lower bound (6.18) for $\mathrm{Re}[h[\Omega(k), \Omega(k)]]$. In fact we prove the stronger estimate

$$\left| h[\Omega(k), \Omega(k)] - \left(E_0 + \frac{U}{\lambda^4}\,G(k)\right) \right| \leqslant \frac{U}{\lambda^4}\left(C_1\,|\rho| + \frac{C_2}{\lambda}\right) G(k) \qquad (8.1)$$

which implies the desired (6.18).

With the goal (8.1) in mind, we will bound the quantity

$$\tilde{G}(k) = \frac{\lambda^4}{U}\,(h[\Omega(k), \Omega(k)] - E_0)$$

$$= 2\,\frac{\lambda^4}{U} \sum_{s \in \Lambda_o} \left(\sin\frac{k\cdot s}{2}\right)^2 \tilde{U}_{s,o;s,o}$$

$$= 2\lambda^4 \sum_{\substack{s \in \Lambda_o \\ x \in \Lambda}} \left(\sin\frac{k\cdot s}{2}\right)^2 \tilde{\varphi}_x^{(o)}(\varphi_x^{(o)})^* \, \tilde{\varphi}_x^{(s)}(\varphi_x^{(s)})^* \qquad (8.2)$$

where we used the expression (6.13) for the matrix element and the representation (5.9) for the effective interaction. Let us introduce

$$\eta_y^{(x)} = (\varphi_y^{(x)})^* - \psi_y^{(x)}, \qquad \tilde{\eta}_y^{(x)} = \tilde{\varphi}_y^{(x)} - \psi_y^{(x)} \qquad (8.3)$$

where $\psi_y^{(x)}$ is the localized basis state (4.23), (4.24) of the flat-band model. Then (8.2) can be written as

$$
\begin{aligned}
\tilde{G}(k) &= 2\lambda^4 \sum_{\substack{s \in A_o \\ x \in A}} \left(\sin \frac{k \cdot s}{2} \right)^2 \\
&\quad \times (\psi_x^{(o)} + \tilde{\eta}_x^{(o)})(\psi_x^{(o)} + \eta_x^{(o)})(\psi_x^{(s)} + \tilde{\eta}_x^{(s)})(\psi_x^{(s)} + \eta_x^{(s)}) \\
&= G_0(k) + G_1(k) + G_2(k) + G_3(k) + G_4(k)
\end{aligned} \tag{8.4}
$$

where $G_i(k)$ denotes the collection of terms which contain the ith power of ψ's when we expand the left-hand side. In the following, we shall control G_i for each $i = 0, 1, 2, 3$, and 4.

We first control $G_0(k)$. It gives the most dominant contribution as

$$
\begin{aligned}
G_0(k) &= 2\lambda^4 \sum_{\substack{s \in A_o \\ x \in A}} (\psi_x^{(o)})^2 (\psi_x^{(s)})^2 \left(\sin \frac{k \cdot s}{2} \right)^2 \\
&= 2 \sum_{f \in \mathscr{F}_o} \sum_{g \in \mathscr{F}_f} \left(\sin \frac{k \cdot (f+g)}{2} \right)^2 = G(k)
\end{aligned} \tag{8.5}
$$

where we used the expression (4.23) of $\psi_x^{(y)}$. See (3.8) and (3.9) for the definitions of $\mathscr{F}_o$ and $\mathscr{F}_f$.

We bound the absolute value of $G_1(k)$. One of the four terms in $G_1(k)$ is bounded as

$$
\begin{aligned}
&\left| 2\lambda^4 \sum_{\substack{s \in A_o \\ x \in A}} \left(\sin \frac{k \cdot s}{2} \right)^2 \tilde{\eta}_x^{(o)} \psi_x^{(o)} (\psi_x^{(s)})^2 \right| \\
&\leqslant 2\lambda^4 \frac{B_1 |\rho| + B_2}{\lambda^2} \frac{1}{\lambda^3} \sum_{f \in \mathscr{F}_o} \sum_{g \in \mathscr{F}_f} \left(\sin \frac{k \cdot (f+g)}{2} \right)^2 \\
&= \frac{B_1 |\rho| + B_2}{\lambda} G(k)
\end{aligned} \tag{8.6}
$$

where we used (4.36) to get the bound $|\tilde{\eta}_x^{(o)}| \leqslant (B_1 |\rho| + B_2)/\lambda^2$. The other three terms in $G_1(k)$ can be bounded similarly, and we get

$$
|G_1(k)| \leqslant \frac{4B_1 |\rho| + 2B_2}{\lambda} G(k) \tag{8.7}
$$

We bound the absolute value of $G_2(k)$. One of the six terms in $G_2(k)$ is bounded as

$$\left| 2\lambda^4 \sum_{\substack{s \in \Lambda_o \\ x \in \Lambda}} \left(\sin \frac{k \cdot s}{2} \right)^2 (\psi_x^{(o)})^2 \eta_x^{(s)} \tilde{\eta}_x^{(s)} \right|$$

$$\leqslant \lambda^4 \sum_{\substack{s \in \Lambda_o \\ x \in \Lambda}} (\psi_x^{(o)})^2 \frac{|k|^2}{2} |s| \cdot |\eta_x^{(s)}| \cdot |s| \cdot |\tilde{\eta}_x^{(s)}|$$

$$\leqslant \lambda^4 \frac{|k|^2}{2} \left(\sum_{s \in \Lambda_o} |s| \cdot |\eta_o^{(s)}| \cdot |s| \cdot |\tilde{\eta}_o^{(s)}| + \frac{1}{\lambda^2} \sum_{f \in \mathscr{F}_o} \sum_{s \in \Lambda_o} |s| \cdot |\eta_f^{(s)}| \cdot |s| \cdot |\tilde{\eta}_f^{(s)}| \right)$$

$$\leqslant \lambda^4 \frac{|k|^2}{2} \left(\sum_{s \in \Lambda_o} |s| \cdot |\eta_o^{(s)}| \right) \left(\sum_{s \in \Lambda_o} |s| \cdot |\tilde{\eta}_o^{(s)}| \right)$$

$$+ 2 |k|^2 \lambda^2 \sum_{f \in \mathscr{F}_o} \left(\sum_{s \in \Lambda_o} |s - f| \cdot |\eta_f^{(s)}| \right) \left(\sum_{s \in \Lambda_o} |s - f| \cdot |\tilde{\eta}_f^{(s)}| \right)$$

$$\leqslant |k|^2 \lambda^4 \left(\frac{1}{2} + \frac{2 |\mathscr{F}_o|}{\lambda^2} \right) \frac{B_1 R |\rho|}{\lambda^2} \frac{B_1 R |\rho| + B_2}{\lambda^2} \tag{8.8}$$

where we used $|\sin(k \cdot s/2)| \leqslant |k| \cdot |s|/2$, $|s| \leqslant 2 |s - f|$, and the bounds (4.28) and (4.39) on the summability of the basis states. Another term

$$\left| 2\lambda^4 \sum_{s \in \Lambda_o, x \in \Lambda} \left(\sin \frac{k \cdot s}{2} \right)^2 (\psi_x^{(s)})^2 \eta_x^{(o)} \tilde{\eta}_x^{(o)} \right|$$

can be bounded by the same quantity as in (8.8).

The remaining four terms in $G_2(k)$ have the common structure

$$\left| 2\lambda^4 \sum_{\substack{s \in \Lambda_o \\ x \in \Lambda}} \left(\sin \frac{k \cdot s}{2} \right)^2 \psi_x^{(o)} \psi_x^{(s)} \bar{\eta}_x^{(o)} \bar{\eta}_x^{(s)} \right|$$

$$\leqslant 2\lambda^2 \sum_{\substack{f \in \mathscr{F}_o \\ g \in \mathscr{F}_f}} |\bar{\eta}_f^{(o)}| \cdot |\bar{\eta}_f^{(f+g)}| \left(\sin \frac{k \cdot (f+g)}{2} \right)^2 \tag{8.9}$$

where $\bar{\eta}$ denotes either η or $\tilde{\eta}$. (The four terms are obtained by assigning η or $\tilde{\eta}$ with each $\bar{\eta}$.) We can bound $|\bar{\eta}_f^{(o)}|$ and $|\bar{\eta}_f^{(f+g)}|$ using (4.25) or (4.36),

depending on whether $\bar{\eta} = \eta$ or $\tilde{\eta}$. By summing the resulting bounds and (8.8), we get

$$|G_2(k)| \leqslant \left\{ \left(1 + \frac{2\,|\mathscr{F}_o|}{\lambda^2} \right) B_1\,R\,|\rho|\,(B_1\,R\,|\rho| + B_2) \right\} |k|^2$$

$$+ \lambda^2 \left(\frac{4B_1\,|\rho| + 2B_2}{\lambda^2} \right)^2 G(k)$$

$$\leqslant \left(C_1\,|\rho| + \frac{B_3}{\lambda^2} \right) G(k) \tag{8.10}$$

with constants C_1 and B_3 depending only on d, v, and R. Here we used the assumed bounds $\lambda \geqslant \lambda_0$ and $|\rho| \leqslant \rho_0$, as well as the bound (6.38) to bound $|k|^2$ by $G(k)$.

The quantities $G_3(k)$ and $G_4(k)$, which contain higher powers of η or $\tilde{\eta}$, can be bounded in a similar (in fact easier) manner, and we get

$$|G_3(k)| \leqslant \frac{B_4\,|\rho|}{\lambda^2}\,G(k) \tag{8.11}$$

and

$$|G_4(k)| \leqslant \frac{B_5}{\lambda^4}\,G(k) \tag{8.12}$$

with constants B_4 and B_5 which depend only on d, v, and R.

By summing up (8.5), (8.7), and (8.10)–(8.12), and comparing the result with (8.2) and (8.4), we finally get

$$|\tilde{G}(k) - G(k)| \leqslant \left(1 - \frac{4B_1\,|\rho| + 2B_2}{\lambda} - C_1\,|\rho| - \frac{B_3}{\lambda^2} - \frac{B_4\,|\rho|}{\lambda^2} - \frac{B_5}{\lambda^4} \right) G(k)$$

$$\leqslant \left(1 - C_1\,|\rho| - \frac{C_2}{\lambda} \right) G(k) \tag{8.13}$$

with a constant C_2 which depends only on d, v, and R. This is nothing but the desired (8.1).

8.2. Bound for $\sum |h[\Omega(k), \Psi]|$

We shall prove the bound (6.19) for the sum of the off-diagonal matrix elements $h[\Omega(k), \Psi]$ stated in Lemma 6.3. We first note that, since Ψ with

$h[\Omega(k), \Psi] \neq 0$ is either of the form $\Phi_{u,r}(k)$ of (6.8) or $\Phi_{u,r,t,s}(k)$ of (6.12), we can write the desired quantity as

$$\sum_{\Psi \in \mathscr{B}_k \backslash \{\Omega(k)\}} |h[\Omega(k), \Psi]|$$

$$= \sum_{\substack{u \in \mathscr{U} \\ r \in \Lambda_o \\ (u,r) \neq (o,o)}} |h[\Omega(k), \Phi_{u,r}(k)]| + \sum_{\substack{u \in \mathscr{U} \\ r \in \Lambda' \\ s,t \in \Lambda_o \\ (s-t \in \Lambda_o^+)}} |h[\Omega(k), \Phi_{u,r,t,s}(k)| \qquad (8.14)$$

To bound the first term in the right-hand side of (8.14), we use the expression (6.14) for the matrix element to get

$$\sum_{\substack{u \in \mathscr{U} \\ r \in \Lambda_o \\ (u,r) \neq (o,o)}} |h[\Omega(k), \Phi_{u,r}(k)]|$$

$$\leqslant \sum_{r \in \Lambda_o} \alpha(k) |(e^{-ik \cdot r} - 1) \, \tau_{r,o}|$$

$$+ \sum_{\substack{u \in \mathscr{U} \\ r,s \in \Lambda_o \\ (u,r) \neq (o,o)}} \alpha(k) |(e^{-ik \cdot r} - e^{-ik \cdot s}) \, \tilde{U}_{s,r;u,s}|$$

$$\leqslant \alpha(k) |k| \sum_{r \in \Lambda_o} |r| \cdot |\tau_{r,o}| + \alpha(k) |k| \sum_{\substack{u \in \mathscr{U} \\ r,s \in \Lambda_o \\ (u,r) \neq (o,o)}} |r-s| \cdot |\tilde{U}_{s,r;u,s}| \qquad (8.15)$$

The first term in the right-hand side is readily bounded by $\alpha(k) \, B_1 Rt \, |\rho| \cdot |k|$ from the summability (4.33) of $\tau_{x,y}$. To bound 'the second term, we use the representation (5.9) for $\tilde{U}_{s,r;u,s}$ and the bound $|r-s| \leqslant |r-x| + |s-x|$ to get

$$\sum_{\substack{u \in \mathscr{U} \\ r,s \in \Lambda_o \\ (u,r) \neq (o,o)}} |r-s| \cdot |\tilde{U}_{s,r;u,s}|$$

$$\leqslant U \sum_{\substack{r,s \in \Lambda_o \\ u \in \mathscr{U} \\ x \in \Lambda}} (|r-x| + |s-x|) \, |\tilde{\varphi}_x^{(s)} \tilde{\varphi}_x^{(r)}| \cdot |\varphi_x^{(u)} \varphi_x^{(s)}|$$

$$\leqslant U \sum_{\substack{x \in \Lambda \\ u \in \mathcal{U}}} \left\{ |\varphi_x^{(u)}| \left(\sum_{s \in \Lambda_o} |\tilde{\varphi}_x^{(s)}| \right) \left(\sum_{r \in \Lambda_o} |r - x| \cdot |\tilde{\varphi}_x^{(r)}| \right) \left(\sum_{s \in \Lambda_o} |\varphi_x^{(s)}| \right) \right\}$$

$$+ U \sum_{\substack{x \in \Lambda \\ u \in \mathcal{U}}} \left\{ |\varphi_x^{(u)}| \left(\sum_{s \in \Lambda_o} |s - x| \cdot |\tilde{\varphi}_x^{(s)}| \right) \left(\sum_{r \in \Lambda_o} |\tilde{\varphi}_x^{(r)}| \right) \left(\sum_{s \in \Lambda_o} |\varphi_x^{(s)}| \right) \right\}$$

$$= 2U \sum_{\substack{x \in \Lambda \\ u \in \mathcal{U}}} \left\{ |\varphi_x^{(u)}| \left(\sum_{s \in \Lambda_o} |s - x| \cdot |\tilde{\varphi}_x^{(s)}| \right) \left(\sum_{s \in \Lambda_o} |\tilde{\varphi}_x^{(s)}| \right) \left(\sum_{s \in \Lambda_o} |\varphi_x^{(s)}| \right) \right\}$$

$$= 2U \sum_{\substack{x \in \Lambda' \\ u \in \mathcal{U}}} \{ \cdots \} + 2U \sum_{\substack{x \in \Lambda_o \\ u \in \mathcal{U}}} \{ \cdots \}$$

$$\leqslant 2U \sum_{\substack{x \in \Lambda' \\ u \in \mathcal{U}}} |\varphi_x^{(u)}| \left(2^v \frac{\sqrt{v}}{2\lambda} + \frac{B_1 R |\rho| + B_2}{\lambda^2} \right)$$

$$\times \left(\frac{2^v}{\lambda} + \frac{B_1 |\rho| + B_2}{\lambda^2} \right) \left(\frac{2^v}{\lambda} + \frac{B_1 |\rho|}{\lambda^2} \right)$$

$$+ 2U \sum_{\substack{x \in \Lambda_o \\ u \in \mathcal{U}}} |\varphi_x^{(u)}| \left(\frac{B_1 R |\rho| + B_2}{\lambda^2} \right) \left(1 + \frac{B_1 |\rho| + B_2}{\lambda^2} \right) \left(1 + \frac{B_1 |\rho|}{\lambda^2} \right)$$

$$\leqslant B_6 \frac{U}{\lambda^2} \tag{8.16}$$

where the constant B_6 depends only on d, v, and R. We have used the expressions (4.23), (4.24) for $\psi_x^{(u)}$ and the bounds (4.27), (4.38), (4.39), and (4.25) for the sum of the basis states.

Next we bound the second term in the right-hand side of (8.14). We again use the expression (6.15) and the representation (5.9) to get

$$\sum_{\substack{u \in \mathcal{U} \\ r \in \Lambda' \\ s,t \in \Lambda_o \\ (s - t \in \Lambda_o^+)}} |h[\Omega(k), \Phi_{u,r,t,s}(k)]|$$

$$\leqslant \alpha(k) \sum_{\substack{u \in \mathcal{U} \\ r \in \Lambda' \\ s,t \in \Lambda_o}} |(e^{-ik \cdot s} - e^{-ik \cdot t}) \tilde{U}_{s,t;u,r}|$$

$$\leqslant \alpha(k) |k| U \sum_{\substack{x \in \Lambda \\ u \in \mathcal{U} \\ r \in \Lambda' \\ s,t \in \Lambda_o}} |s - t| \cdot |\tilde{\varphi}_x^{(s)} \tilde{\varphi}_x^{(t)}| \cdot |\varphi_x^{(u)} \varphi_x^{(r)}| \tag{8.17}$$

To bound the sum, we again use $|s-t| \leqslant |s-x| + |t-x|$, and the symmetry between s and t as we did in (8.16), to get

$$
\sum_{\substack{x \in \Lambda \\ u \in \mathscr{U} \\ r \in \Lambda' \\ s,t \in \Lambda_o}} |s-t| \cdot |\tilde{\varphi}_x^{(s)} \tilde{\varphi}_x^{(t)}| \cdot |\varphi_x^{(u)} \varphi_x^{(r)}|
$$

$$
\leqslant 2 \sum_{\substack{x \in \Lambda \\ u \in \mathscr{U}}} \left\{ |\varphi_x^{(u)}| \left(\sum_{s \in \Lambda_o} |s-x| \cdot |\tilde{\varphi}_x^{(s)}| \right) \left(\sum_{t \in \Lambda_o} |\tilde{\varphi}_x^{(t)}| \right) \left(\sum_{r \in \Lambda'} |\varphi_x^{(r)}| \right) \right\}
$$

$$
= 2 \sum_{\substack{x \in \Lambda' \\ u \in \mathscr{U}}} \{\cdots\} + 2 \sum_{\substack{x \in \Lambda_o \\ u \in \mathscr{U}}} \{\cdots\}
$$

$$
\leqslant 2 \sum_{\substack{x \in \Lambda' \\ u \in \mathscr{U}}} |\varphi_x^{(u)}| \left(2^\nu \frac{\sqrt{\nu}}{2\lambda} + \frac{B_1 R |\rho| + B_2}{\lambda^2} \right) \left(\frac{2^\nu}{\lambda} + \frac{B_1 |\rho| + B_2}{\lambda^2} \right) \left(1 + \frac{B_1 |\rho|}{\lambda^2} \right)
$$

$$
+ 2 \sum_{\substack{x \in \Lambda_o \\ u \in \mathscr{U}}} |\varphi_x^{(u)}| \left(\frac{B_1 R |\rho| + B_2}{\lambda^2} \right) \left(1 + \frac{B_1 R |\rho| + B_2}{\lambda^2} \right) \left(\frac{|\mathscr{F}_o|}{\lambda} + \frac{B_1 |\rho|}{\lambda^2} \right)
$$

$$
\leqslant B_7 \frac{1}{\lambda^2} \tag{8.18}
$$

By combining (8.14)–(8.18), we finally get the desired bound (6.19) with $C_3 = B_6 + B_7$.

8.3. Bounds for the Other Matrix Elements

Here we prove the bounds (6.20) and (6.21) stated in Lemma 6.3.

Instead of proving (6.20) for fixed $u \in \mathscr{U}$ and $r \in \Lambda_o$ with $(u, r) \neq (o, o)$, we prove the bound for their sum

$$
\sum_{\substack{u \in \mathscr{U} \\ r \in \Lambda_o \\ (u,r) \neq (o,o)}} |h[\Phi_{u,r}(k), \Omega(k)]| \leqslant \frac{1}{\alpha(k)} \left(B_1 R |\rho| t + \frac{C_4 U |\rho|}{\lambda^2} + \frac{C_5 U}{\lambda^3} \right) |k| \tag{8.19}
$$

which clearly implies the desired (6.20). By using the expression (6.16) for the matrix element, we have

$$\sum_{\substack{u \in \mathscr{U} \\ r \in \Lambda_o \\ (u,r) \neq (o,o)}} |h[\Phi_{u,r}(k), \Omega(k)]|$$

$$\leqslant \frac{1}{\alpha(k)} \sum_{r \in \Lambda_o} |(e^{ik \cdot r} - 1) \tau_{o,r}| + \frac{1}{\alpha(k)} \sum_{\substack{u \in \mathscr{U} \\ r,s \in \Lambda_o}} |(e^{ik \cdot r} - e^{ik \cdot s}) \tilde{U}_{u,s;s,r}|$$

$$\leqslant \frac{1}{\alpha(k)} B_1 R |\rho| t |k| + \frac{|k|}{\alpha(k)} \sum_{\substack{u \in \mathscr{U} \\ r,s \in \Lambda_o}} |r - s| \cdot |\tilde{U}_{u,s;s,r}| \tag{8.20}$$

where we used the summability (4.33) of $|\tau_{o,r}|$. The second sum can be treated in exactly the same manner as we did for the similar sum in (8.16). As a result, we get

$$\sum_{\substack{u \in \mathscr{U} \\ r,s \in \Lambda_0 \\ x \in \Lambda}} |r - s| \cdot |\tilde{U}_{u,s;s,r}| \leqslant U \sum_{\substack{u \in \mathscr{U} \\ r,s \in \Lambda_o \\ x \in \Lambda}} |r - s| \cdot |\tilde{\varphi}_x^{(u)} \tilde{\varphi}_x^{(s)}| \cdot |\varphi_x^{(s)} \varphi_x^{(r)}|$$

$$\leqslant 2U \sum_{\substack{x \in \Lambda \\ u \in \mathscr{U}}} \left\{ |\tilde{\varphi}_x^{(u)}| \left(\sum_{s \in \Lambda_o} |s - x| \cdot |\varphi_x^{(s)}| \right) \right.$$

$$\left. \times \left(\sum_{s \in \Lambda_o} |\varphi_x^{(s)}| \right) \left(\sum_{s \in \Lambda_o} |\tilde{\varphi}_x^{(s)}| \right) \right\} \tag{8.21}$$

This is the same as the fifth line in (8.16), except that φ and $\tilde{\varphi}$ are switched. Because of the drastic difference in the localization properties of the states φ and $\tilde{\varphi}$, this results in the remarkable difference between $\sum |h[\Omega(k), \Phi_{u,r}(k)]|$ and $\sum |h[\Phi_{u,r}(k), \Omega(k)]|$. Again by decomposing the sum over x as

$$\sum_{x \in \Lambda} \{\cdots\} = \sum_{x \in \Lambda'} \{\cdots\} + \sum_{x \in \Lambda_o} \{\cdots\}$$

and using the expression (4.23) for $\psi^{(u)}$ and the bounds (4.27), (4.28), and (4.38) for the sum of the basis states, we can further bound (8.21) as

$$\sum_{\substack{u \in \mathscr{U} \\ r,s \in \Lambda_o \\ x \in \Lambda}} |r - s| \cdot |\tilde{U}_{u,s;s,r}|$$

$$\leqslant 2U \sum_{\substack{x \in \Lambda' \\ u \in \mathscr{U}}} |\varphi_x^{(u)}| \left(2^\nu \frac{\sqrt{\nu}}{2\lambda} + \frac{B_1 |\rho|}{\lambda^2} \right)$$

$$\times \left(\frac{2^v}{\lambda} + \frac{B_1 \, |\rho|}{\lambda^2} \right) \left(\frac{2^v}{\lambda} + \frac{B_1 \, |\rho| + B_2}{\lambda^2} \right)$$

$$+ 2U \sum_{\substack{x \in A_0 \\ u \in \mathcal{U}}} |\varphi_x^{(u)}| \left(\frac{B_1 R \, |\rho|}{\lambda^2} \right) \left(1 + \frac{B_1 \, |\rho|}{\lambda^2} \right) \left(1 + \frac{B_1 \, |\rho| + B_2}{\lambda^2} \right)$$

$$\leqslant \frac{C_4 \, U \, |\rho|}{\lambda^2} + \frac{C_5 \, U}{\lambda^3} \tag{8.22}$$

where C_4 and C_5 are constants. The desired (8.19) follows from (8.20) and (8.22).

Next we show the bound (6.21) for $h[\Phi_{u,r,t,s}(k), \Omega(k)]$. It is done in a similar manner to the way we bounded $h[\Omega(k), \Phi_{u,r,t,s}(k)]$ in (8.17) and (8.18). From (6.15) and (5.9), we have

$$|h[\Phi_{u,r,t,s}(k), \Omega(k)]|$$

$$\leqslant \frac{1}{\alpha(k)} |e^{ik \cdot s} - e^{ik \cdot t}| \cdot |\tilde{U}_{s,t;u,r}|$$

$$\leqslant \frac{|k| \, U}{\alpha(k)} \sum_{x \in A} |s - t| \cdot |\tilde{\varphi}_x^{(s)} \tilde{\varphi}_x^{(t)}| \cdot |\varphi_x^{(u)} \varphi_x^{(r)}|$$

$$\leqslant \frac{2 \, |k| \, U}{\alpha(k)} \sum_{x \in A} |\varphi_x^{(u)}| \cdot |\varphi_x^{(r)}| \left(\sum_{s \in A_0} |s - x| \cdot |\tilde{\varphi}_x^{(s)}| \right) \left(\sum_{t \in A_0} |\tilde{\varphi}_x^{(t)}| \right)$$

$$\leqslant \frac{2 \, |k| \, U}{\alpha(k)} \sum_{x \in A'} |\varphi_x^{(u)}| \cdot |\varphi_x^{(r)}| \left(2^v \frac{\sqrt{v}}{2\lambda} + \frac{B_1 \, |\rho| + B_2}{\lambda^2} \right) \left(\frac{2^v}{\lambda} + \frac{B_1 \, |\rho| + B_2}{\lambda_2} \right)$$

$$+ \frac{2 \, |k| \, U}{\alpha(k)} \sum_{x \in A_0} |\varphi_x^{(u)}| \cdot |\varphi_x^{(r)}| \left(\frac{B_1 \, |\rho| + B_2}{\lambda^2} \right) \left(1 + \frac{B_1 \, |\rho| + B_2}{\lambda^2} \right)$$

$$\leqslant \frac{1}{\alpha(k)} \frac{C_6 \, U}{\lambda^2} |k| \tag{8.23}$$

which is the desired (6.21).

This completes the proof of Lemma 6.3.

8.4. Proof of Lemma 6.4

We shall prove Lemma 6.4, which controls the sum $\tilde{D}[\Psi_{u,A}(k)]$ of the matrix elements. We recall that the assumptions for this lemma are different from those for Lemma 6.3.

By using the representation (5.8) of the interaction Hamiltonian and the definition (5.26) of the basis state $\Psi_{u,A}(k)$, we find

$$H_{\text{int}} \Psi_{u,A}(k)$$

$$= \sum_{x \in A_o} e^{ik \cdot x} T_x \left[\sum_{y,v,z \in A} \tilde{U}_{y,v;u,z} a^\dagger_{y,\uparrow} a^\dagger_{v,\downarrow} b_{u,\downarrow} b_{z,\uparrow} a^\dagger_{u,\downarrow} \left(\prod_{t \in A} a^\dagger_{t,\uparrow} \right) \Phi_{\text{vac}} \right]$$

$$= \sum_{x \in A_o} e^{ik \cdot x} T_x \left[\sum_{y,v,z \in A} \tilde{U}_{y,v;u,z} \,\text{sgn}[\, y, z; A\,] \, a^\dagger_{v,\downarrow} \left(\prod_{t \in A_{z \to y}} a^\dagger_{t,\uparrow} \right) \Phi_{\text{vac}} \right]$$

$$= \sum_{y,v,z \in A} \text{sgn}[\, y, z; A\,] \, \tilde{U}_{y,v;u,z} \Psi_{v,A_{z \to y}}(k) \tag{8.24}$$

where we have used the translation invariance of $\tilde{U}_{y,v;u,z}$. The set $A_{z \to y}$ is obtained by replacing the sire z in A with y, and $\text{sgn}[\, y, z; A\,] = \pm 1$ comes from the reordering of the fermion oprators. The matrix element $h_{\text{int}}[\, \Psi_{u',A'}(k),$ $\Psi_{u,A}(k)\,]$ can be (in principle) obtained from (8.24) if we take into account the identification (5.27) between the basis states and rewrite $\Psi_{v,A_{z \to y}}(k)$ in terms of some $\Psi_{u',A'}(k) \in \mathcal{B}_k$. But here we take a slightly different strategy.

By $\tilde{h}_{\text{int}}[\cdots, \cdots]$ let us denote the pseudo matrix elements which are directly read off from (8.24) without taking into account the identification (5.27). We immediately find from (8.24) that

$$\tilde{h}_{\text{int}}[\, \Psi_{v,A_{z \to y}}(k), \Psi_{u,A}(k)\,] = \text{sgn}[\, y, z; A\,] \, \tilde{U}_{y,v;u,z} \tag{8.25}$$

and, by a suitable replacement of symbols, that

$$\tilde{h}_{\text{int}}[\, \Psi_{u,A}(k), \Psi_{v,A_{z \to y}}(k)\,] = \text{sgn}[\, y, z; A\,] \, \tilde{U}_{u,z;y,v} \tag{8.26}$$

Since some of the diagonal elements in the true matrix elements $h_{\text{int}}[\cdots, \cdots]$ are treated as off-diagonal elements in the pseudo matrix elements $\tilde{h}_{\text{int}}[\cdots, \cdots]$, we observe that

$$\sum_{\Phi \in \mathcal{B}_k \setminus \{ \Psi_{u,A}(k), \Omega(k) \}} |h_{\text{int}}[\, \Psi_{u,A}(k), \Phi\,]|$$

$$\leqslant \sum_{\Phi \in \mathcal{B}_k \setminus \{ \Psi_{u,A}(k), \Omega(k) \}} |\tilde{h}_{\text{int}}[\, \Psi_{u,A}(k), \Phi\,]| \tag{8.27}$$

and

$$\text{Re}[\, h_{\text{int}}[\, \Psi_{u,A}(k), \Psi_{u,A}(k)\,]\,]$$

$$\geqslant \text{Re}[\, \tilde{h}_{\text{int}}[\, \Psi_{u,A}(k), \Psi_{u,A}(k)\,]\,]$$

$$- \sum_{\Phi \in \mathcal{B}_k \setminus \{ \Psi_{u,A}(k), \Omega(k) \}} |\tilde{h}_{\text{int}}[\, \Psi_{u,A}(k), \Phi\,]| \tag{8.28}$$

for any $u \in \mathcal{U}$, $A \subset \Lambda$ with $|A| = L^d - 1$ such that $(u, A) \neq (o, \bar{\Lambda}_o)$, where $\bar{\Lambda}_o = \Lambda_o \backslash \{o\}$. From (8.27) and (8.28), we can bound the contribution to $\tilde{D}[\Psi_{u,A}(k)]$ of (6.22) from the interaction Hamiltonian as

$$\tilde{D}_{\mathrm{int}}[\Psi_{u,A}(k)]$$

$$= \mathrm{Re}[h_{\mathrm{int}}[\Psi_{u,A}(k), \Psi_{u,A}(k)]] - \sum_{\Phi \in \mathscr{B}_k \backslash \{\Psi_{u,A}(k), \Omega(k)\}} |h_{\mathrm{int}}[\Psi_{u,A}(k), \Phi]|$$

$$\geqslant \mathrm{Re}[\tilde{h}_{\mathrm{int}}[\Psi_{u,A}(k), \Psi_{u,A}(k)]] - 2 \sum_{\Phi \in \mathscr{B}_k \backslash \{\Psi_{u,A}(k), \Omega(k)\}} |\tilde{h}_{\mathrm{int}}[\Psi_{u,A}(k), \Phi]|$$

$$(8.29)$$

for any $(u, A) \neq (o, \bar{\Lambda}_o)$.

By using (8.26), we can evaluate the sum in the right-hand side of (8.29) as

$$\sum_{\Phi \in \mathscr{B}_k \backslash \{\Psi_{u,A}(k), \Omega(k)\}} |\tilde{h}_{\mathrm{int}}[\Psi_{u,A}(k), \Phi]|$$

$$\leqslant \sum_{\substack{y, v, z \in \Lambda \\ (v, y) \neq (u, z)}} |\tilde{U}_{u,z;y,v}|$$

$$\leqslant U \sum_{\substack{x, y, v, z \in \Lambda \\ (v, y) \neq (u, z)}} |\tilde{\varphi}_x^{(u)} \tilde{\varphi}_x^{(z)}| \cdot |\varphi_x^{(y)} \varphi_x^{(v)}| \qquad (8.30)$$

where we used the representation (5.9) for the effective interaction $\tilde{U}_{u,z;y,v}$. We further use the bounds (4.23), (4.24), (4.27), and (4.38) for the sum of the localized basis states to bound (8.30) as

$$\sum_{\Phi \in \mathscr{B}_k \backslash \{\Psi_{u,A}(k), \Omega(k)\}} |\tilde{h}_{\mathrm{int}}[\Psi_{u,A}(k), \Phi]|$$

$$\leqslant U \sum_{x \in \Lambda} \left\{ |\tilde{\varphi}_x^{(u)}| \left(\sum_{z \in \Lambda} |\tilde{\varphi}_x^{(z)}| \right) \left(\sum_{y \in \Lambda} |\varphi_x^{(y)}| \right) \left(\sum_{v \in \Lambda} |\varphi_x^{(v)}| \right) \right\}$$

$$- U(\tilde{\varphi}_u^{(u)})^2 (\varphi_u^{(u)})^2$$

$$\leqslant U \left(\max_{u' \in \mathcal{U}} \sum_{x \in \Lambda} |\tilde{\varphi}_x^{(u')}| \right)^2 \left(\max_{u' \in \mathcal{U}} \sum_{x \in \Lambda} |\varphi_x^{(u')}| \right)^2 - U(\tilde{\varphi}_u^{(u)})^2 (\varphi_u^{(u)})^2$$

$$\leqslant U \left\{ \left(1 + \frac{|\mathscr{F}_o|}{\lambda} + \frac{B_1 |\rho| + B_2}{\lambda^2} \right)^2 \left(1 + \frac{|\mathscr{F}_o|}{\lambda} + \frac{B_1 |\rho|}{\lambda^2} \right) \right.$$

$$\left. - \left(1 - \frac{B_1 |\rho| + B_2}{\lambda^2} \right)^2 \left(1 - \frac{B_1 |\rho|}{\lambda_2} \right)^2 \right\}$$

$$\leqslant B_8 \frac{U}{\lambda} \qquad (8.31)$$

where the constant B_8 depends only on d, v, and R. We have made use of the bounds $\lambda \geqslant \lambda_4$ and $|\rho| \leqslant \rho_0$.

By using (8.25) and the representation (5.9) for $\tilde{U}_{u,y;u,y}$, we write the diagonal element of $\tilde{h}_{\text{int}}[\cdots, \cdots]$ as

$$\tilde{h}_{\text{int}}[\Psi_{u,A}(k), \Psi_{u,A}(k)] = \sum_{y \in A} \tilde{U}_{u,y;u,y}$$

$$= U \sum_{\substack{x \in A \\ y \in A}} \tilde{\varphi}_x^{(u)} \tilde{\varphi}_x^{(y)} (\varphi_x^{(u)} \varphi_x^{(y)})^* \qquad (8.32)$$

Again by using the properties (4.27) and (4.38) of the basis states, we have

$$\text{Re}[\tilde{h}_{\text{int}}[\Psi_{u,A}(k), \Psi_{u,A}(k)]]$$

$$\geqslant \chi[u \in A] \, U \, \text{Re}[(\tilde{\varphi}_u^{(u)})^2 \{(\varphi_u^{(u)})^*\}^2]$$

$$- U \sum_{\substack{x \in A \\ y \in A \\ (x,y) \neq (u,u)}} |\tilde{\varphi}_x^{(u)} \tilde{\varphi}_x^{(y)}| \cdot |\varphi_x^{(u)} \varphi_x^{(y)}|$$

$$\geqslant \chi[u \in A] \, U \, \text{Re}[(\tilde{\varphi}_u^{(u)})^2 \{(\varphi_u^{(u)})^*\}^2]$$

$$- U \left\{ \left(\sum_{x \in A} |\tilde{\varphi}_x^{(u)}| \right) \left(\max_{x' \in A} \sum_{y \in A} |\tilde{\varphi}_{x'}^{(y)}| \right) \left(\sum_{x \in A} |\varphi_x^{(u)}| \right) \right.$$

$$\left. \times \left(\max_{x' \in A} \sum_{y \in A} |\varphi_{x'}^{(y)}| \right) - |\tilde{\varphi}_u^{(u)}|^2 \, |\varphi_u^{(u)}|^2 \right\}$$

$$\geqslant \chi[u \in A] \, U - B_9 \frac{U}{\lambda} \qquad (8.33)$$

where $\chi[\cdots]$ is the indicator function with $\chi[\text{true event}] = 1$ and $\chi[\text{false event}] = 0$.

Substituting (8.31) and (8.33) into (8.29), we get

$$\tilde{D}_{\text{int}}[\Psi_{u,A}(k)] \geqslant \chi[u \in A] \, U - B_{10} \frac{U}{\lambda} \qquad (8.34)$$

Next we examine the matrix elements of the modified hopping Hamiltonian $\tilde{H}_{\text{hop}}$. By using the representation (5.7) and the definition (5.26), we get

$$\tilde{H}_{\text{hop}}\,\Psi_{u,A}(k)$$

$$= \tfrac{3}{4}\lambda^2 t \,|(A \cup \{u\}) \cap \Lambda'|\, \Psi_{u,A}(k)$$

$$+ \sum_{x \in \Lambda_o} e^{ik \cdot x} T_x \left[\sum_{\substack{y,z \in \Lambda_o \\ \sigma = \uparrow,\downarrow}} \tau_{y,z} a_{y,\sigma}^\dagger b_{z,\sigma} a_{u,\downarrow}^\dagger \left(\prod_{l \in A} a_{l,\uparrow}^\dagger \right) \Phi_{\text{vac}} \right]$$

$$= (\varepsilon_0 \,|(A \cup \{u\}) \cap \Lambda_o| + \tfrac{3}{4}\lambda^2 t \,|(A \cup \{u\}) \cap \Lambda'|)\, \Psi_{u,A}(k)$$

$$+ \chi[u \in \Lambda_o] \sum_{y \in \Lambda_o \setminus \{u\}} \tau_{y,u}\Psi_{y,A}(k) + \sum_{\substack{z \in A \cap \Lambda_o \\ y \in \Lambda_o \setminus A}} \tau_{y,z}\,\text{sgn}[\,y, z;\, A\,]\, \Psi_{u,A_{z\rightarrow y}}(k) \tag{8.35}$$

where we wrote $\varepsilon_0 = \tau_{y,y}$ for $y \in \Lambda_o$. Note that $\chi[u \in \Lambda_o] = \delta_{u,o}$ as long as $u \in \mathcal{U}$.

From (8.35), we can read off the matrix elements of $\tilde{H}_{\text{hop}}$ as

$$h_{\text{hop}}[\,\Psi_{u,A}(k),\, \Psi_{u,A}(k)\,] = \varepsilon_0 \,|(A \cup \{u\}) \cap \Lambda_o| + \tfrac{3}{4}\lambda^2 t \,|(A \cup \{u\}) \cap \Lambda'| \tag{8.36}$$

$$h_{\text{hop}}[\,\Psi_{u,A}(k),\, \Psi_{u,A_{z\rightarrow y}}(k)\,] = \text{sgn}[\,y, z;\, A\,]\, \chi[z, y \in \Lambda_o]\, \tau_{z,y} \tag{8.37}$$

and

$$h_{\text{hop}}[\,\Psi_{o,A}(k),\, \Psi_{y,A}(k)\,] = \tau_{o,y} \tag{8.38}$$

where $\Psi_{y,A}(k)$ in (8.38) should be properly interpreted as a state in $\mathcal{B}_k$ using the identification (5.27). We did not define pseudo matrix elements here since $h_{\text{hop}}[\,\Psi_{o,A}(k),\, \Psi_{y,A}(k)\,]$ does not contain any diagonal elements.

Let us use (8.37) and (8.38) to evaluate the sum of the off-diagonal matrix elements as

$$\sum_{\Phi \in \mathcal{B}_k \setminus \{\Psi_{u,A}(k),\, \Omega(k)\}} |h_{\text{hop}}[\,\Psi_{u,A}(k),\, \Phi\,]|$$

$$\leqslant \sum_{\substack{z \in A \cap \Lambda_o \\ y \in \Lambda_o \setminus A}} |\tau_{z,y}| + \delta_{u,o} \sum_{y \in \Lambda_o \setminus \{o\}} |\tau_{o,y}|$$

$$\leqslant (|\Lambda_o \setminus A| + \delta_{u,o})\, B_1 t \,|\rho|$$

$$= (|A \cap \Lambda'| + 1 + \delta_{u,o})\, B_1 t \,|\rho| \tag{8.39}$$

where we used the bound (4.32) for the sum of the effective hopping $\tau_{z,y}$. The identity $|\Lambda_o \setminus A| = |A \cap \Lambda'| + 1$ follows from $|A| + 1 = |\Lambda_o| = L^d$.

By combining (8.36) and (8.39), we can evaluate the contribution of $\tilde{D}[\Psi_{u,A}(k)]$ from the hopping Hamiltonian as

$$\tilde{D}_{\text{hop}}[\Psi_{u,A}(k)] = h_{\text{hop}}[\Psi_{u,A}(k), \Psi_{u,A}(k)]$$

$$- \sum_{\Phi \in \mathscr{B}_k \setminus \{\Psi_{u,A}(k), \Omega(k)\}} |h_{\text{hop}}[\Psi_{u,A}(k), \Phi]|$$

$$\geq \varepsilon_0(L^d - 1 - |A \cap A'| + \delta_{u,o}) + \tfrac{3}{4}\lambda^2 t(|A \cap A'| + 1 - \delta_{u,o})$$

$$- (|A \cap A'| + 1 + \delta_{u,o}) B_1 t |\rho| \qquad (8.40)$$

By summing up the contributions (8.34) and (8.40) from H_{int} and $\tilde{H}_{\text{hop}}$, respectively, we can finally bound the desired quantity $\tilde{D}[\Psi_{u,A}(k)]$ (6.22) as

$$\tilde{D}[\Psi_{u,A}(k)] = \tilde{D}_{\text{int}}[\Psi_{u,A}(k)] + \tilde{D}_{\text{hop}}[\Psi_{u,A}(k)]$$

$$\geq E_0 + \chi[u \in A] U - B_{10} \frac{U}{\lambda}$$

$$+ \left(\frac{3}{4}\lambda^2 t - \varepsilon_0 - B_1 t |\rho|\right)(|A \cap A'| + 1)$$

$$+ \left(-\frac{3}{4}\lambda^2 t + \varepsilon_0 - B_1 t |\rho|\right)\delta_{u,o} \qquad (8.41)$$

where we noted that $E_0 = \sum_{x \in A_o} \tau_{x,x} = L^d \varepsilon_0$. See (5.12).

The desired bounds (6.23)–(6.25) are derived by investigating the bound (8.41) in each situation. We first consider the case $A \cap A' \neq \varnothing$. Noting that $|A \cap A'| \geq 1$, $\delta_{u,o} \leq 1$, $\chi[u \in A] \geq 0$, and $\varepsilon_0 \leq B_1 t |\rho|$, we find from the basic bound (8.41) that

$$\tilde{D}[\Psi_{u,A}(k)] \geq E_0 - B_{10} \frac{U}{\lambda} + \frac{3}{4}\lambda^2 t - 4B_1 t |\rho|$$

$$\geq E_0 + \frac{1}{2}\lambda^2 t \qquad (8.42)$$

which is the desired bound (6.23). To get the final inequality, we have here assumed that

$$B_{10} \frac{U}{\lambda} \leq \frac{1}{8}\lambda^2 t, \qquad 4B_1 t |\rho| \leq \frac{1}{8}\lambda^2 t \qquad (8.43)$$

We then turn to the case $A \cap A' = \varnothing$. Then the state $\Psi_{u,A}(k)$ is nothing but the state $\Phi_{u,r}(k)$ defined in (6.8). When $u \neq o$, the basic bound (8.41) with $|A \cap A'| = 0$ and $\delta_{u,o} = 0$ yields

$$\tilde{D}[\Phi_{u,r}(k)] \geqslant E_0 - B_{10} \frac{U}{\lambda} + \frac{3}{4} \lambda^2 t - 2B_1 t |\rho|$$

$$\geqslant E_0 + \frac{1}{2} \lambda^2 t \tag{8.44}$$

which is the desired bound (6.24). We again used (8.43).

Finally, when $u = o$ and $r \neq o$, we find that $\chi[u \in A] = 1$, since the state $\varphi^{(o)}$ is doubly occupied. Thus the basic bound (8.41) yields

$$\tilde{D}[\Psi_{o,r}(k)] \geqslant E_0 + U - B_{10} \frac{U}{\lambda} - 2B_1 t |\rho|$$

$$\geqslant E_0 + \frac{U}{2} \tag{8.45}$$

which is the desired (6.25). To get the final inequality, we have assumed

$$2B_1 t |\rho| \leqslant \frac{U}{4}, \qquad B_{10} \frac{U}{\lambda} \leqslant \frac{U}{2} \tag{8.46}$$

It only remains to examine the conditions for the parameters. We shall set $K_3 = 8B_1$, $K_4 = (8B_{10})^{-1}$, and

$$\lambda_4 = \max\{\lambda_0, 2B_{10}, \sqrt{32B_1\rho_0}\} \tag{8.47}$$

and make the requirements as in the statement of Lemma 6.4. Then the conditions (8.43) and (8.46) are easily checked to be satisfied. Lemma 6.4 has been proved.

9. UPPER BOUND FOR THE SPIN-WAVE ENERGY

We will here prove Theorem 3.5, which states the upper bound (3.35) for the energy $E_{SW}(k)$ for the elementary spin-wave excitation with the wavenumber vector $k \in \mathscr{K}$. In contrast to the corresponding lower bound, the upper bound can be proved by employing the standard variational argument. The new idea here is to use the state $\Omega(k)$ of (5.28) as a trial state. In the proof, we shall make use of Lemmas 4.2 and 4.4 about the localized basis states, and some estimates about the matrix elements proved

in Section 8 during the proof of Lemma 6.3. The assumption made in the statement of Theorem 3.5 guarantees that we can make use of these results. (See the beginning of Section 8.)

Since we have $\Omega(k) \in \mathcal{H}_k$, the lowest energy $E_{\mathrm{SW}}(k)$ in the space $\mathcal{H}_k$ satisfies the variational inequality

$$E_{\mathrm{SW}}(k) \leqslant \frac{(\Omega(k), H\Omega(k))}{(\Omega(k), \Omega(k))} \tag{9.1}$$

where $(\cdot, \cdot)$ denotes the inner product. Recalling the definitions of $\tilde{H}$ [see (4.21) and (6.1)] and the matrix elements (6.2), we can write

$H\Omega(k) = \tilde{H}\Omega(k)$

$$= h[\Omega(k), \Omega(k)]\,\Omega(k) + \sum_{\substack{u \in \mathcal{U} \\ r \in \Lambda_o \\ (u, r) \neq (o, o)}} h[\Phi_{u, r}(k), \Omega(k)]\,\Phi_{u, r}(k)$$

$$+ \sum_{\substack{u \in \mathcal{U} \\ r \in \Lambda' \\ s, t \in \Lambda_o \\ (s - t \in \Lambda_o^+)}} h[\Phi_{u, r, t, s}(k), \Omega(k)]\,\Phi_{u, r, t, s}(k) \tag{9.2}$$

By noting that $(\Phi_{u, r}(k), \Omega(k)) = 0$ if $r \neq o$, and $(\Phi_{u, r, t, s}(k), \Omega(k)) = 0$, we find

$$(\Omega(k), H\Omega(k)) = h[\Omega(k), \Omega(k)](\Omega(k), \Omega(k))$$

$$+ \sum_{\substack{r \in \Lambda_o \\ r \neq o}} h[\Phi_{o, r}(k), \Omega(k)](\Omega(k), \Phi_{o, r}(k)) \tag{9.3}$$

Recalling the definition (6.8) of $\Phi_{o, r}(k)$, and noting that $\Omega(k) = \alpha(k)^{-1}\Phi_{o, o}(k)$, we have

$$(\Phi_{o, o}(k), \Phi_{o, r}(k)) = \left(\sum_{x \in \Lambda_o} e^{ik \cdot x} a^\dagger_{x, \downarrow} b_{x, \uparrow} \Phi_\uparrow, \sum_{y \in \Lambda_o} e^{ik \cdot y} a^\dagger_{y, \downarrow} b_{y + r, \uparrow} \Phi_\uparrow \right)$$

$$= L^d \sum_{x \in \Lambda_o} e^{-ik \cdot x} (\Phi_\uparrow, b^\dagger_{x, \uparrow} a_{x, \downarrow} a^\dagger_{o, \downarrow} b_{r, \uparrow} \Phi_\uparrow) \tag{9.4}$$

where we made use of the translation invariance to replace y to o. Note that we have encountered the operators a and $b^\dagger$ for the first time in the

present paper. Going back to the definitions (5.1), (5.2), we get the anticommutation relations

$$\{a^\dagger_{x,\sigma}, a_{y,\tau}\} = (G)_{x,y} \delta_{\sigma,\tau} \tag{9.5}$$

and

$$\{b^\dagger_{x,\sigma}, b_{y,\tau}\} = (G^{-1})_{x,y} \delta_{\sigma,\tau} \tag{9.6}$$

where the Gramm matrix G is given by

$$(G)_{x,y} = \sum_{z \in \Lambda} (\varphi^{(x)}_z)^* \varphi^{(y)}_z \tag{9.7}$$

and its inverse is

$$(G^{-1})_{x,y} = \sum_{z \in \Lambda} (\tilde{\varphi}^{(x)}_z)^* \tilde{\varphi}^{(y)}_z \tag{9.8}$$

That (9.8) correctly defines the inverse of G can be easily verified by using the duality relations (4.34) and (4.35). The complicated anticommutation relations (9.5) and (9.6) are major drawbacks of the use of the non-orthogonal basis.

By using (9.5) and (9.6), we can further evaluate (9.4) as

$$(\Phi_{o,o}(k), \Phi_{o,r}(k)) = L^d \left(\sum_{x \in \Lambda_o} e^{-ik \cdot x} (G)_{o,x} (G^{-1})_{x,r} \right) (\Phi_\uparrow, \Phi_\uparrow) \tag{9.9}$$

As for the expectation value in the right-hand side of (9.1), we use (9.3) and (9.9) to get

$$\frac{(\Omega(k), H\Omega(k))}{(\Omega(k), \Omega(k))}$$

$$= h[\Omega(k), \Omega(k)]$$

$$\quad + \frac{\sum_{r \in \Lambda_o(r \neq o)} \alpha(k)^{-1} h[\Phi_{o,r}(k), \Omega(k)](\Phi_{o,o}(k), \Phi_{o,r}(k))}{\alpha(k)^{-2} (\Phi_{o,o}(k), \Phi_{o,o}(k))}$$

$$= h[\Omega(k), \Omega(k)]$$

$$\quad + \frac{\sum_{r \in \Lambda_o(r \neq o)} \alpha(k) h[\Phi_{o,r}(k), \Omega(k)] \sum_{x \in \Lambda_o} e^{-ik \cdot x} (G)_{o,x} (G^{-1})_{x,r}}{\sum_{x \in \Lambda_o} e^{-ik \cdot x} (G)_{o,x} (G^{-1})_{x,o}}$$

$$\tag{9.10}$$

Since the first term in the right-hand side is already controlled by the bound (8.1), we only need to bound the second term.

We start from the denominator of the final term in (9.10). By noting that $\sum_{x \in \Lambda_o} (G)_{o, x} (G^{-1})_{x, o} = 1$, we have

$$\left| 1 - \sum_{x \in \Lambda_o} e^{-ik \cdot x} (G)_{o, x} (G^{-1})_{x, o} \right|$$

$$= \left| \sum_{x \in \Lambda_o} (1 - e^{-ik \cdot x})(G)_{o, x} (G^{-1})_{x, o} \right|$$

$$\leqslant |k| \sum_{x \in \Lambda_o} |x| \cdot |(G)_{o, x} (G^{-1})_{x, o}|$$

$$\leqslant |k| \sum_{\substack{x \in \Lambda_o \\ y, z \in \Lambda}} (|x - y| + |y|) |\varphi_y^{(o)}| \cdot |\varphi_y^{(x)}| \cdot |\tilde{\varphi}_z^{(x)}| \cdot |\tilde{\varphi}_z^{(o)}|$$

$$\leqslant |k| \, 2 \left(\frac{B_1 R |\rho|}{\lambda^2} \right) \left(1 + \frac{|\mathscr{F}_o|}{\lambda} + \frac{B_1 |\rho|}{\lambda^2} \right) \left(1 + \frac{|\mathscr{F}_o|}{\lambda} + \frac{B_1 |\rho| + B_2}{\lambda^2} \right)^2$$

$$\leqslant B_{11} \frac{|\rho|}{\lambda^2} \tag{9.11}$$

where we used (4.23), (4.24), (4.26), (4.25), and (4.36). We also noted that $\lambda \geqslant \lambda_0$, $|\rho| \leqslant \rho_0$, and $|k| \leqslant \sqrt{d} \, \pi$. Thus we get

$$\left| \left(\sum_{x \in \Lambda_o} e^{-ik \cdot x} (G)_{o, x} (G^{-1})_{x, o} \right)^{-1} \right| \leqslant 1 + B_{12} \frac{|\rho|}{\lambda^2} \tag{9.12}$$

We now control the numerator of the final term in (9.10). By noting that $\sum_{x \in \Lambda_o} (G)_{o, x} (G^{-1})_{x, r} = 0$ for $r \neq o$, we get

$$\left| \sum_{r \in \Lambda_o \backslash \{o\}} \alpha(k) \, h[\Phi_{o, r}(k), \Omega(k)] \sum_{x \in \Lambda_o} e^{-ik \cdot x} (G)_{o, x} (G^{-1})_{x, r} \right|$$

$$= \left| \sum_{r \in \Lambda_o \backslash \{o\}} \alpha(k) \, h[\Phi_{o, r}(k), \Omega(k)] \sum_{x \in \Lambda_o} (e^{-ik \cdot x} - 1)(G)_{o, x} (G^{-1})_{x, r} \right|$$

$$\leqslant \alpha(k) \left(\sum_{r \in \Lambda_o \backslash \{o\}} |h[\Phi_{o, r}(k), \Omega(k)]| \right) |k| \left(\sum_{x, r \in \Lambda_o} |(G)_{o, x} (G^{-1})_{x, r}| \right)$$

$$\leqslant \left(B_1 Rt \, |\rho| + \frac{C_4 U |\rho|}{\lambda^2} + \frac{C_5 U}{\lambda^3} \right)$$

$$\times \left(\sum_{x, r \in \Lambda_o} |x| \cdot |(G)_{o, x} (G^{-1})_{x, r}| \right) |k|^2 \tag{9.13}$$

where we used (8.19) to control the sum of the matrix elements $h[\Phi_{o,r}(k), \Omega(k)]$. The remaining factor can be bounded as

$$
\sum_{x, r \in A_o} |x| \cdot |(G)_{o,x}(G^{-1})_{x,r}|
$$

$$
\leqslant \sum_{\substack{x, r \in A_o \\ y, z \in A}} (|x - y| + |y|) |\varphi_y^{(o)}| \cdot |\varphi_y^{(x)}| \cdot |\tilde{\varphi}_z^{(x)}| \cdot |\tilde{\varphi}_z^{(r)}|
$$

$$
\leqslant B_{13} \frac{|\rho|}{\lambda^2} \tag{9.14}
$$

By collecting (9.1), (9.10), and (9.12)–(9.14) and by using the bound (8.1) for the matrix element $h[\Omega(k), \Omega(k)]$ and the bound (6.38) for $G(k)$, we finally get

$$
E_{SW}(k) \leqslant E_0 + \frac{U}{\lambda^4} G(k) + \frac{U}{\lambda^4} \left(C_1 |\rho| + \frac{C_2}{\lambda} \right) G(k)
$$

$$
+ B_{14} \left(B_1 Rt |\rho| + \frac{C_4 U |\rho|}{\lambda^2} + \frac{C_5 U}{\lambda^3} \right) \frac{|\rho|}{\lambda^2} G(k)
$$

$$
\leqslant E_0 + \frac{U}{\lambda^4} \left(1 + \frac{A_4}{\lambda} + A_5 \lambda |\rho| + \frac{A_6 \lambda^2 t |\rho|^2}{U} \right) G(k) \tag{9.15}
$$

which is the desired (3.35).

10. CONSTRUCTION OF THE LOCALIZED BASES

In the present section, we shall explicitly construct the localized bases $\{\varphi^{(x)}\}_{x \in A}$ and $\{\tilde{\varphi}^{(x)}\}_{x \in A}$ and the dispersion relation $\varepsilon_1(k)$ and prove the summability stated in Lemmas 4.2–4.3.

The main problem treated here is a perturbation theory in the finite-dimensional eigenvalue problem (4.9), where the unperturbed problem has an energy gap. It is well established that such a finite-dimensional perturbation theory can be controlled in a perfectly rigorous manner.[19, 38]

However, there are some subtle points specific to the present problem. Here we are treating the set of eigenvalue problems indexed by the parameter $k \in \mathscr{K}$. Moreover, it is essential for us to explicitly construct (unnormalized) eigenvectors which are especially chosen to have a "nice" k dependence. We found that, for this purpose, it is better to directly deal with the Rayleigh–Schrödinger perturbation theory in an explicit manner, rather than to make use of the general theory.[19, 38] Unfortunately, such an

analysis of perturbation theory requires rather involved technical estimates which are summarized in this lengthy section.

10.1. States in the k-Space Representation

The basic starting point in the construction of the bases is the Schrödinger equation written in the form of (4.9), which is

$$\varepsilon \mathbf{v}(k) = (\lambda^2 t\, \mathsf{M}(k) + pt\, \mathsf{Q}(k))\, \mathbf{v}(k) \tag{10.1}$$

where $\mathbf{v}(k) = (v_u(k))_{u \in \mathscr{U}}$ is a b-dimensional vector. The $b \times b$ matrices $\mathsf{M}(k) = (M_{u,u'}(k))_{u,u' \in \mathscr{U}}$ and $\mathsf{Q}(k) = (Q_{u,u'}(k))_{u,u' \in \mathscr{U}}$ are defined in (4.10) and (4.11), respectively. For a fixed $k \in \mathscr{K}$ [see (3.33) for the definition of the space $\mathscr{K}$], (10.1) is an eigenvalue equation of a $b \times b$ matrix. Here $b = |\mathscr{U}| = \binom{d}{v} + 1$ is the number of bands. From a solution $\mathbf{v}(k)$ of (10.1) for some k, we can construct the corresponding Bloch state in the real space by

$$\varphi_x = e^{ik \cdot x} v_{\mu(x)}(k) \tag{10.2}$$

where $\mu(x)$ denotes the unique element in $\mathscr{U}$ such that $x \in \Lambda_{\mu(x)}$. The Bloch state $\varphi = (\varphi_x)_{x \in \Lambda}$ becomes an eigenstate of the original Schrödinger equation (4.2) with the energy eigenvalue ε.

One of our major tasks in the following subsections is to construct, for each $k \in \mathscr{K}$, a vector $\mathbf{v}^{(o)}(k) = (v_u^{(o)}(k))_{u \in \mathscr{U}}$ which satisfies

$$\varepsilon_1(k)\, \mathbf{v}^{(o)}(k) = (\lambda^2 t\, \mathsf{M}(k) + pt\, \mathsf{Q}(k))\, \mathbf{v}^{(o)}(k) \tag{10.3}$$

where $\varepsilon_1(k)$ is the lowest eigenvalue for each k. In other words, $\varepsilon_1(k)$ is the dispersion relation of the lowest band. Thus the Bloch state $\varphi = (\varphi_x)_{x \in \Lambda}$ constructed from $\mathbf{v}^{(o)}(k)$ according to (10.2) is an element of the Hilbert space $\mathscr{H}_{\text{single}}^{(1)}$ [see (4.19)] for the lowest band. In our construction, we do not normalize the vector $\mathbf{v}^{(o)}(k)$. We rather try to get a $\mathbf{v}^{(o)}(k)$ which has a "nice" k dependence so that we finally get sharply localized basis states.

For the moment, we assume that the desired $\mathbf{v}^{(o)}(k)$ is defined, and introduce other related vectors. For each $e \in \mathscr{U}'$ $(= \mathscr{U} \setminus \{o\})$, we define a vector $\mathbf{v}^{(e)}(k) = (v_u^{(e)}(k))_{u \in \mathscr{U}}$ so that the Bloch state (10.2) constructed from $\mathbf{v}^{(e)}(k)$ belongs to the Hilbert space $\mathscr{H}'_{\text{single}}$ [see (4.19)] for the higher bands. For this to be the case, it suffices to have orthogonality[38]

---

[38] $(\cdot, \cdot)$ denotes the standard inner product in the b-dimensional linear space. For $\mathbf{v} = (v_u)_{u \in \mathscr{U}}$ and $\mathbf{w} = (w_u)_{u \in \mathscr{U}}$, we define $(\mathbf{v}, \mathbf{w}) = \sum_{u \in \mathscr{U}} (v_u)^* w_u$.

$(\mathbf{v}^{(e)}(k), \mathbf{v}^{(o)}(k)) = 0$ for each $k \in \mathscr{K}$. The vectors $\mathbf{v}^{(e)}(k)$ are defined in terms of $\mathbf{v}^{(o)}(k)$ as

$$
v_u^{(e)}(k) = \begin{cases} -v_e^{(o)}(k) & \text{if } u = o \\ (v_o^{(o)}(k))^* & \text{if } u = e \\ 0 & \text{otherwise} \end{cases} \tag{10.4}
$$

The required orthogonality is readily verified from the definition. It is also found that, for each k, the vectors $\mathbf{v}^{(e)}(k)$ with $e \in \mathscr{U}'$ are linearly independent of each other. Therefore the collection $\{\mathbf{v}^{(u)}(k)\}_{u \in \mathscr{U}}$ for a fixed k forms a basis of $\mathbf{C}^b$.

We also introduce the dual of the basis $\{\mathbf{v}^{(u)}(k)\}_{u \in \mathscr{U}}$. For each $k \in \mathscr{K}$, we define the Gramm matrix $G(k)$ by

$$
(G(k))_{u, u'} = (\mathbf{v}^{(u)}(k), \mathbf{v}^{(u')}(k)) \tag{10.5}
$$

for $u, u' \in \mathscr{U}$. Since the vectors $\mathbf{v}^{(u)}(k)$ with $u \in \mathscr{U}$ are linearly independent, the corresponding Gramm matrix is invertible. We define the dual vectors by

$$
\tilde{\mathbf{v}}^{(u)}(k) = \sum_{u' \in \mathscr{U}} (G(k)^{-1})_{u', u} \mathbf{v}^{(u')}(k) \tag{10.6}
$$

for each $u \in \mathscr{U}$. We again write the components of the dual vectors as $\tilde{\mathbf{v}}^{(u)}(k) = (\tilde{v}_w^{(u)}(k))_{w \in \mathscr{U}}$. By definition, we have

$$
(\tilde{\mathbf{v}}^{(u)}(k), \mathbf{v}^{(u')}(k)) = \delta_{u, u'} \tag{10.7}
$$

and

$$
\sum_{w \in \mathscr{U}} (\tilde{v}_u^{(w)}(k))^* v_{u'}^{(w)}(k) = \delta_{u, u'} \tag{10.8}
$$

for any $u, u' \in \mathscr{U}$ and for any $k \in \mathscr{K}$.

10.2. Construction of the Localized Basis States

Since we have introduced the vectors (states) in the k-space representation, let us describe how we construct the desired localized basis states. For $x \in \Lambda$, we denote by $\mu(x)$ the unique site in the unit cell $\mathscr{U}$ such that $x \in \Lambda_{\mu(x)}$.

For $x, y \in \Lambda$, we define

$$
\varphi_x^{(y)} = (2\pi)^{-d} \int dk \, e^{ik \cdot (x - y)} v_{\mu(x)}^{(\mu(y))}(k) \tag{10.9}
$$

and

$$\tilde{\varphi}_x^{(y)} = (2\pi)^{-d} \int dk\, e^{ik\cdot(x-y)} \tilde{v}_{\mu(x)}^{(\mu(y))}(k) \tag{10.10}$$

where $\int dk(\cdots)$ is a shorthand for the sum $(2\pi/L)^d \sum_{k\in\mathscr{K}}(\cdots)$.

Let us prove the duality relation (4.34). By using the definition (10.9) and (10.10) and (uniquely) decomposing $y \in \Lambda$ as $y = z + u$ with $z \in \Lambda_o$ and $u \in \mathscr{U}$, we get

$$\sum_{y\in\Lambda} (\tilde{\varphi}_y^{(x)})^* \, \varphi_y^{(x')}$$

$$= \sum_{y\in\Lambda} (2\pi)^{-2d} \int dk\, dk'\, e^{-ik\cdot(y-x)+ik'\cdot(y-x')} (\tilde{v}_{\mu(y)}^{(\mu(x))}(k))^* \, v_{\mu(y)}^{(\mu(x'))}(k')$$

$$= \sum_{\substack{u\in\mathscr{U} \\ z\in\Lambda_o}} (2\pi)^{-2d} \int dk\, dk'\, e^{-i(k-k')\cdot z - i(k-k')\cdot u + ik\cdot x - ik'\cdot x'}$$

$$\times (\tilde{v}_u^{(\mu(x))}(k))^* \, v_u^{(\mu(x'))}(k')$$

$$= (2\pi)^{-d} \int dk\, e^{ik\cdot(x-x')} (\tilde{\mathbf{v}}^{(\mu(x))}(k), \mathbf{v}^{(\mu(x'))}(k))$$

$$= (2\pi)^{-d} \int dk\, e^{ik\cdot(x-x')} \chi[x - x' \in \Lambda_o]$$

$$= \delta_{x,x'} \tag{10.11}$$

where we used the duality relation (10.7) for v and $\tilde{v}$. We have also noted that $\delta_{\mu(x),\mu(x')} = \chi[x - x' \in \Lambda_o]$ with the indicator function $\chi[\text{true}] = 1$, $\chi[\text{false}] = 0$. The other duality relation (4.35) follows from the general argument about the uniqueness of the inverse matrix, or can be shown in a similar manner by using the corresponding relation (10.8).

In Lemma 4.2, we claimed that the sets $\{\varphi^{(x)}\}_{x\in\Lambda_o}$ and $\{\varphi^{(x)}\}_{x\in\Lambda'}$ form bases of the Hilbert spaces $\mathscr{H}_{\text{single}}^{(1)}$ and $\mathscr{H}_{\text{single}}'$, respectively. Note that, in (10.9), $\varphi_x^{(y)}$ is constructed as a superposition of various Bloch states $e^{ik\cdot x} v_{\mu(x)}^{(\mu(y))}(k)$ of the form (10.2). This means that $\varphi^{(x)} \in \mathscr{H}_{\text{single}}^{(1)}$ if $x \in \Lambda_o$ and $\varphi^{(x)} \in \mathscr{H}_{\text{single}}'$ if $x \in \Lambda'$. To prove the completeness of each basis, it therefore suffices to show that the union $\{\varphi^{(x)}\}_{x\in\Lambda}$ is a basis of the whole Hilbert space $\mathscr{H}_{\text{single}}$. But the desired completeness follows readily from the duality relation (4.35). The same argument shows the corresponding claim about the completeness of the dual bases stated in Lemma 4.4.

Finally we investigate the action of the modified hopping matrix $\tilde{T} = (\tilde{t}_{x,y})_{x,y\in\Lambda}$ [see (4.20)] on the basis states. Noting the Bloch state is

given by (10.2), we see that the Schrödinger equation (4.2) and Eq. (4.20) imply

$$\sum_{x' \in \Lambda} \tilde{t}_{x,x'}(e^{ik \cdot x'} v^{(\mu)}_{\mu(x')}(k)) = \begin{cases} \varepsilon_1(k)(e^{ik \cdot x} v^{(o)}_{\mu(x)}(k)) & \text{if } u = o \\ \frac{3}{4}\lambda^2 t(e^{ik \cdot x} v^{(u)}_{\mu(x)}(k)) & \text{if } u \in \mathscr{U}' \end{cases} \quad (10.12)$$

From (10.9) and (10.12), we get for $x \in \Lambda$ that

$$\sum_{z' \in \Lambda} \tilde{t}_{z,z'} \varphi^{(x)}_{z'} = (2\pi)^{-d} \int dk \sum_{z' \in \Lambda} \tilde{t}_{z,z'} e^{ik \cdot (z'-x)} v^{(o)}_{\mu(z')}(k)$$

$$= (2\pi)^{-d} \int dk \, \varepsilon_1(k) \, e^{ik \cdot (z-x)} v^{(o)}_{\mu(z)}(k)$$

$$= \sum_{y \in \Lambda_o} \tau_{y,x} \varphi^{(y)}_z \quad (10.13)$$

which is nothing but (4.30) with $\tau_{y,z}$ defined as in (4.31). The relation (4.29) follows easily from (10.12).

10.3. Basic Setup of Perturbation Theory

In the following construction of various vectors, we treat $k \in \mathscr{K}$ as a fixed parameter. The k dependence of the vectors will play nontrivial roles only in the final Section 10.9.

Let us first set $\rho = 0$ (corresponding to the flat-band model) in the Schrödinger equation (10.1). The eigenvector $\omega(k) = (w_u(k))_{u \in \mathscr{U}}$ with the lowest eigenvalue $\varepsilon = 0$ is given by

$$w_u(k) = \begin{cases} 1 & \text{if } u = o \\ -C_u(k)/\lambda & \text{if } u \in \mathscr{U}' \end{cases} \quad (10.14)$$

where $C_u(k)$ is defined in (4.12). We will construct our $\mathbf{v}^{(o)}(k)$ by the standard Rayleigh–Schrödinger perturbation theory so that it coincides with $\mathbf{w}(k)$ if $\rho = 0$.

For a fixed $k \in \mathscr{K}$, we denote by $\mathbf{P}(k) = (P_{u,u'}(k))_{u,u' \in \mathscr{U}}$ the orthogonal projection (in the linear space $\mathbf{C}^b$) onto the vector $\mathbf{w}(k)$. From (10.14), we explicitly have

$$P_{u,u'}(k) = \left(1 + \frac{A(k)}{\lambda^2}\right)^{-1} \times \begin{cases} 1 & \text{if } u = u' = o \\ -C_u(k)/\lambda & \text{if } u \in \mathscr{U}', \ u' = o \\ -C_{u'}(k)/\lambda & \text{if } u = o, \ u' \in \mathscr{U}' \\ C_u(k) \, C_{u'}(k)/\lambda^2 & \text{if } u, u' \in \mathscr{U}' \end{cases} \quad (10.15)$$

where $A(k)$ is defined in (4.13).

By comparing (10.1) and (4.14), we find that the matrix $M(k)$ (with a fixed k) has simple eigenvalues 0, $1 + A(k)/\lambda^2$, and $(b-2)$-fold degenerate eigenvalue 1. Since $w(k)$ is the eigenvector corresponding to the eigenvalue 0, the matrix $M(k) + P(k)$ has eigenvalues not less than 1, and hence is invertible. We define

$$W(k) = (M(k) + P(k))^{-1} \tag{10.16}$$

From (4.10) and (10.15), we find[39]

$$M(k) + P(k) = 1 + \left(1 + \frac{A(k)}{\lambda^2}\right)^{-1} \mathbf{a}(k) \otimes \mathbf{a}(k) \tag{10.17}$$

where 1 is the identity matrix, and the vector $\mathbf{a}(k) = (a_u(k))_{u \in \mathscr{U}}$ is defined as

$$a_u(k) = \begin{cases} A(k)/\lambda^2 & \text{if } u = o \\ C_u(k)/\lambda & \text{if } u \in \mathscr{U}' \end{cases} \tag{10.18}$$

By using the representation (10.17) and the general formula

$$(1 + \alpha \mathbf{v} \otimes \mathbf{v})^{-1} = 1 - \frac{\alpha}{\alpha(\mathbf{v}, \mathbf{v}) + 1} \mathbf{v} \otimes \mathbf{v} \tag{10.19}$$

we find from (10.16) that

$$W(k) = 1 - \left(1 + \frac{A(k)}{\lambda^2}\right)^{-1} \mathbf{a}(k) \otimes \mathbf{a}(k) \tag{10.20}$$

where we noted $(\mathbf{a}(k), \mathbf{a}(k)) = \{A(k)/\lambda^2\} + \{A(k)/\lambda^2\}^2$.

Following the philosophy of the Rayleigh–Schrödinger perturbation theory, we are going to express the eigenvector of (10.1) (for a fixed k) with the lowest eigenvalue $\varepsilon_1(k)$ as a power series in ρ as

$$\mathbf{v}^{(o)}(k) = \sum_{n=0}^{\infty} \rho^n \mathbf{v}_n^{(o)}(k) \tag{10.21}$$

where $\mathbf{v}_n^{(o)}(k)$ is a vector independent of ρ. We require

$$\mathbf{v}_0^{(o)}(k) = \mathbf{w}(k) \tag{10.22}$$

---

[39] For arbitrary vectors $\mathbf{v} = (v_u)_{u \in \mathscr{U}}$ and $\mathbf{w} = (w_u)_{u \in \mathscr{U}}$, we define their Kronecker product as $\mathbf{v} \otimes \mathbf{w} = (v_u w_{u'})_{u, u' \in \mathscr{U}}$ which can be regarded as a $b \times b$ matrix.

and

$$(\mathbf{w}(k), \mathbf{v}_n^{(o)}(k)) = 0 \tag{10.23}$$

for any $n \geqslant 1$. We also express the eigenvalue as

$$\varepsilon_1(k) = t \sum_{n=1}^{\infty} \rho^n e_n(k) \tag{10.24}$$

where the zeroth order is vanishing since we have $\varepsilon_1(k) = 0$ when $\rho = 0$ (which corresponds to the flat-band model).

By substituting the expression (10.21) into the Schrödinger equation (10.1) and collecting the terms with the nth power of ρ, we get

$$\sum_{\substack{j, l \geqslant 0 \\ (j+l=n-1)}} e_{j+1}(k) \mathbf{v}_l^{(o)}(k) = \lambda^2 \mathsf{M}(k) \mathbf{v}_n^{(o)}(k) + \mathsf{Q}(k) \mathbf{v}_{n-1}^{(o)}(k) \tag{10.25}$$

for any $n \geqslant 1$. In the present and the next subsection, summations like the above are always taken over integers (unless otherwise mentioned). The relation (10.25) is the basis of our perturbation theory. By taking the inner product with $\mathbf{v}_0^{(o)}(k) = \mathbf{w}(k)$ in (10.25), we get

$$e_n(k) = \frac{(\mathbf{w}(k), \mathsf{Q}(k) \mathbf{v}_{n-1}^{(o)}(k))}{(\mathbf{w}(k), \mathbf{w}(k))} \tag{10.26}$$

For $n \geqslant 1$, we have $\mathsf{P}(k) \mathbf{v}_n^{(o)}(k) = 0$ because of (10.23). Thus, by using (10.16), we can write

$$\mathsf{W}(k) \mathsf{M}(k) \mathbf{v}_n^{(o)}(k) = \{\mathsf{M}(k) + \mathsf{P}(k)\}^{-1} \{\mathsf{M}(k) + \mathsf{P}(k)\} \mathbf{v}_n^{(o)}(k) = \mathbf{v}_n^{(o)}(k) \tag{10.27}$$

for $n \geqslant 1$. Applying $\mathsf{W}(k)$ from the left of (10.25) and using (10.27), we get the recursion relation

$$\mathbf{v}_n^{(o)}(k) = -\frac{1}{\lambda^2} \mathsf{W}(k) \mathsf{Q}(k) \mathbf{v}_{n-1}^{(o)}(k)$$

$$+ \frac{1}{\lambda^2} \sum_{\substack{j, l \geqslant 0 \\ (j+l=n-1)}} \frac{(\mathbf{w}(k), \mathsf{Q}(k) \mathbf{v}_j^{(o)}(k))}{(\mathbf{w}(k), \mathbf{w}(k))} \mathsf{W}(k) \mathbf{v}_l^{(o)}(k) \tag{10.28}$$

where we have substituted (10.26) for $e_j(k)$. Since the right-hand side of (10.28) only contains $\mathbf{v}_m^{(o)}(k)$ with $m < n$, we can in principle determine $\mathbf{v}_n^{(o)}(k)$ with any n by using (10.28) recursively.

Let us rewrite the recursion relation (10.28) in a more explicit form. By substituting $(\mathbf{w}(k), \mathbf{w}(k)) = 1 + (A(k)/\lambda^2)$ [which follows from (10.14) and (4.13)] and the explicit form (10.20) of $W(k)$, we find that (10.28) becomes

$$\mathbf{v}_n^{(o)}(k) = \mathbf{V}_n^{(1)} + \mathbf{V}_n^{(2)} + \mathbf{V}_n^{(3)} + \mathbf{V}_n^{(4)} \tag{10.29}$$

with

$$\mathbf{V}_n^{(1)} = -\frac{1}{\lambda^2} Q(k) \, \mathbf{v}_{n-1}^{(o)}(k) \tag{10.30}$$

$$\mathbf{V}_n^{(2)} = \frac{1}{\lambda^2} \sum_{h=0}^{\infty} (h+1) \left(-\frac{A(k)}{\lambda^2} \right)^h (\mathbf{a}(k), Q(k) \, \mathbf{v}_{n-1}^{(o)}(k)) \, \mathbf{a}(k) \tag{10.31}$$

$$\mathbf{V}_n^{(3)} = \frac{1}{\lambda^2} \sum_{h=0}^{\infty} \left(-\frac{A(k)}{\lambda^2} \right)^h \sum_{\substack{j,l \geqslant 0 \\ (j+l=n-1)}} (\mathbf{w}(k), Q(k) \, \mathbf{v}_j^{(o)}(k)) \, \mathbf{v}_l^{(o)}(k) \tag{10.32}$$

and

$$\mathbf{V}_n^{(4)} = -\frac{1}{\lambda^2} \sum_{h=0}^{\infty} \frac{(h+1)(h+2)}{2} \left(-\frac{A(k)}{\lambda^2} \right)^h$$

$$\times \sum_{\substack{j,l \geqslant 0 \\ (j+l=n-1)}} (\mathbf{w}(k), Q(k) \, \mathbf{v}_j^{(o)}(k))(\mathbf{a}(k), \mathbf{v}_l^{(o)}(k)) \, \mathbf{a}(k) \tag{10.33}$$

10.4. Recursive Bounds for the Perturbation Coefficients

Let us construct the vector $\mathbf{v}^{(o)}(k)$ as in the expression (10.21) by using the recursion relations (10.28) and (10.29) along with the initial condition (10.22). The construction proceeds in an inductive manner. We first *assume* that the u component (where $u \in \mathcal{U}$) of the vector $\mathbf{v}_n^{(o)}(k)$ can be written as

$$(\mathbf{v}_n^{(o)}(k))_u = \left(\frac{1}{\lambda^2} \right)^n \sum_{m=0}^{\infty} \left(\frac{1}{\lambda} \right)^m$$

$$\times \sum_{\substack{(s_i, t_i) \in \mathcal{U} \times \mathcal{U} \\ \text{with } i=1,\ldots,n \\ \text{s.t. } (s_i, t_i) \leqslant (s_{i+1}, t_{i+1})}} \sum_{\substack{u_j \in \mathcal{U}' \\ \text{with } j=1,\ldots,m \\ \text{s.t. } u_j \leqslant u_{j+1}}} \alpha_1(u; \{(s_i, t_i)\}, \{u_j\})$$

$$\times \left(\prod_{i=1}^{n} Q_{s_i, t_i}(k) \right) \left(\prod_{j=1}^{m} C_{u_j}(k) \right) \tag{10.34}$$

with k-independent coefficients $\alpha_1(u; \{(s_i, t_i)\}, \{u_j\})$. In (10.34), the summations over $\{(s_i, t_i)\}_{i=1,\ldots,n}$ and $\{u_j\}_{j=1,\ldots,m}$ are restricted to the combinations which satisfy $(s_i, t_i) \leqslant (s_{i+1}, t_{i+1})$ and $u_j \leqslant u_{j+1}$, respectively. Here we have introduced an arbitrary complete ordering in the sets $\mathscr{U} \times \mathscr{U}$ and $\mathscr{U}'$.

Let us define

$$\tilde{\alpha}_1(n, m) = \sup_{\substack{u \\ \{(s_i, t_i)\}_{i=1,\ldots,n} \\ \{u_j\}_{j=1,\ldots,m}}} |\alpha_1(u; \{(s_i, t_i)\}, \{u_j\})| \qquad (10.35)$$

where the sup is taken over all the possible combinations that appear in (10.34) with the given m. The quantity $\tilde{\alpha}_1(n, m)$ plays the essential role in our inductive proof.

From (10.14), it is obvious that $\mathbf{v}_0^{(o)}(k) = \mathbf{w}(k)$ can be written in the form (10.34). We also find that the recursion relation (10.28) "preserves" the form (10.34) since the recursion essentially consists of multiplications by $C_u(k)$ [or $A(k) = \sum_{u \in \mathscr{U}'} \{C_u(k)\}^2$] and the matrix elements of $Q(k)$. See (10.29) and (10.30)–(10.33). This observation determines $\alpha_1(\cdots)$ uniquely, and formally "proves" the validity of the representation (10.34) if one neglects the problem of convergence.

Let us turn to the harder problem of controlling $\tilde{\alpha}_1(n, m)$ inductively and proving convergence of the sum in (10.34). Our strategy is to substitute the expression (10.34) for $\mathbf{v}_1^{(o)}(k),\ldots, \mathbf{v}_{n-1}^{(o)}(k)$ into the right-hand side of the recursion relation (10.28), reorganize the resulting expressions for $\mathbf{v}_n^{(o)}(k)$ so that they become of the form of (10.34), and finally express the coefficients α_1 for $\mathbf{v}_n^{(o)}(k)$ in terms of α_1 for $\mathbf{v}_1^{(o)}(k),\ldots, \mathbf{v}_{n-1}^{(o)}(k)$. The final expression leads us to an upper bound for $\tilde{\alpha}_1(n, m)$ in terms of $\tilde{\alpha}(n', m')$ with $n' < n$ and $m' \leqslant m$. See (10.36), (10.37), (10.39), (10.42), and (10.43).

The above procedure is easy to describe, but is too complicated to be executed explicitly. We shall take a slightly less complicated way, where we skip the intermediate calculations and directly get the final upper bounds for $\tilde{\alpha}_1(n, m)$. To avoid too much complication, we write the desired upper bound as

$$\tilde{\alpha}_1(n, m) \leqslant \sum_{j=1}^{4} \tilde{\alpha}_1^{(j)}(n, m) \qquad (10.36)$$

where $\tilde{\alpha}_1^{(j)}(n, m)$ are suitable upper bounds for the contributions to $\alpha_1(n, m)$ from $\mathbf{V}_n^{(j)}$ in the recursion formula (10.29).

To bound the contribution from $\mathbf{V}_n^{(1)}$ of (10.30) and get an upper bound $\tilde{\alpha}_1^{(1)}$, we assume that $\mathbf{v}_1^{(o)}(k)$ is written as (10.34), and then ask which $Q_{s_i, t_i}(k)$ in (10.34) comes from the $Q(k)$ which explicitly appears in the

right-hand side of (10.30). Since there are at most b^2 different $Q_{s,\,t}(k)$'s, we can set

$$\tilde{\alpha}_1^{(1)}(n, m) = b^2 \tilde{\alpha}_1^{(1)}(n-1, m) \tag{10.37}$$

To bound the contribution from $\mathbf{V}_n^{(2)}$ of (10.31), we note that one of the components of $\mathbf{a}(k)$, (10.18), is $A(k)/\lambda^2 = \sum_{u \in \mathcal{W}} \{C_u(k)/\lambda\}^2$, and $(b-1)$-components of $\mathbf{a}(k)$ are of the form $C_u(k)/\lambda$. By considering all the possible combinations of these components, we can set

$$\tilde{\alpha}_1^{(2)}(n, m) = b^2 \sum_{h=0}^{\infty} (h+1)(b-1)^h \{(b-1)^{2h+4} (b-1)^2 \tilde{\alpha}_1(n-1, m-2h-4)$$

$$+ (b-1)^{2h+3} (b-1)^2 \tilde{\alpha}_1(n-1, m-2h-3)$$

$$+ (b-1)^{2h+2} (b-1) \tilde{\alpha}_1(n-1, m-2h-2)\} \tag{10.38}$$

The prefactor b^2 appears for the same reason as in (10.37). The factors $(b-1)^{2h+4}$, $(b-1)^{2h+3}$, and $(b-1)^{2h+2}$ are the upper bounds for the number of ways to identify $C_u(k)$'s in (10.34) as coming from $A(k)$ or $\mathbf{a}(k)$ in the right-hand side of (10.31). Since $A(k)$ contains products of two $C_u(k)$'s, we have the common factor $(b-1)^{2h}$. For convenience, we reorganize (10.38) as

$$\tilde{\alpha}_1^{(2)}(n, m) = \sum_{h,\, p \geqslant 0} b^2 (b+1)^{h+m-p} (h+1) \tilde{\alpha}_1(n-1, p) \sum_{\mu = 2,\, 3,\, 4} \xi_\mu^{(2)} \delta_{2h+p,\, m-\mu} \tag{10.39}$$

with $\xi_2^{(2)} = b-1$, and $\xi_3^{(2)} = \xi_4^{(2)} = (b-1)^2$.

The next term $\mathbf{V}_n^{(3)}$ of (10.32) contains two $\mathbf{v}_{n'}^{(o)}(k)$ vectors. This means that we need to identify $C_u(k)$'s in $\mathbf{v}_n^{(o)}(k)$ [in the form (10.34)] as either (1) coming from $A(k)$ or $\mathbf{w}(k)$ explicitly contained in (10.32), (2) coming from $\mathbf{v}_j^{(o)}(k)$, or (3) coming from $\mathbf{v}_t^{(o)}(k)$. Identifications of $C_u(k)$'s into the classes 2 and 3 require a new combinatoric estimate. We need to count the number of ways to decompose $(p+q)$ objects into p objects and q objects. There are $(b-1)$ different kinds of objects, and we do not distinguish between objects of the same kind. [Of course, the objects are $C_u(k)$'s]. A crude upper bound for the desired combinatoric number is obtained by considering what are the possible contents of p objects. This observation shows that the desired number is bounded from above by

$$\binom{p+(b-1)-1}{(b-1)-1} \leqslant \frac{\{p+(b-1)-1\}^{(b-1)-1}}{\{(b-1)-1\}!} \tag{10.40}$$

Since there is a similar estimate with p replaced by q, the desired combinatoric number is bounded from above by the quantity $F(b-1; p, q)$, where

$$F(g; p, q) = \min \left\{ \frac{(p+g-1)^{g-1}}{(g-1)!}, \frac{(q+g-1)^{g-1}}{(g-1)!} \right\} \qquad (10.41)$$

An analogous combinatoric problem arises when we identify $Q_{s,l}(k)$'s in $v_n^{(o)}(k)$ as coming from either $Q(k)$, $v_j^{(o)}(k)$, or $v_l^{(o)}(k)$ in the right-hand side of (10.32). Consequently we have the following upper bound for the contribution from $\mathbf{V}_n^{(3)}$:

$$\tilde{\alpha}_1^{(3)}(n, m) = \sum_{h, p, q \geqslant 0} \sum_{\substack{j, l \geqslant 0 \\ (j+l=n-1)}} b^2 (b-1)^{h+m-(p+q)} F(b^2; j, l) F(b-1; p, q)$$

$$\times \tilde{\alpha}_1(j, p) \tilde{\alpha}_1(l, q) \sum_{\mu=0,1} \xi_\mu^{(3)} \delta_{2h+p+q, m-\mu} \qquad (10.42)$$

with $\xi_0^{(3)} = 1$ and $\xi_1^{(3)} = b-1$.

Finally, the contribution from $\mathbf{V}_n^{(4)}$ of (10.33) can be bounded in a similar manner as

$$\tilde{\alpha}_1^{(4)}(n, m) = \sum_{h, p, q \geqslant 0} \sum_{\substack{j, l \geqslant 0 \\ (j+l=n-1)}} b^2 \frac{(h+1)(h+2)}{2} (b-1)^{h+m-(p+q)}$$

$$\times F(b^2; j, l) F(b-1; p, q) \tilde{\alpha}_1(j, p) \tilde{\alpha}_1(l, q)$$

$$\times \sum_{\mu=2, 3, 4, 5} \xi_\mu^{(4)} \delta_{2h+p+q, m-\mu} \qquad (10.43)$$

with $\xi_2^{(4)} = (b-1)$, $\xi_3^{(4)} = b(b-1)$, $\xi_4^{(4)} = b(b-1)^2$, and $\xi_5^{(4)} = (b-1)^3$.

10.5. Upper Bounds for the Coefficients

We shall carry out an inductive proof of upper bounds for $\tilde{\alpha}_1(n, m)$. Our inductive assumption is that there are constants β and γ (determined later) which depend only on the band number b, and we have[40]

$$\tilde{\alpha}_1(n', m') \leqslant \frac{\beta^{n'}}{(n'+1)^{b^2+1}} \frac{\gamma^{m'}}{(m'+1)^b} \qquad (10.44)$$

---

[40] The factors $(n'+1)^{b^2+1}$ and $(m'+1)^b$ are indispensable for carrying out our inductive proof. We do not mean, however, that (10.44) represents the correct asymptotic behavior of $\tilde{\alpha}_1(n', m')$.

for any nonnegative n' and m', such that $n' < n$ and $m' \leqslant m$. Our goal is to prove the same bound for $n' = n$ and $m' = m$. Since $v_0^{(o)}(k) = w(k)$, we have $\tilde{\alpha}_1(0, 0) = 1$ [by comparing (10.14), (10.34), and (10.35)], which clearly satisfies the assumption (10.44) provided that $\beta, \gamma \geqslant 1$.

In what follows, we shall bound each of $\tilde{\alpha}_1^{(j)}(n, m)$ in (10.36) by using the assumption (10.44). We start from $\tilde{\alpha}_1^{(1)}(n, m)$. Since the right-hand side of (10.37) contains only $\tilde{\alpha}_1(n', m')$ with $n' < n$ and $m' \leqslant m$, we can use the assumption (10.44) to get

$$
\begin{aligned}
\tilde{\alpha}_1^{(1)}(n, m) &\leqslant b^2 \frac{\beta^{n-1}}{n^{b^2+1}} \frac{\gamma^m}{(m+1)^b} \\
&= \frac{\beta^n}{(n+1)^{b^2+1}} \frac{\gamma^m}{(m+1)^b} \left\{ \frac{b^2}{\beta} \left(\frac{n+1}{n} \right)^{b^2+1} \right\}
\end{aligned}
\tag{10.45}
$$

Note that we have factored out the desired quantity in front.

Next we investigate $\tilde{\alpha}_1^{(2)}(n, m)$ by substituting the assumption (10.44) into (10.39). Again we factor out the desired quantity to get

$$
\begin{aligned}
&\tilde{\alpha}_1^{(2)}(n, m) \\
&\leqslant \sum_{h, p \geqslant 0} b^2 (b+1)^{h+m-p} (h+1) \frac{\beta^{n-1}}{n^{b^2+1}} \frac{\gamma^p}{(p+1)^b} \sum_{\mu = 2, 3, 4} \xi_\mu^{(2)} \delta_{2h+p, m-\mu} \\
&= \frac{\beta^n}{(n+1)^{b^2+1}} \frac{\gamma^m}{(m+1)^b} \left\{ \frac{b^2}{\beta} \left(\frac{n+1}{n} \right)^{b^2+1} \right\} \\
&\quad \times \left\{ \sum_{h, p \geqslant 0} (h+1) \left(\frac{m+1}{p+1} \right)^b (b-1)^h \left(\frac{b-1}{\gamma} \right)^{m-p} \right. \\
&\quad \times \left. \sum_{\mu = 2, 3, 4} \xi_\mu^{(2)} \delta_{2h+p, m-\mu} \right\}
\end{aligned}
\tag{10.46}
$$

We bound the sum over h and p as

$$
\begin{aligned}
&\sum_{h, p \geqslant 0} (h+1) \left(\frac{m+1}{p+1} \right)^b (b-1)^h \left(\frac{b-1}{\gamma} \right)^{m-p} \sum_{\mu = 2, 3, 4} \xi_\mu^{(2)} \delta_{2h+p, m-\mu} \\
&= \sum_{\mu = 2, 3, 4} \xi_\mu^{(2)} \left(\frac{b-1}{\gamma} \right)^\mu \sum_{h=0}^{[(m-\mu)/2]} \left(\frac{m+1}{m-2h-\alpha} \right)^b (h+1) \left\{ \frac{(b-1)^3}{\gamma^2} \right\}^h
\end{aligned}
$$

$$
\leqslant \sum_{\mu = 2, 3, 4} \xi_\mu^{(2)} \left(\frac{b-1}{\gamma} \right)^\mu \left[\sum_{h=0}^{[(m-2\mu-1)/4]} 2^b (h+1) \left\{ \frac{(b-1)^3}{\gamma^2} \right\}^h \right.
$$

$$
\left. + \sum_{h=[(m-2\mu-1)/4]+1}^{\infty} (m+1)^b (h+1) \left\{ \frac{(b-1)^3}{\gamma^2} \right\}^h \right]
$$

$$
\leqslant \sum_{\mu = 2, 3, 4} \xi_\mu^{(2)} \left(\frac{b-1}{\gamma} \right)^\mu \left(1 - \frac{(b-1)^3}{\gamma} \right)^{-2}
$$

$$
\times \left\{ 2^b + 2^b \left(\frac{m-2\mu-1}{4} + 1 \right) \left(\frac{(b-1)^3}{\gamma^2} \right)^{(m-2\mu+7)/4} \right.
$$

$$
\left. + (m+1)^b \left(\frac{m-2\mu-1}{4} + 2 \right) \left(\frac{(b-1)^3}{\gamma^2} \right)^{(m-2h+3)/4} \right\}
$$

$$
\leqslant 2^{b+1} \sum_{\mu = 2, 3, 4} \xi_\mu^{(2)} \left(\frac{b-1}{\gamma} \right)^\mu \tag{10.47}
$$

where $[\cdots]$ is the Gauss symbol. The final inequality in (10.47) is valid for sufficiently large γ. By substituting (10.47) into (10.46), we get

$$
\tilde{\alpha}_1^{(2)}(n, m) \leqslant \frac{\beta^n}{(n+1)^{b^2+1}} \frac{\gamma^m}{(m+1)^b} \left\{ \frac{b^2}{\beta} \left(\frac{n+1}{n} \right)^{b^2+1} \right\}
$$

$$
\times 2^{b+1} \left\{ \frac{(b-1)^3}{\gamma^2} + \frac{(b-1)^5}{\gamma^3} + \frac{(b-1)^6}{\gamma^4} \right\} \tag{10.48}
$$

We postpone the estimate of $\tilde{\alpha}_1^{(3)}(n, m)$, and treat $\tilde{\alpha}_1^{(4)}(n, m)$. Again by substituting the inductive assumption (10.44) into (10.43), we get

$$
\tilde{\alpha}_1^{(4)}(n, m) \leqslant \sum_{h, p, q \geqslant 0} \sum_{\substack{j, l \geqslant 0 \\ (j+l=n-1)}} \left\{ b^2 \frac{(h+1)(h+2)}{2} (b-1)^{h+m-(p+q)} \right.
$$

$$
\times F(b^2; j, l) \, F(b-1; p, q)
$$

$$
\times \frac{\beta^j}{(j+1)^{b^2+1}} \frac{\gamma^p}{(p+1)^b} \frac{\beta^l}{(l+1)^{b^2+1}} \frac{\gamma^q}{(q+1)^b}
$$

$$
\left. \times \sum_{\mu = 2, 3, 4, 5} \xi_\mu^{(4)} \delta_{2h+p+q, \, m-\mu} \right\}
$$

$$
= \frac{\beta^n}{(n+1)^{b^2+1}} \frac{\gamma^m}{(m+1)^b} \frac{b^2}{\beta} S_1 S_2 \tag{10.49}
$$

with

$$S_1 = \sum_{\substack{j, l \geq 0 \\ (j+l=n-1)}} \left(\frac{n+1}{(j+1)(l+1)} \right)^{b^2+1} F(b^2; j, l) \tag{10.50}$$

and

$$S_2 = \sum_{h, p, q \geq 0} \left\{ \frac{(h+1)(h+2)}{2} (b-1)^j \left(\frac{b-1}{\gamma} \right)^{m-(p+q)} \right.$$
$$\left. \times \left(\frac{m+1}{(p+1)(q+1)} \right)^b F(b-1; p, q) \sum_{\mu=2,3,4,5} \xi_\mu^{(4)} \delta_{2h+p+q, m-\mu} \right\} \tag{10.51}$$

We first bound S_1. By using the symmetry between j and l in (10.50) and in $F(b^2; j, l)$, we have

$$S_1 \leq 2 \sum_{j=0}^{[(n-1)/2]} \left(\frac{n+1}{(j+1)(n-j)} \right)^{b^2+1} \frac{(j+b^2-1)^{b^2-1}}{(b^2-1)!} \tag{10.52}$$

By noting that the bounds $n-j \geq (n+1)/2$ and $(j+b^2-1) \leq (b^2-1)(j+1)$ hold within the range of the summation, we can further bound S_1 as

$$S_1 \leq 2 \left\{ \frac{n+1}{(n+1)/2} \right\}^{b^2+1} \sum_{j=0}^{[(n-1)/2]} \frac{(b^2-1)^{b^2-1} (j+1)^{b^2-1}}{(j+1)^{b^2+1} (b^2-1)!}$$
$$\leq 2 \cdot 2^{b^2+1} \frac{(b^2-1)^{b^2-1}}{(b^2-1)!} \sum_{j=0}^{\infty} (j+1)^{-2}$$
$$= J(b^2) \tag{10.53}$$

where we introduced

$$J(g) = \frac{\pi^2}{3} 2^{g+1} \frac{(g-1)^{g-1}}{(g-1)!} \tag{10.54}$$

The quantity S_2 in (10.51) can be bounded by combining the techniques used in the bounds (10.46), (10.47) and in the bounds (10.52), (10.53). The resulting bound is

$$S_2 \leq J(b-1) \cdot 2^{b+1} \left\{ \frac{(b-1)^3}{\gamma^2} + \frac{b(b-1)^4}{\gamma^3} + \frac{b(b-1)^6}{\gamma^4} + \frac{(b-1)^8}{\gamma^5} \right\} \tag{10.55}$$

By substituting (10.53) and (10.55) into (10.49), we finally get

$$
\tilde{\alpha}_1^{(4)}(n, m) \leqslant \frac{\beta^n}{(n+1)^{b^2+1}} \frac{\gamma^m}{(m+1)^b} \frac{b^2}{\beta} J(b^2) J(b-1)
$$

$$
\times 2^{b+1} \left\{ \frac{(b-1)^3}{\gamma^2} + \frac{b(b-1)^4}{\gamma^3} + \frac{b(b-1)^6}{\gamma^4} + \frac{(b-1)^8}{\gamma^5} \right\} \quad (10.56)
$$

The quantity $\tilde{\alpha}_1^{(3)}(n, m)$ in (10.42) can be bounded in the same manner as $\tilde{\alpha}_1^{(4)}$. The resulting bound is

$$
\tilde{\alpha}_1^{(3)}(n, m) \leqslant \frac{\beta^n}{(n+1)^{b^2+1}} \frac{\gamma^m}{(m+1)^b} \frac{b^2}{\beta} J(b^2) J(b-1) \cdot 2^{b+1} \left(1 + \frac{(b-1)^2}{\gamma} \right)
$$

$$
(10.57)
$$

Finally, by recalling (10.36), we sum up the bounds (10.45), (10.48), (10.56), and (10.57) to bound $\tilde{\alpha}_1(n, m)$ as

$$
\tilde{\alpha}_1(n, m) \leqslant \frac{\beta^n}{(n+1)^{b^2+1}} \frac{\gamma^m}{(m+1)^b}
$$

$$
\times \frac{b^2}{\beta} \left\{ 2^{b^2+1} + 2^{b^2+b+2} \left[\frac{(b-1)^3}{\gamma^2} + \frac{(b-1)^5}{\gamma^3} + \frac{(b-1)^6}{\gamma^4} \right] \right.
$$

$$
+ 2^{b+1} J(b^2) J(b-1)
$$

$$
\times \left. \left[1 + \frac{(b-1)^2}{\gamma} + \frac{(b-1)^3}{\gamma^2} + \frac{b(b-1)^4}{\gamma^3} + \frac{b(b-1)^6}{\gamma^4} + \frac{(b-1)^8}{\gamma^5} \right] \right\}
$$

$$
\leqslant \frac{\beta^n}{(n+1)^{b^2+1}} \frac{\gamma^m}{(m+1)^b} \quad (10.58)
$$

where the final bound holds for sufficiently large β and γ. Note that how large these constants should be depends only on the band number b. Since the bound (10.58) has precisely the same form as the inductive assumption (10.44), we have proved that $\tilde{\alpha}_1(n', m')$ satisfies the bound (10.44) for any $n', m' \geqslant 0$.

10.6. Construction of the Vector $\mathbf{v}^{(o)}(k)$

We are now ready to construct the ground-state vector $\mathbf{v}^{(o)}(k)$, which played an essential role in our construction in Sections 10.1 and 10.2. By

substituting the series (10.34) into (10.21), we get the following power series expression for $v^{(o)}(k) = (v_u^{(o)}(k))_{u \in \mathcal{U}}$:

$$
v_u^{(o)}(k) = w_u(k) + \sum_{n=1}^{\infty} \left(\frac{\rho}{\lambda^2} \right)^n \sum_{m=0}^{\infty} \left(\frac{1}{\lambda} \right)^m
$$

$$
\times \sum_{\substack{(s_i, t_i) \in \mathcal{U} \times \mathcal{U} \\ \text{with } i = 1, \dots, n \\ \text{s.t. } (s_i, t_i) \leqslant (s_{i+1}, t_{i+1})}} \sum_{\substack{u_j \in \mathcal{U}' \\ \text{with } j = 1, \dots, m \\ \text{s.t. } u_j \leqslant u_{j+1}}} \alpha_1(u; \{(s_i, t_i)\}, \{u_j\})
$$

$$
\times \left(\prod_{i=1}^{n} Q_{s_i, t_i}(k) \right) \left(\prod_{j=1}^{m} C_{u_j}(k) \right) \tag{10.59}
$$

See the discussion following (10.34) for the range of the summations. To investigate the convergence of (10.59), we note that (4.12) and (3.9) imply

$$
|C_u(k)| \leqslant |\mathcal{F}_u| = 2^v \tag{10.60}
$$

and

$$
|Q_{s, t}(k)| \leqslant 1 \tag{10.61}
$$

By using the above two bounds, the definition (10.35) of $\tilde{\alpha}_1(n, m)$, and the basic bound (10.44), and by noting that the numbers of possible combinations of $\{(s_i, t_i)\}_{i=1, \dots, n}$ and $\{u_j\}_{j=1, \dots, m}$ are bounded from above by b^{2n} and $(b-1)^m$, respectively, we find that the absolute value of the summand in (10.59) for each pair of n and m is bounded from above by

$$
\left(\frac{|\rho|}{\lambda^2} \right)^n \left(\frac{1}{\lambda} \right)^m b^{2n}(b-1)^m \frac{\beta^n}{(n+1)^{b^2+1}} \frac{\gamma^m}{(m+1)^b} 2^{vm} \tag{10.62}
$$

The quantity (10.62) is summable in n and m provided that $(|\rho|/\lambda^2) b^2 \beta < 1$ and $\lambda^{-1}(b-1)\gamma 2^v < 1$. If this is the case, the infinite sum in (10.59) is absolutely convergent. This completes our construction of the ground-state vector $v^{(o)}(k)$.

Let us summarize the present result as the following lemma.

Lemma 10.1. There exist positive constants β and γ which depend only on the band number b. When the parameters λ and ρ satisfy

$$
\frac{|\rho|}{\lambda^2} < \frac{1}{b^2 \beta} \tag{10.63}
$$

and

$$\lambda > 2^v(b-1)\,\gamma \tag{10.64}$$

the ground-state vector $\mathbf{v}^{(o)}(k)$ [characterized by (10.3)] is expressed by the absolutely convergent sum (10.59). The coefficients $\alpha_1(u, \{s_i, t_i\}, \{u_j\})$ in (10.59) are independent of k, and satisfy the bound

$$\sup_{\substack{u \\ \{(s_i, t_i)\}_{i=1,\dots,n} \\ \{u_j\}_{j=1,\dots,m}}} |\alpha_1(u; \{(s_i, t_i)\}, \{u_j\})| \leqslant \frac{\beta^n}{(n+1)^{b^2+1}} \frac{\gamma^m}{(m+1)^b} \tag{10.65}$$

for any n, m.

10.7. Dispersion Relation

Let us investigate the dispersion relation $\varepsilon_1(k)$ for the lowest band, which appears, e.g., in (10.3). By substituting the expression (10.26) into the formal expansion (10.24) for $\varepsilon_1(k)$, we find

$$\varepsilon_1(k) = t \sum_{n=1}^{\infty} \rho^n \frac{(\mathbf{w}(k), Q(k)\,\mathbf{v}_{n-1}^{(o)}(k))}{(\mathbf{w}(k), \mathbf{w}(k))} \tag{10.66}$$

Since $\mathbf{v}_{n-1}^{(o)}(k)$ is expressed as the convergent expansion (10.34), it is apparent from (10.66) that there is a similar power series expansion for $\varepsilon_1(k)$.

In fact, by substituting the expansion (10.34) into (10.66) and performing some estimates similar to those in Section 10.4, we get the expansion

$$\varepsilon_1(k) = \lambda^2 t \sum_{n=1}^{\infty} \left(\frac{\rho}{\lambda^2}\right)^n \sum_{m=0}^{\infty} \left(\frac{1}{\lambda}\right)^m$$

$$\times \sum_{\substack{(s_i, t_i) \in \mathscr{U} \times \mathscr{U} \\ \text{with } i=1,\dots,n \\ \text{s.t. } (s_i, t_i) \leqslant (s_{i+1}, t_{i+1})}} \sum_{\substack{u_j \in \mathscr{U}' \\ \text{with } j=1,\dots,m \\ \text{s.t. } u_j \leqslant u_j+1}} \alpha_2(\{(s_i, t_i)\}, \{u_j\})$$

$$\times \left(\prod_{i=1}^{n} Q_{s_i, t_i}(k)\right)\left(\prod_{j=1}^{m} C_{u_j}(k)\right) \tag{10.67}$$

where the coefficient $\alpha_2(\{(s_i, t_i)\}, \{u_j\})$ is independent of k. The ranges of the sums over $\{s_i, t_i\}$ and $\{u_j\}$ are the same as those in (10.34), (10.59).

For any $n \geqslant 2$ and $m \geqslant 0$, the coefficient $\alpha_2(\{(s_i, t_i)\}, \{u_j\})$ in (10.67) satisfies the bound

$$
\sup_{\substack{\{(s_i, t_i)\}_{i=1,\ldots,n} \\ \{u_j\}_{j=1,\ldots,m}}} |\alpha_2(\{(s_i, t_i)\}, \{u_j\})|
$$

$$
\leqslant b^2 \sum_{h, p \geqslant 0} (b-1)^h (b-1)^{m-p} \left(\sum_{\mu = 0, 1} \xi_\mu \delta_{2h+p, m-\mu} \right) \tilde{\alpha}_1(n-1, p)
$$

$$
(10.68)
$$

with $\xi_0 = 1$, $\xi_1 = b - 1$. Substituting the bound (10.44) for $\tilde{\alpha}_1$ into the right-hand side of (10.68) and performing estimates similar to those in Section 10.5, we find for $n \geqslant 1$ and $m \geqslant 0$ that[41]

$$
\sup_{\substack{\{(s_i, t_i)\}_{i=1,\ldots,n} \\ \{u_j\}_{j=1,\ldots,m}}} |\alpha_2(\{(s_i, t_i)\}, \{u_j\})| \leqslant \frac{\beta^n}{(n+1)^{b^2+1}} \frac{\gamma^m}{(m+1)^b} \qquad (10.69)
$$

again for sufficiently large β and γ.

The bound (10.69), along with (10.60) and (10.61), proves the convergence of the sum (10.67) for λ and ρ satisfying the conditions (10.63) and (10.64).

10.8. Dual Vectors

We shall develop power series expansions for the dual vectors $\tilde{\mathbf{v}}^{(u)}(k)$ (with $u \in \mathscr{U}$) defined in (10.6). By recalling the definition (10.4) of the vector $\mathbf{v}^{(e)}(k)$ with $e \in \mathscr{U}'$, we can express the components of the Gramm matrix $G(k)$ in (10.5) as $(G(k))_{o, o} = |\mathbf{v}^{(o)}(k)|^2$, $(G(k))_{o, e} = (G(k))_{e, o} = 0$, $(G(k))_{e, e'} = v_e^{(o)}(k) (v_{e'}^{(o)}(k))^*$, and $(G(k))_{e, e} = |v_o^{(o)}(k)|^2 + |v_e^{(o)}(k)|^2$, where $e, e' \in \mathscr{U}'$ and $e \neq e'$. Thus the $b \times b$ matrix $G(k)$ can be compactly written in the form

$$
G(k) = \begin{pmatrix} |\mathbf{v}^{(o)}(k)|^2 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & H(k) & \\ 0 & & & \end{pmatrix} \qquad (10.70)
$$

[41] The estimate for $n = 1$ follows from explicit calculation.

where the $(b-1) \times (b-1)$ matrix $H(k)$ is given by[42]

$$H(k) = |v_o^{(o)}(k)|^2 \, I + g(k) \otimes g(k)^* \tag{10.71}$$

with the $(b-1)$-dimensional vectors $g(k) = (v_e^{(o)}(k))_{e \in \mathscr{U}'}$ and $g(k)^* = ((v_e^{(o)}(k))^*)_{e \in \mathscr{U}'}$.

It is evident from (10.70) that the inverse of the Gramm matrix is written as

$$G(k)^{-1} = \begin{pmatrix} |v^{(o)}(k)|^{-2} & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & H(k)^{-1} & \\ 0 & & & \end{pmatrix} \tag{10.72}$$

As for the inverse of $H(k)$, we use the general formula (10.19) to get

$$H(k)^{-1} = \frac{1}{|v_o^{(o)}(k)|^2} \left(I - \frac{1}{|v^{(o)}(k)|^2} g(k) \otimes g(k)^* \right) \tag{10.73}$$

By substituting (10.72) and (10.73) in the definition (10.6) of the dual vectors, we get

$$\tilde{v}^{(o)}(k) = \frac{1}{|v^{(o)}(k)|^2} v^{(o)}(k) \tag{10.74}$$

and

$$\tilde{v}^{(e)}(k) = \frac{1}{|v_o^{(o)}(k)|^2} v^{(e)}(k) - \frac{(v_e^{(o)}(k))^*}{|v^{(o)}(k)|^2} \sum_{e' \in \mathscr{U}'} v_{e'}^{(o)}(k) \, v^{(e')}(k) \tag{10.75}$$

where $e \in \mathscr{U}'$. We again denote the components of the dual vectors as $\tilde{v}^{(u)}(k) = (\tilde{v}_{u'}^{(u)}(k))_{u' \in \mathscr{U}}$. By using (10.4), we obtain from two equations the following expressions for the components of the dual vectors in terms of the components of the ground-state vector $v^{(o)}(k)$,

$$\tilde{v}_u^{(o)}(k) = |v^{(o)}(k)|^{-2} \, v_u^{(o)}(k) \tag{10.76}$$

for $u \in \mathscr{U}$,

$$\tilde{v}_o^{(e)}(k) = - |v^{(o)}(k)|^{-2} (v_e^{(o)}(k))^* \tag{10.77}$$

$$\tilde{v}_{e'}^{(e)}(k) = - |v^{(o)}(k)|^{-2} (v_o^{(o)}(k))^* (v_e^{(o)}(k))^* \, v_{e'}^{(o)}(k) \tag{10.78}$$

---

[42] $H(k)$ and the identity matrix I in (10.71) and (10.73) are the only $(b-1) \times (b-1)$ matrices that appear in the present paper. Similarly $g(k)$ is the only $(b-1)$-dimensional vector.

and

$$\tilde{v}_e^{(e)}(k) = \frac{1}{v_o^{(o)}(k)} - \frac{(v_o^{(o)}(k))^* \, |v_e^{(o)}(k)|^2}{|\mathbf{v}^{(o)}(k)|^2} \qquad (10.79)$$

where $e, e' \in \mathcal{U}'$ and $e \neq e'$.

Recalling that each $v_u^{(o)}(k)$ admits the power series expansion (10.59), it is clear from the expressions (10.76)–(10.79) that there are similar expansions for the dual vectors. In order to control these expansions, we substitute (10.59) into (10.76)–(10.79) and reorganize the resulting expressions into transparent series expansions. Then by using the bounds (10.65) for the coefficients in (10.59), we can control the coefficients of the new expansions for the dual vectors. Unfortunately this straightforward procedure turns out to be rather tedious to carry out in practice. We shall omit the details here since the required estimates are quite similar to those in Sections 10.4 and 10.5.

The resulting series expansion for the dual vectors can be written as

$$\tilde{v}_{u'}^{(u)}(k) = \tilde{w}_{u'}^{(u)}(k) + \sum_{\substack{g, l \geq 0 \\ (g+l \geq 1)}} \left(\frac{\rho}{\lambda^2}\right)^{g+l} \sum_{m \geq 0} \left(\frac{1}{\lambda}\right)^m$$

$$\times \sum_{\substack{(q_h, r_h) \in \mathcal{U} \times \mathcal{U} \\ \text{with } h = 1,\dots, g \\ \text{s.t. } (q_h, r_h) \leq (q_{h+1}, r_{h+1})}} \sum_{\substack{(s_i, t_i) \in \mathcal{U} \times \mathcal{U} \\ \text{with } i = 1,\dots, l \\ \text{s.t. } (s_i, t_i) \leq (s_{i+1}, t_{i+1})}} \sum_{\substack{u_j \in \mathcal{U}' \\ \text{with } j = 1,\dots, m \\ \text{s.t. } u_j \leq u_{j+1}}}$$

$$\times \alpha_3(u, u'; \{(q_h, r_h)\}, \{(s_i, t_i)\}, \{u_j\})$$

$$\times \left\{\prod_{h=1}^{g} Q_{q_h, r_h}(k)\right\}\left\{\prod_{i=1}^{l} (Q_{s_i, t_i}(k))^*\right\}\left\{\prod_{j=1}^{m} C_{u_j}(k)\right\} \qquad (10.80)$$

We have introduced the dual vectors $\tilde{\mathbf{w}}^{(u)}(k) = (\tilde{w}_{u'}^{(u)}(k))_{u' \in \mathcal{U}}$ with $u \in \mathcal{U}$ for the model with $\rho = 0$ (i.e., the flat-band model). By using (10.76)–(10.79) with $\mathbf{v}^{(o)}(k)$ replaced by $\mathbf{w}(k)$ and the definition (10.14) of $\mathbf{w}(k)$, we find

$$\tilde{w}_{u'}^{(u)}(k) = \begin{cases} (1 + A(k)/\lambda^2)^{-1} \times \begin{cases} 1 & \text{if } u = u' = o \\ -C_{u'}(k)/\lambda & \text{if } u = o, \quad u' \in \mathcal{U}' \\ C_u(k)/\lambda & \text{if } u \in \mathcal{U}', \quad u' = o \\ -C_u(k) \, C_{u'}(k)/\lambda^2 & \text{if } u, u' \in \mathcal{U}', \quad u \neq u' \end{cases} \\ 1 - (1 + A(k)/\lambda^2)^{-1} (C_u(k))^2/\lambda^2 & \text{if } u = u' \in \mathcal{U}' \end{cases}$$

$$(10.81)$$

where we have used (4.13).

The coefficients α_3 in the expansion (10.80) can be shown to satisfy for each $n \geqslant 1$, $m \geqslant 0$ the bound

$$\sup_{\substack{g, l \geqslant 0 \\ (g+l=n)}} \sup_{\substack{u, u' \\ \{q_h, r_h\}_{h=1,\ldots,g} \\ \{s_i, t_i\}_{i=1,\ldots,l} \\ \{u_j\}_{j=1,\ldots,m}}} |\alpha_3(u, u'; \{(q_h, r_h)\}, \{(s_i, t_i)\}, \{u_j\})| \leqslant C_b \tilde{\beta}^n \tilde{\gamma}^m \quad (10.82)$$

where

$$\tilde{\beta} = 8b^4\beta, \qquad \tilde{\gamma} = 8(b-1)\gamma \qquad (10.83)$$

and C_b is a constant which depends only on b.

By using the bounds (10.82), (10.60), and (10.61), we can show that the power series (10.80) for the dual vectors converge provided that

$$\frac{|\rho|}{\lambda^2} \leqslant r_0 = \frac{\theta}{b^2 \tilde{\beta}} \qquad (10.84)$$

and

$$\lambda \geqslant \lambda_0 = \frac{2^v(b-1)\tilde{\gamma}}{\theta} \qquad (10.85)$$

with a constant $0 < \theta < 1$, which we shall now fix. Note that the conditions (10.63) and (10.64) [required for the convergence of the series for $v^{(o)}(k)$ and $\varepsilon_1(k)$] are automatically satisfied if we assume (10.84) and (10.85). This completes our construction of the basis states.

10.9. Summability of the Basis States

It only remains to prove the summability of the basis states $\varphi^{(x)}$, $\tilde{\varphi}^{(x)}$ and the effective hopping $\tau_{y,x}$ stated in Lemmas 4.2, 4.4, and 4.3. It turns out that these bounds are natural consequences of the series expansions (10.59), (10.67), and (10.80).

Let us look at the proofs of the bounds (4.25) and (4.26) in detail. We first recall that, for $x \in \Lambda_o$, the strictly localized basis states $\psi^{(x)}$ [defined by (4.23) and (4.24)] are written in terms of the $w(k)$ of (10.14) as

$$\psi_y^{(x)} = (2\pi)^{-d} \int dk \, e^{-ik \cdot (x-y)} w_{\mu(y)}(k) \qquad (10.86)$$

See Section 10.2 for the notations. Note that (10.86) is a special case of (10.9). From (10.9), (10.86), and the expansion (10.59), we find for $x \in \Lambda_o$ that

$$\varphi_y^{(x)} - \psi_y^{(x)} = (2\pi)^{-d} \int dk \, e^{-ik \cdot (x-y)} (v_{\mu(y)}^{(o)}(k) - w_{\mu(y)}(k))$$

$$= \sum_{n=1}^{\infty} \left(\frac{p}{\lambda^2}\right)^n \sum_{m=0}^{\infty} \left(\frac{1}{\lambda}\right)^m \sum_{\substack{(s_i, t_i) \in \mathscr{U} \times \mathscr{U} \\ \text{with } i=1,\dots,n \\ \text{s.t. } (s_i, t_i) \leqslant (s_{i+1}, t_{i+1})}} \sum_{\substack{u_j \in \mathscr{U}' \\ \text{with } j=1,\dots,m \\ \text{s.t. } u_j \leqslant u_{j+1}}}$$

$$\times \alpha_1(\mu(y); \{(s_i, t_i)\}, \{u_j\}) \, I_{x,y}(\{(s_i, t_i)\}, \{\mu_j\}) \quad (10.87)$$

with

$$I_{x,y}(\{(s_i, t_i)\}, \{u_j\})$$

$$= (2\pi)^{-d} \int dk \, e^{-ik \cdot (x-y)} \left(\prod_{i=1}^{n} Q_{s_i, t_i}(k)\right) \left(\prod_{j=1}^{m} C_{u_j}(k)\right) \quad (10.88)$$

Recalling the definitions (4.11) and (4.12), we find that

$$\sum_{y \in \Lambda} |I_{x,y}(\{(s_i, t_i)\}, \{u_j\})| \leqslant \left\{\max_{z \in \Lambda} \left(\frac{1}{t} \sum_{w \in \Lambda} |t'_{z,w}|\right)\right\}^n (2^\nu)^m \leqslant 2^{\nu m} \quad (10.89)$$

where we used (3.22) and $|\mathscr{F}_f| = 2^\nu$. Similarly we have

$$\sum_{y \in \Lambda} |x - y| \cdot |I_{x,y}(\{(s_i, t_i)\}, \{u_j\})|$$

$$\leqslant n \left\{\max_{z \in \Lambda} \left(\frac{1}{t} \sum_{w \in \Lambda} |t'_{z,w}|\right)\right\}^{n-1} \left\{\max_{z \in \Lambda} \left(\frac{1}{t} \sum_{w \in \Lambda} |w - z| \cdot |t'_{z,w}|\right)\right\} (2^\nu)^m$$

$$+ m \left\{\max_{z \in \Lambda} \left(\frac{1}{t} \sum_{w \in \Lambda} |t'_{z,w}|\right)\right\}^n (2^\nu)^{m-1} \left(\sum_{g \in \mathscr{F}_f} |g|\right)$$

$$\leqslant \left(nR + m \frac{\sqrt{\nu}}{2}\right) 2^{\nu m}$$

$$\leqslant (n+m) R 2^{\nu m} \quad (10.90)$$

where we used (3.23) and noted that $|g| = \sqrt{\nu}/2$ for $g \in \mathscr{F}_f$. In the final step, we used the assumption $\sqrt{\nu}/2 \leqslant R$ introduced right after (3.23).

We substitute the bound (10.89) for $I_{x,y}$ and the bound (10.65) for α_1 to (10.87) to get

$$
\sum_{y \in \Lambda} |\varphi_y^{(x)} - \psi_y^{(x)}|
$$

$$
\leqslant \sum_{n=1}^{\infty} \sum_{m=0}^{\infty} \left(\frac{|\rho|}{\lambda^2} \right)^n \left(\frac{1}{\lambda} \right)^m b^{2n}(b-1)^m \frac{\beta^n}{(n+1)^{b^2+1}} \frac{\gamma^m}{(m+1)^b} 2^{vm}
$$

$$
\leqslant \left\{ \sum_{n=1}^{\infty} \left(\frac{|\rho|}{\lambda^2} b^2 \beta \right)^n \right\} \left\{ \sum_{m=0}^{\infty} \left(\frac{1}{\lambda} 2^v (b-1) \gamma \right)^m \right\}
$$

$$
\leqslant B_1 \frac{|\rho|}{\lambda^2} \tag{10.91}
$$

for ρ and λ satisfying (10.84) and (10.85) [or (10.63) and (10.64)]. The constant B_1 will be fixed later.

Similarly we use (10.90) to get

$$
\sum_{y \in \Lambda} |x-y| \, |\varphi_y^{(x)} - \psi_y^{(x)}| \leqslant \sum_{n=1}^{\infty} \sum_{m=0}^{\infty} \left(\frac{|\rho|}{\lambda^2} \right)^n \left(\frac{1}{\lambda} \right)^m b^{2n}(b-1)^m
$$

$$
\times \frac{\beta^n}{(n+1)^{b^2+1}} \frac{\gamma^m}{(m+1)^b} (n+m) \, R \cdot 2^{vm}
$$

$$
\leqslant B_1 R \frac{|\rho|}{\lambda^2} \tag{10.92}
$$

which is the desired bound (4.26). The bounds (4.25), (4.26) for $x \in \Lambda'$ as well as the remaining bounds (4.27), (4.28) follow in the same manner.

The bounds (4.32) and (4.33) for the effective hopping $\tau_{x,y}$ stated in Lemma 4.3 are proved in exactly the same manner by using the definition (4.31), the expansion (10.67) for $\varepsilon_1(k)$, and the bounds (10.69) for the coefficients.

The bounds (4.36)–(4.39) for the dual basis states can also be shown in the same spirit. A major difference is that $\tilde{\varphi}^{(x)}$ does not coincide with the strictly localized state $\psi^{(x)}$ when $\rho = 0$. To control this situation, we note

$$
|\tilde{\varphi}_y^{(x)} - \psi_y^{(x)}| \leqslant |\tilde{\varphi}_y^{(x)} - \tilde{\psi}_y^{(x)}| + |\tilde{\psi}_y^{(x)} - \psi_y^{(x)}| \tag{10.93}
$$

where $\tilde{\psi}_y^{(x)}$ (which is the dual basis states for $\rho = 0$) is defined as

$$
\tilde{\psi}_y^{(x)} = (2\pi)^{-d} \int dk \, e^{-ik \cdot (x-y)} \tilde{w}_{\mu(y)}^{(\mu(x))}(k) \tag{10.94}
$$

with $\tilde{w}^{(u)}(k)$ defined in (10.81). By using the series expansion (10.80), we can control the term $|\tilde{\varphi}_y^{(x)} - \tilde{\psi}_y^{(x)}|$ in exactly the same way as we controlled $|\varphi_y^{(x)} - \psi_y^{(x)}|$ in the above. Consequently, we get

$$\sum_{x \in \Lambda} |\tilde{\varphi}_y^{(x)} - \tilde{\psi}_y^{(x)}| \leqslant B_1 \frac{|\rho|}{\lambda^2}, \qquad \sum_{y \in \Lambda} |\tilde{\varphi}_y^{(x)} - \tilde{\psi}_y^{(x)}| \leqslant B_1 \frac{|\rho|}{\lambda^2} \qquad (10.95)$$

and

$$\sum_{x \in \Lambda} |x - y| \cdot |\tilde{\varphi}_y^{(x)} - \tilde{\psi}_y^{(x)}| \leqslant B_1 R \frac{|\rho|}{\lambda^2}$$

$$\sum_{y \in \Lambda} |x - y| \cdot |\tilde{\varphi}_y^{(x)} - \tilde{\psi}_y^{(x)}| \leqslant B_1 R \frac{|\rho|}{\lambda^2} \qquad (10.96)$$

At this stage, we fix the constant B_1 so that the bounds (4.25)–(4.28), (10.95), and (10.96) are simultaneously satisfied.[43] Note that B_1 depends only on the band number b.

To control the second term in (10.93), we first note

$$\tilde{\psi}_y^{(x)} - \psi_y^{(x)} = (2\pi)^{-d} \int dk \, e^{-ik \cdot (x-y)} \zeta_{\mu(y)}^{(\mu(x))} \qquad (10.97)$$

with

$$\zeta_{u'}^{(u)} = \begin{cases} -(A(k)/\lambda^2)(1 + A(k)/\lambda^2)^{-1} \times \begin{cases} 1 & \text{if} \quad u = u' = o \\ -C_{u'}(k)/\lambda & \text{if} \quad u = o, \quad u' \in \mathcal{U}' \\ C_u(k)/\lambda & \text{if} \quad u \in \mathcal{U}', \quad u' = o \end{cases} \\ -(C_u(k) \, C_{u'}(k)/\lambda^2)(1 + A(k)/\lambda^2)^{-1} & \text{if} \quad u, u' \in \mathcal{U}' \end{cases}$$

$$(10.98)$$

The expressions (10.97), (10.98) are straightforward consequences of (10.94), (10.81), (10.14), and (10.4). By expanding $\{1 + (A(k)/\lambda^2)\}^{-1}$ in (10.98), we can express $\tilde{\psi}_y^{(x)} - \psi_y^{(x)}$ as a power series of λ^{-2}. By analyzing the series, it is easily shown that, for $\lambda \geqslant \lambda_0$, the summations

---

[43] Of course it is possible to state the bounds (4.25)–(4.28) with smaller B_1 than in (10.95) or (10.96). We have unified the coefficients as much as possible to make the formulas less complicated.

$$\sum_{y \in \Lambda} |\tilde{\psi}_y^{(x)} - \psi_y^{(x)}|, \qquad \sum_{y \in \Lambda} |x - y| \cdot |\tilde{\psi}_y^{(x)} - \psi_y^{(x)}|$$

$$\sum_{x \in \Lambda} |\tilde{\psi}_y^{(x)} - \psi_y^{(x)}|, \qquad \sum_{x \in \Lambda} |x - y| \cdot |\tilde{\psi}_y^{(x)} - \psi_y^{(x)}|$$

are all bounded from above by B_2/λ^2, where B_2 is a constant which depends only on the band number b. By combining these bounds with (10.93), (10.95), and (10.96), we get the desired bounds (4.36)–(4.39).

ACKNOWLEDGMENTS

REFERENCES

1. P. W. Anderson, in *Magnetism I*, G. T. Rado and H. Schul, eds. (Academic Press, New York, 1963).
2. P. W. Anderson, *Phys. Rev. Lett.* **21**:13 (1968).
3. N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Saunders College, 1976).
4. G. Benffatto, G. Gallavotti, A. Procacci, and B. Scoppola, *Commun. Math. Phys.* **160**:93 (1994).
5. F. Bloch, *Z. Phys.* **57**:545 (1929).
6. U. Brandt and A. Giesekus, *Phys. Rev. Lett.* **68**:2648 (1992).
7. J. de Boer and A. Schadschneider, *Phys. Rev. Lett.* **75**:4298 (1995).
8. B. Douçot and X. G. Wen, *Phys. Rev. B* **40**:2719 (1989).
9. F. Dyson, E. H. Lieb, and B. Simon, *J. Stat. Phys.* **18**:335 (1978).
10. J. Feldman, J. Magnen, V. Rivasseau, and E. Trubowitz, in *The LXII les Houches Summer School "Fluctuating Geometries in Statistical Mechanics and in Field Theory," August 2–September 9 1994* [cond-mat/9503047].
11. D. Ghosh, *Phys. Rev. Lett.* **27**:1584 (1971).
12. M. C. Gutzwiller, *Phys. Rev. Lett.* **10**:159 (1963).
13. T. Hanisch and E. Müller-Hartmann, *Ann. Physik* **2**:381 (1993).
14. W. J. Heisenberg, *Z. Phys.* **49**:619 (1928).
15. C. Herring, in *Magnetism IIB*, G. T. Rado and H. Suhl, eds. (Academic Press, New York, 1966).
16. C. Herring, in *Magnetism IV*, G. T. Rado and H. Suhl, eds. (Academic Press, New York, 1966).
17. J. Hubbard, *Proc. Roy. Soc. Lond. A* **276**:238 (1963).
18. J. Kanamori, *Prog. Theor. Phys.* **30**:275 (1963).

19. T. Kato, *A Short Introduction to Perturbation Theory for Linear Operators* (Springer-Verlag, Berlin, 1982).
20. W. Kohn, *Phys. Rev. B* **7**:4388 (1973).
21. T. Koma and H. Tasaki, *Phys. Rev. Lett.* **68**:3248 (1992).
22. K. Kusakabe, Ferromagnetism in strongly correlated electron systems, Ph.D. thesis, University of Tokyo (1993).
23. K. Kusakabe and H. Aoki, *Phys. Rev. Lett.* **72**:144 (1994).
24. K. Kusakabe and H. Aoki, *Physica B* **194–196**:215 (1994).
25. S. Liang and H. Pang, Preprint (1994) [cond-mat/9404003].
26. E. H. Lieb, in *Phase Transitions* (Wiley, Interscience, 1971), p. 45.
27. E. H. Lieb, *Phys. Rev. Lett.* **62**:1201 (1989).
28. E. H. Lieb and D. Mattis, *Phys. Rev.* **125**:164 (1962).
29. D. C. Mattis, *The Theory of Magnetism I* (Springer-Verlag, Berlin, 1981).
30. A. Mielke, *J. Phys. A* **24**:L73 (1991).
31. A. Mielke, *J. Phys. A* **24**:3311 (1991).
32. A. Mielke, *J. Phys. A* **25**:4335 (1992).
33. A. Mielke, *Phys. Lett. A* **174**:443 (1993).
34. A. Mielke and H. Tasaki, *Commun. Math. Phys.* **158**:341 (1993).
35. E. Müller-Hartmann, *J. Low Temp. Phys.* **99**:349 (1995).
36. Y. Nagaoka, *Phys. Rev.* **147**:392 (1966).
37. W. O. Putikka, M. U. Luchini, and M. Ogata, *Phys. Rev. Lett.* **69**:2288 (1992).
38. M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, Vol. IV (Academic Press, New York, 1978).
39. B. S. Shastry, H. R. Krishnamurthy, and P. W. Anderson, *Phys. Rev. B* **41**:2375 (1990).
40. S. Q. Shen, Z. M. Qiu, and G. S. Tian, *Phys. Rev. Lett.* **72**:1280 (1994).
41. J. C. Slater, H. Statz, and G. F. Koster, *Phys. Rev.* **91**:1323 (1953).
42. R. Strack and D. Vollhardt, *Phys. Rev. Lett.* **72**:3425 (1994).
43. R. Strack and D. Vollhardt, *J. Low Temp. Phys.* **99**:385 (1995).
44. A. Sütő, *Phys. Rev. B* **43**:8779 (1991).
45. A. Sütő, *Commun. Math. Phys.* **140**:43 (1991).
46. H. Tasaki, *Phys. Rev. B* **40**:9192 (1989).
47. H. Tasaki, *Phys. Rev. Lett.* **69**:1608 (1992).
48. H. Tasaki, *Phys. Rev. Lett.* **70**:3303 (1993).
49. H. Tasaki, *Phys. Rev. Lett.* **73**:1158 (1994).
50. H. Tasaki, *Phys. Rev. B* **49**:7763 (1994).
51. H. Tasaki, *Phys. Rev. Lett.* **75**:4678 (1995).
52. H. Tasaki, in preparation.
53. D. J. Thouless, *Proc. Phys. Soc. Lond.* **86**:893 (1965).
54. B. Tóth, *Lett. Math. Phys.* **22**:321 (1991).