The Geometry and Topology of Charge-Ordered Quantum Fields in Low-Dimensional Systems

Felix Flicker

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Science. School of Physics, June 2015

Word Count: 39,349

Abstract

Charge ordering is a phenomenon in which the electron field in a material spontaneously breaks the symmetry of the underlying crystal lattice, selecting out a new period and, in dimensions higher than one, a particular direction in space.

In the first half of this thesis I study the consequences of charge ordering in 1D, applying a self-consistent mean-field approach. A system with rational filling p/q forms a period q charge density wave, which I demonstrate exhibits a quantized adiabatic particle transport upon being dragged through a full period. I show that an irrationally-filled system is quasiperiodic, and use the equivalence to show that 1D quasicrystals fit into the topological classification of free fermion systems known as the Tenfold Way. Using a free energy analysis I demonstrate that incommensurate charge order provides a new non-local growth mechanism for 1D quasicrystals, potentially greatly increasing the number of known, naturally-occurring, examples.

In the second half of this thesis I address the question of whether the 1D charge ordering mechanism, the Peierls instability, applies in dimensions higher than one, focussing on the prototypical 2D charge-ordered material niobium diselenide, $NbSe_2$. In this case I definitively rule out such 'weak-coupling' theories, and show that it is necessary to consider a model of a strong electron-phonon coupling dependent on both the ingoing and outgoing electron momenta and the electronic bands scattered between. The model provides the first consistent theoretical account of the full range of experimental results on the system, including a particle/hole asymmetric gap centred above the Fermi level which opens in one band only, the softening of phonon frequencies over a wide range of momenta, and the existence of a pseudogap regime over a range of temperatures, with the latter explained as suppression of charge order through fluctuations of the phonon field.

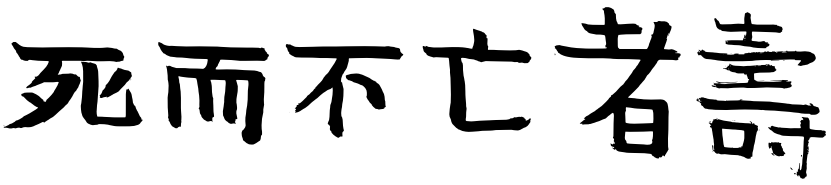
Dedication

David K. Lord, 1989 - 2012

Luke Parsons, 1987 - 2012

Bob Hargrave, 1949 - 2012

Balázs Györffy, 1938 - 2012



Acknowledgments

First and foremost I would like to acknowledge the help of my supervisor Jasper van Wezel, for regular helpful discussions on the many topics covered in this work, guidance in which directions the work should be taken, and for encouraging my interest in new topics which lead to the relative breadth of this Thesis.

I wish to acknowledge a great deal of assistance, both academic and otherwise, from John Hannay. As a former Bristol PhD student (now faculty member elsewhere) put it, “No-one would have received a PhD from the Bristol theory group in the past 30 years if it weren't for John Hannay”. My experience is in accordance with this sentiment.

My entire cohort of friends from PSI 2011 has been invaluable throughout this work, and in my research in general, bringing relevant results to my attention and providing answers to my many questions almost instantaneously at any time of the day.

I gratefully acknowledge advice from Prof. A. V. Chubukov regarding the relative magnitudes of the MMA diagrams, without which I would still be stuck in the nightmarish calculation of vertex corrections. For my work on the Hofstadter butterfly I am indebted to Denis Dalidovich, whose homework question on the Harper equation allowed me to make the link when the same mathematics came up in a seemingly unrelated field. I also wish to thank Michael Wilkinson and Svetlana Jitomirskaya for informative discussions on the same subject, and Nikitas Gidopoulos for many helpful discussions and his continued support and encouragement. Special thanks go to Prof. Eric Weeks of Emory University for the use of his C code to generate images of 2D quasicrystals. Finally, as promised some years ago, I wish to thank J. S. K. Briggs for teaching me to write the letter a, without which 5.6% of this thesis would not have been possible.

Author's Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED:

DATE:

Contents

List of Figures

List of Tables

Definitions Used in this thesis

ARPES Angle Resolved Photoemission Spectroscopy

BCS Bardeen Cooper Schrieffer

CDW Charge Density Wave

DOS Density of States

h.c. Hermitian conjugate

IQHE Integer Quantum Hall Effect

IXS Inelastic X-ray Scattering

LDA Local Density Approximation, a simple Density Functional Theory technique

MFT Mean Field Theory

MMA Mode-Mode Coupling Approximation

(R)BZ (Reduced) Brillouin Zone

RG Renormalization Group

RPA Random Phase Approximation

STM Scanning Tunneling Microscopy

STS Scanning Tunneling Spectroscopy

Throughout this thesis the symbol $\triangleq$ is used to mean 'equal by definition'. The symbol $\equiv$ is used to mean 'exact equivalence'.

I use natural units $\hbar = c = k_B = 1$ everywhere.

Bold symbols such as $\mathbf{k}$ designate vectors with the dimension of the relevant space. In the same context non-bold characters designate $k \triangleq (\omega, \mathbf{k})$, $x \triangleq (t, \mathbf{x})$. Accordingly,

$$\begin{aligned} \sum_k &\triangleq \sum_\omega \sum_{\mathbf{k}} \\ \delta_{kk'} &\triangleq \delta_{\omega\omega'}\delta_{\mathbf{k}\mathbf{k}'}. \end{aligned}$$

The metric $(+,-,-)$ is employed in Part II of this thesis.

The total number of states in the system N is not written explicitly, as it can be inferred from the context of sums and Fourier transforms as appropriate. The symbol β is defined globally to mean inverse temperature, $\beta = 1/T$.

Publications

At the time of writing, parts of Chapters 3 and 4 of this thesis appear in a pre-print

> F. Flicker and Jasper van Wezel, "Quasiperiodicity and 2D Topology in Charge Ordered Materials", arXiv:1408.4735 [cond-mat.str-el].

Some of the work in Part II of this thesis appears in

> F. Flicker and Jasper van Wezel, "Charge Order from Orbital-Dependent Coupling Evidenced by $NbSe_2$", Nature Communications 6, 7034 (2015).

Part I

Charge Ordering in One-Dimensional Systems

1 Introduction

Imagine a periodic array of atoms bonded to form a crystal. Where are the bonding electrons most likely to be found? That is, where is the charge density peaked? Naturally we would expect the answer to be that the charge density peaks at the locations of the ions, forming a periodic modulation maintaining the translational symmetry of the underlying lattice. In this thesis I will be considering what happens when this intuition breaks down, and the charge density spontaneously breaks the symmetry of the lattice to form 'charge order' or, specifically, a 'charge density wave' (CDW).

In fact, quasi-1D materials form CDWs generically by a mechanism known as the Peierls instability [1]. The charge density selects out a period larger than the underlying ion lattice. I treat such quasi-1D systems in a nearest-neighbour model in the first half of this thesis. Despite being perhaps the simplest model in all of condensed matter physics, I will demonstrate that it nevertheless exhibits a remarkable range of complex phenomena, from superconductivity, topological charge transport, and the quantum Hall effect, through to mathematical knots, fractal bandstructures, and quasicrystals. Underlying all of this is a fascinating topological structure.

In dimensions higher than one charge ordering is a rare phenomenon, with only a handful of examples known. The charge density in these cases breaks both the translational and rotational symmetry of the lattice, forming a wave-like striped pattern along a preferential direction. Despite the sparseness of known instances of CDWs in 2D and 3D there is a great deal of interest in them, owing in part to their appearing coincidentally with many cases of high-temperature superconductivity. In the second half of this thesis I will investigate whether the mechanism responsible for CDWs in 1D carries over to higher dimensions, focussing in particular on the prototypical quasi-2D CDW system niobium diselenide.

1.1 Charge Density Waves in 1D

In 1933 Rudolph Peierls demonstrated that (quasi-)1D chains are generically unstable to the formation of Charge Density Waves. Evidently he considered the proof trivial - its first published appearance is in a popular science book [1]. The reasoning is as follows: a 1D system will always benefit from having a gap open at the Fermi level E_F, since the gap will

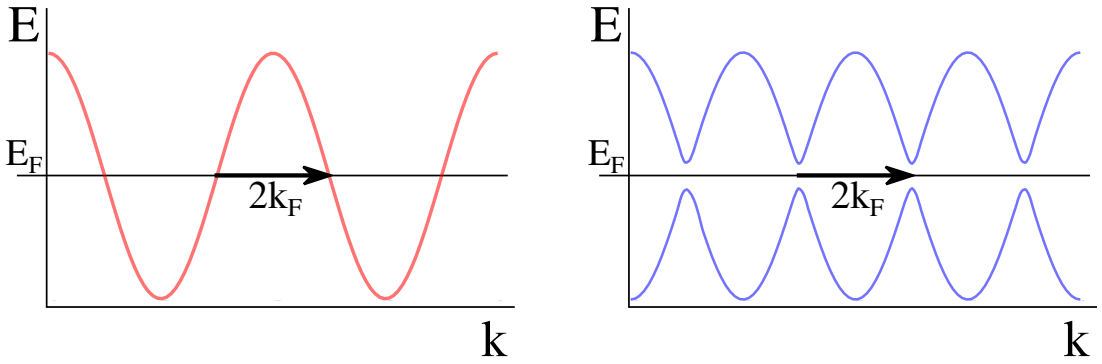


Figure 1: The Peierls instability in 1D. Left: a 1D cosine bandstructure at half filling, with two states separated by $2k_F$ marked. Right: coupling these states causes a CDW gap at E_F, so the occupied states are lowered in energy and the unoccupied states raised.

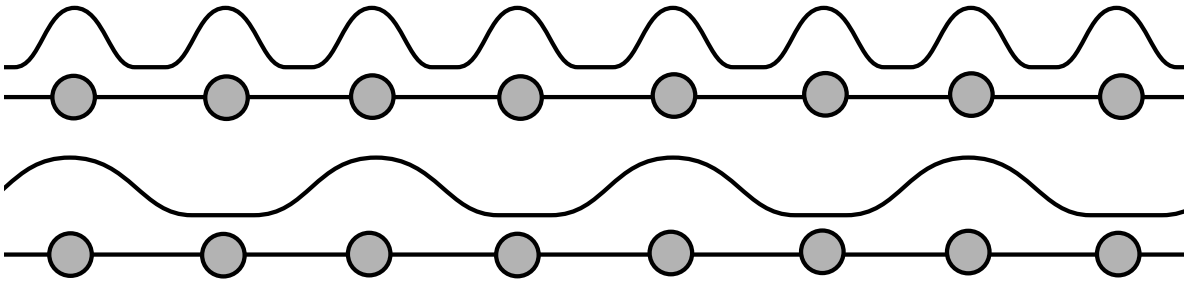


Figure 2: In 1D the Peierls instability causes the period of a crystal's charge density to alter from the lattice spacing a (top) to half the Fermi wavelength $1/2k_F$ (bottom, with $k_F = 1/4a$ here).

push all the occupied states down in energy at the cost of pushing the unoccupied states up - which at zero temperature is no cost at all[1]. Coupling all electron states separated by crystal momentum $2k_F$ opens such a gap and so is energetically favourable (see Figure 1). This process is known as Fermi surface nesting [2].

The effect in real space is to change the period of the CDW from the lattice spacing a to $1/2k_F$. The new charge-ordered pattern's period can either be an integer multiple of the lattice spacing, in which case it is known as a commensurate CDW, or it can be an irrational multiple, with the result known as an incommensurate CDW or ICDW. The situation is shown in Figure 2. In reality the CDW provides a small modulation to the charge density field which is still peaked at the ionic locations due to the many non-bonding electrons, but the system's periodicity is still affected. The CDW period can be either commensurate or incommensurate with the lattice.

[1] In a full treatment the ions will adjust to the new charge configuration, and these adjustments will cost energy.

In dimensions higher than one the generic benefit of nesting no longer exists. With a single fixed wavevector the number of states required to be coupled scales as the power of the dimension, but the energy benefit from the coupling happens only at isolated points. Part II of this thesis concerns the driving mechanism for CDW formation in higher-dimensional systems, where I argue that the Peierls mechanism is the incorrect starting point in the general case.

Whatever mechanism drives a CDW transition, the end result will be a joint distortion of the electron field plus a shifting of the ions. If the driving mechanism is nesting, an electronic effect, the charge field distorts and the ions shift to minimize their energy in the distorted field. If the CDW order is phonon driven the ions shift locations and the electron field distorts to minimize its energy in the new ionic arrangement. A permanently distorted ion lattice can be thought of as a softening of a phonon energy to zero at some nonzero $\mathbf{q}$ vector, referred to as a Kohn anomaly [3].

1.2 Structure of Part I

The first half of this thesis concerns the topology of charge ordering in 1D. In Chapter 2 I give a more detailed account of the Peierls mechanism leading to charge order in a 1D system at half-filling, treated in the mean-field limit, and highlight the well-known mathematical analogy to the BCS theory of superconductivity. In Chapter 3 I extend the working from half-filling to arbitrary rational filling, and find an unexpected equivalence to the 2D quantum Hall effect. I use the result to investigate quantized adiabatic particle transport by CDWs. In Chapter 4 I further extend the results to arbitrary *irrational* fillings and find that the resulting incommensurate CDWs break not only the translational symmetry of the lattice, but in fact *all* translational symmetries, forming so-called 'quasicrystals'. I use the equivalence to address a recent debate regarding whether and how quasicrystals fit into the classification of free fermion systems known as the Tenfold Way. Using a free energy analysis I further demonstrate that incommensurate CDWs can generate a previously unknown nonlocal growth mechanism for quasicrystals, and show that the resulting class of systems could potentially massively increase the number of known natural quasicrystals. In Chapter 5 I consider the results in the wider context and provide some concluding remarks.

2 Half-Filling: BCS Theory

In this Chapter I analyse the Peierls mechanism, in half-filled quasi-1D systems, from the viewpoint of quantum field theory applied to a simple model of spinless fermions with nearest-neighbour interactions. While the working of Part I of this thesis is ostensibly carried out with a Coulomb repulsion between electrons constituting the driving force for the CDW transition, I first establish via a general field theoretical argument in Section 2.1 that in fact it is just as easy to consider a continuous range of possible couplings from purely Coulombic to purely electron-phonon. In Section 2.2 I treat the interaction in the mean field limit, demonstrating the equivalence to the BCS theory of superconductivity. I then solve for the mean field self-consistently in Section 2.3. Finally, I consider a possible fundamental problem regarding the existence of charge order in 1D raised by a renormalization group analysis. Coming full circle I waylay fears by appeal to the work of Section 2.1.

2.1 Inter-Electron Couplings

The overview of CDW formation in Chapter 1 suggests that any arbitrary interaction which couples states at the Fermi energy E_F in 1D will cause a CDW phase transition at sufficiently low temperature. Common interactions to consider are Coulomb repulsion between electrons and electron-phonon coupling. While the precise form of the coupling is indeed somewhat by-the-by in 1D, there are certain arguments which can be made much more precise by establishing here the absolute operational equivalence of the two couplings just mentioned. I will in fact consider the most general form of the argument which applies in arbitrary dimensions; this also pleasingly sets the tone for the field-theoretical analysis returned to repeatedly in the coming chapters.

The Coulomb repulsion and electron-phonon coupling can be conveniently combined into a coupling of free electrons to a single bosonic field, a combined lattice distortion and charge ordering. To see this, consider the action S of a generic field theory of spinless fermions in an arbitrary dimension d with both phonons and an on-site Coulomb interaction:

$$S = \sum_k \psi_k^\dagger G_k^{-1} \psi_k + \sum_q \varphi_q^\dagger D_q^{-1} \varphi_q + \sum_{kq} g_{\mathbf{q}} \varphi_q \psi_{k+q}^\dagger \psi_k + \sum_k h_{\mathbf{k}} n_k n_{-k} \tag{1}$$

where the first term governs the free (non-interacting) electrons, the second governs the

free phonon displacements, the third gives the electron-phonon coupling, and the fourth gives the Coulomb repulsion between electrons. For this section only I will use non-bold k to denote the combined crystal momentum and frequency, $k = (\omega, \mathbf{k})$. The Grassman field $\psi_k^\dagger$ for electrons with energy and momentum $k = (\omega, \mathbf{k})$ is created by the action of the field operator $\hat{\psi}_k^\dagger$ on the vacuum state. The corresponding electron number operator is $\hat{n}_q \triangleq \sum_k \hat{\psi}_{k+q}^\dagger \hat{\psi}_k$, and $\varphi_q^\dagger \equiv \varphi_{-q}$ is a (complex, bosonic) phonon field with energy and momentum $q = (\Omega, \mathbf{q})$. The coupling parameters are static but allowed to vary in space. The electron and phonon propagators are G_k and D_q respectively; their precise forms are not relevant to the present analysis.

The Coulomb term derives from the real space form

$$S_{Coulomb} = \int \mathrm{d}^{d+1} r \mathrm{d}^{d+1} r' h\left(\mathbf{r} - \mathbf{r}'\right) \psi_r^\dagger \psi_r \psi_{r'}^\dagger \psi_{r'}$$

with $r = (t, \mathbf{r})$ and spatial dimension d, which is valid for any off-site[2] interaction $h\left(\mathbf{r} - \mathbf{r}'\right)$. The reciprocal space form in Equation 1 can be made more manageable by use of a Hubbard-Stratanovitch transformation [4, 5]: consider a complex bosonic field Δ_k, obeying $\Delta_k = \Delta_{-k}^\dagger$, which satisfies

$$\frac{1}{\mathcal{N}} \int \mathscr{D}\Delta \exp\left(-\sum_k \left(-\frac{1}{4} h_{\mathbf{k}}\right) \Delta_{-k} \Delta_k\right) = 1.$$

The functional integral is over all possible Δ configurations, so adding a second field fixed with respect to the integral leaves the result unchanged, just as the change of variables $x \to x + c$ leaves infinite integrals $\int_{-\infty}^{\infty} \mathrm{d}x$ unaffected. Shifting $\Delta_k \to \Delta_k - n_{-k}$ gives

$$\frac{1}{\mathcal{N}} \int \mathscr{D}\Delta \exp\left(-\sum_k \left(-h_{\mathbf{k}}\right) \left(\frac{1}{2}\Delta_{-k} - n_k\right) \left(\frac{1}{2}\Delta_k - n_{-k}\right)\right) = 1.$$

Expanding the parentheses and taking the Δ-independent term out of the functional integral gives

$$\frac{1}{\mathcal{N}} \int \mathscr{D}\Delta \exp\left(-\sum_k \left(-h_{\mathbf{k}}\right) \left(\frac{1}{4}\Delta_{-k}\Delta_k - \frac{1}{2}\Delta_k n_k - \frac{1}{2}\Delta_{-k} n_{-k}\right)\right) = \exp\left(-\sum_k h_{\mathbf{k}} n_k n_{-k}\right).$$

[2]The working also holds for on-site interaction, but care has to be taken to include spins explicitly because of the Pauli exclusion principle $\hat{\psi}_{r\sigma}^\dagger \hat{\psi}_{r\sigma}^\dagger \equiv 0$ for equal spins σ. I avoid this case for brevity.

Considering this in the context of the partition function

$$\mathscr{Z} = \frac{1}{\mathcal{N}} \int \mathscr{D}\psi \mathscr{D}\varphi \exp\left(-S\left[\psi, \varphi\right]\right)$$

it is clear that the Hubbard-Stratanovitch transformation has removed the quartic Coulomb term at the expense of introducing a coupling to a new field Δ_k:

$$\begin{aligned} S_{Coulomb}\left[n\right] &= \sum_k h_{\mathbf{k}} n_k n_{-k} \\ &\downarrow \\ S_{Coulomb}\left[n, \Delta\right] &= \sum_k h_{\mathbf{k}} \left(-\frac{1}{4}\Delta_{-k}\Delta_k + \Delta_k n_k\right). \end{aligned}$$

Note that the result is exact. The new bosonic field Δ_k replaces an n_k term in the action and represents the electron charge density. Overall then, expanding $n_q = \sum_k \psi^\dagger_{k+q}\psi_k$, the result is

$$S = \sum_k \psi^\dagger_k G^{-1}_k \psi_k + \sum_q \varphi^\dagger_q D^{-1}_q \varphi_q + \sum_q \Delta^\dagger_q \left(-\frac{1}{4}h_{\mathbf{q}}\right)\Delta_q + \sum_{kq} g_{\mathbf{q}}\varphi_q \psi^\dagger_{k+q}\psi_k + \sum_{kq} h_{\mathbf{q}}\Delta_q \psi^\dagger_{k+q}\psi_k$$

where I have assumed $h_{\mathbf{q}} = h_{-\mathbf{q}}$ for simplicity.

The coupling of the electrons to the fields φ and Δ takes an identical form, and it would seem sensible to write a combined field defined by

$$g_{\mathbf{q}}\varphi_q + h_{\mathbf{q}}\Delta_q \triangleq \mathfrak{g}_{\mathbf{q}} A_q. \tag{2}$$

In fact the substitution proves quite remarkable: it turns out that φ and Δ can both be eliminated entirely, resulting in a theory solely in terms of electrons ψ and a combined charge density/phonon displacement field A. To see this, first use Equation 2 to eliminate Δ:

$$
\begin{aligned}
S\left[\psi,\varphi,A\right] &= \sum_k \psi_k^\dagger G^{-1}\psi_k + \sum_q \left\{ \varphi_q^\dagger \left(D_q^{-1} - \frac{g_{\mathbf{q}}^2}{4h_{\mathbf{q}}} \right) \varphi_q - \frac{\mathfrak{g}_{\mathbf{q}}^2}{4h_{\mathbf{q}}} A_q^\dagger A_q \right. \\
&\left. + \frac{g_{\mathbf{q}}\mathfrak{g}_{\mathbf{q}}}{4h_{\mathbf{q}}} \left(A_q^\dagger \varphi_q + \varphi_q^\dagger A_q \right) + \sum_k \mathfrak{g}_{\mathbf{q}} A_q \psi_{k+q}^\dagger \psi_k \right\}.
\end{aligned}
$$

Now, bearing in mind that the partition function

$$
\mathscr{Z} = \frac{1}{\mathcal{N}} \int \mathscr{D}\psi \mathscr{D}\varphi \mathscr{D}A \exp\left(-S\left[\psi,\varphi,A\right]\right)
$$

is insensitive to linear shifts of the fields, the φ field can be shifted using

$$
\varphi_q' \triangleq \varphi_q - \frac{g_{\mathbf{q}}\mathfrak{g}_{\mathbf{q}}}{4h_{\mathbf{q}}} \left(D_q^{-1} - \frac{g_{\mathbf{q}}^2}{4h_{\mathbf{q}}} \right)^{-1} A_q
$$

which leads, after a bit of algebra, to

$$
\begin{aligned}
S\left[\psi,\varphi',A\right] &= \sum_k \psi_k^\dagger G_k^{-1}\psi_k + \sum_q \left\{ \varphi_q'^\dagger \left(D_q^{-1} - \frac{g_{\mathbf{q}}^2}{4h_{\mathbf{q}}} \right) \varphi_q' \right. \\
&\left. - A_q^\dagger \left(\frac{\mathfrak{g}_{\mathbf{q}}^2}{4h_{\mathbf{q}}} + \left(\frac{g_{\mathbf{q}}\mathfrak{g}_{\mathbf{q}}}{4h_{\mathbf{q}}} \right)^2 \left(D_q^{-1} - \frac{g_{\mathbf{q}}^2}{4h_{\mathbf{q}}} \right)^{-1} \right) A_q + \sum_k \mathfrak{g}_{\mathbf{q}} A_q \psi_{k+q}^\dagger \psi_k \right\}.
\end{aligned}
$$

The functional integral over the φ' field is now a simple Gaussian, and cancels with its normalization in front of the partition function. The final result, then, is

$$
S\left[\psi,A\right] = \sum_k \psi_k^\dagger G_k^{-1}\psi_k + \sum_q A_q^\dagger B_q^{-1} A_q + \sum_{kq} \mathfrak{g}_{\mathbf{q}} A_q \psi_{k+q}^\dagger \psi_k \tag{3}
$$

where the propagator of the A field is

$$
B_q \triangleq \left(\frac{g_{\mathbf{q}}}{\mathfrak{g}_{\mathbf{q}}} \right)^2 D_q - \frac{4h_{\mathbf{q}}}{\mathfrak{g}_{\mathbf{q}}^2}. \tag{4}
$$

The physical significance of this result is that the phonon and charge density fields naturally combine into a single object, the excitations of which are well-defined quasiparticles. An illustration is given in Figure 3.

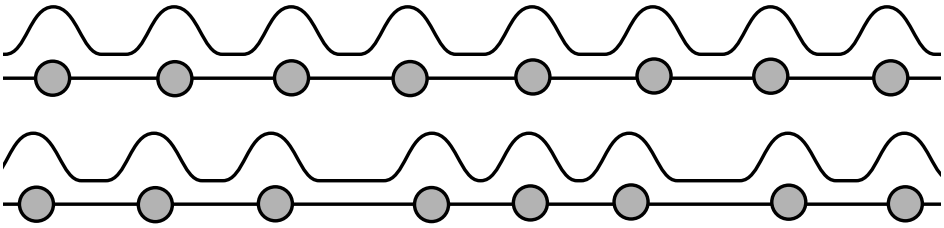


Figure 3: The phonon and charge density fields can be combined into a single field with two degrees of freedom without approximation. The result provides a natural way to describe combined distortions of the lattice and the electron field. Top: undistorted lattice; bottom: lattice with combined charge density/lattice distortions.

This makes sense physically: if a localized lattice distortion were to propagate through the system we would expect it to be accompanied by a localized distortion of the charge, and vice versa. The combined lattice distortion / charge distortion would look like a single object, which we can now describe as a quasiparticle excitation of the A field. Note that no approximations have been introduced - the result is exact.

Various ground states are possible with different values or forms of the couplings g and h. The study of their stability at the mean field level is called 'g-ology' [6]. The name would normally be an amusing play on 'geology' based on the fact that both couplings are traditionally labelled g; I have ruined the joke by calling the Coulomb coupling h. An illustrative schematic is reproduced from Grüner [6] in Figure 4. g-ology has gone out of fashion with the advent of renormalization group analysis, which suggests amongst other things that the mean-field result of a gapped CDW system is incorrect, and the true ground state at sufficiently small couplings is a gapless 'Luttinger Liquid' [7]. I return to this point in Section 2.4. Regardless, g-ology remains relevant today with regard to the (quasi-2D) high temperature superconducting cuprates and pnictides, which contain the possibility that charge density waves, 'spin density waves', and superconductivity all compete for control of the Fermi surface, and that both phonons and Coulomb repulsion may be playing a part (for a thorough review see [8]).

2.2 Mean Field Theory

Having considered a general field-theoretical argument in the previous section I will now focus on the specific case of charge ordering in 1D, by considering a system of spinless

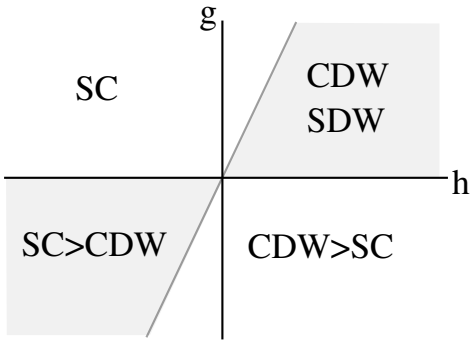


Figure 4: A schematic representation of the possible mean-field ground states of a 1D system as a function of the electron-phonon coupling g and the Coulomb interaction strength h, reproduced from [6] after [9]. The rather nebulous designations in the lower half indicate that the response functions for the 'less than' states are divergent but less so than the other state. 'SC' designates superconductivity and 'SDW' denotes a spin density wave, which is exactly analogous to a CDW but with spins ordering rather than charges.

'electrons' with nearest neighbour Coulomb repulsion[3]. From now on n will designate real-space site n on a 1D lattice, and k will denote a 1D crystal momentum. With $\psi_n^\dagger$ the electron creation operator at lattice site n, the Hamiltonian is

$$\hat{H} = -t \sum_{\langle nm \rangle} \left(\hat{\psi}_n^\dagger \hat{\psi}_m + \hat{\psi}_m^\dagger \hat{\psi}_n \right) - \mu \sum_n \hat{\psi}_n^\dagger \hat{\psi}_n + h \sum_{\langle nm \rangle} \hat{\psi}_n^\dagger \hat{\psi}_n \hat{\psi}_m^\dagger \hat{\psi}_m \tag{5}$$

with $\langle nm \rangle$ indicating the sum is over nearest neighbours only. The hopping parameter t can be taken to be real in the absence of magnetic fields; the chemical potential μ is chosen to be $\mu = h$ to ensure the system stays at half-filling when the interaction h is turned on. The lattice constant is set to unity. For positive h the final term biases against occupation of neighbouring sites, and hence represents the repulsion of the electrons. Since spin is neglected, double occupation of the same site is disallowed.

Fourier transforming by writing

$$\hat{\psi}_n^\dagger = \sum_k \exp\left(-ikn\right) \hat{\psi}_k^\dagger \tag{6}$$

and noting that the sum runs over

$$m = n \pm 1$$

[3]Of course spinless electrons do not exist, but spin plays no part in this story, so to avoid factors of two everywhere I will pretend they do. The name 'Coulomb' is somewhat arbitrary as all repulsions are the same, modulo strength, if the distance is fixed.

gives the result

$$\hat{H} = \sum_k \left(-2t\cos\left(k\right) - \mu\right)\hat{\psi}_k^\dagger\hat{\psi}_k + h\sum_{kk'k''} 2\cos\left(k-k''\right)\hat{\psi}_k^\dagger\hat{\psi}_{k''}\hat{\psi}_{k'}^\dagger\hat{\psi}_{k+k'-k''}. \tag{7}$$

The quartic interaction term cannot be dealt with directly. To proceed it is necessary to decouple it by means of a mean field approximation:

$$\begin{aligned}\hat{\psi}_k^\dagger\hat{\psi}_{k''}\hat{\psi}_{k'}^\dagger\hat{\psi}_{k+k'-k''} &= \left\langle\hat{\psi}_k^\dagger\hat{\psi}_{k''}\right\rangle\hat{\psi}_{k'}^\dagger\hat{\psi}_{k+k'-k''} + \hat{\psi}_k^\dagger\hat{\psi}_{k''}\left\langle\hat{\psi}_{k'}^\dagger\hat{\psi}_{k+k'-k''}\right\rangle \\ &\quad - \left\langle\hat{\psi}_k^\dagger\hat{\psi}_{k+k'-k''}\right\rangle\hat{\psi}_{k'}^\dagger\hat{\psi}_{k''} - \hat{\psi}_k^\dagger\hat{\psi}_{k+k'-k''}\left\langle\hat{\psi}_{k'}^\dagger\hat{\psi}_{k''}\right\rangle\end{aligned} \tag{8}$$

i.e. by considering all possible ways in which two of the four legs in the 4-vertex Coulomb interaction can be replaced by a mean field of the form $\left\langle\hat{\psi}^\dagger\hat{\psi}\right\rangle$. Note that making $\left\langle\hat{\psi}^\dagger\hat{\psi}\right\rangle$ type contractions was not the only possible way to approximate the quartic term by a quadratic; this form was chosen as an ansatz on the basis that $\left\langle\hat{\psi}^\dagger\hat{\psi}\right\rangle$ will have a physical significance as an electron density later on[4]. An alternative pairing of the form $\left\langle\hat{\psi}^\dagger\hat{\psi}^\dagger\right\rangle$ would be considered were we concerned with BCS superconductivity [10], in which case the coupling $\Delta = \left\langle\hat{\psi}^\dagger\hat{\psi}^\dagger\right\rangle$ would be the Cooper pair potential. The BCS case and the CDW case in fact run completely in parallel, and I will point out the analogies when instructive. The quartic term in Equation 7, with the Mean Field Theory (MFT) approximation of Equation 8, after a little neatening, takes the form

$$\begin{aligned}\sum_{kk'k''}\cos\left(k-k''\right)\hat{\psi}_k^\dagger\hat{\psi}_{k''}\hat{\psi}_{k'}^\dagger\hat{\psi}_{k+k'-k''} &= \sum_{kk'k''}\cos\left(k''\right)\left\langle\hat{\psi}_{k'+k''}^\dagger\hat{\psi}_{k'}\right\rangle\hat{\psi}_k^\dagger\hat{\psi}_{k+k''} \\ &\quad + \sum_{kk'k''}\cos\left(k''\right)\hat{\psi}_{k+k''}^\dagger\hat{\psi}_k\left\langle\hat{\psi}_{k'}^\dagger\hat{\psi}_{k'+k''}\right\rangle \\ &\quad - \sum_{kk'k''}\cos\left(k-k'\right)\left\langle\hat{\psi}_{k'+k''}^\dagger\hat{\psi}_{k'}\right\rangle\hat{\psi}_k^\dagger\hat{\psi}_{k+k''} \\ &\quad - \sum_{kk'k''}\cos\left(k-k'\right)\hat{\psi}_{k+k''}^\dagger\hat{\psi}_k\left\langle\hat{\psi}_{k'}^\dagger\hat{\psi}_{k'+k''}\right\rangle\end{aligned}$$

[4] It is a point of interest that this technique corresponds to applying Wick's theorem, or, equivalently, a Hubbard-Stratanovitch transformation combined with the approximation that the Hubbard-Stratanovitch boson operator be replaced by its mean field expectation value. By the method developed in Section 2.1 the current model could be extended to include electron-phonon coupling with no further work.

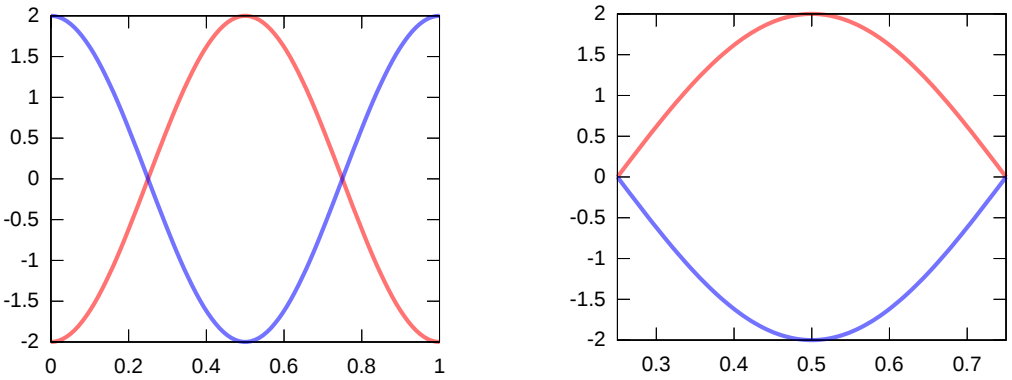


Figure 5: For a half-filled system we anticipate that states k and $k+Q$ with $Q = 2\pi/2$ will be coupled. The original bandstructure $\epsilon_k = -2t\cos(k)$ is shown in red, with ϵ_{k+Q} in blue ($t = 1$). The problem reduces to considering two copies of a reduced Brillouin zone of half the size (shown on the right). The units of k are 2π.

where it is clear that the first and second terms are Hermitian conjugates, as are the third and fourth, so the Hamiltonian is Hermitian. It takes the form

$$\hat{H} \quad = \sum_k \epsilon_k \hat{\psi}_k^\dagger \hat{\psi}_k + \; 2h \sum_{kk'k''} \left\{ \left(\cos\left(k''\right) - \cos\left(k - k'\right)\right) \left\langle \hat{\psi}_{k'+k''}^\dagger \hat{\psi}_{k'} \right\rangle \hat{\psi}_k^\dagger \hat{\psi}_{k+k''} + h.c. \right\} \quad (9)$$

where

$$\epsilon_k \triangleq -2t\cos(k) - \mu.$$

Having approximated the Hamiltonian into a diagonal (quadratic) form I will proceed in the next section to ensure that the mean field theory is self-consistent.

2.3 Self-Consistent Solution

The 1D band considered so far has the chemical potential μ set such that the system is half-filled. For the time being I will *assume* that the system suffers a Peierls instability due to coupling between states at E_F, and will then show this to be a self-consistent solution to the mathematical model. For the half-filling case, then, I take as an ansatz the coupling between states separated by $Q = \pi$, and substitute this fixed wavevector, $q = Q$, into the mean field Hamiltonian of Equation 9. The situation is shown in Figure 5.

Anticipating a doubling of the real-space periodicity, it is convenient to rewrite the Hamiltonian in a 'reduced Brillouin zone' (RBZ) of half the size. The unperturbed Hamiltonian,

for example, is written

$$\begin{aligned}\sum_{k=0}^{2\pi}\epsilon_k\hat{\psi}_k^\dagger\hat{\psi}_k &= \sum_{k=0}^{\pi}\epsilon_k\hat{\psi}_k^\dagger\hat{\psi}_k+\sum_{k=\pi}^{2\pi}\epsilon_k\hat{\psi}_k^\dagger\hat{\psi}_k \\ &= \sum_{k=0}^{\pi}\left\{\epsilon_k\hat{\psi}_k^\dagger\hat{\psi}_k+\epsilon_{k+\pi}\hat{\psi}_{k+\pi}^\dagger\hat{\psi}_{k+\pi}\right\}.\end{aligned}$$

The full Hamiltonian in the RBZ is most conveniently written in matrix form:

$$\hat{H}=\sum_{k\in RBZ}\left(\hat{\psi}_k^\dagger,\hat{\psi}_{k+\pi}^\dagger\right)H_k\begin{pmatrix}\hat{\psi}_k\\ \hat{\psi}_{k+\pi}\end{pmatrix}$$

with

$$\begin{aligned}H_k &= \begin{pmatrix}\epsilon_k & \triangle_k\\ \triangle_k & \epsilon_{k+\pi}\end{pmatrix}\\ \triangle_k &\triangleq \Delta_k+\Delta_k^*=2\mathfrak{Re}\left(\Delta_k\right)\end{aligned}$$

and the self-consistency condition

$$\Delta_k \triangleq -2h\sum_{k'}\left(1+\cos\left(k-k'\right)\right)\left\langle\hat{\psi}_{k'+\pi}^\dagger\hat{\psi}_{k'}\right\rangle. \tag{10}$$

Note that the k' sum is still over the full Brillouin zone. The term $\triangle_k$ is a real-valued function of k, where $2\triangle_k$ gives the CDW energy gap in the system. For a general bandstructure the k-dependence is important, but for the simple case of cosine dispersion considered here it should be possible to find a k-independent solution, since the band only crosses E_F at one point in the RBZ.

The off-diagonal terms do not conserve crystal momentum - they relate to processes in which electrons scatter and lose crystal momentum to the charge density field Δ. Alternatively they can be thought of as corresponding to 'reduced Umklapp processes' - that is, Umklapp processes scattering between RBZs. In BCS theory these terms would be replaced

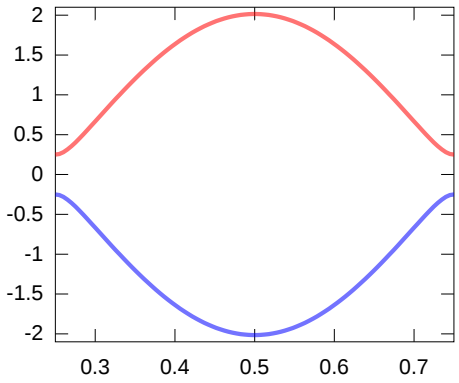


Figure 6: The quasiparticle spectrum in the reduced Brillouin zone, setting a CDW gap of $2\triangle = 0.5t$ (here $t = 1$). The bands split at the degeneracies. The blue states are occupied and the red unoccupied, so opening the CDW gap has lowered the energy of the system. This is the Peierls instability (*cf.* Figure 5 which has $2\triangle = 0$).

by the fermion number non-conserving Cooper pair creation and annihilation operators. The presence of off-diagonal terms in H_k shows that the electrons and holes are no longer good quasiparticles upon turning on the Coulomb interaction. Instead we should rotate to a diagonal basis and find the well-defined quasiparticles of the interacting system. Defining the creation operators of these quasiparticles as $\gamma_k^\dagger$, the desired form is

$$\hat{H} = \sum_k \left(\hat{\gamma}_k^\dagger, \hat{\gamma}_{k+\pi}^\dagger\right) D_k \begin{pmatrix} \hat{\gamma}_k \\ \hat{\gamma}_{k+\pi} \end{pmatrix} \tag{11}$$

$$D_k \triangleq \begin{pmatrix} \sqrt{\epsilon_k^2 + \triangle_k^2} & 0 \\ 0 & -\sqrt{\epsilon_k^2 + \triangle_k^2} \end{pmatrix} \tag{12}$$

$$\begin{pmatrix} \hat{\gamma}_k \\ \hat{\gamma}_{k+\pi} \end{pmatrix} \triangleq U_k \begin{pmatrix} \hat{\psi}_k \\ \hat{\psi}_{k+\pi} \end{pmatrix} \tag{13}$$

with

$$U_k \triangleq \frac{1}{\sqrt{2}} \begin{pmatrix} \frac{\alpha_k}{\sqrt{1+\alpha_k^2}-\sqrt{1+\alpha_k^2}} & \frac{\alpha_k}{\sqrt{1+\alpha_k^2}+\sqrt{1+\alpha_k^2}} \\ \frac{1-\sqrt{1+\alpha_k^2}}{\sqrt{1+\alpha_k^2}-\sqrt{1+\alpha_k^2}} & \frac{1+\sqrt{1+\alpha_k^2}}{\sqrt{1+\alpha_k^2}+\sqrt{1+\alpha_k^2}} \end{pmatrix}$$

$$\alpha_k \triangleq \frac{\triangle_k}{\epsilon_k}.$$

The CDW energy gap $2\triangle$ is apparent from the quasiparticle spectrum in Equation 12, plotted in Figure 6. It is this which must be solved for self-consistently. From Equation 13 we have that

$$\begin{aligned}\hat{\psi}^{\dagger}_{k+\pi} &= \hat{\gamma}^{\dagger}_{k} U_{12} + \hat{\gamma}^{\dagger}_{k+\pi} U_{22} \\ \hat{\psi}_{k} &= \left(U^{\dagger}\right)_{11} \hat{\gamma}_{k} + \left(U^{\dagger}\right)_{12} \hat{\gamma}_{k+\pi}.\end{aligned}$$

The self-consistency condition, Equation 10, needs to be written in terms of these new quasiparticle operators. The relevant term is

$$\begin{aligned}\left\langle \hat{\psi}^{\dagger}_{k+\pi} \hat{\psi}_{k} \right\rangle &= \frac{\alpha_k}{\sqrt{1+\alpha_k^2}} \left(\left\langle \hat{\gamma}^{\dagger}_{k} \hat{\gamma}_{k} \right\rangle - \left\langle \hat{\gamma}^{\dagger}_{k+\pi} \hat{\gamma}_{k+\pi} \right\rangle \right) \\ &\quad + \left(\frac{1}{\sqrt{1+\alpha_k^2}} + 1 \right) \left(\left\langle \hat{\gamma}^{\dagger}_{k} \hat{\gamma}_{k+\pi} \right\rangle + \left\langle \hat{\gamma}^{\dagger}_{k+\pi} \hat{\gamma}_{k} \right\rangle \right). \end{aligned} \tag{14}$$

The power of the diagonalization method derives from the fact that the γ particles are now well-defined fermions, so their occupation numbers are determined by Fermi-Dirac statistics taking as input the energies of D_k in Equation 12. Thus we see that

$$\begin{aligned}\left\langle \hat{\gamma}^{\dagger}_{k} \hat{\gamma}_{k+\pi} \right\rangle &= 0 \\ \left\langle \hat{\gamma}^{\dagger}_{k+\pi} \hat{\gamma}_{k} \right\rangle &= 0 \\ \left\langle \hat{\gamma}^{\dagger}_{k} \hat{\gamma}_{k'} \right\rangle &= f\left(\sqrt{\epsilon_k^2 + \triangle_k^2}\right) \delta_{k,k'} \\ \left\langle \hat{\gamma}^{\dagger}_{k+\pi} \hat{\gamma}_{k'+\pi} \right\rangle &= f\left(-\sqrt{\epsilon_k^2 + \triangle_k^2}\right) \delta_{k,k'} \end{aligned} \tag{15}$$

with $f(E)$ the Fermi-Dirac distribution. Substituting into Equation 14 gives

$$\begin{aligned}\left\langle \hat{\psi}^{\dagger}_{k+\pi}\hat{\psi}_{k}\right\rangle &= \frac{\triangle_k}{\sqrt{\epsilon_k^2+\triangle_k^2}}\left(f\left(\sqrt{\epsilon_k^2+\triangle_k^2}\right)-f\left(-\sqrt{\epsilon_k^2+\triangle_k^2}\right)\right)\\ &\equiv -\frac{\triangle_k}{\sqrt{\epsilon_k^2+\triangle_k^2}}\tanh\left(\frac{1}{2}\beta\sqrt{\epsilon_k^2+\triangle_k^2}\right)\end{aligned}$$

and so (twice the real part of) the self-consistency condition reads

$$\triangle_k = 4h\sum_{k'}\left(1+\cos\left(k-k'\right)\right)\frac{\triangle_{k'}}{\sqrt{\epsilon_{k'}^2+\triangle_{k'}^2}}\tanh\left(\frac{1}{2}\beta\sqrt{\epsilon_{k'}^2+\triangle_{k'}^2}\right).$$

As mentioned above it seems physically intuitive that there should exist a self-consistent gap function independent of wavevector k, since the dispersion only crosses the Fermi level once in the RBZ. Recalling that ϵ_k has period 2π, if $\triangle \neq \triangle_k$ the term multiplied by the cosine drops out since the k' sum is over a full 2π cycle in which the cosine spends half the time positive, half negative. As a result there is indeed a nontrivial ($\triangle \neq 0$) solution of the form

$$1 = 4h\sum_{k}\frac{1}{\sqrt{\epsilon_k^2+\triangle^2}}\tanh\left(\frac{1}{2}\beta\sqrt{\epsilon_k^2+\triangle^2}\right). \tag{16}$$

This can be solved numerically without too much difficulty. Noting the expression's similarity to the BCS gap equation [10, 7], however, it should be possible to find an analytic solution following the standard arguments applied in that case. Taking the continuum limit to turn the sum into an integral, only the energies close to E_F (within a cutoff Λ) contribute significantly. Writing the density of states $g(\epsilon)$ the approximation is:

$$\sum_k \approx \int \mathrm{d}k = \int \mathrm{d}\epsilon\frac{\mathrm{d}k}{\mathrm{d}\epsilon} = \int \mathrm{d}\epsilon g(\epsilon) \approx g(E_F)\int_{-\Lambda}^{\Lambda}\mathrm{d}\epsilon.$$

The 1D density of states at E_F is given by

$$
\begin{aligned}
g\left(E_F\right) &= \frac{1}{2\pi}\left.\frac{1}{\left|\nabla_k \epsilon\right|}\right|_{E_F} \\
&= \frac{1}{2\pi}\frac{2}{\left|2t\sin\left(k_F\right)\right|} \\
&= \frac{1}{2\pi t}\frac{1}{\left|\sin\left(Q/2\right)\right|} \\
&= \frac{1}{2\pi t}.
\end{aligned} \tag{17}
$$

In the zero temperature limit $\beta \to \infty$ it follows that $\tanh\left(\beta x\right) \approx 1$ (x is any finite number with dimensions of energy), and so

$$
\begin{aligned}
\frac{1}{4h} &= g\left(E_F\right)\int_{-\Lambda}^{\Lambda} \mathrm{d}\epsilon \frac{1}{\sqrt{\epsilon^2+\triangle^2}} \\
&= 2g\left(E_F\right)\operatorname{asinh}\left(\frac{\Lambda}{\triangle}\right)
\end{aligned}
$$

or

$$
\begin{aligned}
\triangle &= \Lambda/\sinh\left(1/8g\left(E_F\right)h\right) \\
&\approx 2\Lambda\exp\left(-\frac{\pi}{4}\frac{t}{h}\right)
\end{aligned}
$$

for $h \ll t$. This expression, identical to the BCS gap equation [10, 7], is compared to the numerical result in Figure 7 with good agreement.

In the results so far I have considered the phase of the order parameter to be set to zero. In fact the phase can have a large effect: defining

$$
\begin{aligned}
\Delta &= \left|\Delta\right|\exp\left(i\theta\right) \\
\therefore \triangle &= 2\left|\Delta\right|\cos\left(\theta\right)
\end{aligned}
$$

gap closures occur at $\theta = \left(2p+1\right)\pi/2$ for integer p. This has a significance to the case of

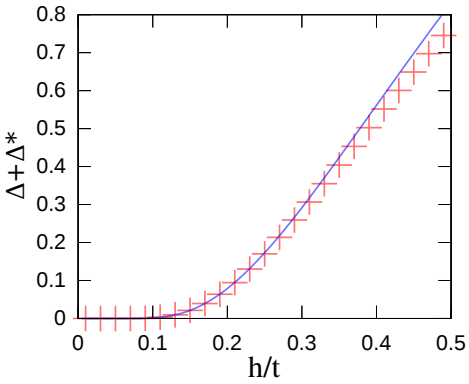


Figure 7: The numerical solution of the self-consistent half-filling gap equation $\triangle(h) \triangleq \Delta(h) + \Delta(h)^*$ (red crosses), with the curve $\triangle(h) = 4\exp\left(-\frac{\pi t}{4h}\right)$ overlayed. This corresponds to the approximate BCS expression $\triangle \approx 2\Lambda \exp\left(-1/\left(8g\left(\epsilon_F\right)h\right)\right)$ with bandwidth $2\Lambda = 4t$ (hopping parameter t) and density of states at the Fermi surface $g\left(E_F\right) \approx 1/2\pi t$. The agreement is best at small h/t, as assumed in the BCS approximation.

general filling, discussed at length in Chapter 3.

2.4 A Brief Comment on Renormalization Group Flow

The theoretical story told so far seems believable, and we know that many real quasi-1D materials such as $NbSe_3$, KCP, and TTF-TCNQ develop CDW instabilities at low temperature [11, 12, 13, 14, 15]. A full renormalization group (RG) analysis, however, suggests that the mean field theory presented here is misleading in its predictions regarding spinless fermions at weak coupling. The argument is thoroughly explained in [7].

We found that an arbitrarily weak Coulomb interaction generates a CDW instability, and I claimed that the working was analogous to BCS superconductivity. In fact, if the arbitrarily weak interaction is attractive rather than repulsive, a superconducting instability develops in exactly the same way. A full RG analysis, however, suggests that in fact the CDW and SC effects cancel at small coupling, with the result being an ungapped Luttinger liquid [7]. It is only at larger couplings that the respective instabilities truly stabilize.

Loosely, the cancellation occurs when making the mean field approximation in Equation 8. There, a choice was made to decouple the quartic fermion interaction term by use of a CDW ansatz $\left\langle \hat{\psi}^\dagger \hat{\psi} \right\rangle$:

$$\begin{aligned}\hat{\psi}_k^\dagger \hat{\psi}_{k''} \hat{\psi}_{k'}^\dagger \hat{\psi}_{k+k'-k''} \approx & \left\langle \hat{\psi}_k^\dagger \hat{\psi}_{k''} \right\rangle \hat{\psi}_{k'}^\dagger \hat{\psi}_{k+k'-k''} + \hat{\psi}_k^\dagger \hat{\psi}_{k''} \left\langle \hat{\psi}_{k'}^\dagger \hat{\psi}_{k+k'-k''} \right\rangle \\ & - \left\langle \hat{\psi}_k^\dagger \hat{\psi}_{k+k'-k''} \right\rangle \hat{\psi}_{k'}^\dagger \hat{\psi}_{k''} - \hat{\psi}_k^\dagger \hat{\psi}_{k+k'-k''} \left\langle \hat{\psi}_{k'}^\dagger \hat{\psi}_{k''} \right\rangle .\end{aligned}$$

Making instead a superconducting ansatz $\left\langle \hat{\psi}^\dagger \hat{\psi}^\dagger \right\rangle$ gives what turns out to be an opposite contribution:

$$\hat{\psi}_k^\dagger \hat{\psi}_{k''} \hat{\psi}_{k'}^\dagger \hat{\psi}_{k+k'-k''} \approx - \left\langle \hat{\psi}_k^\dagger \hat{\psi}_{k'}^\dagger \right\rangle \hat{\psi}_{k''} \hat{\psi}_{k+k'-k''} - \hat{\psi}_k^\dagger \hat{\psi}_{k'}^\dagger \left\langle \hat{\psi}_{k''} \hat{\psi}_{k+k'-k''} \right\rangle .$$

The seemingly miraculous cancellation is guaranteed to all orders in the RG flow by a Ward identity (*i.e.* gauge invariance) [7]. If this claim is correct it invalidates the work in Part I of this thesis, which is based on a mean-field analysis of spinless fermions in 1D.

I will present three ways around this issue. First, we could simply restrict attention to strong enough couplings h, where the RG analysis agrees a nontrivial CDW state can develop. Second, the RG argument holds only for truly 1D systems, whereas in reality what we refer to as a 1D chain is at most quasi-1D: there may be strong coupling between atoms along chains, and weak coupling between the chains, but the inter-chain coupling can never go to zero. Third, the cancellation between BCS and CDW terms works only for the Coulomb-type interaction I have focussed on so far. If I had instead considered an electron-phonon coupling, or indeed any coupling to the ion lattice, the CDW wins the race for control of the Fermi surface. This was Peierls' original formulation, and it withstood the invention of the RG analysis.

The third retort is particularly appealing given the analysis of Section 2.1 because, by the working of that section, we can consider there to *already be* an electron-phonon coupling built into the model. I showed that any number of interaction terms of the form

$$\hat{H}_{int} = \mathfrak{g} \sum_{kk'} \hat{A}_{k'} \hat{\psi}_k^\dagger \hat{\psi}_{k+k'}$$

between bosons created by $\hat{A}^\dagger$ and fermions created by $\hat{\psi}^\dagger$ can be exactly reformulated in terms of a single combined bosonic field. Strictly, then, we can consider the model presented here to include an infinitesimal electron-phonon coupling to stabilize the CDW

over the Luttinger liquid.

3 Rational-Filling: Hofstadter's Butterfly and the Integer Quantum Hall Effect

In the previous chapter I considered a model of a half-filled 1D chain with nearest-neighbour Coulomb interactions. I will now generalize the result to arbitrary rational fillings p/q with $p, q \in \mathbb{N}$ (natural numbers). This requires a more systematic solution, using a physically-motivated ansatz, leading to a theory in terms of a Landau order parameter, which I then solve for self-consistently.

In Section 3.1 I will motivate a choice of order parameter and write it in a form amenable to applying self-consistent mean field theory. In Section 3.2 I take a small diversion to consider the topology of the procedure with regard to Brillouin zone reduction. In Section 3.3 I set up the mean field calculation as a generalization of the half-filling case. Surprisingly, this simple 1D problem turns out to have a very similar structure to the problem considered by Hofstadter [16], of electrons hopping on a 2D lattice in a magnetic field, and in Section 3.4 I investigate the extent of the analogy. Hofstadter's problem, intimately linked to the Integer Quantum Hall Effect (IQHE), provided one of the earliest emergences of topology and topological quantum numbers in condensed matter physics. In Section 3.5 I find the analogous quantities by considering the Berry's phase of the different sub-bands. In Section 3.6 I return to the self-consistent solution and show that it is adiabatically connected to the Hofstadter case.

3.1 Mean Field Order Parameter

I begin by returning to the mean-field Hamiltonian from Equation 9:

$$\begin{aligned}\hat{H} &= \sum_k \left(-2t\cos\left(k\right)-\mu\right)\hat{\psi}_k^\dagger\hat{\psi}_k \\ &\quad +2h\sum_{kk'k''}\left\{\cos\left(k''\right)\left\langle\hat{\psi}_{k'+k''}^\dagger\hat{\psi}_{k'}\right\rangle\hat{\psi}_k^\dagger\hat{\psi}_{k+k''}-\cos\left(k-k'\right)\left\langle\hat{\psi}_{k'+k''}^\dagger\hat{\psi}_{k'}\right\rangle\hat{\psi}_k^\dagger\hat{\psi}_{k+k''}+h.c.\right\}.\end{aligned}$$

In fact for the working of this chapter it will be convenient to rewrite the interaction term as

$$\hat{H}_{int} = 2h \sum_{kk'k''} \left(\cos\left(k''\right), \cos\left(k\right), \sin\left(k\right)\right) \cdot \begin{pmatrix} 1 \\ -\cos\left(k'\right) \\ -\sin\left(k'\right) \end{pmatrix} \left\langle \hat{\psi}^{\dagger}_{k'+k''} \hat{\psi}_{k'} \right\rangle \hat{\psi}^{\dagger}_{k} \hat{\psi}_{k+k''} + h.c. \tag{18}$$

To motivate an ansatz for the order parameter, consider the physical observables of the system:

$$\begin{pmatrix} site\,order \\ bond\,order \\ current \end{pmatrix} \triangleq \begin{pmatrix} \left\langle \hat{\psi}^{\dagger}_{n} \hat{\psi}_{n} \right\rangle \\ \frac{1}{2}\left(\left\langle \hat{\psi}^{\dagger}_{n} \hat{\psi}_{n+1} \right\rangle + \left\langle \hat{\psi}^{\dagger}_{n+1} \hat{\psi}_{n} \right\rangle\right) \\ \frac{i}{2}\left(\left\langle \hat{\psi}^{\dagger}_{n} \hat{\psi}_{n+1} \right\rangle - \left\langle \hat{\psi}^{\dagger}_{n+1} \hat{\psi}_{n} \right\rangle\right) \end{pmatrix}. \tag{19}$$

The top entry, site-centred order, gives the expectation value of the charge density at site n. The second entry, bond-centred order, is a measure of the same quantity on the bond between sites n and $n+1$. It refers to the spin-singlet p-wave CDW in Nayak's generalized classification [17]. The third entry relates to the current in the system (neglecting external currents). This is expected to be zero here, since it would correspond to current flowing into site n, a process which cannot conserve charge in 1D. Zero will indeed turn out to be the self-consistent solution later on. If we were working in 2D the current term would relate to currents running round plaquettes, and would thus measure the presence of a local gauge field $\mathbf{A}$. In this case it would pertain to $d-$wave symmetry, which in the spin-singlet state gives the *staggered flux lattice* [17, 18]. The physical quantities associated with the order parameters are shown schematically in Figure 8.

Applying the Fourier transform

$$\hat{\psi}^{\dagger}_{n} = \sum_{k} \exp\left(-ikn\right) \hat{\psi}^{\dagger}_{k}$$

gives the $k-$space expression

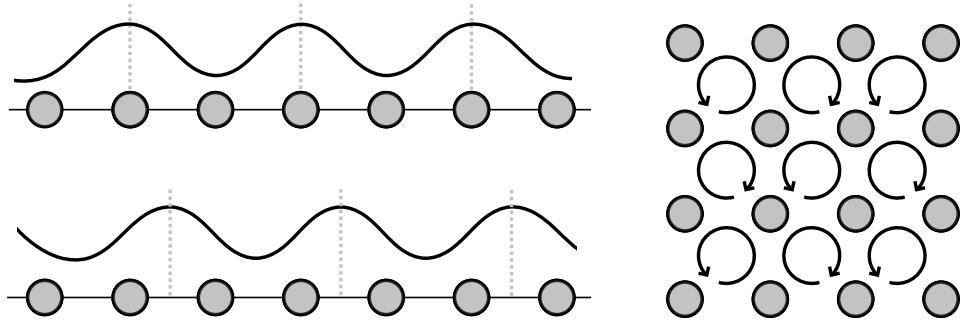


Figure 8: The order parameter of Equation 19 has components corresponding to site order (left, upper), bond order (left, lower), and current order (right). The images show pure examples of each type of order, with a wavelength of two unit cells in each case. Note that current order is undefined in one dimension, and corresponds to magnetic flux through plaquettes in 2D.

$$\begin{pmatrix} site\,order \\ bond\,order \\ current \end{pmatrix} = \sum_{k'} \exp\left(-ik'n\right) \sum_{k} \begin{pmatrix} 1 \\ \exp\left(-ik'/2\right)\cos\left(k+k'/2\right) \\ -\exp\left(-ik'/2\right)\sin\left(k+k'/2\right) \end{pmatrix} \left\langle \hat{\psi}^{\dagger}_{k+k'} \hat{\psi}_{k} \right\rangle . \tag{20}$$

Considering this expression in combination with the interaction Hamiltonian in Equation 18, plus some trial-and-error and guesswork, resulted in my proposing the ansatz

$$2h \sum_{k} \mathbf{R}_{k'/2} \begin{pmatrix} 1 \\ -\cos\left(k\right) \\ -\sin\left(k\right) \end{pmatrix} \left\langle \hat{\psi}^{\dagger}_{k+k'} \hat{\psi}_{k} \right\rangle = \frac{\underline{\Delta}_{Q} \delta_{k',Q} + \underline{\Delta}^{*}_{Q} \delta_{k',-Q}}{2} + \begin{pmatrix} 2h\bar{\rho}\delta_{k',0} \\ 0 \\ 0 \end{pmatrix} \tag{21}$$

with rotation matrix

$$\mathbf{R}_{k} \triangleq \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\left(k\right) & -\sin\left(k\right) \\ 0 & \sin\left(k\right) & \cos\left(k\right) \end{pmatrix} .$$

The three-component complex vector $\underline{\Delta}_{Q}$ provides the order parameter for the phase transition. From the right hand side of Equation 21, it is clear that to implement the ansatz is to make the physical assertion that it is energetically favourable to couple all states k and $k+Q$ with Q fixed. This is simply a statement of the existence of the Peierls instability

in 1D systems.

Substituting the top component of the ansatz into Equation 20 and writing

$$\Delta_j = |\Delta_j| \exp(i\theta_j),$$

with $j \in [1,3]$ the component of the order parameter vector, gives

$$\left\langle \hat{\psi}_n^\dagger \hat{\psi}_n \right\rangle = \frac{|\Delta_1|}{2h} \cos(Qn - \theta_1) + \bar{\rho}. \tag{22}$$

From this it is apparent that the ansatz gives a charge density wave with wavevector Q, where $\overline{\rho}$ is the average charge density of the disordered system (which is located entirely on the sites). Substituting the remaining components gives

$$\begin{pmatrix} site\,order \\ bond\,order \\ current\,order \end{pmatrix} = \frac{1}{2h} \begin{pmatrix} |\Delta_1| \cos(Qn + \phi_1) & +2h\overline{\rho} \\ |\Delta_2| \cos(Qn + \phi_2) & \\ |\Delta_3| \cos(Qn + \phi_3) & \end{pmatrix} \tag{23}$$

with $\phi_1 = -\theta_1$, $\phi_2 = Q/2 + \pi - \theta_2$, and $\phi_3 = Q/2 - \theta_3$. The ansatz therefore leads to a physically sensible charge distribution in real space. Treating n as a parametric variable, Equation 23 parametrizes a closed curve in $\mathbb{R}^3$; in fact it is a planar ellipse[5]. As an interesting aside, if the wavevectors Q were allowed to differ between the components of the vector in Equation 23, the closed curves would generalize to the family of *Lissajous knots [19]*. Such a situation, however, would be rather unphysical.

The ϕ_j come from the complex phase of the order parameter. Without including a coupling to the lattice we expect that the phase of Δ is irrelevant to the free energy of the system, giving a $U(1)$ symmetry. If a coupling to the lattice, or any other mechanism for that matter, preferentially selects a phase on Δ, the $U(1)$ symmetry is broken. From Equation 23 we see that the associated Goldstone mode will correspond to sliding the CDW.

Substitution of the ansatz into the interaction Hamiltonian gives

[5]Strictly the second and third entries in the vector should also have constant terms added, corresponding to the average kinetic energy and the average current in the unperturbed system. Both of these are set to zero in our Hamiltonian - this constitutes a gauge choice $E_0 \triangleq 0$ and $A_\mu \triangleq 0$.

$$\hat{H}_{int} = \sum_k \Xi_k^T \mathbf{R}_{-Q/2} \underline{\Delta}_Q \hat{\psi}_k^\dagger \hat{\psi}_{k+Q} + \Xi_k^T \mathbf{R}_{Q/2} \underline{\Delta}_Q^* \hat{\psi}_k^\dagger \hat{\psi}_{k-Q} + 4h\bar{\rho} \hat{\psi}_k^\dagger \hat{\psi}_k \tag{24}$$

defining

$$\Xi_k^T \triangleq (\cos(Q), \cos(k), \sin(k)). \tag{25}$$

The relatively simple form of Equation 24 makes use of the relation

$$\mathbf{R}_Q \Xi_k = \Xi_{k+Q}.$$

The full Hamiltonian now reads

$$\hat{H} = \sum_{k \in BZ} \epsilon_k \hat{\psi}_k^\dagger \hat{\psi}_k + \Xi_k^T \mathbf{R}_{-Q/2} \underline{\Delta}_Q \hat{\psi}_k^\dagger \hat{\psi}_{k+Q} + \Xi_k^T \mathbf{R}_{Q/2} \underline{\Delta}_Q^* \hat{\psi}_k^\dagger \hat{\psi}_{k-Q} \tag{26}$$

defining the energies

$$\epsilon_k \triangleq -2t\cos(k) + 4h\bar{\rho} - \mu. \tag{27}$$

As in the half-filling case it is easiest to proceed by rewriting the theory in the reduced Brillouin zone. Whereas, in the half-filling case, the reduced Brillouin zone was half the size of the original, in the case of rational filling p/q the vector connecting states at E_F is given by $Q = 2k_F = 2\pi p/q$, and so in this case $BZ \to BZ/q$. The Hamiltonian becomes

$$\begin{aligned} \hat{H} = & \sum_{k \in \frac{BZ}{q}} \sum_{m=1}^{q} \epsilon_{k+mQ} \hat{\psi}_{k+mQ}^\dagger \hat{\psi}_{k+mQ} \\ & + \Xi_{k+mQ}^T \mathbf{R}_{-Q/2} \underline{\Delta}_Q \hat{\psi}_{k+mQ}^\dagger \hat{\psi}_{k+(m+1)Q} + \Xi_{k+mQ}^T \mathbf{R}_{Q/2} \underline{\Delta}_Q^* \hat{\psi}_{k+mQ}^\dagger \hat{\psi}_{k+(m-1)Q} \end{aligned} \tag{28}$$

where $k + qQ \equiv k$. The folding back of the band into the reduced Brillouin zone is shown for different fillings in Figure 9. The simplicity of the back-folding trick hides some interesting topology which it is worth considering in more detail before proceeding.

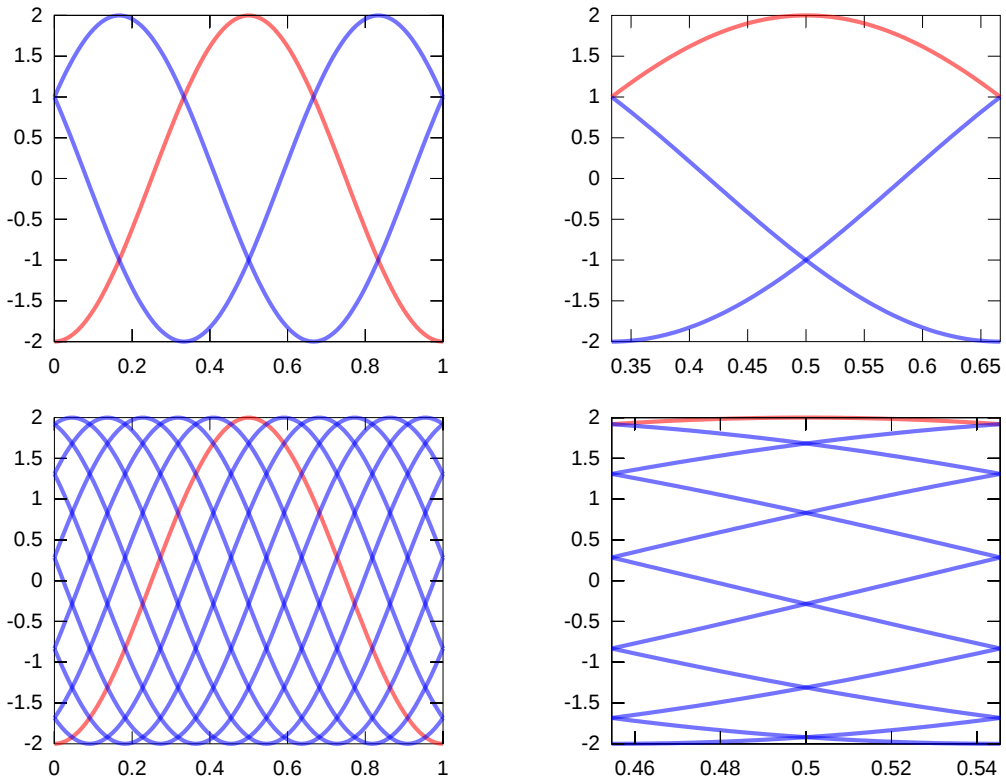


Figure 9: Reducing the Brillouin zone in the case of rational filling $p/q = p/3$ (top) and $p/11$ (bottom). The numerator changes the chemical potential but not the backfolded structure. The original band is shown in red. The figures show the unperturbed band being folded back. This is simply a mathematical rewriting of the problem, but becomes very useful when the interaction is turned on. In that case gaps open up at the crossing points in the reduced zone. The hopping parameter $t = 1$.

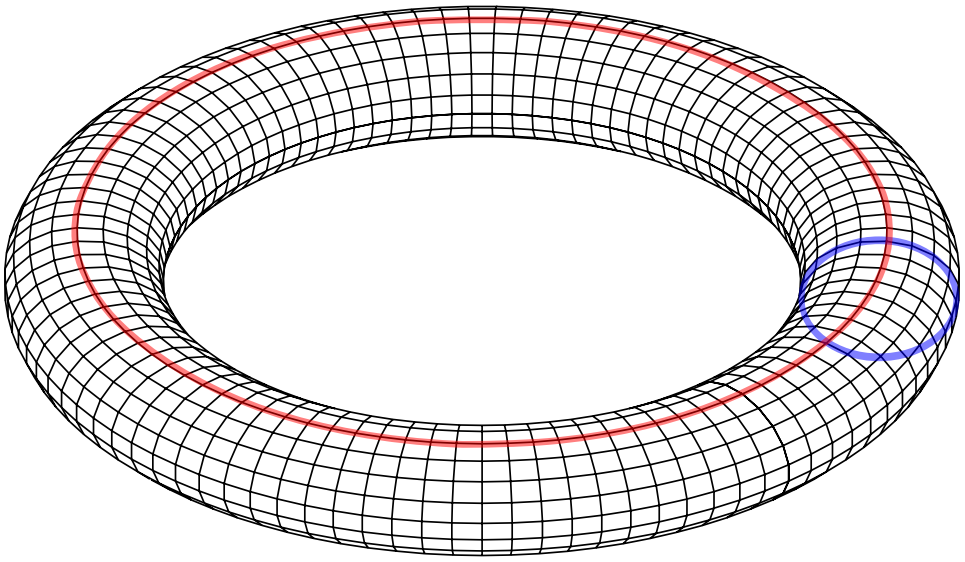


Figure 10: A torus, T^2, is the surface swept out in passing one circle S^1 around another, hence $T^2 \equiv S^1 \times S^1$.

3.2 The Topology of Brillouin Zone Reduction

For matters of topology it will be convenient to again think in the continuum limit of lattice spacing$\rightarrow$ 0, so that the Brillouin zone (BZ) is a continuous line of finite length $[0, 2\pi)$ with the ends of the line equated. Topologically this defines a circle, S^1. In 2D the Brillouin zone can be chosen to be a rectangle with parallel sides equated. Rolling the rectangle up to match the edges appropriately reveals that this is a torus, T^2, often written $S^1 \times S^1$. The latter notation indicates that a torus can be considered to be the surface swept out by a circle carried along a perpendicular circle (see Figure 10). In n dimensions the Brillouin zone has the topology of an $n-$torus, $T^n \equiv S^1 \times S^1 \times \ldots S^1$.

I will restrict attention to 1D here, and will consider the case that it is desirable to express the full BZ as q reduced BZs. Imagine the BZ as a length of string. Now take the string, and make q equally-spaced marks on it. Reducing the BZ means treating any two neighbouring marks as equivalent, so put a loop in the string to make the marks touch. Before the splitting is introduced by $\hat{H}_{int}$ the marks all pile up, giving a bouquet of q circles meeting at the same point. This sounds a bit singular, so to regularize we can consider $\hat{H}_{int}$ to always be turned on at least infinitesimally. Then the intuition afforded by the string is good: neighbouring points touch, but hopping two points is not equivalent to hopping one. The situation is shown in Figure 11.

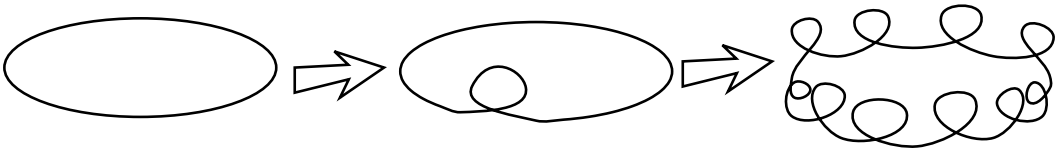


Figure 11: Backfolding the Brillouin zone into the reduced zone scheme, points separated by Q are made equivalent. If the original BZ is topologically a circle, the zone backfolded q times is a circle with q loops, where overlapping lines are considered to be touching (but hopping from one to the other costs energy Δ).

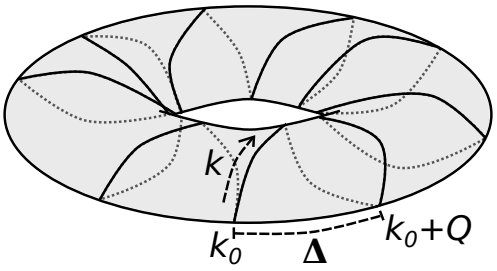


Figure 12: Continuing from Figure 11, our original BZ was a closed loop. Backfolding to q reduced BZs introduces q sub-loops. An electron can either move along the reduced BZ, k, via the field operator $\hat{\psi}^\dagger_{k+k'}\hat{\psi}_k$, or it can scatter between RBZs via a 'reduced Umklapp' process, costing it Δ.

The field operator $\hat{\psi}^\dagger_{k+k'}\hat{\psi}_k$, annihilating an electron at k and creating one at $k+k'$, 'hops' along the RBZ (sub-loop of the string) a distance k'. From $\hat{H}_{int}$ there is now a second allowed process, costing energy $\sim \Delta$, which hops between neighbouring RBZs. Hopping two RBZs costs $\sim \Delta^2$ and so on. These hops correspond to 'reduced Umklapp processes', scattering events between reduced Brillouin zones. A convenient representation is to place the string on a torus, which it winds q times at a constant rate (pitch). This is shown in Figure 12. An electron has two options: to move smoothly along k, along the string, in-keeping with the RBZ story, or to break the illusion and hop off the string and back on a distance Q further along. The latter option costs it Δ.

What about the numerator of the filling fraction, p? The story so far has only concerned the topology of k-space. In fact the BZ has an energy (ϵ_k before perturbation) associated with each k. When backfolding the bands into the RBZ each segment of the full BZ brings its piece of ϵ_k along with it. When the splitting perturbation is turned on the result is q sub-bands in the RBZ, and p of them are filled.

A p/q-filled BZ also has a convenient representation in this scheme, as a string which wraps the torus q times per circuit but which takes p circuits to return to itself. The string forms a *knot,* in the mathematical sense of a closed 1D line embedded in 3D Euclidean space

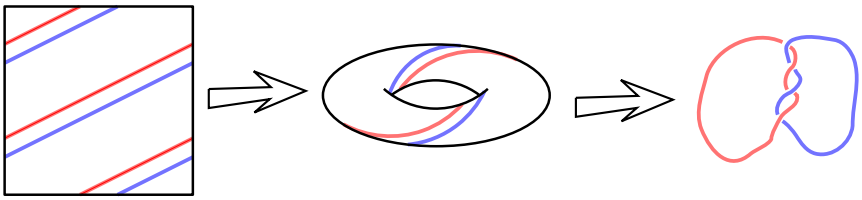


Figure 13: The $(2,4)$ torus knot. A simple way to generate the knots is to draw a square with opposite sides equated; a (p,q) knot hits the right wall q times and the top wall p times before returning to itself. Geometrically our system requires equal spacing between the intersections with the walls (I have relaxed this constraint for the purposes of simplifying the picture). Bringing the equivalent sides into contact forms the torus. Removing the torus leaves the knot. In this case we see that the $(2,4)$ torus knot is actually two linked 'unknots'. In general an (lp,lq) knot decouples to l linked (p,q) knots.

(the same as an everyday knot but with the ends stuck together) [20, 21]. The fact that the knot lies on the surface of a torus defines it to be a (p,q) torus knot. If either p or q are unity the knot is the trivial 'unknot' (a circle with no complications). If the knot is of the form (lp,lq) it decouples into l linked (p,q) torus knots, shown for the particular case of $(2,4)$ in Figure 13.

3.3 The Self-Consistency Condition

Returning to Equation 28 and writing the Hamiltonian as a matrix gives

$$\hat{H} = \sum_{k \in \frac{BZ}{q}} \left(\hat{\psi}^{\dagger}_{k+Q}, \hat{\psi}^{\dagger}_{k+2Q}, \hat{\psi}^{\dagger}_{k+3Q}, \ldots, \hat{\psi}^{\dagger}_{k+qQ} \right) H_k \begin{pmatrix} \hat{\psi}_{k+Q} \\ \hat{\psi}_{k+2Q} \\ \hat{\psi}_{k+3Q} \\ \vdots \\ \hat{\psi}_{k+qQ} \end{pmatrix} \tag{29}$$

with H_k defined by the matrix

$$\begin{pmatrix} \epsilon_{k+Q} & \Xi^T_{k+Q}\mathbf{R}_{-\frac{Q}{2}}\Delta_Q & 0 & \ldots & \left(\Xi^T_{k}\mathbf{R}_{-\frac{Q}{2}}\Delta_Q\right)^* \\ \left(\Xi^T_{k+Q}\mathbf{R}_{-\frac{Q}{2}}\Delta_Q\right)^* & \epsilon_{k+2Q} & \Xi^T_{k+2Q}\mathbf{R}_{-\frac{Q}{2}}\Delta_Q & 0 & 0 \\ 0 & \left(\Xi^T_{k+2Q}\mathbf{R}_{-\frac{Q}{2}}\Delta_Q\right)^* & \epsilon_{k+3Q} & \ddots & 0 \\ \vdots & 0 & \ddots & \ddots & \Xi^T_{k+(q-1)Q}\mathbf{R}_{-\frac{Q}{2}}\Delta_Q \\ \Xi^T_{k}\mathbf{R}_{-\frac{Q}{2}}\Delta_Q & 0 & 0 & \left(\Xi^T_{k+(q-1)Q}\mathbf{R}_{-\frac{Q}{2}}\Delta_Q\right)^* & \epsilon_{k+qQ} \end{pmatrix}.$$

As in the half-filling case, the off-diagonal terms are an indication that electrons and holes do not form well-defined quasiparticles in the presence of the interaction. It should again be possible to form a linear combination of them which is. Diagonalizing the matrix with

$$H_k = U_k D_k U_k^\dagger$$

gives

$$\hat{H} = \sum_{k \in BZ/q} \left(\hat{\gamma}^\dagger_{k+Q}, \hat{\gamma}^\dagger_{k+2Q}, \hat{\gamma}^\dagger_{k+3Q}, \ldots, \hat{\gamma}^\dagger_{k+qQ} \right) D_k \begin{pmatrix} \hat{\gamma}_{k+Q} \\ \hat{\gamma}_{k+2Q} \\ \hat{\gamma}_{k+3Q} \\ \vdots \\ \hat{\gamma}_{k+qQ} \end{pmatrix}$$

where the annihilation and creation operators for the new well-defined quasiparticles are defined as linear combinations of the original particle/hole operators:

$$\begin{aligned} \hat{\gamma}_\alpha &= \sum_\beta U^\dagger_{\alpha\beta} \hat{\psi}^\beta \\ \hat{\gamma}^\dagger_\alpha &= \sum_\beta \hat{\psi}^{\dagger\beta} U_{\beta\alpha} \end{aligned}$$

with the elements $\alpha, \beta \in [1, q]$. Since H_k is Hermitian, U is unitary, *i.e.* $U^\dagger \equiv U^{-1}$. I have dropped the k subscript for clarity. A self-consistent solution involves finding

$$\left\langle \hat{\psi}^\dagger_\alpha \hat{\psi}_\beta \right\rangle = \sum_{\mu\nu} U^\dagger_{\alpha\mu} U_{\nu\beta} \left\langle \hat{\gamma}^{\dagger\mu} \hat{\gamma}^\nu \right\rangle \tag{30}$$

and substituting into Equation 21.

The operator $\hat{\gamma}^\dagger_k$ creates a well-defined fermion with momentum k, governed by the Fermi-Dirac statistics f:

$$\left\langle \hat{\gamma}^{\dagger\mu} \hat{\gamma}^\nu \right\rangle = \delta^{\mu\nu} f\left(D_{\mu\mu}\right)$$

which reduces Equation 30 to

$$\left\langle \hat{\psi}_\alpha^\dagger \hat{\psi}_\beta \right\rangle = \sum_\mu U_{\alpha\mu}^\dagger U_{\mu\beta} f\left(D_{\mu\mu}\right).$$

Returning to the ansatz, Equation 21, and choosing the $k' = Q$ term, gives the self-consistency condition

$$\underline{\Delta}_Q = 4h \sum_{k' \in BZ} \begin{pmatrix} 1 \\ -\cos\left(k' + Q/2\right) \\ -\sin\left(k' + Q/2\right) \end{pmatrix} \sum_\mu U_{1\mu}^\dagger U_{\mu q} f\left(D_{\mu\mu}\right) \tag{31}$$

where 1 and q could be any two neighbouring elements in the U matrices. As a quick check, $Q = 2\pi\frac{1}{2}$ gives

$$\begin{aligned} \underline{\Delta} &= 4h \sum_{k' \in BZ} \begin{pmatrix} 1 \\ \sin\left(k'\right) \\ -\cos\left(k'\right) \end{pmatrix} \left[U_{11}^\dagger U_{12} f\left(D_{11}\right) + U_{12}^\dagger U_{22} f\left(D_{22}\right) \right] \\ &= 4h \sum_{k' \in BZ} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \left[U_{11}^\dagger U_{12} f\left(D_{11}\right) + U_{12}^\dagger U_{22} f\left(D_{22}\right) \right] \end{aligned}$$

where the disappearance of the lower two components follows from the fact that the term in brackets is not only periodic with the full BZ, but in fact has the periodicity of the RBZ (in this case π). The expression agrees with the half-filling case from Chapter 2, where the self-consistency condition read

$$\left(\Delta + \Delta^*\right) = 4h \sum_{k' \in BZ} \left(1 + \cos\left(k - k'\right)\right) \frac{\left(\Delta + \Delta^*\right)}{\sqrt{\epsilon_{k'}^2 + \left(\Delta + \Delta^*\right)^2}} \tanh\left(\frac{1}{2}\beta\sqrt{\epsilon_{k'}^2 + \left(\Delta + \Delta^*\right)^2}\right)$$

and the $\cos\left(k - k'\right)$ term corresponds to what is now the bond order term in the new formalism. Note that the half-filling, 2×2, case is the only one in which element $(1,2) \equiv (1,q)$. This explains why the equation features $\Delta + \Delta^*$ whereas the general case now under

consideration features only Δ.

3.4 Hofstadter's Butterfly

Before solving the rational-filling problem self-consistently, I wish to consider the special case $\underline{\Delta}_Q^T = \frac{t}{\cos(Q)}(1,0,0)$. From Equation 23 this corresponds to a completely site-centred CDW of the form

$$\left\langle \hat{\psi}_n^\dagger \hat{\psi}_n \right\rangle = \frac{t}{2h} \frac{\cos(Qn)}{\cos(Q)}$$

and no current in the system.

In this special case the problem simplifies to finding the eigenvalues of the matrix[6]

$$H_k = t \begin{pmatrix} -2\cos(k+Q) & 1 & 0 & 0 & 1 \\ 1 & -2\cos(k+2Q) & 1 & 0 & 0 \\ 0 & 1 & -2\cos(k+3Q) & 1 & 0 \\ 0 & 0 & 1 & \ddots & 1 \\ 1 & 0 & 0 & 1 & -2\cos(k+qQ) \end{pmatrix} \quad (32)$$

for all possible k and all rational fillings p/q. With the off-diagonal perturbation set to zero the allowed energies all range from $-2t$ to $2t$ (the unperturbed bandwidth). Turning on the interaction opens up a CDW gap at $E_F/t = 4p/q - 2$, but also opens gaps at a total of $q-1$ places in the band, whenever the backfolded bands cross.

For fixed k and $Q = 2\pi p/q$ there will be q eigenvalues for each p, but many of these will be degenerate. To solve numerically, it is necessary to pick a sampling of $k \in [0, 2\pi)$, so I define

$$k = 2\pi \frac{k_{int}}{k_{max}}$$

with k_{int} and k_{max} integers. Let q range from 1 to q_{max}, and p range from 1 to q. In Figure 14 I show the result of plotting the allowed energy states (eigenvalues of the matrix) against

[6] The offset $4h\overline{\rho} - \mu$ was set equal to zero for simplicity, since it causes a linear shift in eigenvalues and has no effect on eigenvectors.

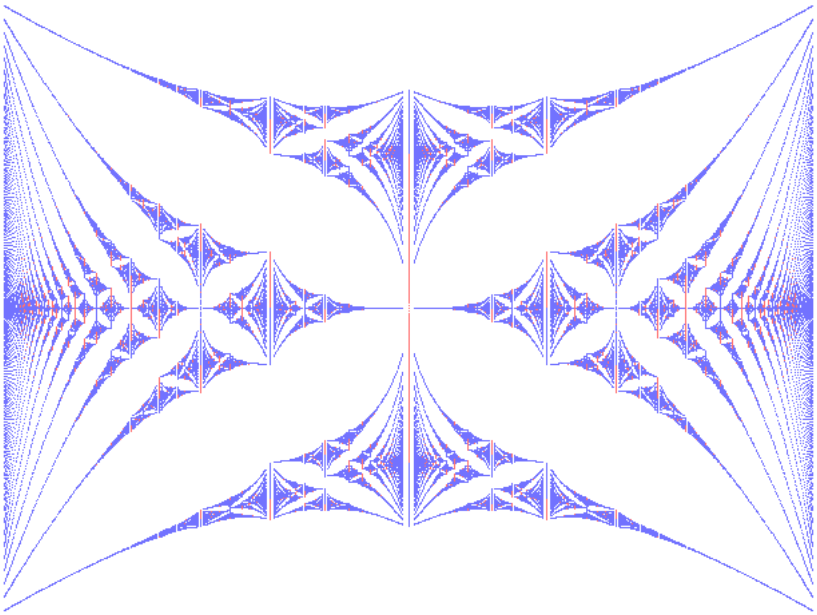


Figure 14: Blue: the allowed energies, sampled at 200 k-points (y-axis from $-4t$ to $4t$), versus filling fraction p/q for integer $p < q$ and $q \leq 50$ (axis from 0 to 1), after forming a purely site-centred CDW with order parameter $\Delta_Q^1 = \frac{t}{\cos(Q)}$. Red: the additional points present in the Hofstadter butterfly.

the filling fraction p/q, for the case $q_{max} = 50$, $k_{max} = 200$. The result appears at first glance to be fractal; in fact it bears a strong resemblance to the famous fractal known as the Hofstadter butterfly [16]. The two are contrasted in the figure.

Hofstadter was considering the allowed energy states of electrons confined to a 2D lattice with a perpendicular magnetic field applied. If a current is present, and the $\mathbf{B}$ field is strong enough, the electrons will move in closed loops, and as they traverse the loops they pick up a phase proportional to $\oint \mathbf{A} \cdot \mathrm{d}\mathbf{l} = \Phi$, the flux enclosed. Destructive interference rules out any orbits other than those enclosing an integer multiple of the flux quantum Φ_0. Say we set up a square lattice and magnetic field such that one flux quantum is enclosed upon hopping around one unit cell. The lattice spacing would be enormous in that case, but a similar set-up has recently been realized through use of a Moiré superlattice in graphene bonded to single-layer Boron Nitride with a different orientation [22]. Now, if we halve the field strength so that $\Phi/\Phi_0 = 1/2$, the electrons must instead hop around two unit cells. For $\Phi/\Phi_0 = p/q$ the electrons must hop around q unit cells, and the area of the Brillouin zone has been reduced by a factor of q (the reduced Brillouin zone being referred to in this case as the magnetic Brillouin zone).

The difference between Hofstadter's case and that considered here comes about because Hofstadter's case, and that of the quantum Hall effect, lives in a two-dimensional space $\mathbf{k} = (k_x, k_y)$, whereas the CDW problem I am considering here is 1D. In fact, there *is* a second degree of freedom hidden in the CDW problem in the form of the complex phase of the order parameter; including this generates the full Hofstadter spectrum, a point returned to in detail in Section 3.5.1.

Note that the strong magnetic field in Hofstadter's case effectively reduces that 2D problem to one dimension by restricting the electrons to move on closed orbits, removing one degree of freedom. This goes some way to explaining how it is that the same spectrum can result from considering only a 1D band[7].

Note that non-coprime filling factors lp/lq, $l \in \mathbb{Z}$ should be excluded from the plot. Including them qualitatively changes the plot obtained. For half-filling, 1/2, the expected result from Chapter 2 is two sub-bands with a gap between them. Including the non-coprime fractions 2/4, 3/6 *etc.* acts to fill the gap in the limit of infinite terms. Of course, $2/4 \equiv 1/2$, since a density of 1 electron in 2 unit cells is the same as 2 electrons in 4 unit cells. Here q_{max} merely refers to the largest denominator of filling fraction considered, but the system could well have a continuum of sites.

Mathematically this works as follows: for the 1/2 filling case

$$\hat{H} = t \sum_{k \in BZ/2} \left(\hat{\psi}^\dagger_{k+\pi}, \hat{\psi}^\dagger_k \right) \begin{pmatrix} 2\cos(k) & 1 \\ 1 & -2\cos(k) \end{pmatrix} \begin{pmatrix} \hat{\psi}_{k+\pi} \\ \hat{\psi}_k \end{pmatrix}$$

and for the 2/4 filling case

$$\hat{H} = t \sum_{k \in BZ/4} \left(\hat{\psi}^\dagger_{k+\pi}, \hat{\psi}^\dagger_k, \hat{\psi}^\dagger_{k+\pi}, \hat{\psi}^\dagger_k \right) \begin{pmatrix} 2\cos(k) & 1 & 0 & 1 \\ 1 & -2\cos(k) & 1 & 0 \\ 0 & 1 & 2\cos(k) & 1 \\ 1 & 0 & 1 & -2\cos(k) \end{pmatrix} \begin{pmatrix} \hat{\psi}_{k+\pi} \\ \hat{\psi}_k \\ \hat{\psi}_{k+\pi} \\ \hat{\psi}_k \end{pmatrix}.$$

[7] The question of the dimension of the system is returned to in detail when considering the case of incommensurate filling in Section 4. In this case it has been argued [23], controversially, that it is conversely the 1D system which takes on 2D properties.

At first it appears that the 4×4 matrix will have different eigenvalues[8] to the 2×2. The trick is that two sets of two of the components in the $\hat{\psi}$ vector are identical, so we have to be careful with diagonalizing the matrix. Whatever method is employed, the result should be identical to the 2×2 case, so the easiest method is simply to neglect non-coprime fractions.

The argument can be cast in the language of RBZ topology developed in Section 3.2. As noted there, the p/q-filling case corresponds to mapping the circular Brillouin zone to a (p, q) torus knot. For lp/lq-filling this gives an (lp, lq) torus knot, which is in fact a torus *link* comprising l copies of a (p, q) knot. Physicality requires that we unlink the knots from one another. The situation is shown in Figure 13.

As a final intriguing note, the image which is generated in the limit of erroneously including all non-coprime fractions is precisely the full Hofstadter spectrum! I am at a loss to explain this result either physically or mathematically.

3.5 The Topology of Hofstadter's Butterfly: Berry's Phase

The Hofstadter spectrum originally arose in the context of the integer quantum Hall effect. The experimental set-up has electrons confined to a thin sheet which I'll define to be perpendicular to $\hat{z}$. This is nowadays generally done by use of a semiconductor heterojunction, although Hall himself used a thin layer of gold leaf on a glass plate [24]. A magnetic field is directed along $\hat{z}$, and an electric field along $\hat{x}$ (pictured in Figure 15). In the standard Hall effect a transverse 'Hall voltage' is detected in the $\hat{y}$ direction owing to the deflection of the current-carrying electrons by the $\mathbf{B}$ field. When the experiment is carried out at sufficiently low temperature in a sufficiently high $\mathbf{B}$ field the 'Hall conductance' in the transverse direction, σ_{xy}, is precisely quantized into integer multiples of $e^2/2\pi\hbar$ [25]. This precise quantization, now used as a standard of conductance, provided the first major instance of topology playing a vital rôle in condensed matter physics [26], although re-interpretations of work on superconductivity suggest that this provided an earlier example [27]. In this section I will investigate whether and how topology enters our case by considering the 'Berry phase' [28] of CDW systems.

[8] In fact John Hannay has pointed out that the eigenvalues of the 4×4 matrix include those of the 2×2 as a subset. More generally the eigenvalues of the lp/lq matrix contain those of the p/q matrix plus $q\,(l-1)$ others.

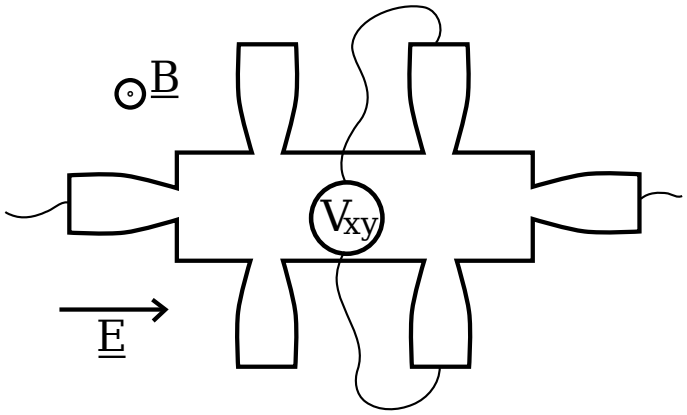


Figure 15: The set-up of the quantum Hall effect. A current enters the sample through the leads on the end; a magnetic field through the plane leads to a transverse voltage V_{xy}, and at sufficiently low T and high $\mathbf{B}$ the transverse (Hall) conductivity σ_{xy} becomes precisely quantized in units of $e^2/2\pi\hbar$.

The Berry phase, an example of a geometric phase, has become so ubiquitous in physics that creating a definition encompassing all the effects associated with it is not an easy task. The key requirement is that the quantum evolution equation (often Schrödinger's equation) is parametrized by some continuous variable. Until the early 1980s it was believed that the vectors of the Hilbert space completely specify the state of the system - overspecify it, in fact, since a global phase cannot be observed. Berry's question was whether, upon adiabatically traversing a closed loop in the parameter space, we necessarily return to the same state in the Hilbert space. Remarkably the answer may be 'no', in the case that the parameter space has some nontrivial geometry or topology.

A classic illustrative example is the Aharonov-Bohm effect, in which a two-slit experiment is carried out with electrons and a solenoid is used as the slit separator [29]. As the magnetic field in the solenoid is increased, the two-slit pattern observed on the screen shifts along, despite the fact that the magnetic field outside an infinite solenoid is zero (so the electrons never move through the $\mathbf{B}$ field itself). The continuous parameter in this case is the magnetic flux enclosed by a real space electron path, and the nontrivial topology arises because the electronic wavefunction is required to be zero at the location of the solenoid. The resulting punctured plane of allowed positions is multiply-connected: winding different numbers of times around the puncture causes the wavefunction to pick up a different phase, and interference of the paths leads to the observed shift in the interference pattern.

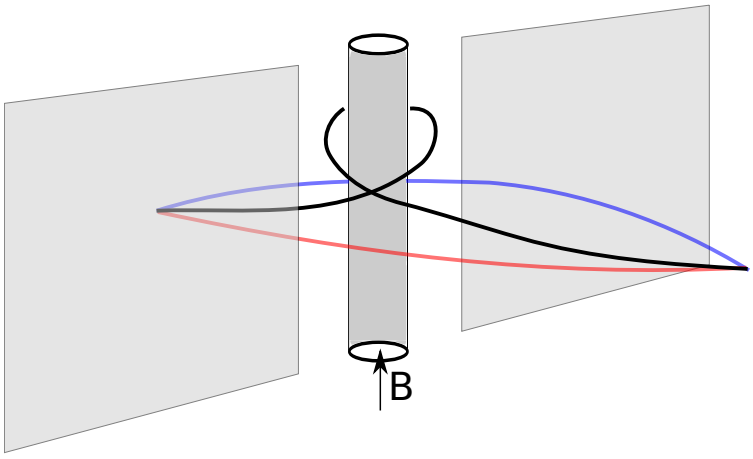


Figure 16: The Aharonov Bohm effect provides an illustrative example of the Berry phase. The red and black paths enclose different amounts of magnetic flux, and interfere with the blue path differently. The problem can be decomposed into the topologies of the paths (specifically their linking numbers with the solenoid).

While one might have expected a unique wavefunction $\langle x|u\rangle$ (I will use ket $|u\rangle$ rather than $|\psi\rangle$ to distinguish from second quantized operators $\hat{\psi}^\dagger$) to be ascribable to each point in space up to overall phase, in fact uniqueness is obtained only when additionally specifying the linking number of the electron's path with the solenoid. The situation is shown in Figure 16. In this case the phase at a point is quantized in units of the flux: quantized because for any definition of 'encircling' the solenoid (any gauge choice) a path must encircle it an integer number of times. The quantization is an indication of a nontrivial topology rather than geometry, but in this case the observable effect is dictated by the geometry, the continuously variable magnetic flux. In other cases the observable relates to a 'topological phase' and leads to quantization of some quantity.

The effect is now commonly explained by borrowing the language of fibre bundles from the mathematical literature. Taking a concrete example, it was noticed by Zak [30] that, in condensed matter, we regularly appeal to Bloch's theorem to label Hamiltonians by wavenumber k, and that k provides a continuously varying parameter of the form required for Berry's construction. At each k we have a different Hilbert space, with a different Hamiltonian $\hat{H}_k$ and eigenkets $|u_k\rangle$[9]. Each Hilbert space forms a 'fibre', and the set of all these Hilbert spaces forms a 'fibre bundle'. Varying k jumps us between fibres, and returning to the original k may end up with us at a different point on the original fibre, depending on the curvature of the underlying manifold. The situation is illustrated

[9]Note that the ket itself if labelled by k; it is not projected into k space.

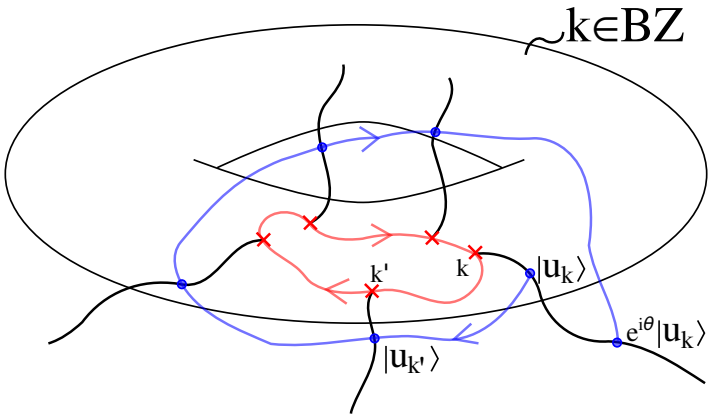


Figure 17: The fibre bundle description of the Berry phase. The Hamiltonian is parametrized by some variable taking a value on a continuous manifold, in this case the Bloch wavevector k in the Brillouin zone. Each point on the manifold has a different Hilbert space associated with it called a 'fibre' (the black lines, spanned by the possible eigenvectors of $\hat{H}_k$). The set of all fibres is a 'fibre bundle'. Traversing the closed red path on the base space (a Brillouin zone torus here) traverses the blue path through the fibre bundle. Despite returning to the original fibre the state may have changed by a phase, in this case $|u\rangle \rightarrow \exp(i\theta)|u\rangle$.

schematically in Figure 17.

The Berry phase acts as a measure of nontrivial geometry [28]. Quantized Berry phases are an indication of nontrivial topology [31]. In all cases the phase can be thought of as being the result of having 'discarded' some information about the system[10], and replacing it with a parameter. In the Aharonov-Bohm effect the discarding of information comes in by treating the electrons as living on a 2D punctured plane, where they really live on a multi-sheeted Riemann surface. The appearance of Hofstadter's butterfly, with its known topological associations, provides a motivation to search for a similar discarding of information in the CDW story considered so far. Indeed, such a discarding does come about during the formation of the CDW. Before discussing this I will first establish some further background on the rôle of phases in the CDW problem, and the quantization of the Berry phase in the integer quantum Hall effect.

3.5.1 The Order Parameter Phase: Sliding the CDW

Let us consider again the Hofstadter case of Section 3.4, but this time explicitly include the CDW phase: $\underline{\Delta}_Q = \frac{t}{\cos(Q)}(\exp(i\theta), 0, 0)$. From Equation 23 this now corresponds to

[10] The idea of throwing away information brings to mind the concept of entropy. In fact the two concepts are intimately related, as shown in reference [32].

a CDW of the form

$$\left\langle \hat{\psi}_n^\dagger \hat{\psi}_n \right\rangle = \frac{t}{2h} \frac{\cos\left(Qn+\theta\right)}{\cos\left(Q\right)}.$$

The matrix requiring diagonalization is this time

$$H_{k,\theta} = t \begin{pmatrix} -2\cos\left(k+Q\right) & \exp\left(i\theta\right) & 0 & 0 & \exp\left(-i\theta\right) \\ \exp\left(-i\theta\right) & -2\cos\left(k+2Q\right) & \exp\left(i\theta\right) & 0 & 0 \\ 0 & \exp\left(-i\theta\right) & -2\cos\left(k+3Q\right) & \exp\left(i\theta\right) & 0 \\ 0 & 0 & \exp\left(-i\theta\right) & \ddots & \exp\left(i\theta\right) \\ \exp\left(i\theta\right) & 0 & 0 & \exp\left(-i\theta\right) & -2\cos\left(k+qQ\right) \end{pmatrix} \quad (33)$$

and I will be considering what happens as θ is caused to vary, *i.e.* the CDW is made to slide. The sliding will be done adiabatically, which in this context means that θ changes on a slow enough timescale that no band mixing occurs. Alternatively first order time-dependent perturbation theory is valid, and the working presented here can be translated directly into that language.

It is now clear that k and θ sensibly make up a two-dimensional parameter space. Accordingly I will temporarily define the 2D vector $\mathbf{k} \triangleq (k_1, k_2) \triangleq (k, \theta)$. The eigenvectors must be 2π periodic in $k_2 \triangleq \theta$, but in fact it turns out that for p/q filling they are $2\pi/q$ periodic, exactly as in $k_1 \triangleq k$. For p/q filling the Brillouin zone splits into q RBZs; splitting the k_2 direction similarly we can define $1/q^2$ of the original Brillouin zone as the *Magnetic Brillouin Zone*, MBZ (see Figure 18 for clarification). The terminology is borrowed from [33] and makes use of the analogy to the quantum Hall effect; there is no true magnetic field in our system.

The Berry phase in band $n \in q$ is defined to be

$$\gamma_n = \oint_{\partial RBZ} \mathbf{A}_n \cdot \mathrm{d}\mathbf{k} = \iint_{RBZ} \Omega_n \mathrm{d}k_1 \mathrm{d}k_2 \quad (34)$$

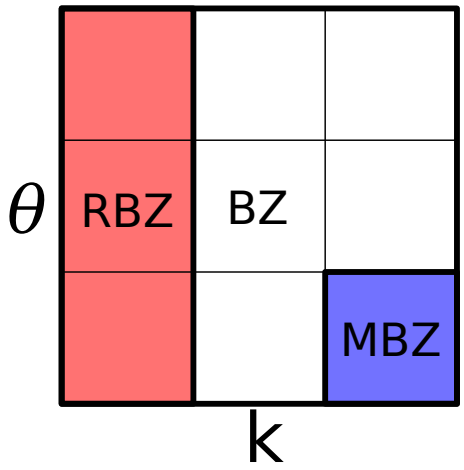


Figure 18: Treating k and θ on the same footing, I maintain the definition of the RBZ as BZ/q in the k direction, but introduce the definition of the magnetic Brillouin zone $MBZ = RBZ/q = BZ/q^2$ where the meaning of this is defined in the image (which has $q = 3$). The total of the 9 squares is the BZ. The construction is convenient since the eigenvalues and vectors have period $1/q$ in both directions.

with the Berry connection $\mathbf{A}_n$ and curvature Ω_n defined to be

$$\begin{aligned} \mathbf{A}_n &= i \langle u_n | \nabla_{\mathbf{k}} | u_n \rangle \\ \Omega_n &= \nabla_{\mathbf{k}} \times \mathbf{A}_n|_{\hat{z}} \end{aligned} \tag{35}$$

with $\nabla_{\mathbf{k}} \triangleq \left(\frac{\partial}{\partial k_1}, \frac{\partial}{\partial k_2} \right)^T$ and $\hat{z}$ defined to be perpendicular to the $\mathbf{k} = (k_1, k_2)$ plane. The functions $|u_n(\mathbf{k})\rangle$ are defined to satisfy

$$H_{\mathbf{k}} |u_n(\mathbf{k})\rangle = \epsilon_n(\mathbf{k}) |u_n(\mathbf{k})\rangle$$

with $H_{\mathbf{k}}$ the $q \times q$ matrix defined in Equation 33 and

$$|u_n(k_1, k_2)\rangle = \exp(i\phi) \left| u_n\left(k_1 + \frac{2\pi}{q}, k_2\right) \right\rangle = \exp(i\phi') \left| u_n\left(k_1, k_2 + \frac{2\pi}{q}\right) \right\rangle$$

with $\mathbf{k}$ considered a parameter to be varied.

Although the Berry phase and curvature are gauge invariant, the connection is not: it encodes a mathematical redundancy in the system in which the combined transformation

$$\begin{aligned} |u\rangle &\rightarrow \exp(i\phi)|u\rangle \\ \mathbf{A} &\rightarrow \mathbf{A} - \nabla_{\mathbf{k}}\phi \end{aligned}$$

leaves the physical system unchanged. Four different methods of calculating the Berry phase are considered in the following section.

3.5.2 Methods of Calculating the Berry Phase

In this section I will outline four methods of calculating the Berry phase of a sub-band. The four methods complement one another in calculations, and I employed all four in producing the results of this thesis.

The primary calculational method was developed by Chang and Niu [33, 34], who considered the topology of the Hofstadter problem in detail. The method employs gauge invariant quantities and involves surface integrals over the Berry curvature. From the definitions of Equation 35 it follows that

$$\begin{aligned} \Omega_n(\mathbf{k}) &= i\left[\left(\frac{\partial}{\partial k_1}\langle u_n|\right)\left(\frac{\partial}{\partial k_2}|u_n\rangle\right) - c.c.\right] \\ &= -2\mathfrak{Im}\left(\frac{\partial}{\partial k_1}\langle u_n|\right)\left(\frac{\partial}{\partial k_2}|u_n\rangle\right). \end{aligned}$$

Inserting a complete set of states gives

$$\Omega_n(\mathbf{k}) = -2\mathfrak{Im}\sum_{n'\neq \mathfrak{n}}\left[\left(\frac{\partial}{\partial k_1}\langle u_n|\right)|u_{n'}\rangle\langle u_{n'}|\left(\frac{\partial}{\partial k_2}|u_n\rangle\right)\right]$$

where the $n' = n$ term is dropped as it doesn't contribute to the sum[11]. After some algebra the expression can be rewritten as

$$\Omega_n(\mathbf{k}) = -2\mathfrak{Im}\sum_{n'\neq n}\frac{\left[\langle u_n|\,\partial_{k_1}H_{\mathbf{k}}\,|u_{n'}\rangle\right]\left[\langle u_{n'}|\,\partial_{k_2}H_{\mathbf{k}}\,|u_n\rangle\right]}{(\epsilon_{n'}-\epsilon_n)^2}.$$

Note that this equation is explicitly gauge invariant. With the Hamiltonian of Equation

[11] The proof follows from $\nabla_{\mathbf{k}}(\langle u_n|u_n\rangle) = 0$.

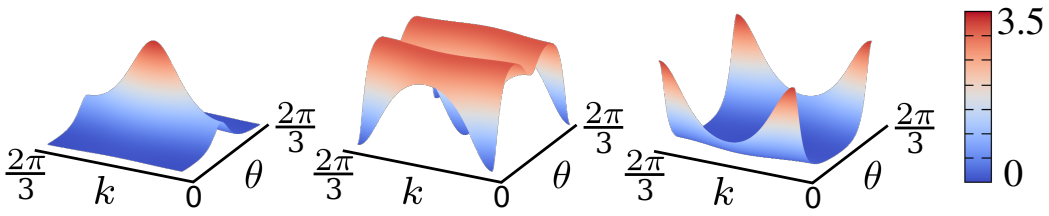


Figure 19: The Berry curvature across the magnetic Brillouin zone $(k, \theta) \in \left[\frac{BZ}{3}, \frac{BZ}{3}\right]$ in sub-bands $\{1, 2, 3\}$ for 1/3 filling. The integral across the RBZ (which happens to equal the integral across the pictured regions in this case) gives the Berry phase γ, which comes out as $\gamma/2\pi = \{1, -2, 1\}$. At 2/3 filling we find $\gamma/2\pi = \{-1, 2, -1\}$.

33 it is a simple matter to take the derivatives analytically. Integrating the resulting Berry curvature numerically over the RBZ to find the Berry phase gives 2π times an integer for each sub-band. This is exactly what is expected by analogy to the quantum Hall effect, discussed in more detail in Section 3.5.3. The sum of the phases over all bands is zero point-by-point in $\mathbf{k}$, since consideration of all sub-bands does not constitute a discarding of information. For $p/q = 1/3$ the bands have Berry phases $\gamma_n/2\pi = \{1, -2, 1\}$. This is shown in Figure 19.

To motivate the second Berry phase calculational method consider the following paradox. From Stokes' theorem it seems a line integral of the connection around the boundary of the RBZ should give the same result as a surface integral of the curvature across it - this was assumed in Equation 34. However, the RBZ is topologically a torus, so has no boundary, and an integral of the connection around the edge of the areas in Figure 19 does indeed give zero in each sub-band individually.

To see what goes wrong it is necessary to make a gauge choice for $\mathbf{A}$. Numerically, the Hamiltonian is diagonalized by the Lapack ZHEEV routine, which has no specification for complex phase of the returned eigenvectors. This is a problem when dealing with the connection since it is a gauge dependent quantity, and a randomized phase of the returned eigenvectors causes randomized jumps in $\mathbf{A}$ across the Brillouin zone. To remedy this, whenever ZHEEV is called we can multiply each eigenvector by $\exp(-i \arg(u_1))$, where u_1 is the first component of $|u\rangle$, making this component real. This is a form of gauge choice. The resulting gauge-fixed connection should be well-defined everywhere.

What happens when $u_1 = 0$, though? The answer is that there results a singular pole in the connection. The location of the pole in the RBZ depends on the choice of gauge.

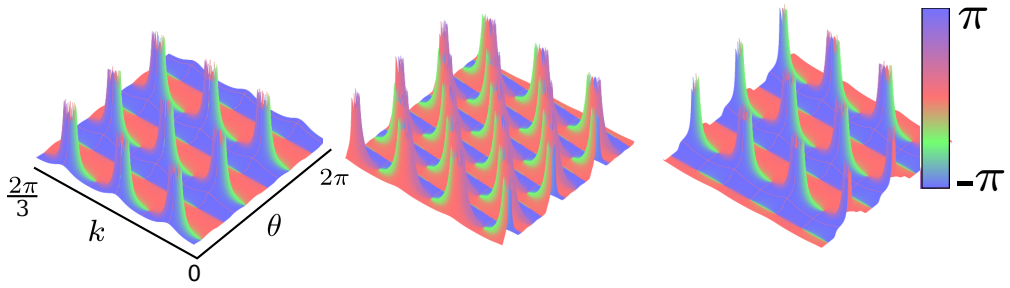


Figure 20: The Berry connection $\mathbf{A}_n$, a 2-component vector in each of the n sub-bands, across the reduced Brillouin zone for the three sub-bands at 1/3 filling. The height gives the $\mathbf{A}$ vector's modulus, the colour its argument $\operatorname{atan}(A_2/A_1)$. The corresponding Berry phases are $\gamma_n/2\pi = \{1, -2, 1\}$, which can be seen from the total 'charge' (phase winding, colour scale) of the poles. Note that singularities either correspond to zeroes or poles of the connection, and it is the latter which dictate the Berry phase. The Brillouin zone is slightly offset from the origin for clarity.

Having made a choice the pole can be located, and a closed line integral of the connection around this pole will give 2π times an integer. This is pictured in Figure 20.

In fact, if Chang's and Niu's method gives a result $2\pi m$ ($m \in \mathbb{Z}$), there must be exactly m poles in the connection per RBZ. These poles cancel the curvature arising from the nonsingular part of the field. Stokes' theorem is recovered. The singularities will move depending on the chosen gauge, but there must be m in total[12].

This resolves the paradox of how the line integral of a field (in this case the connection) around the reduced Brillouin zone can give zero when a surface integral of the curl of that field is nonzero. A nice way to think of it is in terms of Riemann sheets. Assume for a second there is exactly one pole in the connection per RBZ. Taking the line integral of $\mathbf{A}$ on a path close to and surrounding the singularity gives 2π. Expanding the area of the loop to include more of the reduced Brillouin zone, the value decreases as it picks up the curvature in the rest of the field, until the loop hits the RBZ edge and the total exactly hits zero (shown in Figure 20). The problem is, when the singularity is encircled, the line jumps onto a different Riemann sheet. To avoid this it is necessary to introduce a branch cut, and it's this which provides the boundary needed to get a nonzero result from Stokes' theorem. This is shown in Figure 21, along with a similar but distinct argument due to Kohmoto [35].

Taking the correct contour gives the same result as the method of Chang and Niu, but

[12]Note that the zeroes of the wavefunction in real space, being observable, are fixed.

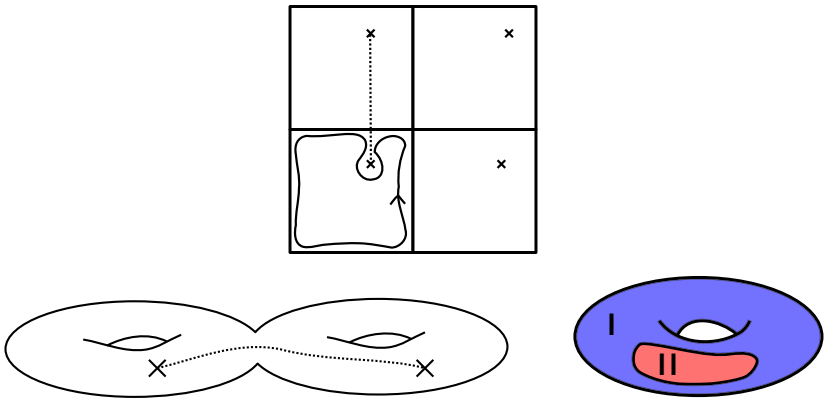


Figure 21: Stokes' theorem states that the integral of the curl of a vector field across a surface is equal to the line integral of that vector field around the boundary of the surface. This suggests the surface integral of the Berry curvature across the RBZ should equal the line integral of the Berry connection around the boundary of the RBZ - but the latter *has* no boundary as it is a torus. A singular pole in the connection introduces a branch cut, giving the required boundary. Top: at p/q filling the space seen by the connection is not a torus but q tori joined at the edges of the reduced Brillouin zones. Applying Stokes' theorem to one torus, it is necessary to avoid the singularity but otherwise take the obvious boundary. Bottom left: the situation with the two relevant reduced Brillouin zones. Bottom right: Kohmoto's argument [35] is different but related. He defines two regions with different gauge choices so that the phase is everywhere defined. The join between the regions forms the boundary required, and the matching condition for the gauges on the boundary gives the desired result.

an easier method in this case would be simply to count the singularities, and keep track of whether they are positive or negative (depending on the direction of phase winding). As expected for the 1/3 filling case, in the three sub-bands there are $\{1, -2, 1\}$ (positive) singularities.

As an interesting aside, note that the argument just given applies to any vector field on a closed manifold: if the field is to have a curl, it must also have singularities. An everyday corollary of this is what John Hannay refers to as the 'no wind theorem': given that the Earth has no boundary (it's topologically a sphere, S^2), and that the velocity field of the wind is not everywhere zero (and therefore has a curl by the 'hairy ball theorem' [21]), there must be a point on the Earth's surface where the amplitude of the wind's velocity field goes to zero in a singular manner (*i.e.* a vortex exists). The argument holds layer-by-layer in the atmosphere.

The third method of Berry phase calculation, due to King-Smith and Vanderbilt [36], allows calculation of closed line integrals of the connection without having to introduce gauge dependent quantities. Their formula is

$$\gamma_n = -\Im\mathfrak{m} \ln \prod_{s=0}^{M-1} \langle u_n(\mathbf{k}_s) \, | u_n(\mathbf{k}_{s+1}) \rangle$$

where s labels the sites along a closed path with the points $s = 0$, M equated. The phase introduced by a gauge transformation at site $\mathbf{k}_s$ gets cancelled by an exactly opposite contribution from the next term on the path.

This method is computationally efficient, but requires one to pick routes through the RBZ based on the unknown pole structure of $\mathbf{A}$. The logarithmic branch cut is quite apparent, but the formula is clearly only valid on the first Riemann sheet (numerical logarithm routines return the principal part of their kernel). A solution is to combine this method with a choice of gauge. By examining the buildup of the Berry phase γ along the path before it closes it is apparent whether the value is x or $x + 2\pi$.

The final method is extremely computationally efficient but is specific to the problem considered here, and the equivalent problem of the quantum Hall effect. It was first stated in reference [26], popularly known as TKNN according to the authors' initials. By the admission of the first author, T (David Thouless), the derivation proceeds "with some difficulty" [37], and I will omit the proof here. The result is a *Diophantine Equation*: an equation containing integer coefficients where integer solutions are sought.

In the present case, for filling fraction p/q, the Berry curvature s_n of the n^{th} sub-band is found from the equation

$$n = s_n p + t_n q \tag{36}$$

where $|s_n| \leq q/2$ [26, 37, 31]. Solving the equation is extremely quick numerically. There is a nice geometrical interpretation of Equation 36 which can be seen by defining vectors

$$\mathbf{s}_n \triangleq \begin{pmatrix} s_n \\ t_n \end{pmatrix}, \qquad \mathbf{p} \triangleq \begin{pmatrix} p \\ q \end{pmatrix}$$

in which case the equation can be rewritten

$$\frac{n}{\sqrt{p^2 + q^2}} = \mathbf{s}_n \cdot \hat{\mathbf{p}}$$

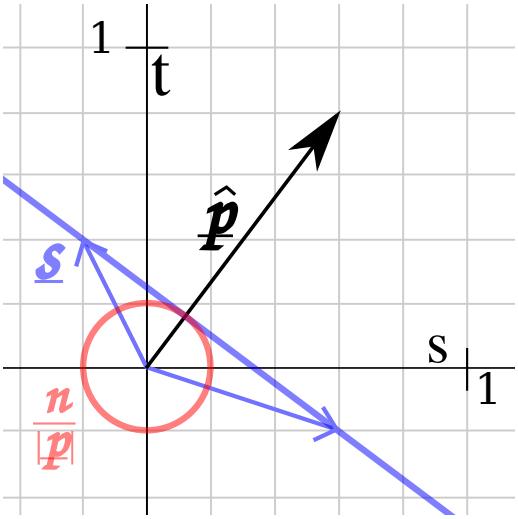


Figure 22: A geometrical interpretation of the Diophantine equation, Equation 36, for the case $p/q = 3/4$ and $n = 1$. A vector $\mathbf{s}_n = (s_n, t_n)$ is sought with projection $n/|\mathbf{p}| = 1/5$ along $\mathbf{p} = (p, q)$. The vector is required to hit a site of the 2D (s, t) lattice, and the vector with the smallest-magnitude s component is taken, in this case $s_1 = -1$.

so solutions are vectors $\mathbf{s}_n$ with a fixed projection $n/\sqrt{p^2+q^2}$ along the line $\mathbf{p}$. Note that $n \leq q$ so the desired projection is always less than one. The situation is illustrated in Figure 22.

3.5.3 The Berry Phase of CDWs: Quantized Particle Transport

In the previous section I demonstrated that when a p/q-filled chain develops a period q CDW, each of the resulting q sub-bands has a quantized Berry phase. Where is information being discarded for this to be happening, and why is the phase topological rather than geometrical?

When the Brillouin Zone, BZ, is restricted to a reduced Brillouin zone RBZ=BZ/q, there are $1/q$ as many k states in the RBZ, but there are q times as many sub-bands. As this is simply a mathematical reformulation there is no loss of information. Upon introducing a coupling between states at E_F, via the interaction Hamiltonian H_{int}, the bands split at the degeneracies into q sub-bands, p of which are filled. At this stage in the calculation, however, we assign a label $n \in [1, q]$ to each sub-band, and treat these sub-bands as independent. It is at this stage that information is lost: attention is restricted only to the occupied sub-bands, and information about the unoccupied sub-bands is discarded. An intuitive way to see how the Berry phase can enter is this: k-space is periodic with the period

of the BZ, but upon forming a CDW it *appears* that the periodicity increases to BZ/q, and therefore that $|u\,(k=0)\rangle = |u\left(k=\frac{2\pi}{q}\right)\rangle = |u\left(k=2\cdot\frac{2\pi}{q}\right)\rangle = \ldots$.; in fact these states are only equal up to a phase, and the curvature of the sub-bands necessitates these phases be nonzero. In this case the correct relation is $|u\,(k=0)\rangle = \exp(2\pi im/q)\,|u\left(k=m\cdot\frac{2\pi}{q}\right)\rangle$, so that the kets are still truly periodic over the full BZ.

In the IQHE the quantization of the Berry phase is observable either as a quantized Hall conductivity σ_{xy}, seen by measuring the transverse voltage upon application of a longitudinal current, or as an integer number of current-carrying 1D edge states on the boundary of the 2D sample [25]. In the case of CDWs in 1D, the θ component of the parameter space $\mathbf{k} = (k, \theta)$ is rather abstract, and the observable quantity associated with the quantized Berry phase is less clear.

Bearing in mind $k_2 = \theta$ has been shown to give the phase of the CDW, it seems the problem has to do with sliding a CDW and getting a current. This turns out to be the case [31]; varying θ over one cycle leads to a 'quantized adiabatic particle transport'. The analogues to the two currents j_x, j_y in the quantum Hall effect are now a current in real space j and a current in the space of the CDW phase, j_θ. By this it is meant that if the CDW were dragged along the lattice through one wavelength (implementing a j_θ current) a transport of electrons (j current) would be induced from one end of the crystal to the other. The number of electrons transported would be an integer given by the sum of the Berry phases of the occupied sub-bands.

This result is surprising for a number of reasons. First, given that the system develops an energy gap upon formation of the CDW, no conductance is expected at all. Second, there is no *a priori* reason to expect an integer number of electrons to be transferred when the electrons form a continuous probability distribution. Third, the sum of Berry phases of occupied sub-bands up to a given gap can be negative, as in the upper gap at 1/3 filling which has sum of Berry phases equal to -1; in this case the net result of dragging the CDW left through a full period is the transfer of one electron to the right! Perhaps most fundamentally, by analogy to normal water waves, it is as if the movement of the wave results in a net transfer of water molecules, which we know is not usually the case.

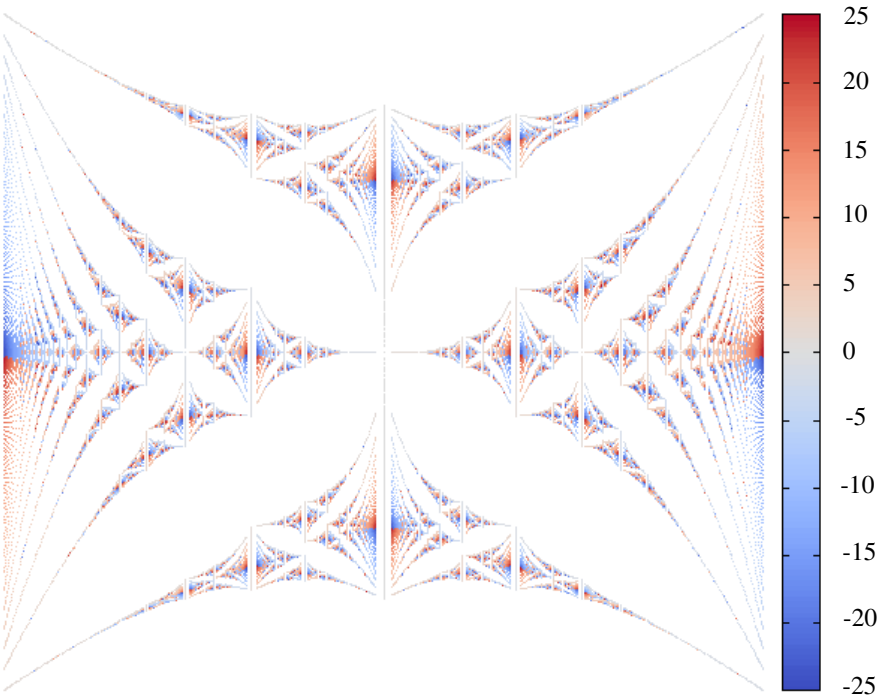


Figure 23: Allowed energies (y-axis, range $[-4t, 4t]$) against filling fraction $p/q \in [0, 1]$. The colour indicates the sum of the Berry phases up to and including each sub-band, which gives the conductivity of the relevant gap (the number of electrons transferred across the system upon dragging the CDW through a full period). The plot settings are as in Figure 14.

3.5.4 Results: How the Butterfly Got its Colours

The methods of calculation of the Berry phases of the different sub-bands have been explained in detail. For the purposes of colouring the butterfly with these phases, Figure 23, the diophantine equation is the quickest, with a verification of particular cases using the method of Chang and Niu.

The colour of each sub-band in Figure 23 corresponds to the sum of the Berry phases of the sub-bands up to and including that one. This corresponds to the Hall conductance of the gap lying immediately above the sub-band, which is equal to the number of electrons which would be transferred across the system if the CDW were dragged through one full period. Of particular interest is the fact that the undersides of the wings of the butterfly have a constant colour, indicating that the wing itself (band gap) has a constant conductance for varying filling fraction.

3.6 Self-Consistent Solution

The results presented so far for rational filling have been based on a carefully chosen order parameter corresponding to a purely site-centred CDW. Mean field theory derives its utility from its self-consistency. I will now consider the self-consistent solution, for the formation of CDWs in 1D bands of rational filling, by solving Equation 31 for each p and q at a given h. As we don't meet true fractals in everyday life it would seem likely that the self-consistent solution will lose its approximately scale-invariant quality. Hofstadter himself considered this issue by allowing for finite instrumental resolution, and showed that the fine-grained structure is washed out [16]. In fact the self-consistent butterfly loses its wings even before appeal to sensible considerations such as Hofstadter's.

3.6.1 Varying Interaction Strength

The Coulomb interaction strength enters the Hamiltonian as the real variable h. I found that for all fillings and interaction strengths the third component of the order parameter, corresponding to a current, is zero: $|\Delta_3| = 0$. This confirms that there is no current in the self-consistent 1D solution, as expected. Whatever complex phase is given to Δ as a seed is the result found self-consistently, so the $U(1)$ symmetry of the problem is maintained. This is again to be expected without a coupling to the lattice, as the CDWs have no knowledge of absolute position and can slide without resistance.

The remaining components of the order parameter, corresponding to site- and bond-order, are a good fit to

$$|\Delta_{1,2}(h)| \approx 2\Lambda \exp(-1/N_Q h)$$

where the cutoff energy Λ (*cf.* Section 2) is of the order of the bandwidth. The 'generalized BCS' guess for N, naïvely extrapolating the half-filling approximate solution, following Equation 17, would be

$$N_Q = 8g(E_F) = \frac{4}{\pi t\,|\sin(Q/2)|}.$$

For $Q = 2\pi \cdot \frac{1}{6}$ this predicts $N_Q = 8/\pi t$, which is a surprisingly good match to the numerical

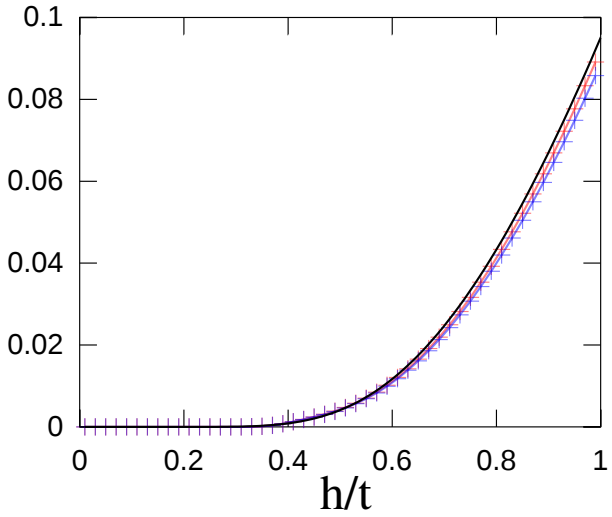


Figure 24: Self-consistent solution for 1/6-filling. Red: site-order$|\Delta_1(h)|$, blue: bond-order $|\Delta_2(h)|$. The black line indicates the line $2.2\exp(-\pi t/h)$ which is of the form of the 'generalized BCS' guess. The hopping parameter $t = 1$ here.

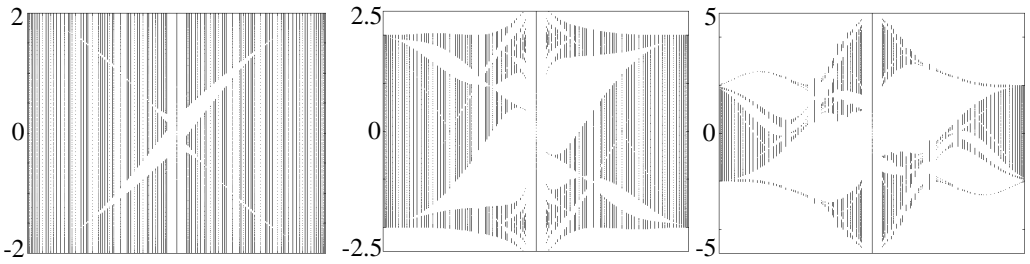


Figure 25: The self-consistent solution for different interaction strengths h, with $q_{max} = 18$ and $k_{max} = 200$. Left, $h/t = 0.5$, middle, $h/t = 1$, right, $h/t = 2$. The abscissa is the filling fraction from 0 to 1, the ordinate the energy in units of the hopping parameter t. Note that the unperturbed bandwidth is $4t$, so these h values are unphysically high. A point of interest is that even for large denominator q, say 10/17, a gap still opens up, so $\Delta > 0$.

result as shown in Figure 24.

In Figure 25 I show the self-consistent solution for the butterfly (allowed energies versus filling fraction) at different couplings h. The symmetry about $p/q = 1/2$ is lost as a result of including the Fermi function in the calculation: a significant gap now opens up only at the CDW wavevector, which is a more physical solution. For example, for 3/7 filling, gaps previously opened at $2\pi \cdot p/7$ for all $p < 7$, whereas now a significant gap opens only at $p = 3$.

Varying both the filling and interaction strength it is evident that the site- and bond-order parameters behave somewhat independently. Plots of $|\Delta_{1,2}(Q, h)|$ are given in Figure 26. Convergence of the self-consistent series requires about 20 iterations, so I applied a series-

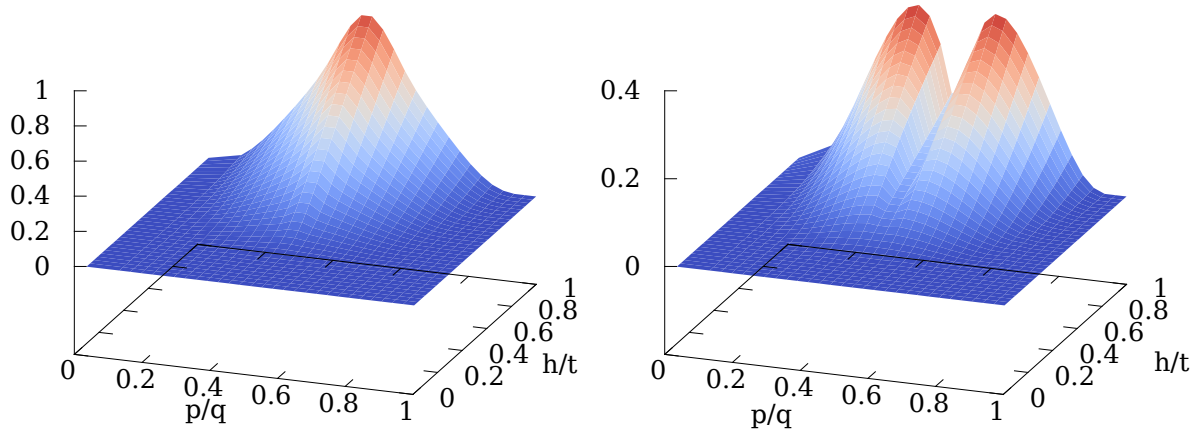


Figure 26: The self-consistent gap function as a function of Coulomb interaction strength h/t and filling fraction p/q. Left: site-order parameter $|\Delta_1|/t$; Right: bond-order parameter$|\Delta_2|/t$.

acceleration method (specifically a Shanks transformation [38]). For all the cases I checked this gave in 3 iterations the result of around 100 iterations of the original series. The cases in which the Shanks converges more slowly are those where the self-consistent result is zero. In these cases taking 3 iterations gives $\mathcal{O}\left(10^{-5}\right)$, which is still negligible compared to the $\mathcal{O}\left(10^{-1}\right)$ results of the other order parameter components.

Varying the magnitude of the start seed in the range $[0, 1]$ has little to no effect on the resultant magnitudes of the vectors, and the relative magnitudes of the three components have similarly little effect. In all cases $|\Delta_3| = 0$ up to the error given above. The phase of the start seed exactly dictates the phase of the result. Using a seed $(1, 1, 1)\exp(i\phi)$ and varying $\phi \in [0, 2\pi]$ gave $< 2\%$ variation in magnitude of $\Delta_{1,2}$, but $\phi_{1,2} \approx \phi$ in each case (with a small wobble).

3.6.2 Topology of the Self-Consistent Solution

Figure 27 shows the results for the self-consistent solution with $h/t = 1$, coloured by the sum of Berry phases of the occupied sub-bands. Although the value of $|\Delta_Q(h)|$ varies with filling fraction $Q = 2\pi p/q$, the self-consistent value is never zero, meaning that no gaps close up relative to the Hofstadter case of Figure 23. This means the two cases are adiabatically connected, and the same diophantine equation can be used to colour Figure 27.

In reality the values of $|\Delta|$ may become arbitrarily close to zero, and gaps should be

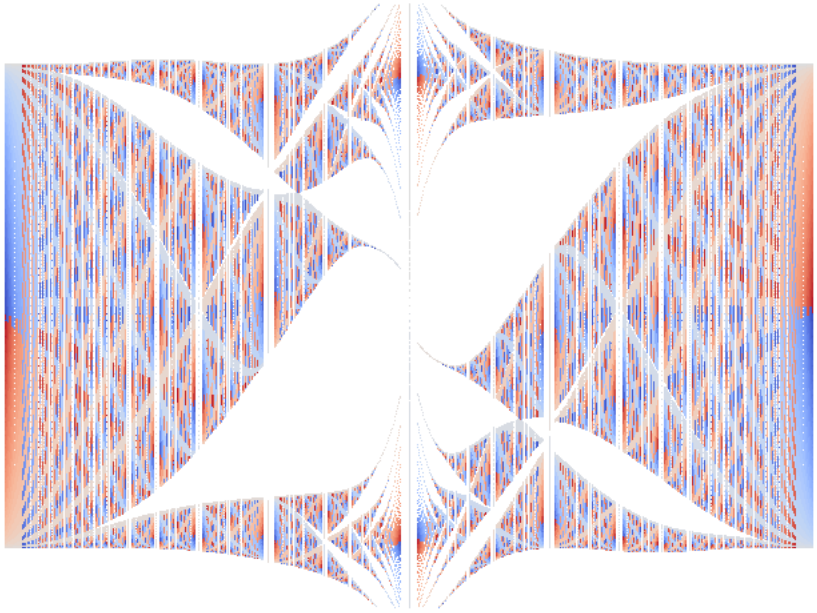


Figure 27: The allowed energy levels (range $\pm 2.5t$) against filling fraction $p/q \in (0, 1)$, $q \in [0, 50]$ for the self-consistent solution with $h = t$. The colour of a sub-band indicates the sum of the Berry phases of sub-bands up to and including that one. The full colour scale $C_1 \in [-25, 25]$ is included (darker blue being more negative).

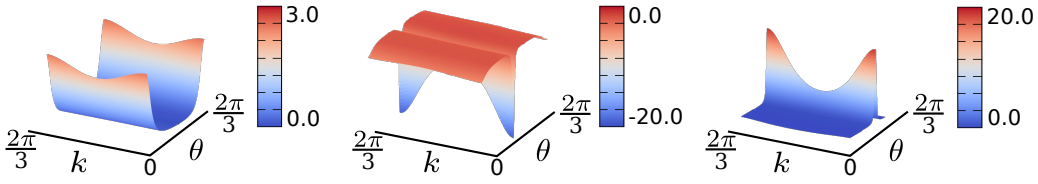


Figure 28: The Berry curvature across the magnetic Brillouin zone for the case of 1/3 filling, for the self-consistent solution with $h = t$. Although the values are different to the Hofstadter case of Figure 19, the integrals over the RBZ give the same results $\{1, -2, 1\}$. This is expected since the value of Δ is nonzero and the two cases are therefore adiabatically connected.

considered to be 'closed' when of the order of thermal or disordering effects. When two sub-bands meet, their Berry phases must add (proven in the next chapter). For a reasonable choice of parameters this still allows at least the large gaps to remain open.

4 Irrational Filling: Incommensurate Charge Order and Quasicrystals

In previous chapters I have considered commensurate CDWs, where a rational filling of the system leads to a CDW with a period which is an integer multiple of the underlying lattice. In this chapter I will consider the case of incommensurate charge density waves (ICDWs). The filling, being an irrational number, is no longer expressible as a fraction, and the period of the resulting CDWs is an irrational multiple of the lattice spacing. I will argue that such states are 'quasicrystalline', in a sense to be described below, and will use this fact to address a certain controversy which has arisen regarding the dimension of quasicrystals as measured by the types of quantum number they are able to exhibit.

In Section 4.1 I provide some background on quasicrystals and a recent classification scheme of the possible topologies of free fermion theories known as the 'Tenfold Way'. In Section 4.2 I investigate claims that quasicrystals are manifestations of higher-dimensional crystals which, nevertheless, exist in our world. The argument is intricately tied to topology and the aforementioned topological classification scheme. I rephrase the issue in terms of ICDWs and find agreement with the counter-argument - that it is only whole families of quasicrystals which demonstrate a higher-dimensional topology. Finally, in Section 4.3, I show how ICDWs can lead to a new, non-local, growth mechanism for quasicrystals, potentially increasing the set of naturally-occuring real-world cases from two to infinity.

4.1 Background

4.1.1 Quasicrystals

Figure 29 shows two examples of 2D quasicrystals. The definition of a quasicrystal is a matter of some debate [39]. Loosely, the idea is this: a crystal is a periodic repetition of a certain basic unit, the unit cell. The unit cell is tessellated to fill all of space without any gaps. A quasicrystal has two or more unit cells[13]. These are arranged in such a fashion

[13]Even this is debated. The Socolar-Taylor tile has been suggested as a one-tile quasicrystal in 2D, but the tile requires markings and matching rules dictating which edges can touch which others. It is not representable simply as a shape. In 3D the Schmitt-Conway-Danzer tile can form (something closely related to) the brilliantly-named quasiperiodic structure Gyrobifastigium [39]. The task of finding a single-tile quasicrystal is known as the 'einstein problem' (German for 'one stone') and is closely related to Hilbert's 18^{th} Problem.

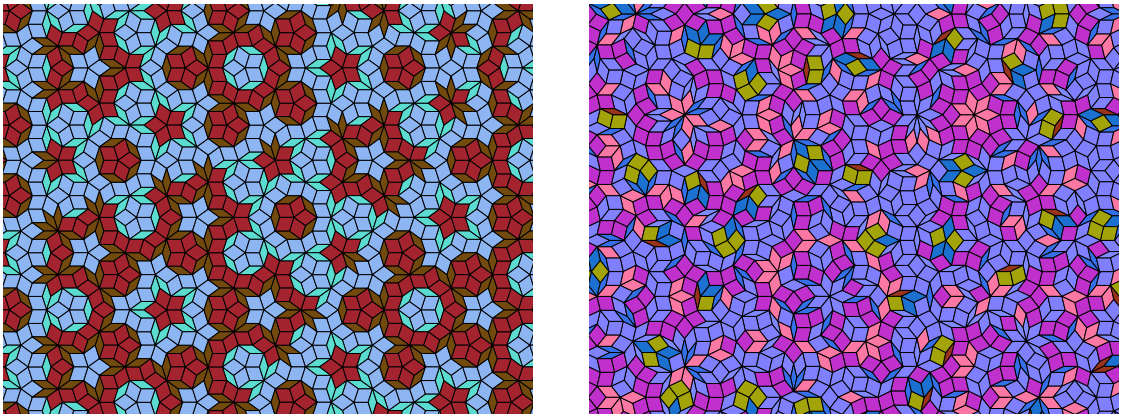


Figure 29: Two examples of quasicrystals in 2D: the left image has centres of 5-fold symmetry, the right centres of 7-fold symmetry (this pattern has three unit cells). The colours are purely aesthetic. I am indebted to Prof. Eric Weeks at Emory University for the use of his C code to generate these images.

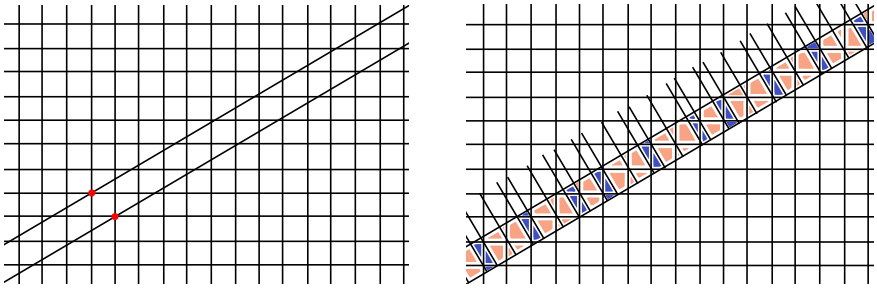


Figure 30: Generating a 1D quasicrystal by projecting a 2D crystal. Left: a 2D lattice is drawn, in this case a square lattice. A line is drawn at angle θ, with $\tan(\theta)$ irrational such that the line hits precisely one lattice point. A second line is drawn parallel to the first intersecting the opposite vertex of the same unit cell. Whenever a point of the 2D lattice falls between the lines the point is projected perpendicularly. Right: the result of the construction is a sequence of two unit cell lengths, $\cos(\theta)$ and $\sin(\theta)$, which is never-repeating. The different unit cells have been coloured differently for clarity.

that they fill space, again without any gaps, but this time not in a periodic manner: there is no minimal unit which can be tessellated periodically to generate the structure. The quasicrystal is not periodic, but nor is it aperiodic: aperiodic tilings are random, but there is an underlying structure to the arrangement of quasicrystal cells.

In fact, quasicrystals can be generated by the projection of *crystals* from a higher dimension [40, 39]. This provides an elegant way to see the underlying order in the patterns. The method is most easily demonstrated in the case of 1D quasicrystals formed by projection of 2D crystals, shown in Figure 30.

In this case, a square lattice of unit side lengths is chosen in 2D. A line is drawn at an angle θ to the lattice such that $\tan(\theta)$ is an irrational number. It therefore hits precisely one point of the 2D lattice. A second parallel line is drawn to intersect the opposite

corner of the same unit cell. Whenever a point of the 2D lattice falls between the lines it is projected perpendicularly onto them. It is easy to see that all distances between the projected points are either $\cos(\theta)$ or $\sin(\theta)$, *i.e.* the construction generates precisely two unit cell types. The sequence of Longs (L) and Shorts (S) contains no infinitely repeating finite subsequence. This is guaranteed by the irrational $\tan(\theta)$. The result is therefore a quasicrystal as defined above. We also see that although the sequence of Ls and Ss is not periodic it is still predictable, since we can appeal to the parent 2D lattice which *is* periodic.

What would happen if we translated the parallel lines along the perpendicular connecting them? There is a periodicity given by the 2D unit cell: once the lines are translated across one cell's diagonal the original quasicrystal is returned. For each infinitesimal shift *en route,* however, points of the 2D lattice will drop in and out of the region between the lines, changing the sequence of Ls and Ss. The lines will still hit precisely one lattice site each. The original sequence of Ls and Ss has therefore been translated along the line some amount. The family of quasicrystals generated by such a translation of the lines is known as a 'Local Isomorphism Class' [39, 40]. Two quasicrystals are locally isomorphic if and only if every finite string of Ls and Ss appearing in one appears in the other [40]. A physical consequence is that two locally isomorphic quasicrystals have identical diffraction patterns [39, 41].

The diffraction patterns of quasicrystals are particularly interesting. In real space a quasicrystal is really a slice, rather than a projection, through a higher-dimensional crystal. In reciprocal space, however, a quasicrystal is a true projection of a higher-dimensional crystal. In the 1D case just considered, the 2D crystal's reciprocal lattice is formed in the usual manner from the real-space lattice. All the points of this 2D reciprocal lattice are then projected perpendicularly onto a line parallel to the slice direction. The result is a dense covering of the space (similar to the diffraction pattern of a disordered system), but a dense covering containing sharp, well-defined peaks (similar to the diffraction pattern of a crystal). The situation is shown in Figure 31.

The dimension of the quasicrystal's reciprocal lattice is higher than that of its real-space lattice. This is actually fairly intuitive: we now have two unit cell types, so the appearance of two different reciprocal lattice vectors is perhaps unsurprising. The diffraction pattern of

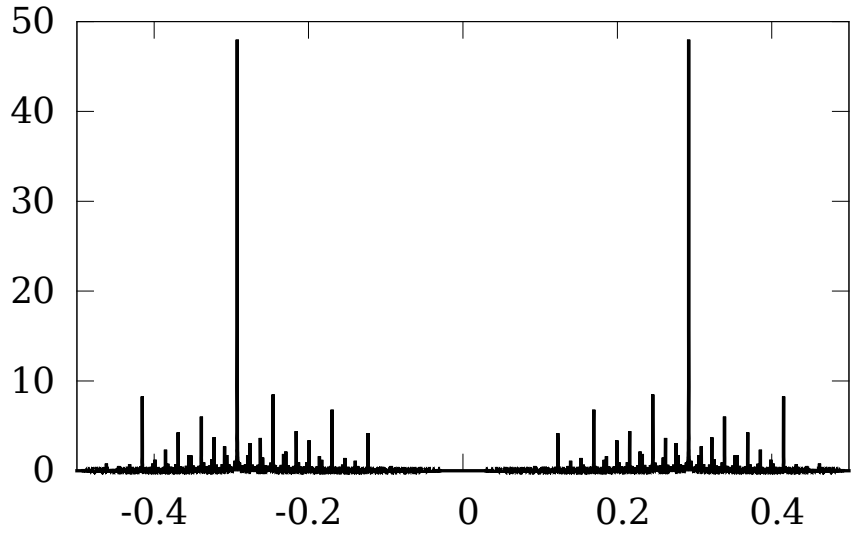


Figure 31: The calculated Fourier transform of a 1D quasicrystal (arbitrary units) over a range of viewing angles. The spectrum is dense, similar to disordered materials, but also has sharp peaks similar to the Bragg peaks seen in crystals.

the 1D quasicrystal could just as easily be generated by considering the set of points which can be reached by combinations of these vectors along a line - the covering will be dense because the relative lengths of the vectors are incommensurate. The higher dimensionality of the reciprocal lattice follows from rotating one set of vectors relative to the other, which is a lot easier to visualise, but the true situation is returned to by projecting back onto the one dimension which physically exists. Senechal argues for this as a definition of quasicrystallinity: a quasicrystal is a system where the dimension of the reciprocal lattice is higher than the dimension of the real-space lattice [39].

Some authors prefer a definition of quasicrystallinity which appeals to the violation of the 'crystallographic restriction theorem'. This theorem states that crystals in two or three dimensions can have only 2-, 3-, 4-, or 6-fold symmetry [42]. In contrast, the quasicrystals in Figure 29 show local regions of 5- and 7-fold symmetry respectively. The corresponding diffraction patterns in fact have 5- or 7-fold global symmetry, following again from the fact that the reciprocal lattice is a true projection of the higher-dimensional crystalline reciprocal lattice [39]. In Figure 32 I reproduce the fivefold-rotationally symmetric diffraction pattern of the first known quasicrystal, which was found by D. Schechtman and others [43]. A definition of quasicrystallinity in terms of rotational symmetries rules out the possibility of one-dimensional quasicrystals since there is no sense of rotational symmetry in 1D. As

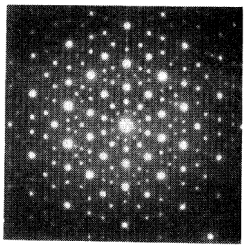


Figure 32: X-ray diffraction pattern of the first known quasicrystal, an alloy of Aluminium and Manganese. The pattern has 10-fold symmetry, forbidden by the crystallographic restriction theorem. Reproduced from [43].

I will be concerned *solely* with one-dimensional systems I will instead define a 1D quasicrystal to be a sequence of long and short unit cells generated by the projection method in Figure 30. I will avoid further comment on higher-dimensional quasicrystals.

4.1.2 The Tenfold Way

In what has come to be known as Wigner's Theorem, the symmetries of a Hamiltonian fall into two classes: unitary, $\mathcal{U}$, or antiunitary [44]. An antiunitary operator can be written as $\mathcal{U}K$ where K denotes complex conjugation:

$$\begin{aligned} KiK^{-1} &= -i \\ &\downarrow \\ \{K, i\} &= 0. \end{aligned}$$

As an example, the time-reversal operator $\mathcal{T}$ in a basis of spin J can be written [45]:

$$\mathcal{T} = \exp\left(-i\pi J_y\right) K$$

where the following relations hold:

$$\begin{aligned} \{\mathcal{T}, i\} &= 0 \\ \{\mathcal{T}, J\} &= 0 \\ \{K, J\} &= 0. \end{aligned}$$

The square of an antiunitary operator is unitary; for $\mathcal{T}$, noting that the above relations imply $[K, iJ] = 0$, we have

$$\begin{aligned}\mathcal{T}^2 &= \exp(-i\pi J_y) K \exp(-i\pi J_y) K \\ &= \exp(-2\pi i J_y) K^2 \\ &\equiv \exp(-2\pi i J_y).\end{aligned}$$

The eigenvalues of J_y move in half-integer steps from zero, so the eigenvalues of $\mathcal{T}^2$ are ± 1 depending on the spin. We can therefore classify all possible gapped Hamiltonians[14] into three sets: symmetric under $\mathcal{T}^2$, antisymmetric under $\mathcal{T}^2$, and those with no $\mathcal{T}$ symmetry, which I will label as $\mathcal{T}^2 = 0$ in-keeping with convention. This classification, developed by Freeman Dyson in 1962, is known as the 'Threefold Way' [46, 47]. The reasoning is based on a general anti-unitary operator, not necessarily $\mathcal{T}$. Dyson's work showed that[15], writing the Hamiltonian in a second-quantized basis,

$$\hat{H} = \sum_k \left(\hat{\psi}_k^\dagger, \hat{\psi}_{k+Q}^\dagger, \ldots, \hat{\psi}_{k+qQ}^\dagger\right) H_k \begin{pmatrix} \hat{\psi}_k \\ \hat{\psi}_{k+Q} \\ \vdots \\ \hat{\psi}_{k+qQ} \end{pmatrix}$$

the elements of the Hermitian matrices H_k are either real $\mathbb{R}$, quaternionic $\mathbb{H}$, or complex $\mathbb{C}$, for the three cases $\mathcal{T}^2 = 1, -1, 0$ respectively. More precisely:

1. if $\mathcal{T}^2 = 1$ there exists a basis e_i such that $\mathcal{T} e_i = e_i$, and in this basis H_k is real and symmetric.

2. As an example of a $\mathcal{T}^2 = -1$ situation consider the interaction Hamiltonian with a disordered potential U_{dis} and spin-orbit coupling: $\hat{H}_{int} = U_{dis}(\hat{\mathbf{x}}) + \sum_{ijk} \epsilon_{ijk} V_{SO}^i \sigma^j \hat{\mathbf{p}}^k$ with σ the vector of Pauli matrices, ϵ the Levi-Civita symbol and $\{i, j, k\} \in [1, 3]$. The Pauli matrices form a basis for $SU(2)$, isomorphic to the group of quaternions

[14] Gapped Hamiltonians, with a nonzero energy gap between the ground and first excited state, are assumed throughout this chapter.

[15] This is actually a re-interpretation of Dyson's work in the modern framework.

(set $\mathbb{H}$) [47].

3. If no symmetry is present, '$\mathcal{T}^2 = 0$', the Hamiltonian is unrestricted and is therefore complex Hermitian.

The disordering potential in (2) is important. Dyson was working with random matrices, and assumed disordered systems. In any ordered systems there are many more symmetries present, and while the above general reasoning should hold the three classes may not be of much use.

Dyson's working, and the earlier work of Wigner on which it is based, makes heavy use of group theory, and the technical aspects are beyond the limits of this thesis (and my understanding) [46, 44]. It suffices to say that Dyson's three classes can be given labels according to the Killing-Cartan classification of simple Lie groups. In particular,

1. $\mathcal{T}^2 = 1$, group AI (orthogonal)
2. $\mathcal{T}^2 = -1$, group AII (symplectic)
3. $\mathcal{T}^2 = 0$, group A (unitary).

In 1996 Altland and Zirnbauer, also considering random matrices and disordered systems, extended the classification by adding a second distinct anti-unitary operator[16], charge conjugation symmetry $\mathcal{C}$. As before, if $\mathcal{C}$ symmetry is present all H_k matrices are classified by $\mathcal{C}^2 = \pm 1$, otherwise the convention is to label the set $\mathcal{C}^2 = 0$. Among the physical motivations for extending the scheme included the observation of Andreev reflection, which did not fit into the existing classes [47]. Choosing $\mathcal{C}$ as the second anti-unitary extends Dyson's work defined on Hilbert space to the general multi-particle Fock space [49, 47].

The combinations of $\mathcal{C}$ and $\mathcal{T}$ give nine possible classes. A tenth comes about because a system can have neither $\mathcal{C}$ nor $\mathcal{T}$ symmetry but can still have the combination $\mathcal{S} \triangleq \mathcal{C}\mathcal{T}$. Note that $\mathcal{S}$ is unitary, and if present as a symmetry its eigenvalue is unity. The full set of ten possible symmetry classes is given in Table 1, where the corresponding simple Lie groups which lend the classes their names are also listed.

The additional restrictions allow a more precise specification of the possible forms of H_k. The remaining classes fall into two categories. In the first set, $\mathcal{C}$ symmetry is present,

[16] In particle physics $\mathcal{C}$ is defined to be unitary [48]. The K is added here to give the correct conjugation properties of BdG Hamiltonians in the usual form.

Class	diagonal	off-diagonal	simple Lie Group	
A	$\mathbb{C}$	$\mathbb{C}$	$U(N)$	Unitary
AI	$\mathbb{R}$	$\mathbb{R}$	$U(N)/O(N)$	Orthogonal
AII	$\mathbb{H}$	$\mathbb{H}$	$U(2N)/Sp(2N)$	Symplectic
D	$\mathbb{C}$	antisymmetric $\mathbb{C}$	$SO(2N)$	BdG
C	$\mathbb{C}$	symmetric $\mathbb{C}$	$Sp(2N)$	BdG
CI	0	symmetric $\mathbb{C}$	$Sp(2N)/U(N)$	BdG
DIII	0	antisymmetric $\mathbb{C}$	$SO(2N)/U(N)$	BdG
AIII	0	rectangular $\mathbb{C}$	$U(N+M)/U(N)\times U(M)$	Chiral Unitary
BDI	0	rectangular $\mathbb{R}$	$O(N+M)/O(N)\times O(M)$	Chiral Orthogonal
CII	0	rectangular $\mathbb{H}$	$Sp(N+M)/Sp(N)\times Sp(M)$	Chiral Symplectic

Table 1: The classification of possible quadratic Hamiltonians according to their behaviour under anti-unitary symmetries, after [50]. 'Class' is the Killing-Cartan label for the appropriate Lie group, the group itself listed in the rightmost two columns. This restricts the possible forms of the on-diagonal and off-diagonal parts of the Hamiltonian matrix H_k as discussed in the text. The top three classes form Dyson's threefold way. The next four fall into the Bogoliubov-de Gennes formalism, and the final three form the chiral classes (except for AIII these are also BdG). The total result is the Altland-Zirnbauer tenfold way [49, 50, 51, 47].

so H_k can be decomposed into 2×2 blocks corresponding to particles and antiparticles. This is just the Bogoliubov-de Gennes basis used in superconductivity, or charge density waves at half-filling as we saw in Chapter 2. The on-diagonal blocks are either general Hermitian or zero. The off-diagonal blocks are either symmetric, $M^\dagger = M$, or antisymmetric, $M^\dagger = -M$. This gives four of the classes, listed as BdG in Table 1. The remaining three cases are known as the chiral classes, where the operator $\mathcal{S}$ is known as the chiral operator. These cases have zeroes for diagonal entries and rectangular off-diagonal matrices. The elements of these matrices follow those of the threefold way. Note that any entry on the table with a nonzero $\mathcal{C}^2$ element can be written in the BdG form, so of the additional seven entries introduced by Altland and Zirnbauer six are BdG type. The remaining case has $\mathcal{C}^2 = \mathcal{T}^2 = 0, \mathcal{C}\mathcal{T} = 1$.

Although I will not go into details regarding the simple Lie groups, an important result of Altland and Zirnbauer was to observe that the first homotopy groups of these simple Lie groups give the sets of possible edge states / topological phases in the different classes in different dimensions. The first homotopy group can be thought of as the set of inequivalent 1D loops one can draw in a space, where two loops are equivalent if they can be continuously deformed into one another. For example, the first homotopy group of the circle S^1 is $\mathbb{Z}$, since the loop can wind the circle any integer number of times (the set of integers being

Class	$\mathcal{T}^2$	$\mathcal{C}^2$	$\mathcal{S}^2$	$d=0$	1	2	3	4	5	6	7
A	0	0	0	$\mathbb{Z}$	0	$\mathbb{Z}$	0	$\mathbb{Z}$	0	$\mathbb{Z}$	0
AIII	0	0	1	0	$\mathbb{Z}$	0	$\mathbb{Z}$	0	$\mathbb{Z}$	0	$\mathbb{Z}$
AI	1	0	0	$\mathbb{Z}$	0	0	0	$\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$
BDI	1	1	1	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0	0	$\mathbb{Z}$	0	$\mathbb{Z}_2$
D	0	1	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0	0	$\mathbb{Z}$	0
DIII	-1	1	1	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0	0	$\mathbb{Z}$
AII	-1	0	0	$\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0	0
CII	-1	-1	1	0	$\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0
C	0	-1	0	0	0	$\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0
CI	1	-1	1	0	0	0	$\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$

Table 2: The 'Tenfold Way', after [50]. The rows are re-ordered relative to Table 1. Given the dimension of the parameter space d there are three possible topological classifications: 0 for no non-trivial phases; $\mathbb{Z}_2$ for one non-trivial phase and one trivial phase; and $\mathbb{Z}$ for any number of non-trivial phases. These sets are the homotopy groups of the corresponding simple Lie groups, in the stated dimension [51, 52].

denoted $\mathbb{Z}$). Thus $\pi_1\left(S^1\right) = \mathbb{Z}$. On a sphere S^2, however, all loops can be contracted to a point, so $\pi_1\left(S^2\right) = 0$.

Although the working is complicated, the result is both very elegant and very relevant to this thesis. It is shown in Table 2. The dimensions in question are the dimensions of the parameter space rather than the physical space: to return to the familiar example of rationally-filled CDWs in 1D, while the physical space is 1D the parameter space, spanned by k and θ, is 2D.

To illustrate some well-known examples from Tables 1 and 2: if $d = 2$ and no symmetries are assumed (class A) we have the standard quantum Hall effect. In this case the dimension of the parameter space matches that of real space. In the same system, if we add in time reversal symmetry by removing the external magnetic field, but maintain $\mathcal{T}^2 = -1$ by having permanent spin-up and spin-down edge currents running in opposite directions (by some mechanism), we have instead the quantum spin Hall effect (class AII) with its famous $\mathbb{Z}_2$ topological number found by Kane and Mele [53]. In the same symmetry class but with $d = 3$ is the $\mathbb{Z}_2$ topological insulator of Fu, Liang and Kane [54]. Other examples are the topological superconductors in $d = 2$, with $\mathcal{T}^2 = 0$ due to the non-zero angular momentum of the Cooper pairs: $\mathcal{C}^2 = 1$, class C, $d + id$; and $\mathcal{C}^2 = -1$, class D, $p + ip$ [55].

As a final note in this introduction, although the work to date has largely concerned disordered systems, recent advances have been made in extending the classification to include

lattice symmetries (space groups) [56, 57, 58, 59]. In this thesis I will only be concerned with 1D systems, which have only two possible space groups: trivial, and dihedral (mirrored). In fact it will turn out that even in the dihedral case the symmetry is spontaneously broken by the CDW. This means that no complications arise due to the space group in 1D, and the Tenfold Way applies even in the crystalline case.

4.2 Are 1D Quasicrystals 2D?

A debate has arisen recently as to whether quasicrystals fit into the Tenfold Way directly, or whether they instead adopt the topological properties of the higher-dimensional parent crystal from which they are a mathematical projection (or, more precisely, a slice). In this section I will outline both sides of the argument, before demonstrating the equivalence between 1D quasicrystals and incommensurate CDWs. The mapping gives a new perspective from which to view the debate, and from this perspective it would seem it is only families of quasicrystals, rather than individual members of the family, which demonstrate the topological properties in question.

4.2.1 The Harper Equation

Kraus *et al.* consider a 1D tight-binding model with a periodic on-site modulation potential [23]. The Hamiltonian takes the form

$$H(\theta)\,\psi_n = t\,(\psi_{n+1} - \psi_{n-1}) + \lambda \cos(2\pi b n + \theta)\,\psi_n \tag{37}$$

with lattice sites n, hopping parameter t, and modulation amplitude λ. In this case the periodic modulation is considered to be externally enforced somehow, so the period b and phase θ can be chosen arbitrarily. The similarity of the expression to the formulæ appearing throughout Chapter 3, such as Equation 29, is no coincidence. In fact they are all instances of the Harper equation [16], considered by Hofstadter in his study of the quantum Hall effect which was discussed in Section 3.4. The parameter b, for example, appears in Hofstadter's work as the filling fraction Φ/Φ_0 (the ratio of the magnetic flux through a plaquette to the flux quantum), or in Chapter 3 of this thesis as the filling fraction of the band.

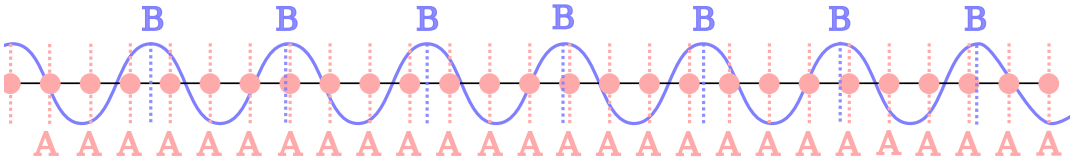


Figure 33: A lattice with a periodically modulated on-site potential, where the periods are incommensurate. The combination is quasiperiodic in the sense that, if we label the sites A and the peaks of the potential B, the sequence of As and Bs forms a quasicrystal. In this case the two unit cell types are BAAA and BAAAA.

If b is chosen to be irrational the period of the modulation is incommensurate with the lattice spacing. The combined system, with its two incommensurate periods, can be considered quasicrystalline. In particular, if we assign an A to each lattice site and a B to each peak in the on-site modulation, the sequence of As and Bs forms a quasicrystal, as shown in Figure 33. There is some sleight of hand here: this labelling procedure removes information about the proximity of the peaks, and in a true quasicrystal it is sometimes (although not always) required that there be a shortest distance between peaks [39]. However, Kraus and Zilberberg show separately that the topological properties of Equation 37 are identical to those of true quasicrystals [60].

As explained in Section 4.1.1, 1D quasicrystals can be generated as projections of 2D crystals by drawing two parallel lines which intersect the 'parent' 2D lattice once each, and projecting all points of the parent lattice which fall between the lines (Figure 30). By way of close analogy, in Equation 37 we see that varying the parameter b has the effect of rotating the parallel lines used in the projection. The rôle of θ, which gives the relative offset of the lattice and the periodic potential, is less clear. In fact varying θ corresponds to translating the parallel lines perpendicularly to their direction. In terms of the quasicrystal this generates different quasicrystals in the same local isomorphism class. In the 1D picture of Figure 33, if we vary θ and change the relative offset of the lattice and potential, we mix up the sequence of As and Bs, but all the generated sequences can be seen to be closely related as the period of the potential is unchanged. In the 2D picture of Figure 30, if we vary θ and translate the parallel lines, some of the points of the 2D lattice drop out of the projection region and others drop in, mixing the sequence of quasicrystal unit cells, but the sequences can be seen to be closely related as the angle of the lines is unchanged.

Kraus *et al.* used optical waveguide arrays to realize Equation 37. The spacing of the

waveguides was made quasiperiodic along a line, then neighbouring waveguide arrays had orderings of cells so as to effectively sweep through θ. The entire apparatus thus recognised a whole 'family' (local isomorphism class) of quasicrystals. The setup was used to demonstrate topological optical pumping: injecting light into one end of the first quasicrystal leads to its transference to the other end as the family of quasicrystals is swept though [23, 61]. The pumping is quantized in precisely the same manner as the transverse Hall conductivity in the quantum Hall effect, or the quantized adiabatic particle transport in commensurate CDW systems given in Chapter 3.

When considering this particle transport in 1D CDW systems it was important to note that there are two degrees of freedom in the problem: wavevector k and CDW offset from the lattice, θ. This explained the mathematical link to the 2D quantum Hall effect: the second degree of freedom comes from θ rather than a second spatial dimension. The same is true in the case of optical pumping in quasicrystals, this time with the degrees of freedom being movement along the quasicrystal (translations generated by the wavevector k coming from the Fourier transform of Equation 37) and θ switching between members of the local isomorphism class. Physically, in the CDW systems I found that dragging the CDW through a full period lead to the transfer of an integer number of electrons across the sample. The equivalent here is that re-ordering the cells of the quasicrystal in the correct manner, so as ultimately to return to the initial configuration, transfers an integer number of electrons across the system[17]. Mathematically the calculation proceeds in a similar manner to that considered in Chapter 3, with an integral of the Berry curvature over the 2D parameter space.

The claims so far are uncontroversial. The disputed argument is that, while the curvature must be integrated over the full 2D space (k, θ) to find an integer Berry phase, for quasicrystals the calculation is independent of θ, allowing us to assign a 'Chern density' to each individual quasicrystal[18]. The statement sounds innocuous enough; certainly there exist functions of two variables which are in fact independent of one. The problem comes in when considering the Tenfold Way. A 2D family of quasicrystals has no $\mathcal{C}$, $\mathcal{T}$, or $\mathcal{S}$

[17] The experimental set-up involves observation of localized photonic edge states and does not involve electrons, so I am imagining the quasicrystals being realized as 1D tight-binding systems.

[18] The quantized Berry phase of a sub-band is an example of the first Chern class, or more loosely the Chern number.

symmetries, placing the family in class A alongside the quantum Hall effect[19]. Families of quasicrystals have two degrees of freedom, meaning the Tenfold Way predicts a set of $\mathbb{Z}$ possible edge states (Table 2). This agrees with the waveguide experiment. If an individual quasicrystal already 'knew about' the total integer Berry phase (*a.k.a.* Chern number) of the family, via its Chern density, this would mean a truly 1D system in class A was also demonstrating knowledge of the $\mathbb{Z}$ classification - but the Tenfold Way states that 1D systems in class A have trivial topologies.

If we take the Tenfold Way to be a statement of physics, the assignment of a Chern density to an individual quasicrystal would mean that 1D quasicrystals are *physically* 2D objects, as opposed to simply being mathematical projections of 2D crystals. This is extremely exciting - the same group has subsequently proposed an experiment to measure 4D quantum numbers in 2D quasicrystals in a bid to show the first evidence of higher-dimensional objects in the lab [62, 63].

An objection to the claim was stated clearly in a response by Madsen *et al.*, who showed that quasicrystals in fact have no special topological properties compared to ordinary crystals, and that the two cases can be adiabatically connected without any gap closures [64]. It is only *families* of 1D quasicrystals which can demonstrate 2D topological behaviour. Rather than reproduce the convincing arguments of reference [64] I will rephrase the debate as a question about CDWs.

4.2.2 Quasicrystals as Incommensurate CDWs

The quasicrystalline systems in the previous section are closely related to incommensurate CDWs. Kraus *et al.* consider a 1D lattice with a periodic on-site potential. ICDWs naturally generate their own periodic modulation with period $2k_F$, so provide a real physical example of such a system.

A clear way to see that ICDWs in combination with their parent lattices form quasiperiodic systems is via a projection method, shown in Figure 34. Say we wish to make an ICDW with period between one and two lattice lengths, such as $\sqrt{2}$. In this case we draw a second copy of the lattice but with twice the spacing. We then rotate the new lattice relative to the first until the sites, projected perpendicularly onto the original lattice, have

[19] Proven in Section 4.2.2.

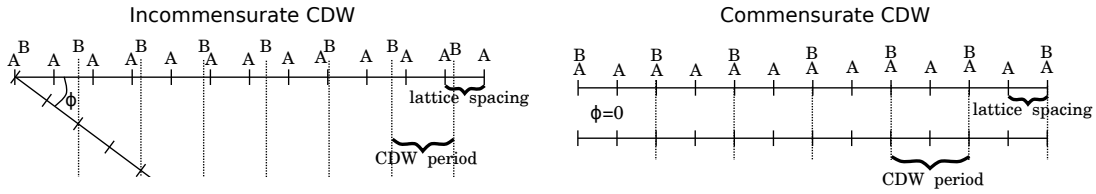


Figure 34: CDWs can be generated by a projection method. If the period of the CDW is between one and two (left) we take a lattice with period two then rotate through ϕ such that the projections onto the ion lattice gives the desired period. Provided $\cos(\phi)$ is irrational the generated CDW is incommensurate. The method demonstrates that the system of ICDW-plus-lattice is quasicrystalline, since the sequence of CDW peaks (B) and atomic sites (A) is never-repeating, but is 'long-range ordered'. Commensurate CDWs can trivially be made by the same method for example by setting $\phi = 0$ (right).

the correct spacing of $\sqrt{2}$. Wherever there is a site of this projected lattice we place a peak of the ICDW. Labelling these peaks B and the original lattice sites A the sequence is quasiperiodic, just as in Figure 33. The sequence of As and Bs is never-repeating, as guaranteed by the incommensurate periods, but it is 'long-range ordered' in the sense that the cell-type any distance along the line can be inferred by considering the second lattice being projected onto the first.

The working of Chapter 3 demonstrates that commensurate CDWs support quantized adiabatic particle transport. It is interesting to consider if and how they fit into the Tenfold Way. Although the classification was created for disordered systems, crystalline systems in 1D can have only two space groups: trivial or dihedral. Varying the offset of the CDW from the lattice, θ, so that the CDW peaks move off the lattice sites, any dihedral symmetry which did exist is lost, leaving only the trivial space group. The Tenfold Way applies.

In terms of topology the maths of commensurate 1D CDWs is identical to that of the quantum Hall effect, coming from Harper's equation, suggesting they fit in class A with no symmetries. The quickest way to verify this is via Table 1, which gives the symmetry class based on the elements of the matrix H_k. In the case of commensurate CDWs H_k is given in Equation 29 of Chapter 3. The matrix is dimension $q \times q$ for p/q filling, meaning that it cannot be represented as a 2×2 matrix except in the special case of half-filling (Chapter 2). This rules out the four BdG cases D, C, CI, DIII. The diagonal elements are non-zero, which additionally rules out classes AIII, BDI, CII. Of the remaining three classes, the fact that the off-diagonal elements are necessarily complex places us in class A, with no $\mathcal{C}$, $\mathcal{T}$,

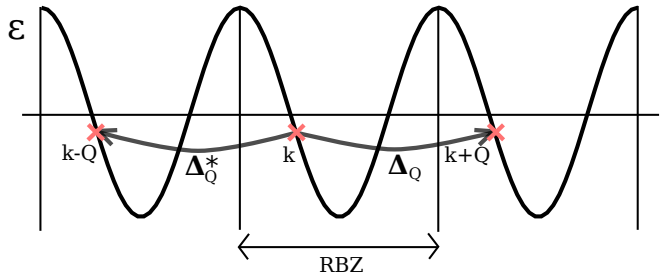


Figure 35: The order parameter Δ_Q corresponds to scattering between reduced Brillouin zones ('reduced Umklapp' processes). The complex phase of Δ_Q gives a current in reciprocal space, breaking time reversal symmetry. Destructive interference with Δ_Q^* means there is no corresponding current in real space.

or $\mathcal{S}$ symmetries, except at half-filling where the additional particle/hole symmetry places the CDW in class C. Note that with a two-dimensional parameter space (k, θ) both classes A and C are expected from Table 2 to show a set of $\mathbb{Z}$ possible edge states, in agreement with the findings of Chapter 3.

Physically the absence of particle/hole symmetry is clear enough in CDWs away from half-filling. Why time reversal symmetry is absent is less clear. From Equation 21 (slightly simplified) we see that the CDW order parameter in reciprocal space is

$$\left\langle \hat{\psi}^\dagger_{k+Q} \hat{\psi}_k \right\rangle = \Delta_Q$$

at filling fraction Q. It is the complex phase of Δ_Q which leads to the breaking of $\mathcal{T}$. The form of this order parameter suggests a current in k-space via scattering events $k \to k+Q$. The change in momentum Q corresponds to a scattering across the reduced Brillouin zone. These 'reduced Umklapp processes' are shown schematically in Figure 35 (*cf.* Figure 12). The current in k-space breaks time reversal symmetry, but for every scattering $k \to k+Q$ there is a reverse process $k \to k-Q$ guaranteed by the Hermiticity of H_k. This causes destructive interference and cancels the corresponding real-space current.

The reasoning so far demonstrates that commensurate CDWs fit firmly into the Tenfold Way. Incommensurate CDWs cannot be treated by the methods of Chapter 3, but it is possible nevertheless to deduce certain properties of their topological behaviour. Consider the gap which forms the lower left wing of the Hofstadter butterfly, Figure 23, or equivalently the lower left wing of the self-consistent CDW solution (equivalent as the cases are adiabatically connected). A reproduction of the wing in the Hofstadter case is given in

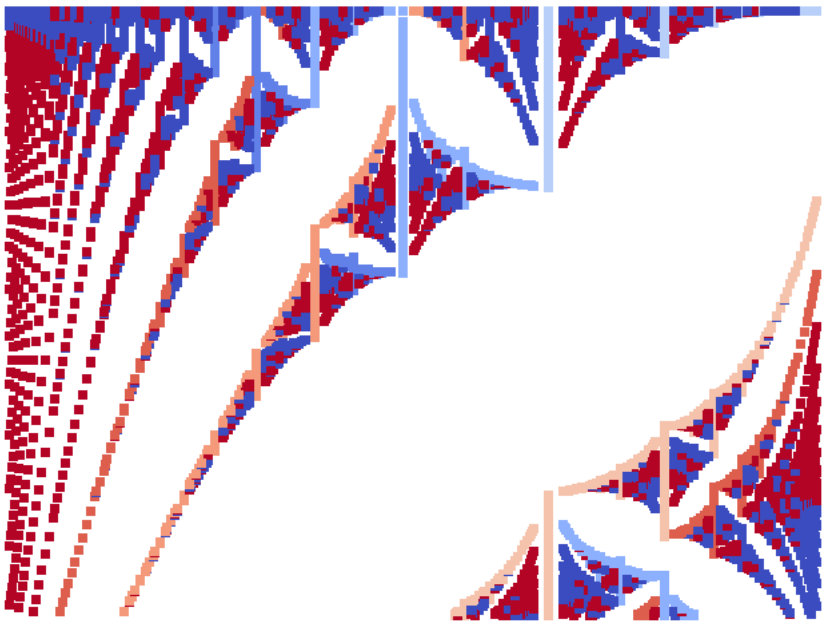


Figure 36: A zoomed look at the lower left wing of the butterfly from Figure 23 (allowed energies plotted against filling fraction). The colour of each point indicates the Hall conductivity of the gap above it (scale truncated at ± 4). The important point to note is that the gap formed by the wing itself is assigned a single conductivity, $+1$. Each gap can be similarly labelled, and several more can be seen in the image.

Figure 36. The conductivity of the gap is given by the sum of Berry phases (Chern numbers) of the sub-bands below the gap. This conductivity is independent of filling fraction, as evidenced by the fact that the underside of the wing is 'painted' a single colour (light pink, $C_1 = 1$). In fact this is necessary for the self-similarity of the fractal. Thus the conductivity of an incommensurate CDW with a filling which places it in the same gap will demonstrate the same quantization. The ICDW at filling $p/q + \eta$, with infinitesimal irrational η, has the same topological properties as the CDW at filling p/q, and the cases are again adiabatically connected by tuning the chemical potential ($\propto \eta$). The argument is made somewhat more mathematically precise in the next section.

4.2.3 The Gap Labelling Theorem

The arguments of the previous section give a physical intuition that the topological properties of incommensurate CDWs can be adiabatically connected to those of the commensurate case. This implies that, since the latter obey the Tenfold Way, so must the former, and since the former are 1D quasicrystalline systems it follows that quasicrystals obey the Tenfold Way. This is in agreement with Madsen *et al.* [64], who similarly give an adiabatic

Figure 37: A Cantor set can be formed by removing the central third of a line segment, then removing the central third of the remaining line segments, and so on. The set is what remains after an infinite number of iterations. The sub-band distribution of CDW systems at irrational fillings is homeomorphic to the Cantor set as demonstrated by Hofstadter and proven rigorously by Avila and Jitomirskaya [16, 65].

connection to commensurate systems. In particular, individual 1D quasicrystals cannot show a nontrivial topology; it is only families of locally isomorphic quasicrystals which can.

The reasoning of the previous section was based on the self-similarity of the Hofstadter spectrum, but it could be argued that the incommensurate cases are strictly absent from this construction. Hofstadter himself considers the limit of incommensurability by a recursive argument, and establishes, for example, that the sub-band distribution at irrational fillings is itself fractal[20], and is homeomorphic to the Cantor set (shown in Figure 37) [16]. Hofstadter did not factor in Berry phases to his calculations (he was working eight years before its development), so strictly his arguments do not concern the coloured-in butterfly which includes such effects [28].

There is a wide literature associated with the mathematical properties of the energy spectra of quasicrystals. A key result is the 'gap labelling theorem' attributed to J. Bellissard [66, 67]. The gap labelling theorem states that there is a consistent way to label each of the infinity gaps in the irrational filling spectrum with a unique integer, and that this labelling is robust to perturbation of the energy levels. The integer in the case of the Harper equation is precisely the sum of Berry phases of the sub-bands below the gap - the quantized Hall conductance [66, 67].

As an interesting aside, it follows from the Diophantine equation used to generate the Berry phases of the sub-bands (Equation 36) that there are only two possible Berry phases for

[20]The general proof that the bands of the almost-Matthieu operator (of which the Harper equation is a special case) form a Cantor set, and that all the band gaps except the central one are open, is known as the 'Ten Martini Problem' after M. Kac promised that set of beverages to anyone able to prove it. It took until 2009 until A. Avila and S. Jitomirskaya solved the problem [65]. Kac died in 1984 and failed to make appropriate provisions in his will. Perhaps as part-compensation, Prof. Avila received the 2014 Fields medal. Prof. Jitomirskaya informs me that she received an olive. This reveals the drinks to be dry Martinis, but it remains unclear whether they were intended to be gin- or vodka-based.

any of the sub-bands at a given filling [26]. For incommensurate fillings there are infinitely many sub-bands and infinitely many gaps. Each sub-band's Berry phase is taken from a set of two integers which depends on the filling fraction. Bearing in mind that the label assigned to each gap is the sum of Berry phases of the sub-bands below it, and that all gap labels are unique, it follows that the sequence of phases of the sub-bands must itself be quasiperiodic. That the sequence can be neither aperiodic nor periodic is fairly intuitive, both cases following from the fact that the lowest gap always has a finite Chern number, and the highest 'gap' (the energies above the highest sub-band) always has Chern number zero. This is too much structure for an aperiodic series, and no repeating sequence can fulfill the conditions[21].

Importantly, a perturbation to which the gap labels are invariant, where they continue to exist, is changes to the filling, including passing through rational values. When two sub-bands meet, their Berry phases must add in order to keep the label of the highest (still unclosed) gap constant. This makes rigorous the claim that all charge transport in the gap of the lower left butterfly wing, for example, is adiabatically connected to the simplest case of 1/3 filling. In terms of topological charge transport there is nothing special about ICDWs compared to commensurate CDWs, or, correspondingly, quasicrystals compared to crystals. It is therefore only families of quasicrystals, generated by varying θ, which exhibit topological non-triviality, but the families are 2D (existing in a space spanned by (k, θ)), and so quasicrystals behave like crystals in the Tenfold Way.

As a final point, note that the invariance of the gap labels to shifts in the filling does not apply at the singular point of half-filling, which is topologically distinct from the rest of the butterfly. This is because it is in class C, rather than A, of the Tenfold Way, or equivalently because half-filling has particle/hole symmetry. This too can be inferred from requiring the coloured butterfly to be self-similar.

4.3 Are Quasicrystals Generic in 1D? Free Energy Analysis

The systems considered so far in this chapter are 'quasicrystalline' if one considers the combination of atomic sites plus the incommensurate CDW sat on top of them. To recap,

[21] Further motivation that the sequence is quasiperiodic comes from considering the geometrical 'projection method' of solving the Diophantine equation in Figure 22 of Chapter 3.

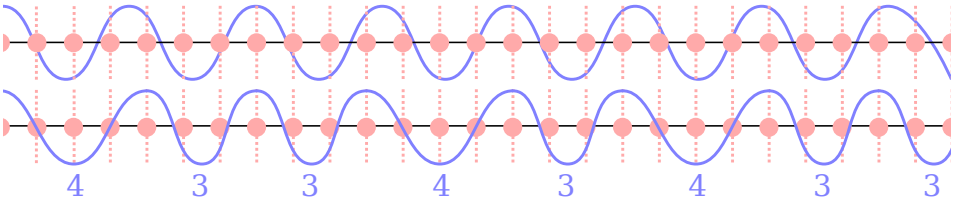


Figure 38: An incommensurate CDW (top row) might be expected to adjust infinitesimally such that each peak in the charge density shifts to the nearest positive ion. In this case the ICDW has period $\frac{7}{2} - \eta$, for irrational η, so the resulting sequence has cells of length 3 or 4. The lattice would similarly be expected to adjust into the same quasiperiodic sequence (not shown). The sequence is now a true quasicrystal.

if we move along the system and write an A every time we find a lattice site (ion), and a B every time we find a peak in the CDW, the sequence of As and Bs will be quasiperiodic. At the same time this seems to fall short of a true 'quasicrystal', which would ideally have the CDW peaks separated by one of two different distances *Long*, L, and *Short*, S, with the sequence of Ls and Ss being quasiperiodic. Better still would be to have the ions themselves in such an arrangement.

Until now I have not considered the reaction of the positive ions to the presence of the modified charge density, nor the back-reaction of the charge density in the presence of these modified ion locations. In fact I have implicitly assumed the Born-Oppenheimer approximation: that the slow and fast degrees of freedom, the movement of the ions and electrons, are separated[22].

Going beyond the Born-Oppenheimer approximation analytically is a difficult matter. Physically, though, the following scenario seems plausible: each positive ion will shift infinitesimally towards the nearest region of high charge density, and the regions of high charge density will similarly move towards the nearest ions. The situation is shown in Figure 38, where the case of an incommensurate CDW of period between three and four is demonstrated.

What is the resulting structure of the charge density field? The sequence of left/right shifts is quasiperiodic, since the sequence of As (ions) and Bs (peaks in the CDW) was quasiperiodic. The end result then is a quasiperiodic sequence of long and short unit cells, in this case period three and period four. The charge density now forms a truly quasicrystalline pattern. After the ion lattice adjusts to this charge distribution the entire

[22] I thank Nikitas Gidopoulos for making this clear.

system is precisely what is required of a true 1D quasicrystal.

The argument seems plausible, but to test whether the system really prefers the quasicrystalline state over incommensurate order it will be necessary to consider the free energies of the respective states. In fact, there is a third option which I will introduce in the next section.

4.3.1 Discommensurations in ICDWs

The standard story of incommensurate charge ordering, dating to the 1970s, is due to a series of papers by McMillan [68, 69, 70]. The incommensurate nature of the CDWs in question guarantees that the charge density can be maximum on at most one lattice site. This seems like an untenable situation, and one would imagine that the energy might be lowered by shifting to a commensurate period, in which an infinite number of CDW peaks can match with lattice points. McMillan's solution is more elegant still. He proposes that the ICDW adopts the closest commensurate period available to it over a large range of unit cells, before phase slipping, and repeating. The phase slip, termed a 'discommensuration', is such that the average period over the enlarged unit cell is the incommensurate period $(2k_F)^{-1}$. This way most of the negative charge sits at regions of high positive charge, but the energy gain due to the Peierls instability is also present. There is an energy cost associated with the discommensurations, and the period of the discommensuration lattice comes about through a balancing of these terms. Figure 39 shows the four possible cases: incommensurate CDW, commensurate CDW, discommensuration state, and the quasicrystal I proposed in the previous section.

McMillan made the argument quantitative by considering a phenomenological free energy expansion for the ICDW system, and was able to show that the discommensuration state is indeed of a lower free energy than either the commensurate or incommensurate states, interpolating between these two extremes. Note that all three cases McMillan considered are periodic, despite potentially having a very large unit cell. Periodicity was a natural assumption to make when the theory was developed in the late 1970s as quasicrystals were yet to be discovered. In the next section I will extend the free energy analysis to include the possibility of quasicrystallinity, where the periodicity is lost.

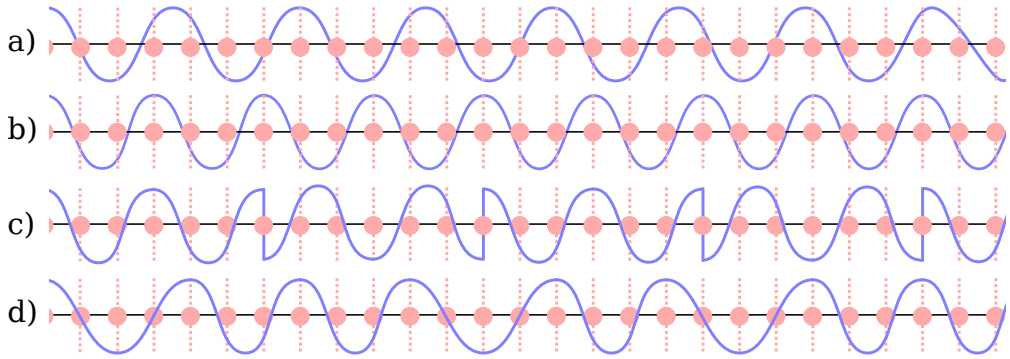


Figure 39: The four possible low energy states when the Fermi wavevector is incommensurate with the lattice spacing. (a) an incommensurate CDW with period $\frac{7}{2} - \eta$; (b) a commensurate CDW with period 3; (c) a discommensuration state (period 3 CDW with phase slips); (d) a quasicrystal.

4.3.2 Free Energy Analysis

For a quantitative investigation of the relative stabilities of the quasicrystal and the four periodic states outlined in the previous section it will be necessary to specify a free energy functional. The majority of the calculations of this section were carried out by my supervisor Jasper van Wezel, and for this reason I will provide less calculational detail than elsewhere in this thesis.

McMillan suggested a phenomenological free energy functional for investigating the CDW phase transition as well as the incommensurate-to-commensurate 'lock-in' transition [68, 69, 70]. The motivation is that many compounds which have $(2k_F)^{-1}$ incommensurate with their lattice spacing initially develop incommensurate CDWs upon cooling, but have their CDW wavevector shift towards a commensurate value as temperature decreases within the ordered state. At a lower temperature the (inverse of the) CDW wavevector locks into a rational multiple of the lattice spacing, forming a commensurate CDW. Examples include $TaSe_2$ [71, 68, 69] and TTF-TCNQ [72].

Positing an ansatz for the CDW order parameter of the form $\Delta \cos(\theta(x))$, with fixed magnitude Δ, McMillan's free energy functional reads

$$F[\theta] = \int \mathrm{d}x \left\{ a\Delta^2 \cos^2(\theta) - b\Delta^3 \cos^3(\theta) + c\Delta^4 \cos^4(\theta) + eQ^2\Delta^2 |\nabla\theta - Q|^2 \right\} \qquad (38)$$

where the terms a, b, and c form the standard Landau free energy expansion, and the e

term favours the incommensurate CDW with wavevector $Q = 2k_F$. McMillan's expression was created for $TaSe_2$ which has a discommensurate period lying close to period three; in this case the period-three lock-in is accounted for by including the lattice structure in the third-order coefficient:

$$b = b_0 + b_1 \cos(Kx) + \ldots$$

with reciprocal lattice vector K. In general, including such structure in the n^{th}-order coefficient allows for a period n lock-in. Intuitively the quasicrystal is likely to form when the incommensurate CDW wavevector sits close to half way between two commensurate periods, where discommensurations will be tightly packed into the lattice causing a significant energy cost. A generalization of Equation 38 is necessary which accounts for at least two commensurate lock-ins.

The $\mathcal{O}(\Delta^2)$ terms are more naturally written in k-space, as it is a divergence in the electronic susceptibility $\chi(k)$ at the point $k = Q$ which leads to the Peierls instability. The expression is given by

$$\chi_k \triangleq -\sum_{k'} \mathfrak{Re}\left(\frac{f(\xi_{k'}) - f(\xi_{k'+k})}{\xi_{k'} - \xi_{k'+k} + i\Omega}\right)$$

regularized by the infinitesimal real number Ω, and $f(\xi_k)$ the Fermi-Dirac distribution at electronic energy ξ_k. The expression diverges whenever k is such that $\xi_k = \xi_{k+k'}$. This is exactly the Fermi surface nesting condition, and happens generically in 1D whenever $k = 2k_F$. Note that χ_k is always positive.

A suitable generalization of McMillan's free energy is

$$\begin{aligned} F &= -\sum_k a_0 \hat{\chi}_k \tilde{\Delta}_k^2 + \int \mathrm{d}x c_0 \Delta^4 \cos^4(\theta(x)) \\ &\quad - \int \mathrm{d}x \sum_n b_n \Delta^n \cos(n\theta(x) - Kx) \end{aligned} \tag{39}$$

with

$$\begin{aligned}\hat{\chi}_k &\triangleq \frac{\chi_k}{\sum_k \chi_k} \\ \tilde{\Delta}_k &\triangleq \Delta \int \mathrm{d}x \exp(ikx) \cos(\theta(x)).\end{aligned}$$

The first term favours a CDW with the period dictated by $2k_F$, which is chosen here to be incommensurate with the lattice. The b_n term favours a lock-in to the lattice by a commensurate CDW with period n. The charge density field Δ is defined to be dimensionless so that all coefficients a_0, c_0, b_n have dimensions of energy.

Returning to the form of the ansatz for the CDW order parameter, $\Delta\cos(\theta)$, if $\theta(x) = Qx$ this is simply a sinusoidal CDW with period $2\pi/Q$, covering both commensurate and incommensurate CDWs by different choices of Q. On the other hand the discommensuration state, in the form of a period q commensurate CDW with periodic phase slips, can be written in the form

$$\theta_{DC}(x) = \delta x + \sum_{n=1}^{N} A_n \sin(qn\delta x)$$

where δ and A_n are variational parameters found by minimizing the free energy, and N is a chosen cut-off. Finally, the quasicrystal can be constructed piecewise out of single cells of two different commensurate CDW states. The cases are shown graphically in Figure 39.

Cooling from high temperature, Δ becomes nonzero at T_{CDW}. This happens smoothly so that, immediately below the transition, the quadratic term in the free energy dominates and an incommensurate CDW forms. As the temperature lowers further, Δ increases, allowing the higher-order terms to have an effect. These solely benefit commensurate CDWs, and the dominant contributions can be expected to come from those commensurate CDWs with periods either side of $(2k_F)^{-1}$. To take a concrete example, if $(2k_F)^{-1} = \sqrt{11.7} \approx 3.42$, only period three and four CDWs need be considered ($b_{n>4} = b_{n<3} = 0$).

Figure 40 shows the free energies of the different states for this filling as a function of the (dimensionless) cubic coefficient b_3/a_0 which multiplies the term favouring a period three commensurate CDW. In the figure, $b_3 > b_4 (> 0)$, so the period three CDW is always energetically favourable to the period four. As expected, the incommensurate CDW is

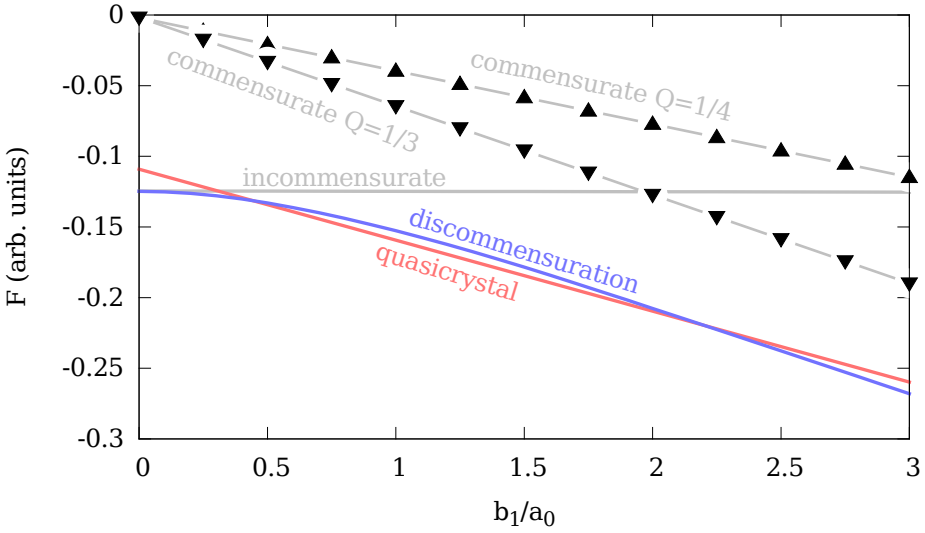


Figure 40: The free energies of different orderings for $(2k_F)^{-1} = \sqrt{11.7}$ as a function of the parameters b_3/a_0, a dimensionless form of the period 3 lock-in coefficient (Equation 39). Red: quasicrystal; blue: discommensuration; grey solid: incommensurate CDW period $1/\sqrt{11.7}$; grey downward triangles: commensurate CDW period 3; grey upward triangles: commensurate CDW period 4. The quasicrystal has the minimum energy configuration from $0.5 \lesssim b_3/a_0 \lesssim 2$.

favoured over both at low values of the lock-in coefficient, but for a sufficiently large value of b_3/a_0 the commensurate period three CDW stabilises. At even larger values the commensurate period four CDW also becomes more stable than the incommensurate state as the ratio b_4/b_3 is held constant. The discommensuration state, being able to adapt and interpolate between the three states, always has a lower free energy than any of them. At low b_3/a_0 it can simply take $A_n = 0 \forall n$ and $\delta = Q$, mimicking the incommensurate state. As b_3/a_0 increases, phase slips begin to nucleate to take advantage of the lock-in. It begins to mimic the commensurate period three CDW, but maintains an advantage over that state through its global agreement with $Q = 2k_F$. Note that the gradient of the discommensuration line tends asymptotically to that of the period three CDW line at high b_3/a_0.

The gradient of the quasicrystal line in Figure 40 is a weighted average (weighted by the proximity of $(2k_F)^{-1}$ to each commensurate value) of those of the two commensurate CDWs' lines, which is expected since it is locally pieced together from bits of each. With a global CDW wavevector of $Q = 2k_F$ the quasicrystal beats both commensurate states at low b_3/a_0 where lattice coupling is unimportant. Failing to match the $2k_F$ condition lo-

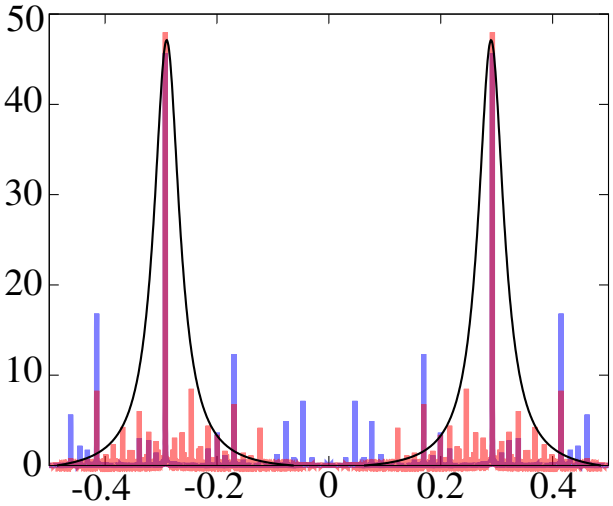


Figure 41: The diffraction intensity (arbitrary units) of the quasicrystal (red) and discommensuration state (blue) for $2k_F = 1/\sqrt{11.7} \approx 0.29$. In black is a Lorentzian fit to the susceptibility χ_k. Both diffraction patterns are peaked at $2k_F$, but the quasicrystal has more Bragg (-like) peaks close to $2k_F$ giving it a lower free energy over a range of parameters.

cally, however, the quasicrystal loses out to both the incommensurate state and its mimic, the discommensuration state, in this regime. As b_3/a_0 is increased both the commensurate parts of the quasicrystal begin to benefit locally, and it quickly stabilises relative to the incommensurate CDW. The competition is now between the quasicrystal and the discommensuration. At very large b_3/a_0 the discommensuration state takes on the character of the period three CDW, while the quasicrystal still has a large component of period four which is not as effective at taking advantage of the cubic lock-in. The discommensuration wins. At intermediate values of b_3/a_0, however, the quasicrystal is able to benefit relative to the discommensuration state by taking advantage of the finite spread of wavevectors covered by the susceptibility χ_k. This can be seen most easily from the diffraction patterns of the two types of order, shown in Figure 41. While both have a Bragg(-like) peak at $2k_F$, where χ_k peaks, the quasicrystal has more peaks clustered in the *vicinity* of $2k_F$. This set of secondary peaks stabilises the quasicrystal as the lowest-energy state over a range of b_3/a_0. A phase diagram is given in Figure 42.

The result shows that it is theoretically possible for quasicrystals to form naturally in quasi-1D systems. The requirements for this to happen[23] are:

1. a quasi-1D material with $(2k_F)^{-1}$ incommensurate with the lattice spacing and close

[23] Originally laid out by Jasper van Wezel.

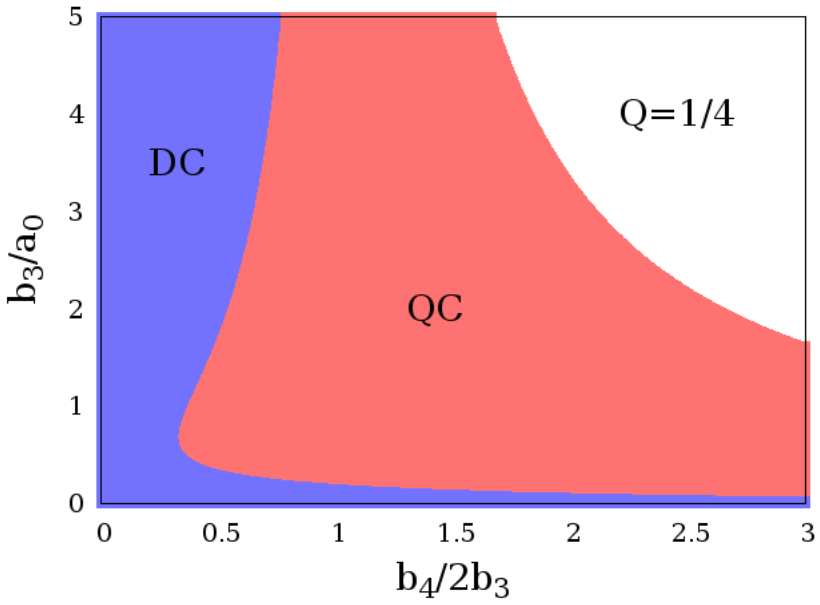


Figure 42: The phase diagram for the free energy of Equation 39 as a function of the dimensionless parameters b_3/a_0 and $b_4/2b_3$. Blue: discommensurations, red: quasicrystal, white: period four CDW (this would actually be a discommensuration state very close in character to a period four CDW).

to half way between two commensurate values

2. a sharply-peaked electronic susceptibility, nevertheless broadened over a range of k

3. a momentum-independent electron-phonon coupling (necessary to ensure that the CDW lattice coupling does not favour any particular commensurate value of the CDW wavevector).

Quasi-1D materials occur whenever there is strong bonding along chains relative to the coupling between them. In this case point (2) is guaranteed, and the incommensurate wavevector required in point (1) is the natural expectation given that the cardinality of the irrational numbers is greater than that of the rationals[24]. The question then arises as to why we are not overwhelmed with naturally-occuring 1D quasicrystals in quasi-1D materials. First, the charge density modulation Δ is small compared to the unaltered charge distribution, suggesting that (having not known to look for it) the quasiperiodicity may easily have been overlooked in existing data. Second, despite the fact that the inter-electron coupling (say electron-phonon coupling g or Coulomb repulsion h) is often assumed

[24]Strictly a multi-band model is required for a CDW period incommensurate with the lattice.

independent of the electrons' momenta, this is rarely an accurate simplification. As I show thoroughly in Part II of this thesis, except in cases where the 1D approximation is particularly good, or where electronic nesting is very strong, a momentum- and orbital-dependent electron-phonon coupling is generally a *necessary* condition for the formation of CDWs.

In all, then, it seems there may be a large class of 1D quasicrystals in existence which have previously gone un-noticed, and which a directed search of the known incommensurate CDW systems should reveal.

5 Summary of Part I

In this first part of the thesis I have considered the formation of charge density waves in quasi-1D systems of different fillings, from half-filling (Chapter 2), through arbitrary rational filling (Chapter 3) to arbitrary irrational filling (Chapter 4).

CDWs at half-filling are a shadowy reflection of BCS superconductivity; aside from being mathematically analogous, a renormalization group analysis suggests the two perfectly cancel if the instabilities are driven by a weak Coulomb repulsion. Aside from known ways around this issue, I provided a new explanation by way of a general field-theoretical argument, showing that Coulomb repulsion and electron-phonon coupling can be conveniently amalgamated into a single coupling to a combined charge/lattice distortion field. Considering the electron-phonon coupling to be turned on infinitesimally is sufficient to stabilize the CDW state and explain the observation of CDWs in real quasi-1D materials.

In the case of rational filling I demonstrated that CDWs are inherently topological in nature. Choosing a purely site-centred CDW with a particular magnitude, a plot of the allowed energies against filling fraction is identical to the famous fractal image of Hofstadter's butterfly. The reproduction of this image is no coincidence; rather, it is a consequence of the mathematical description of 1D CDWs matching that of the 2D integer quantum Hall effect. The second degree of freedom derives from the CDWs' ability, in the mean field model, to slide relative to the underlying lattice at no energy cost. A testable consequence of this topological phenomenon is a quantized adiabatic particle transport, *i.e.* the transfer of an integer number of electrons across the system upon sliding the CDW through a full period. The physical, self-consistent solution to this mean-field treatment is adiabatically connected to the Hofstadter case, in the sense that the two results can be interchanged by varying the gap magnitudes without the magnitudes passing through zero. While real systems cannot be expected to maintain an infinitely-fine fractal structure, the topology of the Hofstadter butterfly is unaffected by gap closures so long as the gap of interest remains open. The largest gap in the Hofstadter spectrum of Figure 14, and probably the easiest-implemented, occurs at 1/3 filling. Physical CDW systems would therefore be expected to exhibit quantum transport, closely related to what has become known as 'Thouless Pumping' [37].

Some comment is necessary on the possibility of seeing this topological charge transport in experiment. There are at least two major obstacles to dragging a CDW along the lattice. First, although not included directly in the mean-field treatment in this thesis, there will be a coupling between the CDW and the lattice which will attempt to keep the peaks of the CDW matched to the ionic locations. Second, pinning to defects is often an insurmountable barrier in quasi-1D systems, as can be seen on purely topological grounds: whereas in 2D and higher a point defect can be 'got around', in 1D a point-like defect blocks off any passage along the line [73].

The first point can be overcome by use of incommensurate CDWs, which lack a predilection for aligning with a certain lattice site. Various experiments have established the possibility of generating charge transport by sliding incommensurate CDWs with an applied voltage, although these did not distinguish the contributions from the topological charge transport, considered here, and a non-adiabatic pushing of charge through the system, which would involve effects of band mixing [74, 75, 76, 2]. Rather than attempt to fight against the two problems it may be possible to work with them. Recent work has established the possibility of realizing CDWs, the Hofstadter spectrum, and 1D quasicrystals, in cold atom condensates in optical lattices [77, 78, 79]. Such an experimental set-up allows for much more control of the system than would be obtained by using real quasi-1D materials. Imagine purposefully introducing a defect into the optical lattice, with an energy scale large enough that it can be assumed to be the dominant pinning centre in the system, but small enough that it does not cause band mixing between the filled and empty sub-bands. By construction the CDW will peak on this defect, so by moving the defect it should be possible to move the CDW.

When considering irrational fillings I showed the equivalence of the resulting incommensurate CDWs to 1D quasicrystals, and used the relation to provide some insight into the question of whether quasicrystals fit into the Tenfold Way classification of free fermion systems. I found that the irrational case is again adiabatically connected to the rational case, made rigorous by appeal to Bellissard's Gap Labelling Theorem [80, 66], and so the topological properties of incommensurate CDWs are the same as those of commensurate CDWs. This suggests that quasicrystals fit into the Tenfold Way in the same manner as normal crystals.

Considering incommensurate CDWs in combination with their underlying lattices, the resulting system is quasiperiodic, although the claim seems something of a cheat. In fact any system containing two incommensurate periods could be argued to exhibit the same quasiperiodicity. A simple example from quantum mechanics could be a 2D quantum harmonic oscillator with potential chosen such that the (dimensionless) eigenenergies take the form

$$E_{n_x,n_y} = n_x + \phi n_y + \frac{1}{2}(1+\phi)$$

with n_x, n_y integers and where ϕ is some irrational number such as the golden mean. A simpler example still, of course, could come from looking at a row of gardens with picket fences, viewed from a distance such that the spacing of one row of pickets appears to be an irrational multiple of another. An example receiving some mainstream coverage recently was so-called 'Golden Stars', whose luminosity is seen to vary with two incommensurate periods related by the golden mean, argued to provide the first example of a natural non-chaotic strange attractor [81].

In the case of ICDWs, however, I demonstrated through a free energy analysis that this system of two incommensurate periods can, under the correct circumstances, form a true quasicrystal, through adjustments of the charge density profile, and corresponding adjustments of the ion locations. There are precedents in the development of (what can now be seen as) quasiperiodic ground states in both the quantum Hall effect [82] and the Hubbard model [83]. It would be interesting to consider what would happen when taking into account a similar coupling between the incommensurate periods in other systems such as the Golden stars, or even the incommensurate 2D harmonic oscillator suggested above.

The growth mechanisms of quasicrystals are an interesting topic in their own right, since quasiperiodicity is a global topological property of a system. Even in 1D where the quasicrystal simply takes the form of a sequence of long and short cells, when adding an additional cell on the end it is frequently necessary to know about all the other cells to know which of the two types to place. In 2D it appears even harder. Proposed growth mechanisms in 2D and higher include: recognising vertex matching rules which imply 'forced tiles', random tiling models, and relaxation processes from non-quasiperiodic sys-

tems [41]. A neat and realistic growth mechanism is based on nucleation from a defect, much like crystal growth: given a certain type of defective start cluster (a 'decapod' in a Penrose tiling, for example [41]), there will always be at least one forced tile on the surface. A tile is forced if there is only one possible option as to what can be placed, based only on the immediate neighbours (rather like a confidently-flagged square in minesweeper). By always filling in the forced tile first the entire quasicrystal can be filled in very efficiently. The mechanism was shown by Janot to be similar to, but more efficient than, a screw dislocation nucleating a crystal [41].

All these growth mechanisms are local, circumventing the perceived difficulty of needing to know the global structure. In the development of the quasicrystal out of incommensurate charge order, although the adjustments to charge density happen locally, the order develops in a co-ordinated manner simultaneously along the length of the system, providing the first example of a realistic non-local quasicrystal growth mechanism. It is relatively easy to think of generalizations of the method to higher dimensions, but thinking of physically plausible scenarios in which they might occur is much more difficult.

Despite the plethora of artificial quasicrystals developed in the lab since Schechtman's original Al-Mn alloy [43], to date only two naturally occurring quasicrystals have been found, both in the same Siberian meteorite [84, 85]. The new method of quasicrystal growth from incommensurate charge order I have outlined in this thesis promises to greatly increase the number of naturally-occuring quasicrystals we know about, albeit with the new examples all taken from 1D.

In summary, despite being perhaps the simplest system in all of condensed matter physics, a model of spinless fermions hopping on a 1D chain in a self-consistent mean-field treatment demonstrates a truly remarkable level of complexity. The resulting charge-ordered states provide a naturally occurring analogue to a number of complicated physical phenomena, from superconductivity, through topological charge transport, to quasicrystals, and, in so doing, provide a different perspective from which to think about these systems. In the second half of this thesis I will consider the development of charge order in dimensions higher than one, where it turns out the ordering mechanism is in general quite different.

Part II

The Charge Ordering Mechanism in Dimensions Higher Than One: the Necessity of Including Strong Coupling Effects Evidenced by Niobium Diselenide

6 Introduction

6.1 Charge Density Waves in Higher Dimensions

In Part I of this thesis I considered quasi-1D systems, in which the system can always be expected to develop charge order through the Peierls mechanism. Nesting of the Fermi surface (matching a significant amount of the states at E_F onto themselves under translation through a fixed wavevector $\mathbf{Q}$) is generic in 1D, since the 'surface' is 0D. Translating one point onto another with a fixed 1D vector is always possible. In dimensions higher than one the argument does not hold: in quasi-2D systems the Fermi surface is one-dimensional, and shifting the entire surface through a single fixed wavevector is still expected to cause it to map onto itself only at isolated 0D points. The situation is shown in Figure 43. In three dimensions nesting is less likely still.

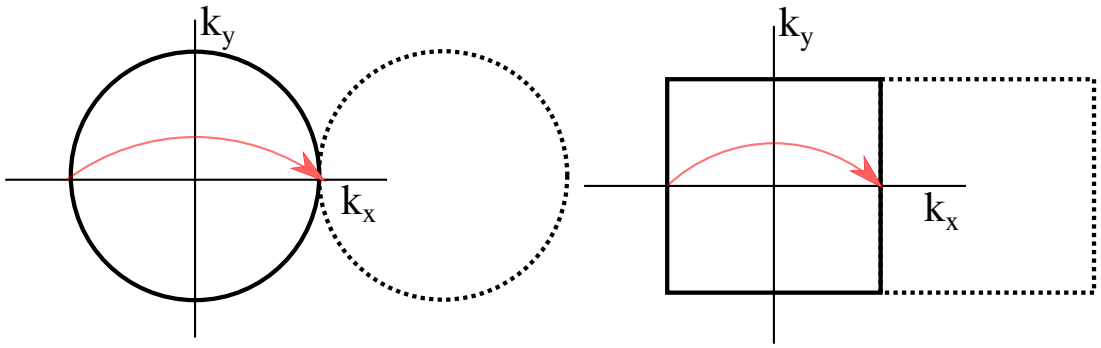


Figure 43: In quasi-2D systems Fermi surface nesting is not generic. A circular Fermi surface, left, meets itself only at an isolated point under coupling all states $\mathbf{k}$ to $\mathbf{k}+\mathbf{Q}$ for fixed $\mathbf{Q}$. There is unlikely to be an energy gain from such a coupling. In some cases, like the square on the right, nesting can occur, but it requires a matching of both the line of the Fermi surface and its gradient.

At the same time, charge density waves *are* seen in dimensions higher than one. An early 2D CDW system to be found was niobium diselenide, considered at length in the remainder of this thesis [86]. Additional 2D examples include various related transition metal dichalcogenides such as TaS_2, $TaSe_2$, NbS_2 and $TiSe_2$ [87, 88, 89, 90], and, of much current interest, the layered high-T_C superconductors (cuprates and pnictides) such as Bi2122 [91, 92, 93], $Ba_{1-x}Na_xTi_2Sb_2O$ [94], and YBCO [95, 96]. Perhaps the most famous three-dimensional density wave system is Chromium, which supports a *spin* density wave through Fermi surface nesting as indicated in Figure 44 (for a review see [97]).

In dimensions higher than one the CDW not only breaks translational symmetry but

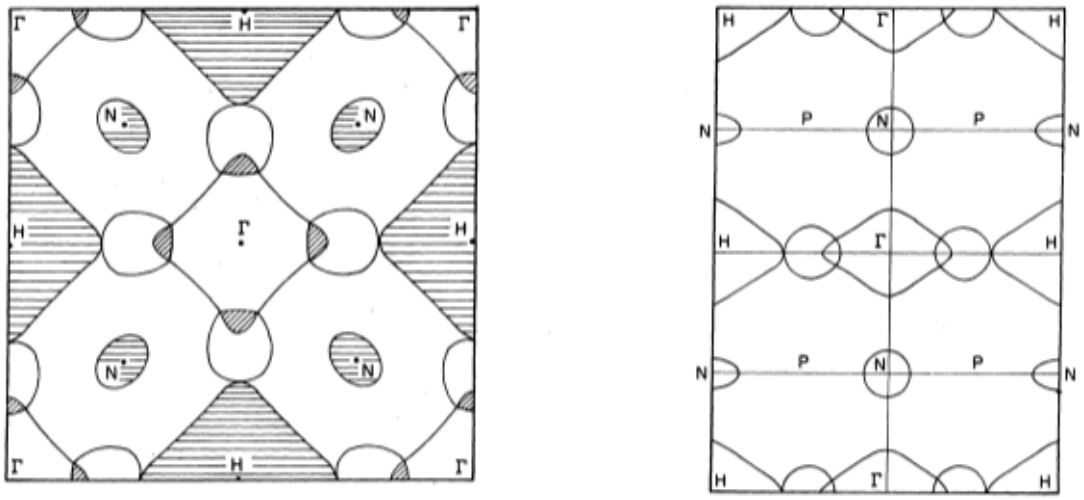


Figure 44: Cross sections of the Chromium Fermi surface (left (001), right (011)). Chromium demonstrates a density wave instability driven by Fermi surface nesting, in this case between the flat faces of the pockets around Γ and those around H. Figure reproduced from Laurent *et al.* [98].

also rotational symmetry of the underlying lattice, selecting out a preferential direction in space. The name charge density *wave* makes a little more sense in this case - charge density always looks wavelike in 1D. Aside from forming a CDW with one fixed wavevector ('1Q' CDW) it may be possible to form multiple coexisting CDWs. In the layered cuprate superconductors, which have a square lattice within the layers, it may be possible to form a 2Q CDW, with CDWs along both in-plane lattice directions [5]. In hexagonal layered materials it is possible to form 3Q CDWs - that is, three superposed 1Q CDWs at angles of $2\pi/3$.

Given that nesting is not generic in two- and three-dimensional systems, but that CDW order is nevertheless observed experimentally in a number of cases, it is natural to ask whether the CDW order derives from nesting. In the coming chapters I will consider the charge ordering mechanism in the specific quasi-2D material niobium diselenide ($NbSe_2$).

6.2 Niobium Diselenide

The $2H$-$NbSe_2$ polytype of Niobium Diselenide has a simple hexagonal crystal structure, space group $P6_3/mmc$ (D^4_{6h}) [86], with two niobium atoms per primitive unit cell [71, 87]. The material is layered, with a small ratio of interlayer to intralayer coupling, meaning quasi-2D models can be expected to capture most if not all of the important physics [87, 99].

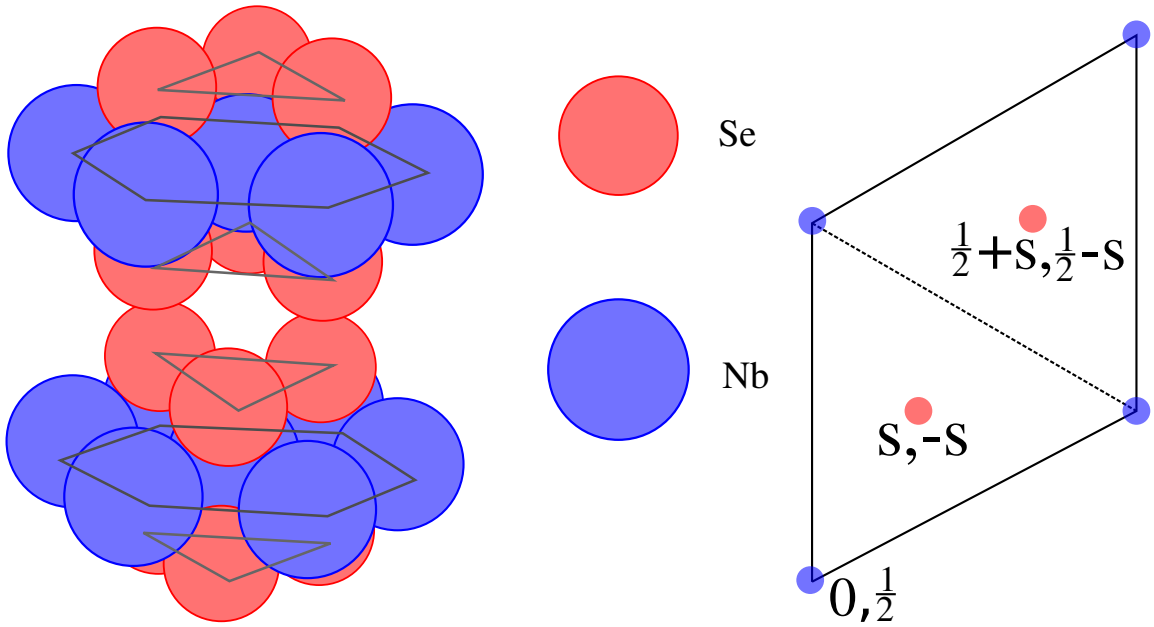


Figure 45: Left: the crystal structure of $2H$-$NbSe_2$. Niobium atoms are shown in blue and selenium atoms in red. Two formula units fill one primitive unit cell, since the relative positions of the seleniums and niobiums differ between the two pictured 'sandwiches'. Right: the 2D unit cell used in this thesis, with $S \approx 0.13c$.

The unit cell, shown in Figure 45, has two 'sandwiches' where each has a layer of selenium atoms as the bread and a layer of niobium atoms as the filling. In the lower sandwich three of the six interstices in the 2D hexagonal niobium structure have selenium atoms above and below them, and in the upper sandwich it is the other three.

$NbSe_2$ undergoes a phase transition to a 3Q incommensurate CDW state at 33.5 K. From neutron diffraction the CDW wavevectors are seen to sit along the $\overrightarrow{\Gamma M}$ directions at $\mathbf{Q}_{CDW} = (1-\delta)\,\frac{2}{3}\overrightarrow{\Gamma M}$, with $\delta \approx 0.014$ [71, 100]. There is a second phase transition, to a superconducting state, at 7.2 K [101]. In a recent scanning tunneling spectroscopy (STS) study [102] it was found that $NbSe_2$ can also support a 1Q CDW along the same directions, with wavevector selected from $\mathbf{Q}_{CDW} = (1-\delta)\,\frac{2}{3}\overrightarrow{\Gamma M}$ this time with $\delta \approx 0.143$ (still incommensurate, but close to $\mathbf{Q}_{CDW} \approx \frac{2}{7}\overrightarrow{\Gamma M}$). An STS image featuring domains of both symmetries is given in Figure 46. In X-ray diffraction studies of the phonon dispersion [103, 104] it is found that a Kohn anomaly develops in the longitudinal acoustic phonons over a broad range of momenta around the aforementioned locations in the Brillouin zone, shown in the same figure.

The breadth of the phonon softening in k-space for $NbSe_2$ is one of a number of anomalies

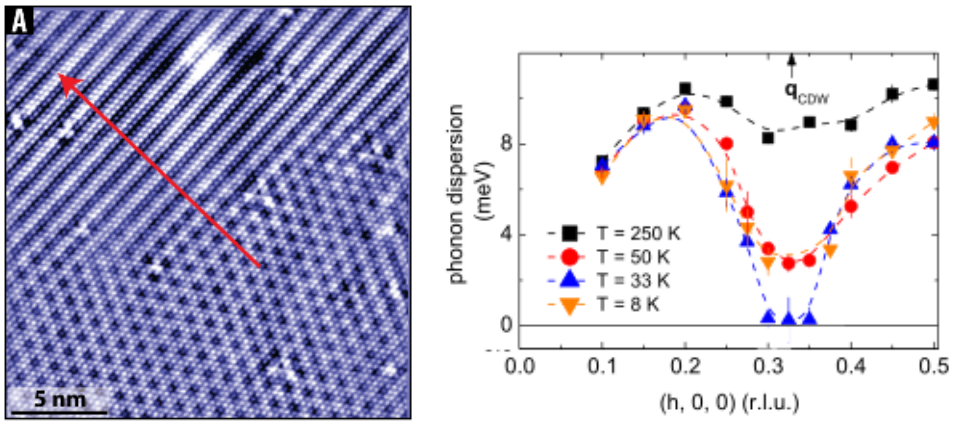


Figure 46: Experimental results for $NbSe_2$. Left: STM image of the surface of the material cleaved along the (001) direction. The top half of the image shows a 1Q CDW domain, whereas the lower half shows a 3Q domain (from [102]). Right: the dispersion of the longitudinal acoustic phonon which develops the Kohn anomaly. The mode softens to zero around $\mathbf{Q}_{CDW}$ at 33.5 K as expected. The plot, obtained via X-ray diffraction, is from [103].

associated with the CDW state in the material. In Angle Resolved Photo-Emission Spectroscopy (ARPES) the gap is seen to open only at select points on the Fermi surface, on the inner pockets around K, meaning the material remains metallic and conducting below the CDW transition [105, 106, 95, 101]. Why only one band develops a gap is currently a mystery. There is large disagreement regarding the gap size, with estimates ranging from 0 meV [105] to 35 meV [107]. There is also disagreement regarding a possible offset of the gap from E_F [108], and the extent of the asymmetry in the particle and hole states near the gap [102].

Finally, it has been postulated [105, 109] that $NbSe_2$ demonstrates a pseudogap in analogy[25] to that seen in the layered high-T_C superconductors [93]. The pseudogap is characterized by a reduction in the *observed* density of states at E_F at discrete regions of the Brillouin zone. The reduction could be due either to an actual reduction in the DOS, or a mechanism which stops the observation of these states in available experiments [110]. The pseudogap occurs in all hole-doped cuprate superconductors; a particular subset of these also form what appears to be a fluctuating CDW under the same conditions [93]. In $NbSe_2$ a reduced DOS is seen over an energy range of ≈ 35 meV for a range of temperatures both above and below T_{CDW} [107, 105]. In the temperatures above T_{CDW} large fluctuations in the phonon

[25] I use the word 'analogy' as the mechanism in $NbSe_2$ may be unrelated to that in the cuprates; I later claim to explain the former but do not address the latter.

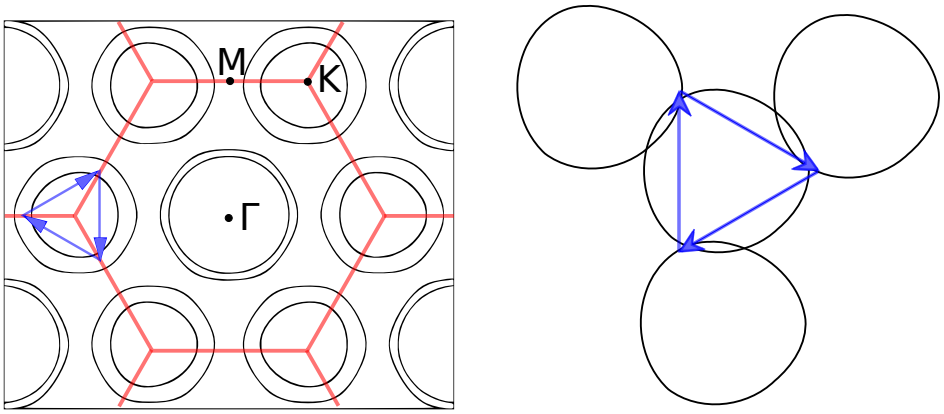


Figure 47: The Fermi surface of $NbSe_2$. On the left is the Brillouin zone used in this thesis (marked in red), with high symmetry positions indicated. In blue are the three experimentally observed CDW wavevectors. They are placed at the locations considered (and largely dismissed) as candidates for nesting in previous studies [106, 86, 95]. A detail of the mapped states is shown on the right. The CDW gap is seen in ARPES to open at three points in the inner pocket around K [105, 106], at approximately the locations the blue arrows intersect the pocket.

field are observed [109].

A number of bandstructure fits exist for $NbSe_2$ of both theoretical and numerical origin [86, 87] and as phenomenological fits to experimental data [106]. The latter paper considers a tight binding fit to a 2-independent band model constrained with ARPES data, and provides a particularly useful expression in this thesis; the Fermi surface is shown in Figure 47.

It is notable from Figure 47 that there is no clear nesting exhibited by the $NbSe_2$ Fermi surface. The right hand image depicts the inner K pockets, in which the CDW gap opens, translated through the observed CDW wavevectors. The unconvincing nesting has lead to a number of alternative proposals for the driving mechanism of the CDW transition, including nested saddle-points in the electronic dispersion [111], local field effects [112], or a combination of weak nesting with momentum-dependent electron-phonon coupling [99, 103]. The lack of agreement regarding the ordering mechanism provides a convenient opportunity to examine the rôle of nesting in CDW transitions more generally in systems of dimension higher than one.

In subsequent chapters I will set up a field theoretical description of the CDW transition in $NbSe_2$. For this reason I will briefly comment on the relevant couplings between electrons[26]. $NbSe_2$ is a metal above T_{CDW}, and in fact maintains its metallic character below T_{CDW}

[26] *cf.* Section 2.1.

as large parts of the Fermi surface remain ungapped [105]. For this reason there is a high density of states at E_F, leading to a large screening of the Coulomb repulsion [42]. I will therefore take the Coulomb interaction, denoted h in Section 2.1, to be zero in the following working.

In the closely related system of $TiSe_2$ both h and the electron-phonon coupling g are of importance as the system is believed to be a semiconductor (with an indirect band gap) [113, 114, 115] or a semimetal [116, 117]. The system orders to a chiral CDW state [90] at around 200 K [113]. In that case, which of g and h drive the phase transition is debated, and two recent papers have suggested that it is in fact a combination of electron-phonon coupling and exciton condensation which leads to ordering [89, 118]. The method of combining the terms g and h provided in Section 2.1 gives a convenient rephrasing of the results of these papers in terms of a well-defined emergent quasiparticle: a hybrid phonon-exciton.

6.3 Honorable Mention of Relevant Experimental Techniques

When comparing my calculations to experimental results in later chapters I will employ various abbreviations. I will very briefly outline the experimental procedures to which these abbreviations refer.

In Angle Resolved Photoemission Spectroscopy, ARPES, the photoelectric effect is utilized as a direct probe of the electronic bandstructure of a material in reciprocal space. Incoming photons cause the ejection of valence electrons from the material, and by knowing the ingoing photons' energies and momenta, and measuring those of the ejected electrons, the energies and crystal momenta of the allowed states in the material can be deduced [119]. In particular ARPES can provide an accurate image of the Fermi surface of the material, or the band energies away from the Fermi level. A drawback to the technique is that the states must be occupied in order for them to be detected, meaning that, for example, states significantly above the Fermi level are ruled out [120].

Scanning Tunneling Microscopy (STM) and Scanning Tunneling Spectroscopy (STS) are two closely related experimental techniques. In the former, a sharp metallic tip is held close to a sample, with a potential difference maintained between the sample and tip. The two are close enough for significant overlap of the electronic wavefunctions, allowing

electrons to tunnel through the gap. The tip is moved across the sample, and in 'constant current topography mode' the height of the tip is adjusted so as to maintain a constant current [121]. The tip height as a function of position on the sample can give information about atomic configurations, or, more relevantly to this thesis, the local electron density in the sample. The image of 1Q and 3Q domains in Figure 46 was made using STM [102]. Limitations of the STM apparatus include poor temporal resolution of around 1ms (significant for CDW studies as charge order fluctuations are faster than this [101, 108]) and the flatness of sample required, although the latter is not a significant hindrance to the layered dichalcogenides with their easy cleavage planes [108].

STS is an alternative mode of operation of the STM in which the bias voltage is varied; $\mathrm{d}I/\mathrm{d}V$ is measured, which is proportional to the local density of states [121]. Important later on is the fact that while the proportionality constant is unknown, the zero of the density of states is fixed to zero potential difference between sample and tip. A major advantage to STS is that the states do not have to be occupied to be probed, meaning information about the states above the Fermi level can be obtained. A disadvantage is that contributions to the density of states may not come purely from electrons, but could also come for example from phonons coupled to the electrons, giving an artificially large reading. Fortunately in the results relevant to the later chapters of this thesis the contributions are clear.

Inelastic X-ray Scattering (IXS) consists of X-ray scattering processes in which the energy of the photons is not conserved. As opposed to elastic X-ray scattering, which samples the crystal structure, inelastic scattering acts as a probe of excitations in the system. Relevant here is the procedure's use in revealing phonon dispersions. The procedure has been carried out in a range of CDW studies to date [122, 103, 104].

6.4 Format of Part II

The Peierls instability, ubiquitous in quasi-1D systems and investigated throughout the first half of this thesis, is a 'weak coupling' mechanism, meaning that the CDW order it leads to is guaranteed by the divergent electronic susceptibility, and it is merely a matter of finding out *which* coupling (Coulomb repulsion, electron-phonon coupling, *etc.*) drives the transition.

In this half of the thesis I will demonstrate that such weak-coupling models of the CDW transition in $NbSe_2$ are doomed to fail. Instead I present a 'strong-coupling' model: while still a perturbative quantum field theory, it is necessary to include information about the specific form of the electron-phonon coupling, in particular including information about the ingoing and outgoing electron momenta, and the orbital character of the electronic bands scattered between.

In Chapter 7 I carry out a Slater-Koster tight binding fit in order to establish both the bandstructure and orbital make-up of the bands in $NbSe_2$, and to establish a physically motivated form for the electron-phonon coupling. In Chapter 8 I use these results to develop a field theoretical description of the CDW order in $NbSe_2$ via the free energy. I rule out weak-coupling mechanisms in Section 8.2. Treating the free energy in the Random Phase Approximation (RPA), and including uniaxial strain in the model, I develop a temperature/strain phase diagram in Section 8.5. In Chapter 9 I extend the model to include fluctuations of the phonon field in the so-called Mode Mode Coupling Approximation (MMA), and find that this extension is necessary to obtain the correct temperature dependence. I discuss the relevance of the RPA and MMA transitions to the existence of the $NbSe_2$ pseudogap and extend the phase diagram to include this regime. In Chapter 10 I employ the Nambu Gor'kov technique to investigate properties of the CDW gap within the ordered state, and demonstrate that the strong-coupling model reproduces a range of experimental results including STM/STS and ARPES measurements. In Chapter 11 I conclude by discussing the relevance of the $NbSe_2$ results to the bigger picture of charge ordering in dimensions higher than one. I argue that strong-coupling models are in fact the natural starting consideration whenever CDW order is seen.

7 The Electronic Structure of $NbSe_2$

In this chapter I establish the bandstructure of $NbSe_2$, including the bands' orbital composition, via a Slater-Koster tight-binding calculation. The orbital make-up turns out to be the key to resolving the mystery of why the CDW gap in $NbSe_2$ opens in one band only. In Section 7.2 I use the electronic properties to deduce a form for the electron-phonon coupling, which is a vital part of the model developed in later chapters.

7.1 Slater-Koster Tight-Binding Fit

7.1.1 22 Band Model

The primitive unit cell of $NbSe_2$, shown in Figure 45, contains two niobium atoms and four selenium atoms. The electronic structure of these elements prior to bonding is $(...)\,5s^1 4d^4$ and $(...)\,4s^2 4p^4$ respectively [123]. This means the niobium atoms contribute five d-orbitals, traditionally labelled d_{xy}, d_{yz}, d_{xz}, $d_{x^2-y^2}$, $d_{3z^2-r^2}$, and the seleniums three p-orbitals, labelled p_x, p_y, p_z. This gives 22 orbitals in total, which will mix to form 22 bands. The probability densities of the orbitals are plotted in real space in Figure 48.

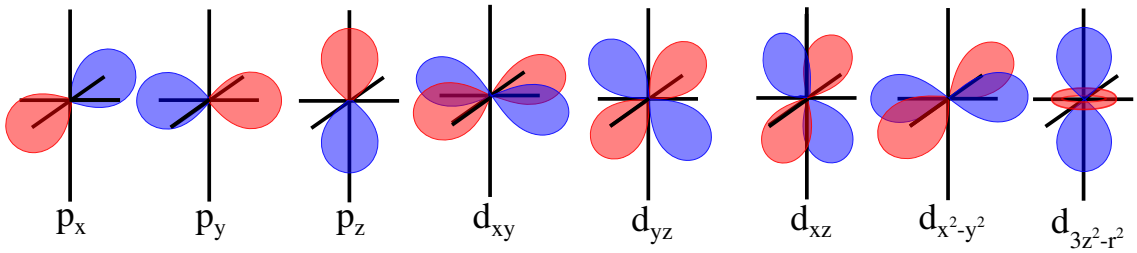


Figure 48: The probability density of the p and d orbitals (constant probability surfaces). The diagrams are schematic since, for instance, a different basis could be picked in which the shapes are different. The two colours indicate the signs of the orbital wavefunctions.

The tight binding method consists of solving the Schrödinger equation

$$H|n\rangle = E_n S|n\rangle \tag{40}$$

where the overlaps S are introduced because of the non-orthogonality of the orbitals: the set of orbitals on the same atom form an orthonormal basis, but orbitals on neighbouring atoms obey no such constraint. The starting point is the pair of 22×22 matrices

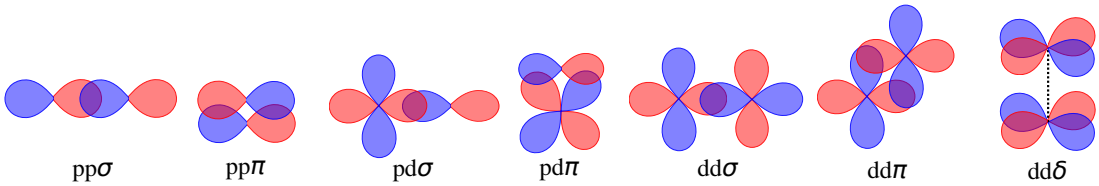


Figure 49: The orbital overlaps of the Slater-Koster fit, taken as free parameters.

$$H_{nm} = \langle \phi_m | H | \phi_n \rangle$$
$$S_{nm} = \langle \phi_m | \phi_n \rangle$$

with H_{nm} the hopping amplitudes, and S_{nm} the overlaps, between orbitals $|\phi_n\rangle$. The method of Slater and Koster [124] amounts to treating orbital overlaps as free parameters in a numerical minimization of the difference in energies between a model tight-binding bandstructure and experimental band data. For example, the overlap between a d-orbital and a p-orbital could be either $pd\sigma$ or $pd\pi$ depending on the orientation of the bonds (see Figure 49); $pd\sigma$ and $pd\pi$ are taken as free parameters, since the quantum mechanical calculation required to find their true values would be too complex and too dependent on the specific crystal-field environment to make it feasible. The full set of seven possible overlaps is shown in Figure 49.

Slater and Koster tabulated the overlap functions between each pair of orbitals in terms of the direction cosines between them. For example if a p_x orbital and a d_{xy} are separated by a vector $\mathbf{r}$ the direction cosines are

$$l = \hat{\mathbf{r}} \cdot \hat{\mathbf{x}}$$
$$m = \hat{\mathbf{r}} \cdot \hat{\mathbf{y}}$$
$$n = \hat{\mathbf{r}} \cdot \hat{\mathbf{z}}$$

and the corresponding element in the Slater-Koster table [124] is

$$E_{x,xy} = \sqrt{3} l^2 m \, (pd\sigma) + m \left(1 - 2l^2\right) (pd\pi) \, .$$

This expression forms element $n = p_x, m = d_{xy}$ of the matrices H_{nm} and S_{nm}, after Fourier transform (in this case a simple multiplication by $\exp(i\mathbf{k}\cdot\mathbf{R})$ with $\mathbf{R}$ the vector connecting the unit cells of the two orbitals) [124]. The $pp\sigma$ (*etc.*) seen by H_{nm} may be different to that seen by S_{nm}, giving 14 free parameters if all first-nearest neighbours are considered. This approximation will be employed here. The overlaps within a sandwich in the unit cell may be different to those between sandwiches, giving 28 free parameters. Additionally the diagonal entries of the Hamiltonian contain the chemical potential associated with each orbital, giving another 8 free parameters. However, considering the form of the orbitals in real space, and the crystal structure of $NbSe_2$, there are certain symmetry constraints on the chemical potentials[27]:

$$\begin{aligned} \mu_x &= \mu_y \\ \mu_{xy} &= \mu_{x^2-y^2} \\ \mu_{xz} &= \mu_{yz}. \end{aligned}$$

This gives a total of 33 degrees of freedom considering only the first-nearest neighbours.

For a given set of parameters the eigenvalues and eigenvectors of Equation 40 can be found. This gives the band energies, and orbital contributions to each band, respectively. The problem is known as the *generalized eigenvalue problem*, and is readily solved numerically[28]. By comparing the calculated E_n to experimental values the accuracy of the free parameters can be determined.

To constrain the bandstructure fit I used two sets of data. The first set was ARPES data from Rahn *et al.* [106] which gives accurate information about the two bands crossing the Fermi level. The data come with a phenomenological tight-binding fit, to fifth-nearest neighbour in-plane, giving an analytic expression for the two bands across the ΓMK plane of the Brillouin zone. As ARPES is generally only capable of probing states near E_F, other data were required for the remaining bands. For these I used the results of a previous LDA

[27] The chemical potential of orbitals with angular momentum $l \leq 2$ varies at most as $\cos^2(\theta)$. If the lattice has any rotational symmetry greater than twofold the only compatible symmetry of the potential is full circular symmetry. I thank John Hannay for pointing this out.

[28] Specifically I employed the routine ZGGEV in the LAPACK library for Fortran 90.

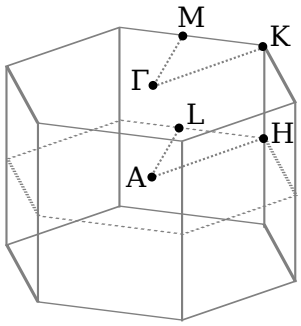


Figure 50: The Brillouin zone with high symmetry points indicated. The image is stretched disproportionately along ΓA for clarity - in fact the crystal's c/a ratio is 18.1/6.5 [86, 87].

calculation by Rossnagel *et al.* [95] which are likely less accurate than the direct ARPES measurement but have data available for all 22 bands. The band energies give 22 constraints per point in the Brillouin zone.

Initially I used data from high symmetry points in the Brillouin zone (Γ, M, K, A, L, H; see Figure 50), but later included additional points along the line $\Gamma MK\Gamma$, giving a total of 18 sampling locations and therefore $18 \times 22 = 396$ constraints. Note that ARPES data are only available in the ΓMK plane; I assumed the behaviour to be identical in the ALH plane - an accurate assumption in the limit of no interlayer coupling.

Defining a parameter matrix $P = (pp\sigma, \ldots, \mu_{3z^2-r^2})$ the problem amounts to minimizing the functional

$$D\left[E_{tb}\left(P\right)\right] \triangleq \sum_{j=1}^{18} \sum_{n=1}^{22} f_n \cdot \left(E_{expt}^{nj} - E_{tb}^{nj}\left(P\right)\right)^2$$

where E_{tb}^{nj} are the tight-binding energies of band n at point j , E_{expt}^{nj} are the energies measured in ARPES or derived from LDA, and f_n is a weight function, used to increase the accuracy of the fit for the two bands at E_F. For the fit itself I employed the Monte Carlo algorithm plus a 'point dragging' routine explained shortly.

The Monte Carlo algorithm is as follows: randomly generate a P. If the resulting D is smaller than the previous guess accept the new P. Otherwise accept the new P with probability $\exp\left(-\sqrt{D}/T\right)$ where T is an artificial 'temperature'. At higher temperatures it is more likely worse configurations will be accepted. The purpose of the temperature is to allow the search to jump out of local minima in the energy landscape of $\sqrt{D}$. For the

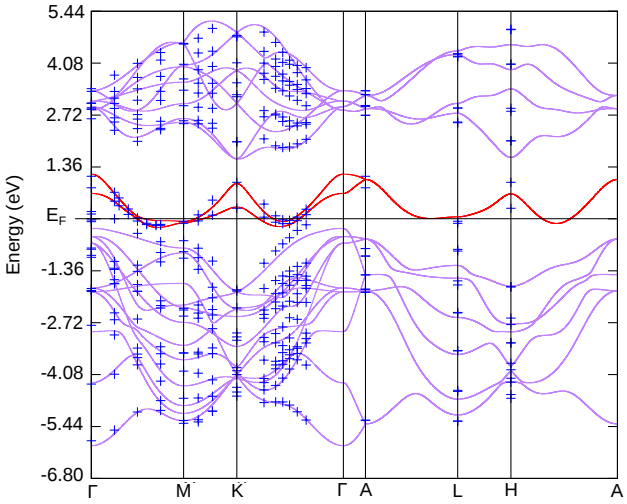


Figure 51: The result of the 22-band Slater-Koster fit. The blue crosses indicate the data being fit to, taken from ARPES for the two red bands crossing E_F, and LDA for the remaining 20 bands. The red bands were given approximately ten times the weight in the fit (approximate since the weighting was increased as the fit progressed). The energy tics correspond to tenths of a Rydberg.

seed P I used the values of a previous tight binding fit by Doran *et al.* [87].

The naïve way to apply the fit would be to use the seed value of P and run the Monte Carlo routine at high temperature down to low temperature (annealing). The problem with this is that between the band configuration generated by the seed, and the best fit values, there may be introduced unphysical band crossings. To account for this I instead introduced a 22-component alteration vector mapping E_{expt} onto E_{tb}:

$$\begin{aligned} E^{nj}_{expt} + \delta^{nj} &= E^{nj}_{tb} \\ \text{let} \qquad E^{nj}_{expt} &\rightarrow E^{nj}_{expt} + \alpha\delta^{nj} \end{aligned}$$

with $\alpha = 1$ initially. The Monte Carlo routine was then applied to the altered energies until the system had equilibrated, with equilibration being defined to be no change in D in 10^4 steps. An exponential temperature decrease was used to anneal the system. Once equilibrated, α was decreased, so that the experimental bands moved slightly towards their true values. The process was repeated until $\alpha = 0$. The advantage of this band dragging method is that the minimization can be adjusted if a band crossing is developing. The results of the fit are given in Figure 51.

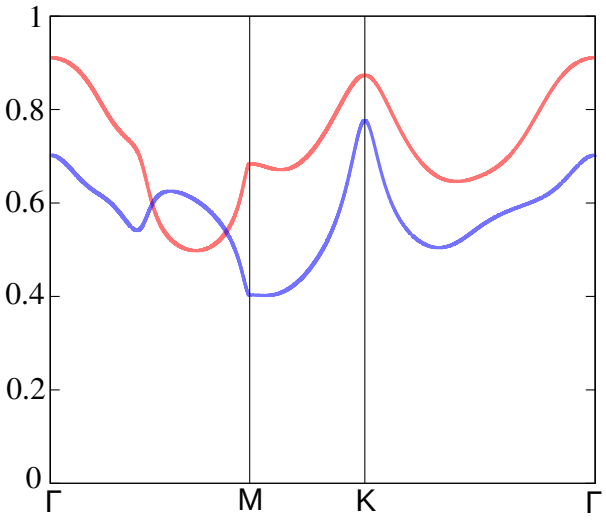


Figure 52: The percentage of the two bands crossing E_F being contributed by the two niobium $d_{3z^2-r^2}$ orbitals. Red is the lower energy (inner) band, blue the outer. As noted by Doran [99] the contribution is approximately 60%. Given that there are 22 orbitals which could potentially contribute, a two-band model is surprisingly accurate.

The orbital make-up of the bands can be deduced from the eigenvectors. A given eigenvector $|n\rangle$ corresponds to one band E_n, and the 22 components of $|n\rangle$ give a measure of the contribution from each orbital to that band. Inspecting the components of eigenvectors $|13\rangle$ and $|14\rangle$, *i.e.* the two bands at E_F, I found that throughout the Brillouin zone these two bands are composed primarily of the two niobium $d_{3z^2-r^2}$ orbitals. As shown in Figure 52 I find agreement with the approximate figure of 60% stated in a previous study by Doran [99]. Calling these two orbitals $|d^1_{3z^2-r^2}\rangle$ and $|d^2_{3z^2-r^2}\rangle$ the contribution to the two bands coming from these orbitals is precisely

$$\begin{aligned} |13\rangle &= \frac{1}{\sqrt{2}}\left(|d^1_{3z^2-r^2}\rangle + |d^2_{3z^2-r^2}\rangle\right) \\ |14\rangle &= \frac{1}{\sqrt{2}}\left(|d^1_{3z^2-r^2}\rangle - |d^2_{3z^2-r^2}\rangle\right). \end{aligned}$$

That the relation is exact follows from the $d_{3z^2-r^2}$ orbitals seeing identical environments to first-nearest neighbour.

Given this discovery the physics of the system will be well-described by a model of two non-interacting bands. This more restricted model should give a clearer physical picture as to the behaviour of the system.

7.1.2 2 Band Model

Having established that a model of two independent bands is a good approximation to the behaviour of $NbSe_2$, I will employ the phenomenological tight-binding fit to ARPES data of Rahn *et al.* [106] for the bandstructure, taking the two bands to be derived from the two $d_{3z^2-r^2}$ orbitals of the previous section. That fit considered in-plane hoppings to fifth-nearest neighbour, and no inter-plane hopping. Having an analytic form for the energies is convenient, but a second tight binding fit must be carried out for this simplified model.

The method proceeds as in the previous section[29]. The generalized eigenvalue problem now reads

$$\left[\begin{pmatrix} H_{11} & H_{12} \\ H_{12}^* & H_{22} \end{pmatrix} - E_n \begin{pmatrix} S_{11} & S_{12} \\ S_{12}^* & S_{22} \end{pmatrix}\right] |n\rangle = 0. \tag{41}$$

In the absence of time reversal symmetry breaking it is always a legitimate choice of gauge to take the elements of H and S, the hopping amplitudes and orbital overlaps, to be real. The two sandwiches of niobium atoms in the $NbSe_2$ unit cell are identical (as we are neglecting the selenium orbitals). Therefore we have that $H_{11} = H_{22}$ and $S_{11} = S_{22}$. With these restrictions in place the problem is

$$\begin{pmatrix} H_{11} - S_{11}E_n & H_{12} - E_nS_{12} \\ H_{12} - E_nS_{12} & H_{11} - S_{11}E_n \end{pmatrix} |n\rangle = 0.$$

The eigenvectors are

$$|\pm\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \pm 1 \end{pmatrix} \tag{42}$$

completely independent of $\mathbf{k}$.

The fitting problem rapidly becomes over-constrained when increasing the number of neighbours included in the fit, meaning it is possible to get an exact match to the experimental band energies for a range of (potentially unphysical) parameters. A physical restriction

[29] The Slater-Koster fit of this section was carried out by my supervisor Jasper van Wezel.

is necessary. To this end we included terms to fifth-nearest neighbour in-plane but only second-nearest neighbour out of plane. This allows for a very good fit to the band energies with a physically motivated choice of parameters. In fact the bandstructure fit is nearly indistinguishable, so I continued to use the expression of [106].

Discussion of the best-fit values of the matrices is saved for the following section where they can be considered in the proper context of their contributions to the (measurable) electron-phonon coupling.

7.2 Electron-Phonon Coupling From the Electronic Bandstructure

As mentioned in Chapter 6 the Coulomb interaction is small compared to the electron-phonon coupling in $NbSe_2$. I will therefore restrict attention to a model with Coulomb coupling $h = 0$ from now on. In this section I will deduce a functional form for the electron-phonon coupling g.

7.2.1 General d-orbital Transition Metal Compounds

Varma *et al.* [125] derived an expression for the electron-phonon coupling in transition metal compounds with a predominantly d-orbital character at the Fermi level. Since the d-orbitals are well-localized to their parent atoms the (non-orthogonal) tight-binding picture is a good approximation. The expression is

$$\mathbf{g}^{\mu,\nu}_{\mathbf{k},\mathbf{k}'} = \mathbf{v}^{\mu}_{\mathbf{k}} \left[A^{\dagger}_{\mathbf{k}} S_{\mathbf{k}} A_{\mathbf{k}'} \right]^{\mu\nu} - \left[A^{\dagger}_{\mathbf{k}} S_{\mathbf{k}'} A_{\mathbf{k}'} \right]^{\mu\nu} \mathbf{v}^{\nu}_{\mathbf{k}'} \tag{43}$$

with

$$\mathbf{v}^{\mu} \triangleq \frac{\partial \xi^{\mu}_{\mathbf{k}}}{\partial \mathbf{k}} \tag{44}$$

the electron velocity in band μ. An overall (purely imaginary) prefactor has been omitted. The vector nature of $\mathbf{v}$ and $\mathbf{g}$ accounts for the Cartesian real space directions. The matrix A is the matrix of eigenvectors solving the generalized eigenvalue problem of Equations 40 and 41.

Equation 43 is quite remarkable. Naturally one would assume that any expression for the electron-phonon coupling would require a knowledge of both the electron and phonon fields

separately, whereas this expression refers only to electronic properties. Nevertheless it has been well-tested in various transition metals and their compounds [125]. The catch is that knowledge of the orbital composition is required, which is difficult to ascertain with any certainty in general systems. In our case we have this information with some confidence by the working of the previous sections.

7.2.2 The Specific Case of $NbSe_2$

I will now apply Equation 43 to the specific case of the two-band model of $NbSe_2$ developed in Section 7.1. From Equation 42 we know that in this case we have

$$A_{\mathbf{k}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \tag{45}$$

completely independent of $\mathbf{k}$. The overlap matrix S takes the general form

$$S_{\mathbf{k}} = \begin{pmatrix} \alpha_{\mathbf{k}} & \beta_{\mathbf{k}} \\ \beta_{\mathbf{k}} & \alpha_{\mathbf{k}} \end{pmatrix}$$

with $\alpha, \beta \in \mathfrak{Re}$. The expression for the electron-phonon coupling in the two bands therefore simplifies drastically to

$$\mathbf{g}_{\mathbf{k},\mathbf{k}'} = (\alpha_{\mathbf{k}} \pm \beta_{\mathbf{k}})\, \mathbf{v}_{\mathbf{k}} - (\alpha_{\mathbf{k}'} \pm \beta_{\mathbf{k}'})\, \mathbf{v}_{\mathbf{k}'} \tag{46}$$

with the $+$ corresponding to the lower energy inner band and the $-$ to the outer band. The two extremes are $\alpha = 0$ xor $\beta = 0$, in which case the coupling in the two bands is identical, and $\alpha = \beta$, in which case the coupling in the outer band is zero. In fact, as a general measure across the Brillouin zone, $2\alpha \approx \beta$, meaning the coupling is around three times stronger in the inner band than the outer by this loose measure.

Plotting $\mathbf{g}_{\mathbf{k},\mathbf{k}'}$ directly would be difficult and not especially desirable. A physically measurable quantity, explained in detail in Chapter 8, is the generalized electronic susceptibility $D_2(\mathbf{k} - \mathbf{k}')$. A plot of this is shown in Figure 53 for the two bands, where it can be seen that the coupling in the inner band (*cf.* Figures 47 and 51) leads to approximately three to four times the susceptibility of that in the outer band around the CDW wavevector.

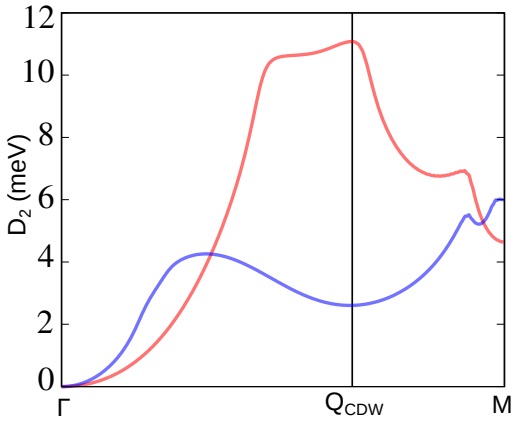


Figure 53: The generalized susceptibility D_2 in the two bands, using the analytically calculated electron-phonon coupling. Red corresponds to the inner band in Figure 47 or equivalently the lower red band in Figure 51, *i.e.* the band in which the CDW gap opens. The vector coupling has been projected along the direction of momentum transfer. Note that the susceptibility is peaked at precisely the experimentally observed CDW wavevector. This point is returned to in Section 8.3.

This resolves one of the controversies regarding $NbSe_2$, introduced in Chapter 6: why does the CDW gap open in one band only? If the CDW order were created through nesting it is not clear why a gap opens in the inner band but not the outer band, given that the CDW wavevectors lie somewhere between the two. If the CDW transition is in fact driven by electron-phonon coupling, and Equation 43 is trusted to give the form of that coupling, it is now clear it follows that the orbital composition of the bands can lead directly to the observed relative gap magnitudes.

Equation 46 is a vector expression with components corresponding to the Cartesian directions in space. The CDW is seen in X-ray scattering to correspond to a softening of the longitudinal acoustic phonon mode, with little to no softening of the transverse modes [103, 104]. For this reason it is convenient to project into a basis parallel and perpendicular to the momentum transfer at each point in reciprocal space. The scalar quantity of interest is therefore the longitudinal projection

$$g^{\mu}_{\mathbf{k},\mathbf{k}'} \triangleq \frac{\mathbf{k}-\mathbf{k}'}{|\mathbf{k}-\mathbf{k}'|} \cdot \mathbf{g}^{\mu}_{\mathbf{k},\mathbf{k}'}. \tag{47}$$

The overall scale of g is the only remaining free parameter in the theory. An order of magnitude estimate for the magnitude of the coupling is $\sim 100\,\text{meV}$ [126]; the value will be constrained by the transition temperature in the next chapter.

8 Modelling $NbSe_2$ in the Random Phase Approximation

In this chapter I apply the bandstructure fit, orbital band composition, and electron-phonon coupling, deduced in Chapter 7, to the modelling of the CDW transition in $NbSe_2$ using quantum field theory. In Section 8.1 I set up the basic field theory, then use it in Section 8.2 to eliminate the hypothesis that Fermi surface nesting drives the CDW transition. Having established the need for a strong-coupling theory, I re-introduce into the model the full electron-phonon coupling deduced in the previous chapter, and treat the model in the Random Phase Approximation in Section 8.3. I show that this approximation gives an accurate phonon dispersion compared to that seen in IXS. Finally in Sections 8.4 and 8.5 I add uniaxial strain into the model and find a phase diagram for the different CDW geometries (3Q and 1Q) over a range of strains and temperatures.

8.1 Field Theory Preliminaries

I begin by defining a partition function

$$\mathscr{Z} = \int \mathscr{D}\psi\mathscr{D}\varphi \exp\left(-S\left[\psi,\varphi\right]\right)$$

with the action S being a functional of the electron field ψ_k and phonon displacement field φ_q. Note that the phonon field is real in real space, meaning that $\varphi_q^\dagger = \varphi_{-q}$. As explained in Chapter 6, $NbSe_2$ is well-approximated as a quasi-2D layered material, and I will assume a two-dimensional space from here on in. Defining '3-momenta' $k = (i\omega_n, \mathbf{k})$ and $q = (i\Omega_n, \mathbf{q})$ the precise form of the action is

$$S = \sum_{k\nu} \psi^\dagger_{k\nu} G^{-1}_{k\nu} \psi_{k\nu} + \sum_q \varphi^\dagger_q D^{-1}_q \varphi_q + \sum_{kq\mu\nu} g^{\mu\nu}_{\mathbf{k},\mathbf{k}+\mathbf{q}} \varphi_q \psi^\dagger_{k+q,\nu} \psi_{k,\mu}$$

with electronic band indices μ, ν. The electron propagator $G_{k\nu}$ is of the standard [127] Schrödinger form:

$$G_{k\nu} \triangleq \left(i\omega_n - \xi^\nu_{\mathbf{k}} + \mu\right)^{-1} \tag{48}$$

with fermionic Matsubara[30] frequencies $\omega_n = (2n+1)\,\pi/\beta$ and electron dispersion $\xi^\nu_{\mathbf{k}}$ in

[30] Matsubara frequencies are probably most conveniently thought of simply as a mathematical abstraction

band ν. The symbol μ appears both as a band index and to signify the chemical potential; both notations are standard, and band indices appear only as subscripts or superscripts. The phonon propagator

$$D_q = \frac{-2\Omega_{\mathbf{q}}}{(i\Omega_n)^2 - \Omega_{\mathbf{q}}^2} \tag{49}$$

with bosonic Matsubara frequencies $\Omega_n = 2\pi n/\beta$ is derived from first principles in Appendix A. The bare phonon frequency is $\Omega_{\mathbf{q}}$. The final term in the action couples the electron field to the phonon field. Without it the model would contain only free, non-interacting particles, but with it the model contains the possibility of an electron being scattered from state k in band μ to state $k+q$ in band ν, via a phonon which conserves energy and momentum.

I will proceed by integrating out the electron field. To 'integrate out' a field means to carry out the functional integral over that field, and arrive at an effective action in terms of the remaining degrees of freedom, giving a low-energy effective field theory which hopefully captures the relevant physics of the system. In our case the nonlinear interaction term must be dealt with perturbatively via a Taylor expansion of the exponential:

$$\mathscr{Z} = \int \mathscr{D}\psi\mathscr{D}\varphi \exp\left(-S_\psi\right) \exp\left(-S_\varphi\right) \left(1 - S_{int} + \frac{1}{2}S_{int}^2 - \dots\right) \tag{50}$$

with S_ψ and S_φ the free electron and phonon actions respectively. The lowest-order non-vanishing contribution comes from the quadratic term:

$$S_{int}^2 = \sum_{\mu\nu\rho\sigma} \sum_{kk'qq'} g^{\mu\nu}_{\mathbf{k},\mathbf{k}+\mathbf{q}} g^{\rho\sigma}_{\mathbf{k}',\mathbf{k}'+\mathbf{q}'} \varphi_q \varphi_{q'} \psi^\dagger_{k+q\nu} \psi_{k\mu} \psi^\dagger_{k'+q'\sigma} \psi_{k'\rho}. \tag{51}$$

A functional average over the electron field is defined as

$$\langle \mathscr{O}[\psi] \rangle_\psi \triangleq \mathscr{Z}^{-1} \int \mathscr{D}\psi \exp\left(-S_\psi\right) \mathscr{O}[\psi] \tag{52}$$

(I will drop the subscript ψ from now on with the understanding that functional averages

introduced through the Wick rotation $t \to i\tau$, wherein the imaginary time dimension is seen to be cyclic, allowing a Fourier decomposition of functions in the τ coördinate. The different frequencies decomposed into are designated ω_n. See for example [128].

are always taken with respect to the electron field). The electrons, being fermions, are described by a Grassman field, defined by the anticommutation relation

$$\left\{\psi^{\dagger}_{k\nu}, \psi_{q\mu}\right\} \triangleq 0 \tag{53}$$

and the two-point correlator is derived from the action in Appendix B, taking the form

$$\left\langle \psi^{\dagger}_{k\nu}\psi_{q\mu}\right\rangle_{\psi} \quad = \quad G_{k\mu}\delta_{kq}\delta_{\mu\nu}. \tag{54}$$

Finally we need Wick's theorem, which states that (time-ordered) averages of products of fields can be replaced by the sum of possible pairwise contractions of those fields [127]. Combining these results we have for the expansion of Equation 50, after carrying out the integral over the ψ field,

$$\mathscr{Z} \quad \approx \quad \int \mathscr{D}\varphi \exp\left(-S_{\varphi}\right)\left(1+\frac{1}{2}\sum_{\mu\nu\rho\sigma}\sum_{kk'qq'} g^{\mu\nu}_{\mathbf{k},\mathbf{k}+\mathbf{q}} g^{\rho\sigma}_{\mathbf{k}',\mathbf{k}'+\mathbf{q}'}\varphi_q\varphi_{q'}\left\langle \psi^{\dagger}_{k+q\nu}\psi_{k\mu}\psi^{\dagger}_{k'+q'\sigma}\psi_{k'\rho}\right\rangle\right)$$

or finally

$$\mathscr{Z} \quad = \quad \int \mathscr{D}\varphi \exp\left(-S_{\varphi}\right)\left(1-\frac{1}{2}\sum_{\mu\nu}\sum_{kq} g^{\mu\nu}_{\mathbf{k},\mathbf{k}+\mathbf{q}} g^{\nu\mu}_{\mathbf{k}+\mathbf{q},\mathbf{k}}\varphi_q\varphi_{-q}G_{k+q\nu}G_{k\mu}\right).$$

Hermiticity of the interaction Hamiltonian requires $g^{\nu\mu}_{\mathbf{k}+\mathbf{q},\mathbf{k}} = g^{\mu\nu *}_{\mathbf{k},\mathbf{k}+\mathbf{q}}$. Finally we consider the term in the second set of parentheses to be the truncation of an exponential[31], giving the result

$$\mathscr{Z} \approx \int \mathscr{D}\varphi \exp\left(-S_{eff}\left[\varphi\right]\right)$$

with

[31] This 'reëxponentiation' always works; the rigorous proof is via the moment generating function.

$$S_{eff}\left[\varphi\right]=\sum_{q}\varphi_{q}^{\dagger}\left(D_{q}^{-1}+\frac{1}{2}\sum_{\mu\nu}\sum_{k}g_{\mathbf{k},\mathbf{k}+\mathbf{q}}^{\mu\nu}g_{\mathbf{k}+\mathbf{q},\mathbf{k}}^{\nu\mu}G_{k+q\nu}G_{k\mu}\right)\varphi_{q}.$$

The 'generalized electronic susceptibility', referred to in Chapter 7, is now properly defined through the final term:

$$D_{2}^{\mu\nu}\left(\mathbf{q},\Omega\right)\triangleq-\sum_{k}g_{\mathbf{k},\mathbf{k}+\mathbf{q}}^{\mu\nu}g_{\mathbf{k}+\mathbf{q},\mathbf{k}}^{\nu\mu}G_{k+q\nu}G_{k\mu}.$$

Following Doran [99] I have named this D_2, where the symbol indicates the second term in a series of increasing number of external phonon legs, starting with the one-legged bare phonon propagator D. The name correctly implies that the usual 'electronic susceptibility' in linear response theory [127] is given by the case $g\equiv 1$:

$$\chi_{\mu\nu}\left(\mathbf{q},\Omega\right)\triangleq-\sum_{k}G_{k+q\nu}G_{k\mu}.$$

The Matsubara sum over electronic frequencies is carried out in Appendix C, giving

$$D_{2}^{\mu\nu}\left(\mathbf{q},\Omega\right) = -\sum_{\mathbf{k}}g_{\mathbf{k},\mathbf{k}+\mathbf{q}}^{\mu\nu}g_{\mathbf{k}+\mathbf{q},\mathbf{k}}^{\nu\mu}\frac{f\left(\xi_{\mathbf{k}}^{\mu}\right)-f\left(\xi_{\mathbf{k}+\mathbf{q}}^{\nu}\right)}{\xi_{\mathbf{k}}^{\mu}-\xi_{\mathbf{k}+\mathbf{q}}^{\nu}+i\Omega} \tag{55}$$

$$\chi_{\mu\nu}\left(\mathbf{q},\Omega\right) = -\sum_{\mathbf{k}}\frac{f\left(\xi_{\mathbf{k}}^{\mu}\right)-f\left(\xi_{\mathbf{k}+\mathbf{q}}^{\nu}\right)}{\xi_{\mathbf{k}}^{\mu}-\xi_{\mathbf{k}+\mathbf{q}}^{\nu}+i\Omega} \tag{56}$$

with f the Fermi-Dirac distribution function and Ω the frequency of an externally applied phonon used to probe the system in linear response. Although I'll be considering the DC response of the system, a small nonzero Ω acts to regularize the numerical sum over $\mathbf{k}$. In this case the real part of the susceptibility should be taken, although it has been argued that for some systems the imaginary part also has relevance to the investigation of electronic nesting [112]. The negative sign in front ensures that the expression is always positive.

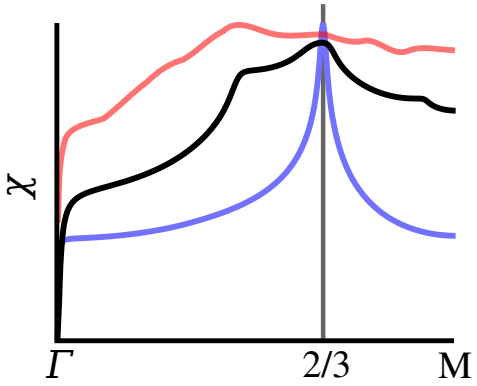


Figure 54: The electronic susceptibility, χ, for three cases. The magnitude is arbitrary in each case - only the shape is relevant. Blue: 1D bandstructure with cosine dispersion and nesting vector $\frac{2}{3}\overrightarrow{\Gamma M}$. Red: $NbSe_2$ bandstructure assuming a 2-band model with equally weighted inter- and intra-band scattering. Black: $NbSe_2$ with inner band scattering only. All plots at $T = 33\,\mathrm{K}$ with $\Omega = 5\,\mathrm{meV}$ regularization.

8.2 The Extent of Nesting in $NbSe_2$

Nesting is only guaranteed in 1D, and, while some higher-dimensional density wave systems such as chromium certainly exhibit it, there is no reason to believe it is the generic ordering mechanism in such cases. In Chapter 6 I showed that qualitatively $NbSe_2$ does not appear likely to exhibit nesting. To make this statement quantitative it is convenient to consider a restricted electron-phonon coupling which is a function only of the magnitude of the momentum transfer: $g_{\mathbf{k},\mathbf{k}+\mathbf{q}} \rightarrow g_{|\mathbf{q}|}$. In this case the generalized susceptibility is simply the product of the square of the electron-phonon coupling with the standard electronic susceptibility: $D_2(\mathbf{q}) = g^2(|\mathbf{q}|)\,\chi(\mathbf{q})$. From Equation 56 it is clear that the susceptibility χ should diverge if the Fermi surface nests, since this means that $\xi^{\mu}_{\mathbf{k}} = \xi^{\mu}_{\mathbf{k}+\mathbf{q}}$ for a range of possible $\mathbf{k}$, sending the denominator close to zero.

In Figure 54 I show the susceptibility for three cases: a prototype 1D bandstructure; equally-weighted inter- and intra-band scattering for $NbSe_2$ ($g^{\mu\nu} \equiv 1$); and intra-band scattering for the $NbSe_2$ inner band only ($g^{\mu\nu} = \delta^{\mu,inner}\delta^{\nu,inner}$). The last case is of interest since the true $NbSe_2$ coupling is known from Chapter 7 to be approximately of this form owing to the orbital make-up of the bands themselves. The case of equally-weighted inter- and intra-band scattering has been the focus of previous studies [112, 106, 102], and would be the natural consideration without the additional knowledge of the orbital composition of the bands gleaned from Chapter 7.

The 1D bandstructure has a divergence in the susceptibility at the nesting vector, chosen to be $\frac{2}{3}\overrightarrow{\Gamma M}$. The $NbSe_2$ bandstructures, on the other hand, are rather flat. The susceptibility of the inner band is maximum around the CDW ordering vector, which is suggestive, but comparing with the true divergence of the 1D case it appears qualitatively that $NbSe_2$ does not exhibit nesting. For a quantitative assessment I now turn to the free energy.

Expanding the inverse phonon propagator the effective action now reads

$$\begin{aligned} S_{eff}\left[\varphi\right] &= \sum_q \varphi_q^\dagger \left(\frac{\left(i\Omega_n\right)^2 - \Omega_{\mathbf{q}}^2}{-2\Omega_{\mathbf{q}}} - \frac{1}{2}\sum_{\mu\nu} \left(g^{\mu\nu}\right)^2 \chi^{\mu\nu} \right) \varphi_q \\ &= \sum_q \varphi_q^\dagger \left(\frac{\left(i\Omega_n\right)^2 - \Omega_{\mathbf{q}}^2 + \Omega_{\mathbf{q}} \sum_{\mu\nu} \left(g^{\mu\nu}\right)^2 \chi^{\mu\nu}}{-2\Omega_{\mathbf{q}}} \right) \varphi_q. \end{aligned} \tag{57}$$

The term in parentheses can be considered to be a renormalized inverse phonon propagator, with renormalized frequency

$$\Omega_{RPA}^2 = \Omega_{\mathbf{q}}^2 - \Omega_{\mathbf{q}} \sum_{\mu\nu} \left(g^{\mu\nu}\right)^2 \chi^{\mu\nu}\left(\mathbf{q}, \Omega\right) \tag{58}$$

with RPA indicating that this constitutes the 'Random Phase Approximation'. The name is largely historical, but an alternative derivation gives some motivation for it while revealing the physical meaning: neglecting indices for clarity, the free phonon propagator is corrected by an infinite series of electron-hole loops

$$D_{RPA} = D + Dg^2\chi D + Dg^2\chi Dg^2\chi D + \ldots$$

where the behaviour of each loop is uncorrelated with the others, hence 'random phase'. Taking out a common factor on the right gives

$$\begin{aligned} D_{RPA} &= D + Dg^2\chi\left(D + Dg^2\chi D + \ldots\right) \\ &= D + Dg^2\chi D_{RPA} \end{aligned}$$

and rearranging

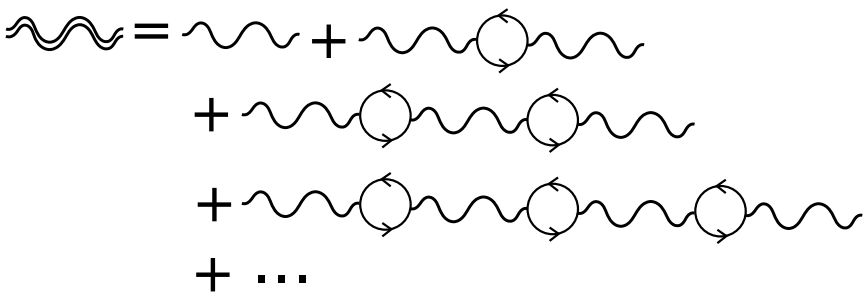


Figure 55: The Random Phase Approximation (RPA). The bare phonon is renormalized by a process of splitting into an electron-hole pair and recombining to a phonon. It can do this an infinite number of times, but each splitting introduces two factors of the coupling g. The infinite sum can be carried out exactly provided the loops are uncorrelated.

$$D_{RPA} = \left(D^{-1} - g^2\chi\right)^{-1}.$$

Inserting Equation 49 for the bare phonon propagator D gives

$$D_{RPA} = \frac{-2\Omega_{\mathbf{q}}}{(i\Omega_n)^2 - \Omega_{\mathbf{q}}^2 + \Omega_{\mathbf{q}} g^2 \chi}$$

and Equation 58 follows from consideration of the new pole location. The approximation is shown diagrammatically in Figure 55.

A few words on the philosophy of renormalization are in order. The programme just carried out assumed that 'bare' phonons existed in the material and had dispersion relation $\Omega_{\mathbf{q}}$. When the phonon and electron fields interact via electron-phonon coupling, neither bare phonons nor bare electrons are any longer good quasiparticles. The well-defined excitations of the system are a mix of the two. However, by integrating out the electron field we can approximate the new field as a perturbation to the original phonon field - it is still possible to define a propagator, but the effect of the electrons is to renormalize the dispersion $\Omega_{\mathbf{q}} \rightarrow \Omega_{RPA}(\mathbf{q})$ (at this order of approximation). In fact, since the electron-phonon coupling was always present, it is the renormalized field we can probe in experiment, and the bare field is something of a mathematical abstraction.

The phonon spectrum in $NbSe_2$ has been measured directly by Weber *et al.* using inelastic X-ray scattering (IXS) [103, 104]. The group measured the energies of longitudinal acoustic phonons at a range of temperatures (250 K, 50 K, 33 K, 8 K) and found a softening of the frequencies at a range of wavevectors around the CDW ordering vector, dipping to zero at

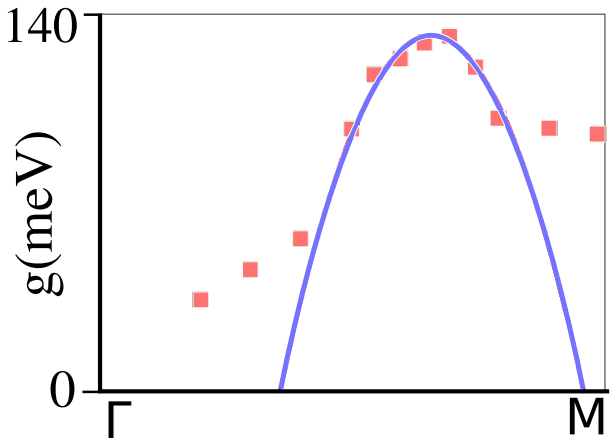


Figure 56: Using the phonon dispersion data of [103], from X-ray diffraction, and the bandstructure of [106], estimated from ARPES, in combination with Equation 58, I estimated the electron-phonon coupling g point by point. I then made a least-squares fit to a parabola.

33 K as expected. Some of the results are reproduced in Figure 46.

Despite the flat electronic susceptibility χ of Figure 54 the RPA could still account for the data in Figure 46 if the electron-phonon coupling $g_{|\mathbf{q}|}$ itself has a shape. In fact, using Equation 58, it is possible to work out what the shape must be for consistency. The bare phonon spectrum is a mathematical abstraction, but we can imagine that in the limit of infinite temperature the fields become decoupled. We therefore take the highest temperature (250 K) data as an approximation to the bare phonon dispersion. Using the bandstructure of Rahn *et al.* to calculate χ, there remain two sets of data (50, 33 K) to constrain $g_{|\mathbf{q}|}$.

I extracted the data points from the plot of reference [103] reproduced in Figure 46. For each I calculated the required g at both temperatures. The value of g varied very little between the two, suggesting it is largely temperature independent in the temperature range of interest. I therefore focussed on the 33 K data, for which the most data points are available. Figure 56 shows the calculated coupling, plus a parabolic fit of the form

$$g_{|\mathbf{q}|} = -a\left(q_{peak} - |\mathbf{q}|\right)^2 + g_{max} \tag{59}$$

found using GNUPlot's least-squares fitting algorithm applied to the points around $\mathbf{Q}_{CDW}$. The best-fit values were $a = 5813.8\,\text{meV}\left|\overrightarrow{\Gamma\Gamma}\right|^{-2}$, $g_{max} = 131.9\,\text{meV}$ with q_{peak} constrained to the value observed at T_{CDW} in IXS and STM [129] of $q_{peak} = 0.986 \cdot \frac{2}{3}\left|\overrightarrow{\Gamma M}\right|$.

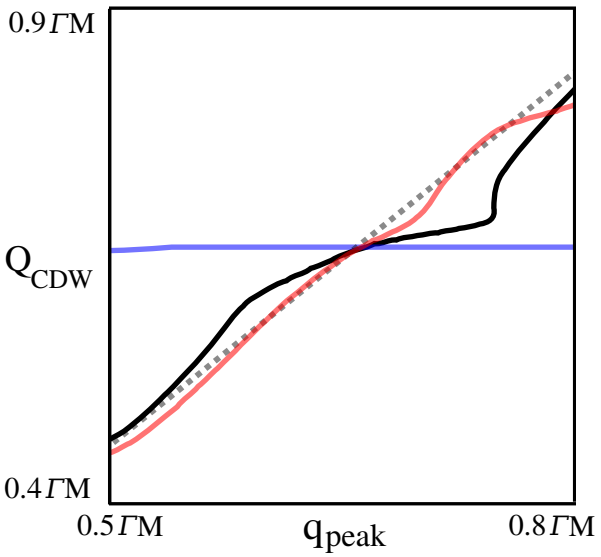


Figure 57: The RPA prediction of the CDW ordering vector Q_{CDW} as a function of the peak location of the electron-phonon coupling. Blue: a 1D bandstructure with nesting vector $\frac{2}{3}\left|\overrightarrow{\Gamma M}\right|$. Red: $NbSe_2$ bandstructure including inter- and intra-band scattering in the two-band model. Black: $NbSe_2$ bandstructure with scattering in the inner band only. Grey (dotted) is a line at 45°. Compare to Figure 54 for the corresponding susceptibilities χ.

Upon cooling the system the RPA predicts the development of a CDW with wavevector $\mathbf{Q}$ when $\Omega_{RPA}(\mathbf{Q})$ first softens to zero.

The CDW ordering vector is given in the RPA by the convolution of g^2 with χ. In a nested bandstructure the divergence in χ would set $\mathbf{Q}_{CDW}$, and the result would be independent of q_{peak}. In a flat bandstructure q_{peak} would set $\mathbf{Q}_{CDW}$.

In Figure 57 I show the $\mathbf{Q}_{CDW}$ predicted from RPA as a function of q_{peak}, with q_{peak} varied over a wide range about the best-fit value. In the case of a 1D bandstructure there is a wide plateau at the nesting vector $\frac{2}{3}\left|\overrightarrow{\Gamma M}\right|$, indicating that the CDW ordering vector is dictated by the peak in χ and is ambivalent to the form of the electron-phonon coupling. This is to be expected since the phase transition is certainly driven by nesting in 1D.

Two cases are again considered for $NbSe_2$: all inter- and intra-band scatterings, and inner band scatterings only. In both cases the lines are approximately diagonal. In the inner band there is a slight plateau around $\frac{2}{3}\left|\overrightarrow{\Gamma M}\right|$, due to the small bump (or 'nubbin') in that band's susceptibility at that wavevector.

In the combined case given by the red line (the case considered in previous studies [112, 106, 102]), it is clear there is no nesting. The diagonal line indicates that q_{peak} completely

dictates $\mathbf{Q}_{CDW}$, and consequently the electron-phonon coupling completely dictates the CDW order. From the working of Chapter 7, however, we know that a model of two independent bands is a more accurate description of $NbSe_2$, and so the inner-band only (black line) is the more appropriate consideration. In this case the line is still approximately diagonal, ruling out true nesting, but there is a slight levelling-off which is best inferred from the small jump around $0.75\overrightarrow{\Gamma M}$. This implies some limited matching of states on the Fermi surface, certainly believable from Figure 47, which ultimately acts to cause the nubbin at $\mathbf{Q}_{CDW}$ evident in the electronic susceptibility χ in Figure 54.

It is therefore a combination of the electronic contribution and electron-phonon coupling which drives the CDW transition in $NbSe_2$. The orbital content of the bands, a strong-coupling consideration introduced via $g^{\mu\nu}$, isolates the bands' contributions to the generalized susceptibility $g^2\chi$. The shape of $g_{|\mathbf{q}|}$ provides the peak in the generalized susceptibility necessary to drive the CDW phase transition, but it is the limited matching of states separated by $\mathbf{Q}_{CDW}$ on the inner K pockets of the Fermi surface which causes a nubbin in χ and selects the precise CDW wavevector.

8.3 The CDW Wavevector in the RPA

Having used the reduced form $g_{|\mathbf{q}|}$ to investigate the extent of nesting I now return to the full expression found in Section 7.2 and particularly Equation 47. The only free parameter is the overall magnitude of g. The RPA was introduced in Section 8.2, and from Equation 58 it follows that the renormalized phonon frequency takes the form

$$\Omega^2_{RPA}(\mathbf{q}) = \Omega^2_{\mathbf{q}} - \Omega_{\mathbf{q}} D_2(\mathbf{q}, \Omega). \tag{60}$$

For the bare phonon frequency $\Omega_{\mathbf{q}}$ I used a phenomenological fit to X-ray scattering results using Brillouin functions, as proposed in [104]. Specifically the form I employed is

$$\Omega_{\mathbf{q}} = 11.2\text{meV} \cdot \left[\frac{3}{2}\coth\left(\frac{3}{2}\cdot 12\,|\mathbf{q}|\right) - \frac{1}{2}\coth\left(\frac{1}{2}\cdot 12\,|\mathbf{q}|\right)\right]. \tag{61}$$

Figure 58 shows D_2 for the longitudinal phonon mode in the inner band at 33.5 K, where the magnitude of g has been set to give the RPA phase transition at this temperature. Figure 59 shows the corresponding renormalized phonon frequency.

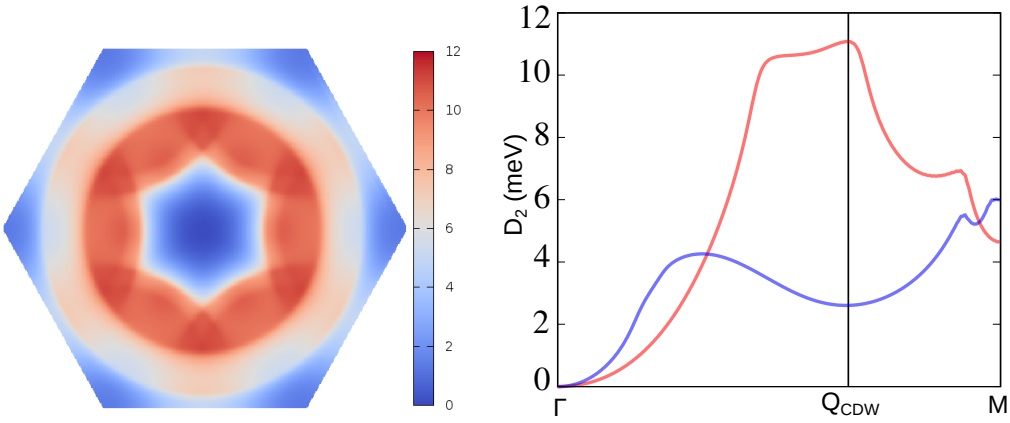


Figure 58: Left: generalized susceptibility D_2 in the RPA across the Brillouin zone for longitudinal phonons in the (inner) band forming the CDW. The maxima lie along ΓM (units meV). Right (reproduction of Figure 53): a cut along ΓM reveals that the peak exactly corresponds to the observed CDW ordering vector (red line). The blue line shows the corresponding expression for longitudinal phonons in the outer band. The plots are at 33.5 K. The numerical sums have approximately 200 steps in each direction and a 10 meV regularization is employed.

The plot across the Brillouin zone shows that the phonon mode first softens to zero along the lines $\overrightarrow{\Gamma M}$. Taking a cut in this direction reveals a remarkable agreement with experiment: the point at which Ω_{RPA} first touches zero, which dictates the CDW ordering vector, exactly corresponds to the $3Q$ ordering vector $\mathbf{Q}_{CDW} = 0.986 \cdot \frac{2}{3}\overrightarrow{\Gamma M}$ seen, for example, in neutron scattering [71, 100], or X-ray diffraction [103, 104] (see Figure 46). Additionally there is a plateau encompassing lower Q values, which allows for the $1Q$ ordering value of $\mathbf{Q}_{1Q} \approx \frac{2}{7}\overrightarrow{\Gamma M}$ considered shortly. The broadness of the phonon softening was considered anomalous in $NbSe_2$, but is in fact quite reasonable now that nesting has been eliminated as the driving mechanism.

The shape of the RPA curves is impressively accurate. The weakness of the method lies in the temperature dependence: since only uncorrelated electron-hole loops are included in the phonons' renormalization there is no contribution from the phonon field itself. As it is clear at this stage that phonons, via the electron-phonon coupling, are playing a decisive rôle in driving the phase transition, a proper account of their entropy should be included through additional diagrams [130, 131]. The relevant terms form the so-called Mode-Mode Coupling Approximation, MMA, and are dealt with in Chapter 9.

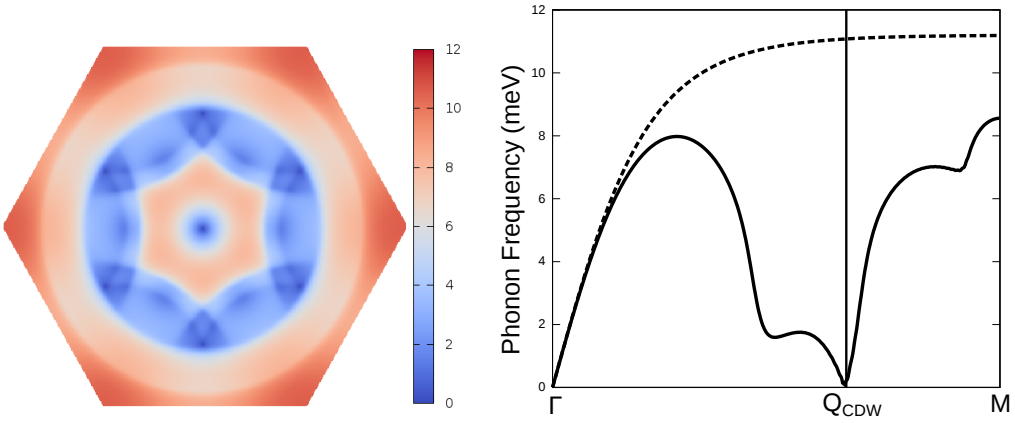


Figure 59: Left: the RPA renormalized phonon frequency across the Brillouin zone, in units of meV. Right: a cut along ΓM. The dashed line is the bare phonon frequency $\Omega_{\mathbf{q}}$ and the solid line is the RPA renormalized frequency Ω_{RPA}, seen to soften to zero at the experimental CDW ordering vector.

8.4 Higher Order Diagrams

8.4.1 Free Energy Expansion

The Random Phase Approximation alters the form of the two-point correlator for the phonon field. Provided the phase transition is second order this term (order φ^2) dictates the transition. To include the possibility of a first-order transition, and to gain knowledge of the form of the ordered phase, I will now consider higher order terms in the free energy expansion.

The free energy of the system, F, is defined through the relation

$$\begin{aligned} \exp\left(-\beta F\right) &\triangleq \mathscr{Z} \\ &= \int \mathscr{D}\varphi \exp\left(-S_{eff}\left[\varphi\right]\right) \end{aligned} \tag{62}$$

with partition function $\mathscr{Z}$ and effective action (after integrating out the electron field perturbatively) S_{eff}. I relegate the mathematical details of the perturbative expansion to Appendix E. Aside from more complicated combinatorics the method is identical to that used for the φ^2 term already considered. Diagrammatically the Feynman rules are shown in Figure 60 and the free energy expansion in Figure 61.

Expanding to fourth order in number of external legs (phonon propagators) the result is

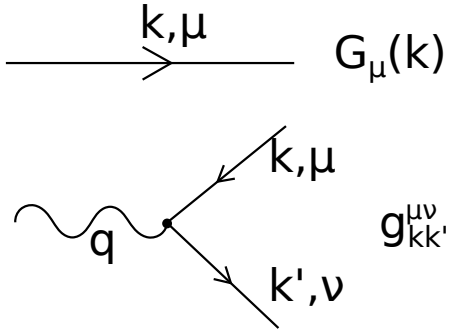


Figure 60: The Feynman rules for the perturbative expansion of the field theory.

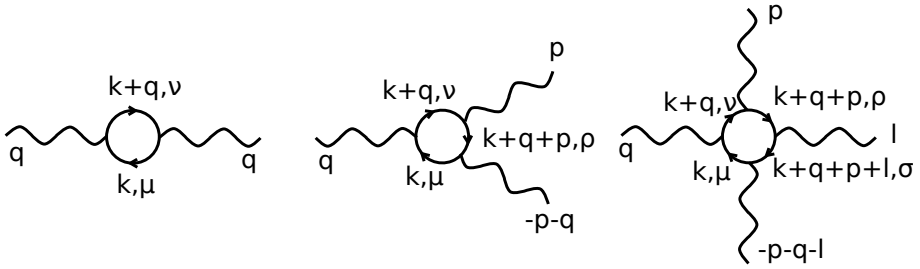


Figure 61: The Feynman diagrams constituting the weak coupling field theory. The expansion is in terms of number of external legs (phonon fields).

$$
\begin{aligned}
S_{eff}\left[\varphi\right] &= \frac{1}{2}\sum_{q}\varphi_{q}\left(\frac{\Omega_{n}^{2}}{\Omega_{\mathbf{q}}}+\Omega_{\mathbf{q}}+\sum_{\mu\nu}\sum_{k}g_{\mathbf{k},\mathbf{k}+\mathbf{q}}^{\mu\nu}g_{\mathbf{k}+\mathbf{q},\mathbf{k}}^{\nu\mu}G_{k\mu}G_{k+q\nu}\right)\varphi_{-q}\\
&+\frac{1}{3}\sum_{\mu\nu\rho}\sum_{kqp}g_{\mathbf{k},\mathbf{k}+\mathbf{q}}^{\mu\nu}g_{\mathbf{k}+\mathbf{q},\mathbf{k}+\mathbf{q}+\mathbf{p}}^{\nu\rho}g_{\mathbf{k}+\mathbf{q}+\mathbf{p},\mathbf{k}}^{\rho\mu}\varphi_{p}\varphi_{q}\varphi_{-p-q}G_{k}^{\mu}G_{k+q}^{\nu}G_{k+p+q}^{\rho}\\
&+\frac{1}{4}\sum_{\mu\nu\rho\sigma}\sum_{kpql}\left(g_{\mathbf{k},\mathbf{k}+\mathbf{q}}^{\mu\nu}g_{\mathbf{k}+\mathbf{q},\mathbf{k}+\mathbf{q}+\mathbf{p}}^{\nu\rho}g_{\mathbf{k}+\mathbf{q}+\mathbf{p},\mathbf{k}+\mathbf{q}+\mathbf{p}+\mathbf{l}}^{\rho\sigma}g_{\mathbf{k}+\mathbf{q}+\mathbf{p}+\mathbf{l},\mathbf{k}}^{\sigma\mu}\cdot\right.\\
&\left.\cdot\varphi_{q}\varphi_{p}\varphi_{l}\varphi_{-l-p-q}G_{k}^{\mu}G_{k+q}^{\nu}G_{k+p+q}^{\rho}G_{k+p+l+q}^{\sigma}\right).
\end{aligned}
$$

The Matsubara sums over electron frequency are calculated in Appendix C. Recalling the definition of the generalized susceptibility

$$
D_{2}^{\mu\nu}\left(\mathbf{q},\Omega\right)\triangleq-\sum_{\mathbf{k}}g_{\mathbf{k},\mathbf{k}+\mathbf{q}}^{\mu\nu}g_{\mathbf{k}+\mathbf{q},\mathbf{k}}^{\nu\mu}\frac{f\left(\xi_{\mathbf{k}}^{\mu}\right)-f\left(\xi_{\mathbf{k}+\mathbf{q}}^{\nu}\right)}{\xi_{\mathbf{k}}^{\mu}-\xi_{\mathbf{k}+\mathbf{q}}^{\nu}+i\Omega}
$$

I will also define similar terms for loops coupling three and four phonons, D_3 and D_4, respectively:

$$
\begin{aligned}
D_3^{\mu\nu\rho}(\mathbf{q},\mathbf{p}) &\triangleq \sum_k g^{\mu\nu}_{\mathbf{k},\mathbf{k+q}} g^{\nu\rho}_{\mathbf{k+q},\mathbf{k+q+p}} g^{\rho\mu}_{\mathbf{k+q+p},\mathbf{k}} G^{\mu}_{k} G^{\nu}_{k+q} G^{\rho}_{k+p+q} \\
&= \sum_{\mathbf{k}} g^{\mu\nu}_{\mathbf{k},\mathbf{k+q}} g^{\nu\rho}_{\mathbf{k+q},\mathbf{k+q+p}} g^{\rho\mu}_{\mathbf{k+q+p},\mathbf{k}} \cdot \\
&\quad \cdot \left[\frac{f\left(\xi^{\mu}_{\mathbf{k}}\right)}{\left(\xi^{\mu}_{\mathbf{k}} - \xi^{\nu}_{\mathbf{k+q}}\right)\left(\xi^{\mu}_{\mathbf{k}} - \xi^{\rho}_{\mathbf{k+p+q}}\right)} + 2\, cyclic perms. \right]
\end{aligned} \tag{63}
$$

and

$$
\begin{aligned}
D_4^{\mu\nu\rho\sigma}(\mathbf{q},\mathbf{p},\mathbf{l}) &\triangleq \sum_k g^{\mu\nu}_{\mathbf{k},\mathbf{k+q}} g^{\nu\rho}_{\mathbf{k+q},\mathbf{k+q+p}} g^{\rho\sigma}_{\mathbf{k+q+p},\mathbf{k+q+p+l}} g^{\sigma\mu}_{\mathbf{k+q+p+l},\mathbf{k}} G^{\mu}_{k} G^{\nu}_{k+q} G^{\rho}_{k+p+q} G^{\sigma}_{k+p+q+l} \\
&= \sum_{\mathbf{k}} g^{\mu\nu}_{\mathbf{k},\mathbf{k+q}} g^{\nu\rho}_{\mathbf{k+q},\mathbf{k+q+p}} g^{\rho\sigma}_{\mathbf{k+q+p},\mathbf{k+q+p+l}} g^{\sigma\mu}_{\mathbf{k+q+p+l},\mathbf{k}} \cdot \\
&\quad \cdot \left[\frac{f\left(\xi^{\mu}_{\mathbf{k}}\right)}{\left(\xi^{\mu}_{\mathbf{k}} - \xi^{\nu}_{\mathbf{k+q}}\right)\left(\xi^{\mu}_{\mathbf{k}} - \xi^{\rho}_{\mathbf{k+q+p}}\right)\left(\xi^{\mu}_{\mathbf{k}} - \xi^{\sigma}_{\mathbf{k+q+p+l}}\right)} + 3\, cyc.\, perm. \right].
\end{aligned} \tag{64}
$$

As D_2 corresponds to the generalized susceptibility in the linear response regime, Ω can be interpreted as the frequency of an externally applied probe phonon. There is no similar interpretation available in the nonlinear response $D_{>2}$, so I set $\Omega = 0$ and regularized the numerical calculation analytically. When two energies ξ meet, the expressions appear to diverge, but in fact there is a cancellation between the cyclically permuted terms [4] (see Appendix E for details).

Taking the mean field solution for φ, and so neglecting the functional integral in Equation 62, the zero frequency case becomes

$$
\begin{aligned}
\beta F[\varphi] = S_{eff}[\varphi] &= \frac{1}{2}\sum_{\mathbf{q}} |\varphi_{\mathbf{q}}|^2 \left[\Omega_{\mathbf{q}} - D_2(\mathbf{q})\right] \\
&\quad + \frac{1}{3}\sum_{\mathbf{qp}} \varphi_{\mathbf{q}}\varphi_{\mathbf{p}}\varphi_{-\mathbf{p}-\mathbf{q}} D_3(\mathbf{q},\mathbf{p}) \\
&\quad + \frac{1}{4}\sum_{\mathbf{pql}} \varphi_{\mathbf{q}}\varphi_{\mathbf{p}}\varphi_{\mathbf{l}}\varphi_{-\mathbf{l}-\mathbf{p}-\mathbf{q}} D_4(\mathbf{q},\mathbf{p},\mathbf{l})
\end{aligned}
$$

where I have suppressed band indices for clarity. The system will order into the field

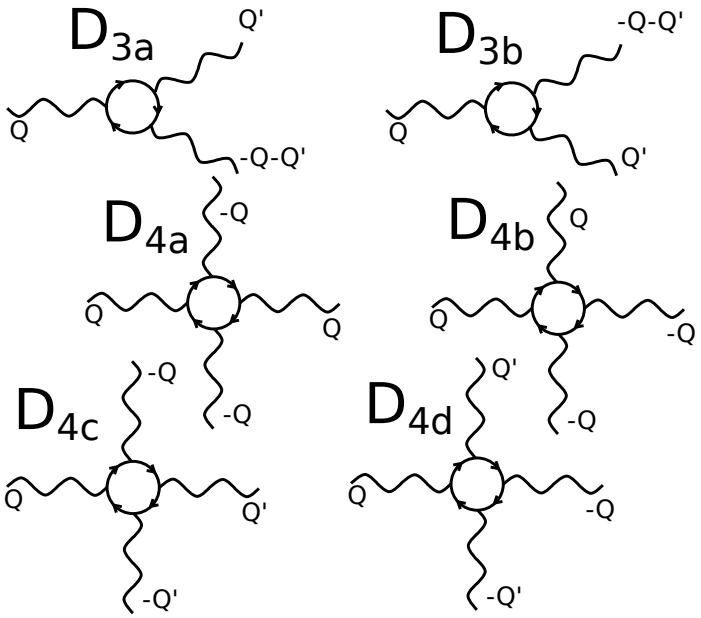


Figure 62: The Feynman diagrams in the free energy expansion after restricting attention to the CDW vectors $\mathbf{q} \in \mathbf{Q}_i$ (Equations 65 and 66). The labels Q and Q' indicate different members of the set of six vectors $\{\mathbf{Q}_i\}$. External phonon legs should be considered to be amputated.

configuration of minimum free energy. Considering all possible phonon momenta is too computationally demanding, and is not necessary - as we will only be concerned with the first ordered state the system drops into, all the phonon order parameters are zero for all $\mathbf{q}$ except the first one in which order develops. We can therefore neglect the sums over momenta in what follows. I consider instead only the fixed momenta of the CDW ordering vectors seen in experiment, along the $\overrightarrow{\Gamma M}$ directions, $\pm\mathbf{Q}_i$ $i \in [1,3]$. Writing $\varphi_i \triangleq \varphi(\mathbf{Q}_i)$ and considering all possible momentum conserving combinations this gives

$$\begin{aligned}
\beta F &= \frac{1}{2} \times 2 \times [\Omega_{\mathbf{q}} - D_2] \left(|\varphi_1|^2 + |\varphi_2|^2 + |\varphi_3|^2 \right) \\
&\quad -\frac{1}{3} \times 6 \times (D_{3a} + D_{3b}) \varphi_1 \varphi_2 \varphi_3 \\
&\quad +\frac{1}{4} \times 2 \times (D_{4a} + 2D_{4b}) \left(|\varphi_1|^4 + |\varphi_2|^4 + |\varphi_3|^4 \right) \\
&\quad +\frac{1}{4} \times 8 \times (D_{4d} + 2D_{4c}) \left(|\varphi_1|^2 |\varphi_2|^2 + |\varphi_1|^2 |\varphi_3|^2 + |\varphi_2|^2 |\varphi_3|^2 \right)
\end{aligned}$$

where the negative sign on the cubic term is added since the minimum energy solution will always have a negative cubic part. The terms are defined in Figure 62. Defining

$$
\begin{aligned}
r &= -\Omega_{\mathbf{q}} + D_2 \\
a &= 2\left(D_{3a} + D_{3b}\right) \\
b &= D_{4a} + 2D_{4b} \\
bc &= 4\left(D_{4d} + 2D_{4c}\right)
\end{aligned} \tag{65}
$$

the final expression for the free energy of the system is

$$
\beta F = -r\sum_{i=1}^{3}|\varphi_i|^2 - a\varphi_1\varphi_2\varphi_3 + \frac{1}{2}b\left(\sum_{i=1}^{3}|\varphi_i|^4 + c\left(|\varphi_1|^2|\varphi_2|^2 + |\varphi_1|^2|\varphi_3|^2 + |\varphi_2|^2|\varphi_3|^2\right)\right). \tag{66}
$$

Minimizing this expression is still too arduous a task. Instead we can put in more physical information, namely that the system is seen to order into one of the two states:

$$
\begin{aligned}
3Q: &\qquad \varphi_1 = \varphi_2 = \varphi_3, &&\qquad \mathbf{Q} = 0.986 \cdot \frac{2}{3}\overrightarrow{\Gamma M} \\
1Q: &\qquad \varphi_1 = \varphi,\ \varphi_2 = \varphi_3 = 0, &&\qquad \mathbf{Q} = \frac{2}{7}\overrightarrow{\Gamma M}
\end{aligned}
$$

giving the results

$$
\begin{aligned}
\beta F_{3Q}[\varphi] &= -3r\varphi^2 - a\varphi^3 + \frac{3}{2}b\left(c+1\right)\varphi^4 \\
\beta F_{1Q}[\varphi] &= -r\varphi^2 + \frac{1}{2}b\varphi^4.
\end{aligned} \tag{67}
$$

Note that if $b < 0$ then strictly the sixth-order term should be included. Alternatively it could be argued that the regions with $b < 0$ are too far away from the minimum in $F[\varphi]$; in either case, b is greater than zero in all the cases I consider in this thesis. Minimizing the functionals of Equation 67 gives the minimum energy solutions. For the case of the 1Q CDW the result is

$$\beta F_{1Q}^{min} = \begin{cases} -r^2/2b, & r > 0 \\ 0, & r \leq 0 \end{cases}$$

where order develops through a second order phase transition at $r = 0$ when approaching from the disordered regime with $r < 0$. For the 3Q case the situation is complicated by the cubic term. The free energy of the ordered system is

$$\begin{aligned} \beta F_{3Q}^{min} &= \frac{-1}{64b^3 (c+1)^3} \left(a^4 + 24ra^2b (c+1)\right. \\ &\quad \left. +96r^2b^2 (c+1)^2 + |a| \left(a^2 + 16rb (c+1)\right)^{\frac{3}{2}} \operatorname{sgn}(c+1)\right) \end{aligned}$$

and the order develops via a second order phase transition if

$$r > 0,\, c > -1 \tag{68}$$

or a first order phase transition if

$$r < 0,\, c > -1,\, b (c+1) < -a^2/18r. \tag{69}$$

8.4.2 Proximity of Phase Transitions

The question of the order of the phase transition in $NbSe_2$ is an old one. Early work assumed a (bulk) 3Q ordered state; in 1975 McMillan used a free energy calculation to argue that the transition should be first-order or 'weakly first-order' in the case that the third-order term (a in our case) is small [68, 69, 70]. Weakly first-order means there exists a small but noticeable discontinuity in C_V. Moncton, Axe, and DiSalvo predicted a weakly first-order transition, and found a lack of evidence for a strongly first-order transition experimentally, so deduced a small a term via McMillan's argument [71, 100]. They also predicted a second-order transition to a hypothetical 1Q state, although they considered such a state merely as a mathematical simplification rather than a stable CDW geometry in $NbSe_2$. Harper *et al.* inspected the heat capacity of the system and deduced that the transition is either weakly first-order or second-order [132]. In a 1980 review Rice states

that from free energy considerations the transition is second order to $\mathcal{O}\left(\varphi^4\right)$ but first order if $\mathcal{O}\left(\varphi^6\right)$ terms are kept [133].

Inspecting the ordering conditions of Equations 68 and 69 an interesting fact presents itself. Say the system is disordered, cooling towards a second-order phase transition via an increasing $r\left(T\right)$ (recalling that $r\left(T\right) < 0$ in the disordered phase). As r passes through zero the system orders to either a 1Q state, or a 3Q state if $c > -1$. Which one wins requires an examination of the free energies themselves, but for the sake of argument assume the state will be of the 3Q geometry. Just before $r = 0$ it must be the case that $r \to 0^-$. In this limit, for the case $c > -1$ and any nonzero a, the term a^2/r diverges, and the *first*-order condition is fulfilled. The second-order 3Q transition can never occur.

This leaves three options as temperature is lowered: a second-order phase transition to the 1Q geometry; a first-order phase transition to the 3Q geometry; or a weakly first-order phase transition to the 3Q geometry, in which a would-be second-order phase transition is 'intercepted' by a first-order one. The second-order transition to 3Q always occurs at a lower temperature than the first-order, although the two will be infinitesimally close in any case where the ordering is due to a decreasing $r\left(T\right)$ as opposed to a changing $a\left(T\right)$ or $b\left(T\right)$.

8.4.3 Including Strain

Once the magnitude of the electron-phonon coupling g is constrained by requiring $T_{RPA} = 33.5\,\mathrm{K}$ the only parameter in the theory is temperature. In order to make new predictions for experiment it is necessary to consider an additional degree of freedom. I will consider uniaxial strain, which one might intuitively expect to stabilize a 1Q CDW over 3Q or disorder.

To include strain it is necessary to break the sixfold symmetry which is present in both the disordered and $3Q$ states. One way to do this is by giving the bare phonons different frequencies in the different directions: physically, stretching the lattice in one direction might be expected to lower the frequency of vibrations in that direction, just as cosmological redshift decreases the energy of photons traversing an expanding universe. If we assume as a first approximation that the volume of the system is conserved in the stretch, the perpendicular direction will contract and the corresponding phonon energies will increase.

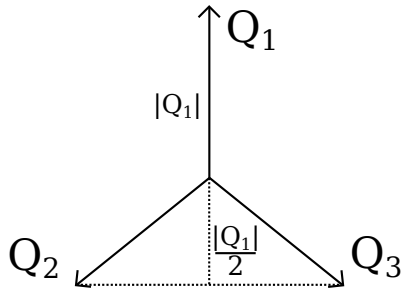


Figure 63: As the three **Q** vectors making up the 3Q state add to zero it is natural, when increasing the phonon energies by $\Omega_1 \to \Omega_1\,(1+\sigma)$ in one direction, that half the amount should be subtracted from the other two (to a first approximation): $\Omega_{2,3} \to \Omega_{2,3}\left(1-\frac{\sigma}{2}\right)$.

The uniform volume assumption corresponds to a Poisson's ratio of unity; this is likely too large, but will work as a proof of concept. If we assume further that phonon energies are linear in extension[32], the result is that if $\Omega_1 \to \Omega_1\,(1+\sigma)$ with σ the dimensionless strain parameter (positive for 'blue shift' in the cosmology analogy). The constant volume condition then leads to $\Omega_{2,3} \to \Omega_{2,3}\left(1-\frac{\sigma}{2}\right)$, shown visually in Figure 63.

The effect on the free energy is to give different quadratic coefficients in the now-inequivalent directions:

$$\begin{aligned} r_1 &= -\Omega_{\mathbf{q}}\,(1+\sigma) + D_2 \\ r_{2,3} &= -\Omega_{\mathbf{q}}\left(1-\frac{\sigma}{2}\right) + D_2. \end{aligned}$$

Since none of the $D_{2,3,4}$ diagrams contain internal phonon lines they remain unaffected.

8.5 RPA Phase Diagram

The phase diagram for $NbSe_2$ treated in the RPA as a function of temperature and strain is given in Figure 64. While the 3Q geometry is stable at zero strain it only takes around 0.1% anisotropic reduction in the bare phonon energy to break the symmetry down to 1Q. An estimate for the corresponding lattice strain would require knowledge of the in-plane Grüneisen parameter[33] $-\mathrm{d}\ln\Omega/\mathrm{d}\ln a$ with in-plane lattice parameter a. In the layered

[32] Thinking about a spring we might be tempted to assume instead that $\Omega \propto x^2$, but the spring in the analogy is under tension and not sat around the minimum energy.

[33] I am indebted to Bruno Amorim for bringing the concept of the Grüneisen parameter to my attention, and for pointing me to the listed references.

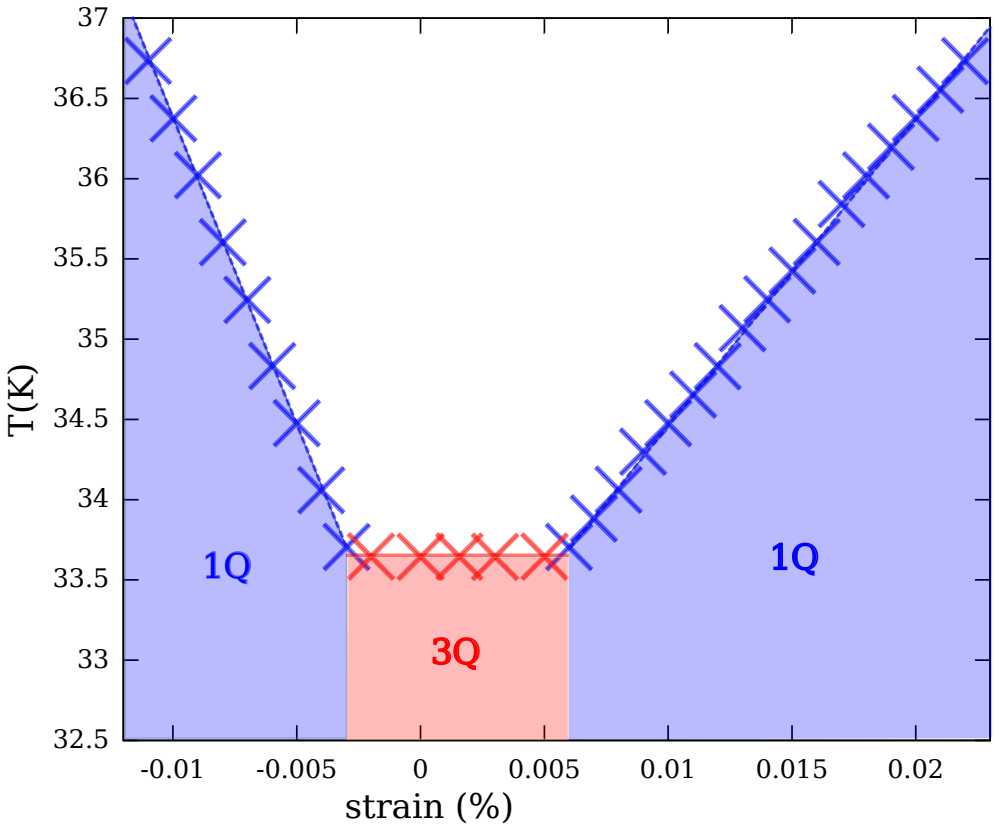


Figure 64: The RPA phase diagram as a function of temperature and strain. The one free parameter in the theory, the overall strength of the electron-phonon coupling, has been set to give $T_{CDW} = 33.5\,\mathrm{K}$ at zero strain. Dashed lines bounding the 1Q regions indicate second-order phase transitions whereas the solid line bounding 3Q indicates a weakly-first-order transition.

hexagonal materials MoS_2, BN, and graphite, the parameter is of order unity throughout most of the Brillouin zone [134, 135]. Assuming the trend holds for $NbSe_2$ the percentage change in the phonon energy can be taken to be the percentage change in the lattice spacing, *i.e.* the more standard definition of strain. In that case, to an order of magnitude, it takes around 0.1% uniaxial strain to stabilize the 1Q geometry.

The transition temperature increases with increasing strain, which is expected since the symmetry is broken in favour of stabilizing a 1Q CDW along a given direction. The slopes of the transition lines at positive and negative strain differ since it is $-\sigma/2$ which lowers the free energy at positive strain and $+\sigma$ which lowers the free energy at negative strain.

Although seemingly quite unphysical it could be the case that a 2Q state, with 1Q CDWs developing simultaneously along *two* of the $\overrightarrow{\Gamma M}$ directions but not all three, may stabilize under strain. The free energies of all three possibilities are

$$\begin{aligned}
\beta F_{3Q}[\varphi] &= -3r\varphi^2 - a\varphi^3 + \frac{3}{2}b(c+1)\varphi^4 \\
\beta F_{2Q}[\varphi] &= -2r\varphi^2 + \frac{b}{2}(c+2)\varphi^4 \\
\beta F_{1Q}[\varphi] &= -r\varphi^2 + \frac{1}{2}b\varphi^4.
\end{aligned}$$

Minimizing the 2Q and 1Q free energies (taking the nontrivial solutions $\varphi \neq 0$) reveals that

$$F_{1Q}^{min} = (c+2)\, F_{2Q}^{min}$$

and given that the free energy is negative in the ordered regime we see that the 1Q geometry is energetically favourable to the 2Q geometry for any $c > -1$. This was already a requirement for the 3Q expression to be valid (it is unbounded for $b > 0$, $c < -1$), so provided we are in the region of validity of the field theory the 1Q state is always preferred to the 2Q, and the 2Q never forms.

9 The CDW Pseudogap

In the previous chapters I have built a case for a CDW transition in $NbSe_2$ driven by a strong electron-phonon coupling dependent on the electronic bands scattered between and the ingoing and outgoing electron momenta. In the Random Phase Approximation considered in Chapter 8, the phonons themselves did not contribute to the renormalization of the phonon propagator. The RPA did an excellent job of predicting the CDW Q vector, but to get the correct temperature dependence it is necessary to add the entropy of the phonons into the free energy calculation. In this chapter I will consider such effects by the introduction of higher order diagrams. Since I will be working with a model of two independent bands I will neglect band indices from now on. The result is a range of temperatures in which the CDW order predicted by RPA is suppressed, which I will argue models the pseudogap state seen for example in ARPES [105] and STM [108].

9.1 The Mode-Mode Coupling Approximation

The simplest diagrams accounting for renormalization contributions from the phonon field constitute the so-called Mode Mode coupling Approximation, MMA [136, 130, 131]. The diagrams are given in Figure 65.

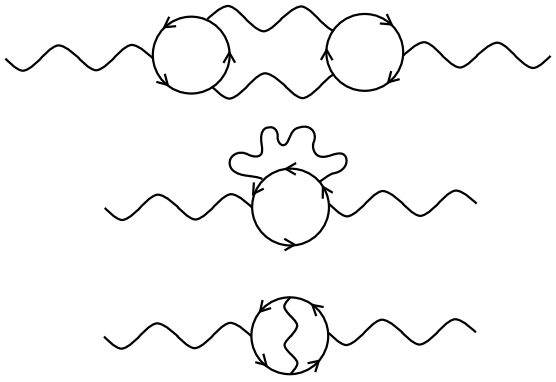


Figure 65: The Feynman diagrams constituting the MMA. The middle diagram dominates in this system.

Of these, the middle diagram dominates in our system. The top diagram can be neglected simply because it is of order g^6 where the others are order g^4. The reason it is sometimes considered (for example in [130] but not in [131]) is that in some systems Landau damping [137] suppresses a factor of g^2. This is the case for example in spin density waves in

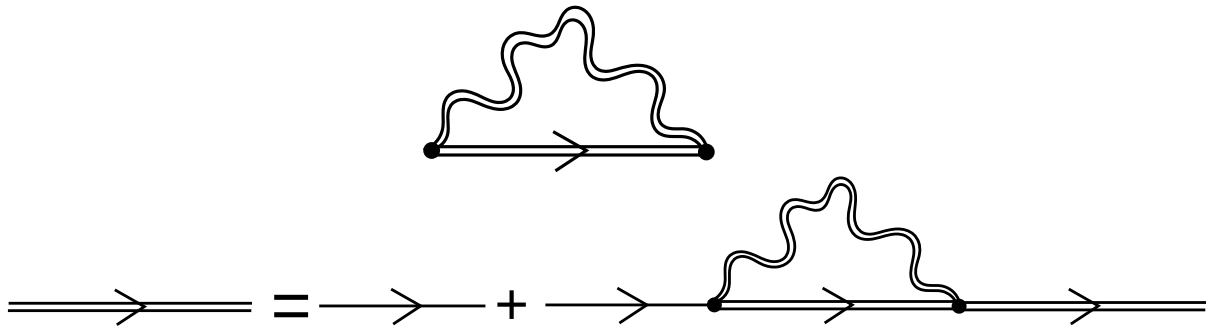


Figure 66: Top: the Feynman diagram corresponding to the electron self-energy. The double lines indicate renormalized propagators (in the case of the phonon line the RPA renormalized value). Bottom: the renormalized electron propagator is defined self-consistently through a Dyson series.

2D metals [138]. The internal phonon line in the bottom diagram constitutes a vertex correction, and can be dismissed by appeal to Migdal's theorem, which states that such corrections are suppressed by an additional factor of $\sqrt{electron\,mass/ion\,mass}$ [139, 10]. This dismissal is legitimate if the momentum transfer is large. Since the contribution is dominated by phonons with momenta $|\mathbf{Q}_{CDW}| \approx \frac{2}{3}\left|\overrightarrow{\Gamma M}\right|$ the appeal is well-founded[34].

The middle diagram can be thought of as the RPA renormalization but with the self-energy of one of the electrons included. The self-energy and its effect on the electron propagator are given diagrammatically in Figure 66.

The corresponding analytic expression for the self-energy is

$$\Sigma\left(\mathbf{k}, i\omega_n\right) \triangleq -\sum_{\mathbf{q}, i\Omega_n} \left|g_{\mathbf{k},\mathbf{k}-\mathbf{q}}\right|^2 G\left(\mathbf{k}-\mathbf{q}, i\omega_n - i\Omega_n\right) D\left(\mathbf{q}, i\Omega_n\right) \tag{70}$$

where the negative sign is included so that the real part of the expression acts as a positive energy in the Green's function (Equation 72). The phonon propagator D can either be bare or the RPA renormalized expression

$$D\left(\mathbf{q}, i\Omega_n\right) = \frac{-2\Omega_{\mathbf{q}}}{\left(i\Omega_n\right)^2 - \Omega_{\mathbf{q}}^2 + \Omega_{\mathbf{q}} D_2\left(\mathbf{q}, i\Omega\right)}. \tag{71}$$

Relabelling the bare electron Green's function as G_0 to avoid confusion, the renormalized expression G takes a particularly simple form:

[34] I wish to thank Prof. A. V. Chubukov for making these points clear to me while walking between talks at the APS March Meeting.

$$\begin{aligned} G &= G_0 + G_0 \Sigma G \\ \therefore\ G^{-1} &= \left(G_0^{-1} - \Sigma\right)^{-1} \end{aligned}$$

and so

$$G\left(\mathbf{k}, i\omega_n\right) = \left(i\omega_n - \xi_{\mathbf{k}} - \Sigma_k + \mu\right)^{-1}. \tag{72}$$

The effect of the self-energy on the electrons is most clearly seen by considering the 'spectral function'

$$A\left(\mathbf{k}, \epsilon\right) \triangleq -\frac{1}{\pi} \Im \mathfrak{m}\left(G\left(\mathbf{k}, \epsilon + i\delta\right)\right) \tag{73}$$

where a Wick rotation has been carried out from Matsubara frequency ω_n to energy ϵ with an infinitesimal imaginary part: $i\omega_n \rightarrow \epsilon + i\delta$. The spectral function can be thought of as the probability of finding a quasi-electron with energy ϵ and crystal momentum $\mathbf{k}$ [128, 10]. For the bare Green's function G_0 it is simply given by

$$A\left(\mathbf{k}, \epsilon\right) = -\frac{1}{\pi} \frac{\delta}{\left(\epsilon - \xi_{\mathbf{k}} + \mu\right)^2 + \delta^2}.$$

I prove in Appendix C that this is a delta function (in the limit $\delta \rightarrow 0^+$) located at the on-shell condition $\epsilon = \xi_{\mathbf{k}} - \mu$. Splitting the self-energy into its real and imaginary parts

$$\Sigma_{\mathbf{k}} = \Sigma'_{\mathbf{k}} + i\Sigma''_{\mathbf{k}}$$

the renormalized spectral function is given by

$$A\left(\mathbf{k}, \epsilon\right) = -\frac{1}{\pi} \frac{\delta - \Sigma''_{\mathbf{k}}}{\left(\epsilon - \xi_{\mathbf{k}} - \Sigma'_{\mathbf{k}} + \mu\right)^2 + \left(\delta - \Sigma''_{\mathbf{k}}\right)^2}$$

where the self-energy's effect on the bare electrons is seen to be twofold: the real part $\Sigma'_{\mathbf{k}}$ shifts the energy $\xi_{\mathbf{k}}$, and the imaginary part $\Sigma''_{\mathbf{k}}$ broadens the delta function into a Lorentzian. The process is clear from a physical point of view: before interaction the electron is a well-defined and infinitely long-lived quasiparticle on the mass shell. When

the interaction is turned on the resultant quasiparticles have a finite lifetime, which smears the energy distribution into a Lorentzian of finite width. There is also a shift in the energy from taking into account the contributions from the phonon field. I also show in Appendix C that

$$\int_{-\infty}^{\infty} A\left(\mathbf{k},\epsilon\right) \mathrm{d}\epsilon = \mathrm{sign}\left[\mathfrak{Im}\left(\Sigma\right) - \delta\right]$$

which constitutes one of Mahan's sum rules [128]. Physically the statement is that although the energy of the quasiparticle may have shifted, and become less sharply defined, the total probability of finding the quasiparticle across all energies is still unity[35]. Mathematically the sum rule provides a useful check that enough energies are covered in numerical sums. For the purposes of evaluating the self-energy via Equation 70 it is convenient to use the so-called *Lehmann representation* of the interacting Green's function [140, 128], proven in Appendix D:

$$G\left(\mathbf{k},i\omega_n\right) = -\frac{1}{\pi}\int \mathrm{d}\epsilon \frac{\mathfrak{Im}\left(G\left(\mathbf{k},\epsilon+i\delta\right)\right)}{i\omega_n - \epsilon} = \int \mathrm{d}\epsilon \frac{A\left(\mathbf{k},\epsilon\right)}{i\omega_n - \epsilon}. \tag{74}$$

The advantage of Equation 74 is that the locations of the Green's function's poles[36] in the complex plane are apparent, which is not true of Equation 72 when the self-energy is only known numerically. Note that as $\Sigma \rightarrow 0$ the spectral function reduces to a delta function $A\left(\mathbf{k},\epsilon\right) \rightarrow \delta\left(\epsilon - \xi_{\mathbf{k}} + \mu\right)$ so the Lehmann representation reduces correctly to the expression for the bare Green's function.

The Matsubara sum for the self-energy is also carried out in Appendix C. The result with RPA-renormalized internal phonons is

$$\begin{aligned}\Sigma\left(\mathbf{k},\epsilon+i\delta\right) &= -\frac{1}{\pi}\sum_{\mathbf{q}} \left|g_{\mathbf{k},\mathbf{k}-\mathbf{q}}\right|^2 \left(\frac{\Omega_{\mathbf{q}}}{\Omega_{RPA}}\right) \int \mathrm{d}\epsilon' \mathfrak{Im}\left[G\left(\mathbf{k}-\mathbf{q},\epsilon'+i\delta\right)\right] \cdot \\ &\quad \cdot \left\{\frac{n_B\left(\Omega_{RPA}\right)+1-f\left(\epsilon'\right)}{\epsilon-\epsilon'-\Omega_{RPA}+i\delta} + \frac{n_B\left(\Omega_{RPA}\right)+f\left(\epsilon'\right)}{\epsilon-\epsilon'+\Omega_{RPA}+i\delta}\right\}.\end{aligned} \tag{75}$$

To calculate the generalized susceptibility including the self-energy, which I will label

[35] The sign difference amounts to inspecting the quasiparticle/quasihole states, or the retarded/advanced Green's function.

[36] Strictly the poles condense into a branch cut if ϵ is continuous.

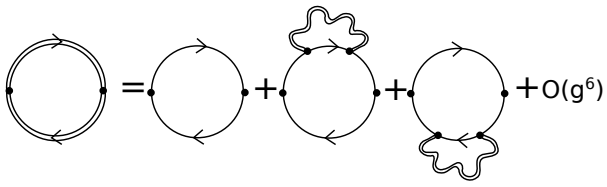


Figure 67: The generalized susceptibility including electron self-energy. The diagrams with phonon loops contribute equally.

D_2^{Σ}, both electron lines in the D_2 diagram must be renormalized separately (Figure 67), although these two diagrams turn out to be identical.

The resulting expression for D_2^{Σ} is

$$D_2^{\Sigma}(\mathbf{q}, i\Omega) = -\frac{1}{\pi}\sum_{\mathbf{k}} |g_{\mathbf{k},\mathbf{k}-\mathbf{q}}|^2 \int \mathrm{d}\epsilon \mathfrak{Im}\left(G(\mathbf{k}, \epsilon + i\delta)\right) \mathfrak{Re}\left[\frac{f(\epsilon) - f(\xi_{\mathbf{k}-\mathbf{q}})}{\epsilon - \xi_{\mathbf{k}-\mathbf{q}} - i\Omega}\right] \tag{76}$$

which reduces to the usual D_2 expression, Equation 55, as $\Sigma \to 0$.

At zeroeth order the self-consistent calculation for D_2^{Σ} of Equation 76 contains three nested sums. Each additional order introduces three additional sums through Equation 75. First order captures the physics of the problem, and given processing constraints I restrict attention to this approximation. The calculation would converge perfectly in this one step given the perfect seed, since such a seed would sit at a non-trivial fixed point of the flow of iterations. This point is returned to in Chapter 10 when the self-consistent gap is calculated. The reason I postpone a full discussion is that the argument is much neater in the case of the gap, which is effectively real-valued. The seed choice for the self-energy, $\Sigma^{(0)}$, suffers from the fact that the self-energy is complex, meaning each iteration of the calculation takes a point in $\mathbb{C}$ to another point in $\mathbb{C}$ for each wavevector. Nevertheless, a choice of $\Sigma^{(0)} = 7\,\mathrm{meV}$ (purely real) lies in close proximity to the fixed point in $\mathbb{C}$ at the range of $\mathbf{k}$ points I tested. Whenever the self-consistent calculation gave an unphysical result (the value of $D_2^{\Sigma}(\mathbf{q})$ jumping up at an isolated $\mathbf{q}$, for example) I was able to regularize by taking instead $\Sigma^{(0)} = 2\,\mathrm{meV}$ at that point.

9.2 MMA Renormalized Phonon Spectrum

Figure 68 shows the renormalized phonon frequency in both the Mode-Mode coupling and Random Phase Approximations. The coupling strength g has this time been set to give the MMA phase transition at 33.5 K, the transition temperature seen in experiment. Strictly

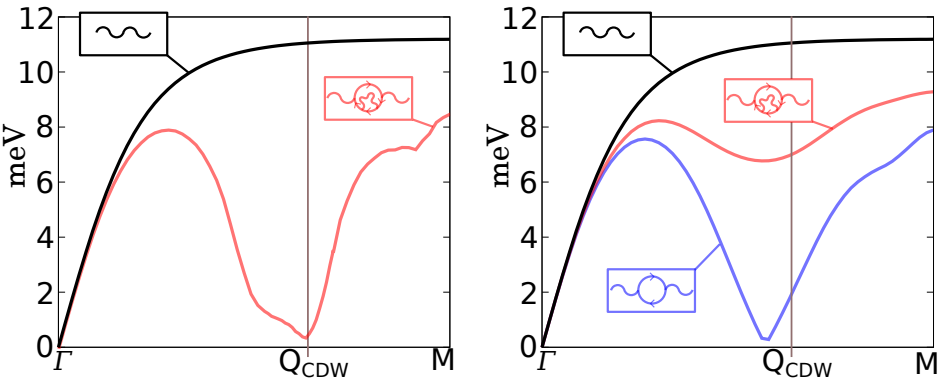


Figure 68: The renormalized phonon dispersions in the MMA (red) and RPA (blue). Left: $T_{MMA} = 33.5\,\mathrm{K}$, but g is set slightly too small so that the dispersion has not broken down; right: $T_{RPA} = 303.5\,\mathrm{K}$ at this coupling. In the left image the RPA phonon dispersion is deep into the ordered phase so is not shown. The experimentally observed ordering vector is marked. *cf.* Figure 59 which shows the RPA result where it is T_{RPA} which has been set instead to 33.5K through a different choice of the magnitude of g.

g has been set slightly smaller than the required value to give a more intuitive plot - at the transition itself the dispersion goes imaginary around $\mathbf{Q}_{CDW}$. Also shown are the same dispersions at 303.5 K where the RPA prediction for the transition temperature lies (at this coupling). It is clear that the MMA suppresses the renormalization of the phonon frequency relative to the RPA.

Setting g precisely to give $T_{MMA} = 33.5\,\mathrm{K}$, rather than the near miss shown in Figure 68, actually results in $T_{RPA} = 416.4\,\mathrm{K}$; the MMA phonon frequencies are a nonlinear function of g and a small change in the coupling can cause a large change in ordering temperature. The physical interpretation is as follows: the RPA predicts an ordering temperature of around 400 K. The system attempts to order at this temperature, with the amplitude of the phonon field taking a non-zero value $\sqrt{\left\langle |\varphi|^2 \right\rangle} > 0$. However, φ is a complex quantity, and fluctuations of the phase suppress the order itself: $\langle \varphi \rangle = 0$. These fluctuations are taken into account by the MMA. In fact the system has to be cooled to 33.5 K before the fluctuations are damped sufficiently to allow the system to order, giving $\langle \varphi \rangle > 0$ at the true, MMA, transition temperature. A simple phase diagram is given in Figure 69 to illustrate the point.

This theory is in complete agreement with the recent results of X-ray scattering, ARPES, and tunneling (STS) measurements as reported in [109]. Specifically the paper notes the following: a broadening of the peak in the Fourier transform of the STM signal above

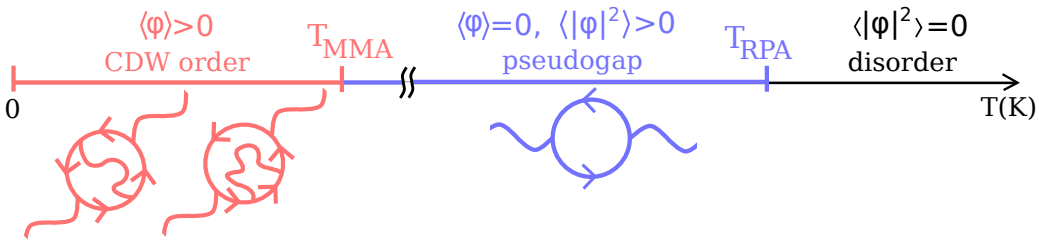


Figure 69: The phase diagram of $NbSe_2$. The system attempts to order when the RPA predicts a phase transition at $T_{RPA} = 416.4\,\mathrm{K}$ (blue). In fact phonon fluctuations suppress the order down to the the MMA phase transition at $T_{MMA} = 33.5\,\mathrm{K}$ (red). The intermediate regime is labelled *pseudogap*.

T_{CDW} indicating a shift from long range order $\langle\varphi\rangle > 0$ to short range order $\sqrt{\left\langle|\varphi|^2\right\rangle} > 0$; the loss of the sharp edge in the tunneling profile but the persistence of a reduced DOS above T_{CDW}; and the loss of a coherent peak in ARPES intensity above T_{CDW} but the persistence of an energy gap. The persistent features are related to the idea of a *pseudogap* in Section 9.2.1.

Reference [109] also gives the same theoretical interpretation as I have provided here. In fact, the idea of phase incoherence suppressing CDW order was proposed for 1D CDW systems as early as 1973 [141]. The same idea was used by McMillan in 1977 to explain discrepancies between theory and experiment in the closely related system of $TaSe_2$ [70].

9.2.1 The Pseudogap Regime

In the phase diagram of Figure 69 the intermediate regime between disorder and CDW order is labelled the 'pseudogap'. The properties of pseudogaps are a controversial topic. The universally agreed upon feature is a reduction in the observed density of states for a range of temperatures above the ordering temperature [110]. Of course we cannot tell without a firmer theoretical understanding whether the density of states is actually reduced or whether experimental probes are simply unable to detect the states. The seemingly reduced DOS occurs generically in hole-doped cuprate superconductors [93]. In many of these systems fluctuating stripes of charge order are observed to develop simultaneously with the reduced DOS, although it is debated whether the two always coincide and, if they do, which is the cause and which the effect [93, 110, 142, 143, 144, 145]. It is further debated whether it is fluctuations of the stripes which link with the pseudogap or whether the static stripe structure itself is the relevant property [5]. The term pseudogap *regime*

(as opposed to 'state' or 'phase') tends to be used because it is an important open question as to whether the pseudogap constitutes a phase of matter - that is, whether there exists long-range order which can be characterized by a Landau order parameter.

In $NbSe_2$ a reduced density of states is seen over a range of about 35 meV around E_F, starting below T_{CDW} and showing no signs of abating by 119 K [105]. It is the prevailing view at present that this does not constitute a true CDW gap[37], but rather a pseudogap similar to that seen in the layered high-T_C compounds [109, 105].

In light of the previous discussion I am confident in making the following claims regarding the origin of the pseudogap in $NbSe_2$:

1. the pseudogap is an expression of a nonzero CDW amplitude, which can be quantified by short-range fluctuations of the phonon field $\sqrt{\left\langle |\varphi|^2 \right\rangle}$.

2. As a corollary of (1), there is no long-range order parameter characterizing the pseudogap, so it does not form a new state of matter in the Landau symmetry breaking sense.

3. As a second corollary of (1), it is indeed fluctuations of CDW order which characterize the pseudogap state rather than the static stripe structure. This is clear in $NbSe_2$ since it is the ordered state which is characterized by a static structure.

4. Phase incoherence between the phonon modes suppresses true CDW order down to $T_{CDW} = 33.5$ K.

5. The pseudogap will disappear at around 400 K.

As further evidence for these claims, real-space STM imaging has detected the formation of small, uncorrelated regions of CDW order nucleating around surface defects at 96 K [108]. The defects stabilize a given CDW phase, breaking the $\langle\varphi\rangle = 0$ symmetry locally (but in an uncorrelated manner, so maintaining the symmetry globally). When the temperature is decreased to T_{CDW} long range order has set in and the islands of stability become correlated.

[37] In fact the 35 meV suppression is most likely an artefact of STS, with the increase in DOS outside this range due to phonon contributions [109]. Regardless, there is a consensus that some form of reduced electronic DOS persists above T_{CDW}.

Points $(1-4)$ are in accord with existing experimental results as described above. Point (5) provides a further test of the theory. To date, the highest-temperature measurement of the $NbSe_2$ pseudogap is the 119 K result of Borisenko *et al.* [105]. At that temperature the gap magnitude has nearly doubled relative to that at T_{CDW}, making the pseudogap's fate at high temperatures unclear. It should be noted however that at present it seems unlikely that the pseudogap could be probed at temperatures as high as 400 K.

It remains an open question as to whether the theory I have given for the $NbSe_2$ pseudogap is relevant to the story of the pseudogap in the the high-T_C superconductors. In the high-T_C literature [110] this story is known as the 'pre-formed pairs theory': Cooper pairs form at a high temperature T^* (*cf.* T_{RPA}) but have uncorrelated phases. As the temperature approaches T_C (in our case $T_{MMA} \equiv T_{CDW}$) the pairs become coherent, and long-range order develops. There is certainly no consensus as to whether this is the correct description of the cuprate pseudogap.

9.3 MMA Phase Diagram

9.3.1 Free Energy Including Strain

The MMA diagrams act to further renormalize the phonon 2-point correlator (propagator) beyond the RPA renormalization. They have no effect on the 3- and 4-point correlators, so these remain the same in the free energy expression. Recalling the definitions from Equation 67 of Chapter 8 we have

$$\begin{aligned} \beta F_{3Q}[\varphi] &= -3r\varphi^2 - a\varphi^3 + \frac{3}{2}b(c+1)\varphi^4 \\ \beta F_{1Q}[\varphi] &= -r\varphi^2 + \frac{1}{2}b\varphi^4 \end{aligned}$$

with

$$\begin{aligned} r_j &= -\left(\Omega_{\mathbf{q}}\left(1+\tfrac{\sigma}{2}\left(3\delta_{j1}-1\right)\right) - D^{\Sigma}_{2j}(\sigma)\right) \qquad j \in [1,3] \\ a &= 2\left(D_{3a}+D_{3b}\right) \\ b &= D_{4a}+2D_{4b} \\ bc &= 4\left(D_{4d}+2D_{4c}\right). \end{aligned}$$

Strain, σ, again only affects the quadratic term r, which now gains an asymmetry between the three $\overrightarrow{\Gamma M}$ directions (although two remain the same). Unlike in the RPA the renormalizing term is itself affected by the strain, since it now includes an internal phonon line. The required generalization of D_2^{Σ} from Section 9.1 is

$$\begin{aligned} D_{2j}^{\Sigma}\left(\mathbf{q}, i\Omega\right) &= -\frac{1}{\pi}\sum_{\mathbf{k}}\Bigg(\left|g_{\mathbf{k},\mathbf{k}-\mathbf{q}}\right|^2 \cdot \\ &\quad \cdot \int \mathrm{d}\epsilon \Im\mathfrak{m}\left(G_0\left(\mathbf{k}, \epsilon + i\delta\right)^{-1} - \Sigma_j\left(\mathbf{k}, \epsilon + i\delta, \sigma\right)\right)^{-1} \Re\mathfrak{e}\left[\frac{f\left(\epsilon\right) - f\left(\xi_{\mathbf{k}-\mathbf{q}}\right)}{\epsilon - \xi_{\mathbf{k}-\mathbf{q}} - i\Omega}\right]\Bigg) \end{aligned}$$

with a different self-energy in the three CDW directions, j, as before:

$$\begin{aligned} \Sigma_j\left(\mathbf{k}, \epsilon + i\delta, \sigma\right) &= -\frac{1}{\pi}\sum_{\mathbf{q}}\Bigg(\left|g_{\mathbf{k},\mathbf{k}-\mathbf{q}}\right|^2 \frac{\Omega_{\mathbf{q}}}{\Omega_{RPA}}\left(1 + \frac{\sigma}{2}\left(3\delta_{j1} - 1\right)\right) \cdot \\ &\quad \cdot \int \mathrm{d}\epsilon' \Im\mathfrak{m}\left[G_0\left(\mathbf{k} - \mathbf{q}, \epsilon' + i\delta\right)\right]\left\{\frac{n_B\left(\Omega_{RPA}\right) + 1 - f\left(\epsilon'\right)}{\epsilon - \epsilon' - \Omega_{RPA} + i\delta} + \frac{n_B\left(\Omega_{RPA}\right) + f\left(\epsilon'\right)}{\epsilon - \epsilon' + \Omega_{RPA} + i\delta}\right\}\Bigg). \end{aligned}$$

The calculation otherwise remains the same as in the RPA case.

9.3.2 The Phase Diagram

The phase diagram as a function of temperature and strain is given in Figure 70. This time both the RPA and MMA transitions are included. In Section 8.5 of Chapter 8 a similar diagram was produced with the magnitude of the electron-phonon coupling g set to give $T_{RPA} = T_{CDW} = 33.5\,\mathrm{K}$. This time g is chosen so as to give $T_{MMA} = T_{CDW} = 33.5\,\mathrm{K}$. In this case at zero strain T_{RPA} occurs at around $416\,\mathrm{K}$.

A brief comment on the validity of the method is in order. In either the MMA or RPA, the field theory is a perturbative expansion around a disordered Gaussian fixed point. Lowering temperature, as soon as the order parameter $\varphi_{\mathbf{q}}$ becomes non-zero at any point in the Brillouin zone a structural instability is triggered, and the system develops CDW order. The original theory becomes invalid, and a new theory must be constructed consisting of an expansion about the new Gaussian fixed point (which is then again treated as 'disordered' with respect to a new order parameter). In constructing Figure 70 I have used a much

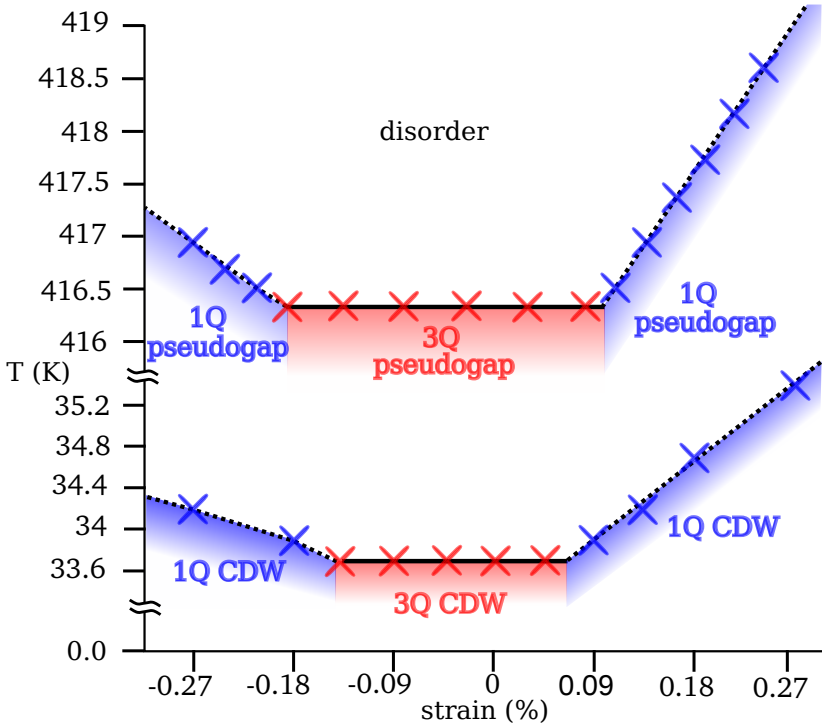


Figure 70: The CDW phase diagram in both the RPA and MMA as a function of temperature and uniaxial strain. Cooling from high temperature, the initially disordered system develops fluctuating islands of stability when the RPA predicts a phase transition. These islands have a 3Q character at zero strain, but it takes only around 0.1% anisotropic change in the phonon energies to stabilize the 1Q geometry. As the system cools further the true CDW phase transition occurs, at the MMA transition. As in RPA the geometry of the CDW is 3Q type, but as little as 0.1% strain stabilizes the 1Q state. The MMA includes internal phonon contributions to the electron self-energy. These are themselves affected by the strain, with the result being that the 1Q state stabilizes at lower strain in the gapped state than the pseudogap regime. Dashed black lines indicate second-order phase transitions whereas solid black lines indicate weakly-first-order transitions.

more naïve approach, considering the two field theories, RPA and MMA, independently. Strictly only the phase transition lines appearing in the figure can be trusted. According to RPA, CDW order develops below T_{RPA}, but nothing more can be said about lower temperatures. The MMA makes the same prediction about T_{MMA}. The statement that the range of temperatures between these two transitions corresponds to a pseudogap state, where phonon fluctuations suppress the CDW order, is therefore an additional physical assertion, but a sensible one given that it was the basis for the original choice of the diagrams making up the MMA (the lowest-order diagrams including the entropy of the phonon field) [136, 130, 131].

Cooling from high temperature at zero strain the system undergoes a weakly first-order phase transition at T_{RPA} to the pseudogap regime characterized by fluctuating short-range CDW order. The fluctuations of the phonon field suppress true CDW order down to 33.5 K, at which point the ordered state develops again via a weakly first-order phase transition. In both transitions it takes only around 0.1% anisotropic shift in the phonon energies, corresponding to a similar percentage lattice distortion, to stabilize the 1Q geometry. There is an experimental precedent for this in the ordered regime, in the form of STM measurements on the surface of $NbSe_2$ which show domains of 1Q order believed to be stabilized by local strain effects. These effects have an estimated upper bound of around 0.45% lattice strain [102] in agreement with the results of this model. Similar measurements on the geometry of the fluctuating CDWs in the pseudogap regime have not been carried out.

In MMA, D_2^{Σ} itself depends on strain via the internal phonons in the electronic self-energy. The result is a reduction in the range of strains over which the 3Q geometry is stable, as evident in Figure 70. The phase transition to the 1Q state in both the RPA and MMA cases is second-order, which constitutes a testable prediction of the theory.

10 The CDW Gap Equation

In Chapter 8 I developed a quantum field theoretical description of $NbSe_2$ using the Random Phase Approximation. Strictly the theory holds only in the disordered regime, $T > T_{RPA}$. In Chapter 9 I extended the model to include fluctuations of the phonon field via the Mode-Mode Coupling Approximation, and deduced a range of temperatures over which a pseudogap regime exists, $T_{MMA} < T < T_{RPA}$. In this chapter I will extend the field theory to look into the ordered state itself.

10.1 The Nambu-Gor'kov Method

10.1.1 Use in Superconductivity

Perturbative QFT, in essence, consists of finding a stable fixed point of a physical system (the mean field solution) then perturbing about this fixed point in powers of a small parameter (in this case the electron-phonon coupling g). So far I have worked entirely in the disordered regime. With the methods employed so far, I was able to demonstrate the renormalization of quantities such as the phonon dispersion. As soon as the energy of a $\mathbf{q} \neq \mathbf{0}$ phonon hits zero, however, a permanent lattice distortion is induced. The symmetry of the system is reduced, and the disordered fixed point is no longer the lowest energy fixed point in the problem. A new field theory should be constructed by making perturbations about the new fixed point.

There is, however, an ingenious method to establish properties of the ordered phase using the field theory of the disordered phase. The method was developed in the context of superconductivity by Nambu [146] and Gor'kov [147], building on foundations laid by Bogoliubov [148], and was popularized in the West by de Gennes [127]. I will give a brief account of the superconductivity case before generalizing the method to the study of charge density waves.

The BCS superconductivity story is these days well-known [149]. The two facts relevant here are:

1. electrons of opposite spin pair to form composite bosons

2. the superconducting condensate is not an eigenstate of the particle number operator.

First I will define Green's functions for the particles and holes separately:

$$\begin{aligned} G_0 &= \left\langle \psi_{k\sigma} \psi^\dagger_{k\sigma} \right\rangle = (i\omega_n - \xi_k)^{-1} \\ \tilde{G}_0 &= \left\langle \psi^\dagger_{k\sigma} \psi_{k\sigma} \right\rangle = (i\omega_n + \xi_k)^{-1} \end{aligned}$$

with σ a spin index[38], although the two spins behave identically. Note the particle-hole symmetry, absent from the case of CDWs except at half-filling. I will additionally define particle number non-conserving *anomalous Green's functions*

$$\begin{aligned} F^\dagger &= \left\langle \psi^\dagger_{k\uparrow} \psi^\dagger_{k\downarrow} \right\rangle \\ F &= \langle \psi_{k\uparrow} \psi_{k\downarrow} \rangle \,. \end{aligned}$$

The operator $F^\dagger$ creates a Cooper pair, and F annihilates one. Before the introduction of the F operators the electron self-energy renormalizes the electron Green's function as

$$G = G_0 + G_0 \Sigma G \tag{77}$$

but after the introduction we get the additional term

$$G = G_0 + G_0 \Sigma G + G_0 \Delta F^\dagger$$

shown diagrammatically in Figure 71. The symbol Δ represents the superconducting gap function, and in general is a complex-valued function of momentum and energy. I will take it to be a constant scalar for the sake of this exposition.

Also shown in Figure 71 are the diagrams which can be constructed out of the same elements. In equation form:

[38] I maintain spin indices should be called *spindices*. Since I will never get this past an editor am I stating it here.

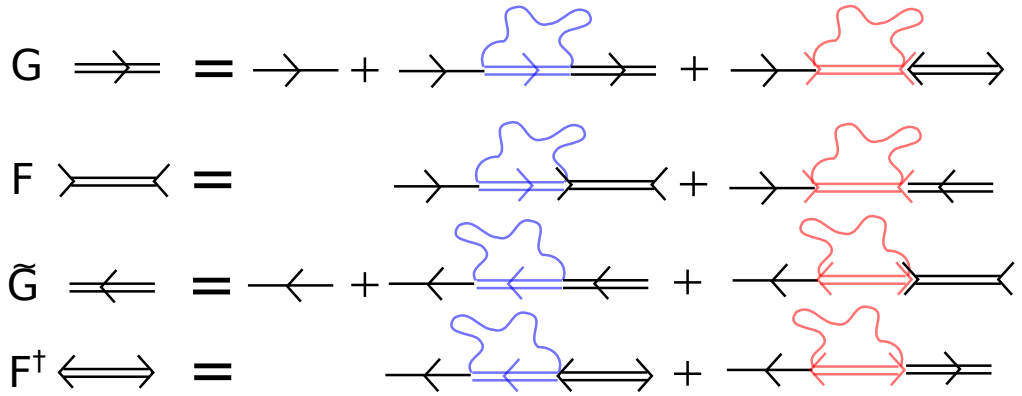


Figure 71: Nambu-Gor'kov Feynman diagrams including the pair creation/annihilation operators in superconductivity. In blue is the self-energy Σ, in red the gap function Δ. The vertex symbols have been omitted for clarity. The double-headed arrows do not conserve particle number.

$$\begin{aligned} G &= G_0 + G_0\Sigma G + G_0\Delta F^\dagger \\ F &= G_0\Sigma F + G_0\Delta\tilde{G} \\ \tilde{G} &= \tilde{G}_0 + \tilde{G}_0\tilde{\Sigma}\tilde{G} + \tilde{G}_0\Delta^* F \\ F^\dagger &= \tilde{G}_0\tilde{\Sigma}F^\dagger + \tilde{G}_0\Delta^* G. \end{aligned}$$

These can be recast as a simple self-energy equation in terms of a 2×2 matrix:

$$\begin{pmatrix} G & F \\ F^\dagger & \tilde{G} \end{pmatrix} = \begin{pmatrix} G_0 & 0 \\ 0 & \tilde{G}_0 \end{pmatrix} + \begin{pmatrix} G_0 & 0 \\ 0 & \tilde{G}_0 \end{pmatrix} \begin{pmatrix} \Sigma & \Delta \\ \Delta^* & \tilde{\Sigma} \end{pmatrix} \begin{pmatrix} G & F \\ F^\dagger & \tilde{G} \end{pmatrix}. \tag{78}$$

Note that Equation 78 is simply Equation 77 with the various symbols now representing 2×2 matrices. Setting the self-energy to zero for simplicity, the $(1,1)$ element of the Green's function matrix - the propagator for the 'Bogoliubov quasiparticle' in the ordered state - can be evaluated using the usual Dyson series re-arrangement of Equation 78 in matrix form. The result is

$$G = \frac{i\omega_n + \xi_{\mathbf{k}}}{(i\omega_n)^2 - \xi_{\mathbf{k}}^2 - |\Delta|^2}.$$

Thus we have a Green's function for quasiparticles valid in the ordered regime, which reduces smoothly to the bare electron propagator as the gap Δ vanishes above T_C. This

is the essence of the Nambu-Gor'kov method.

10.1.2 Generalization to Charge Order

In a commensurate CDW system with rational period $2\pi n$ in real space we instead have $n \times n$ matrices of the form

$$G_0^{nn} = \left\langle \psi_{k+nQ} \psi_{k+nQ}^\dagger \right\rangle = (i\omega - \xi_{k+nQ})^{-1}.$$

Unlike in the BCS theory, in CDW systems the particle number is still well-defined. However, crystal momentum is not, since the increase in the real space lattice size induced by the CDW transition causes a reduction of the Brillouin zone, making states $\mathbf{k}$ and $\mathbf{k}+\mathbf{Q}_{CDW}$ equivalent (discussed at length in Section 3.2). Accordingly I will define crystal momentum non-conserving operators

$$\begin{aligned} F_m^n &\triangleq \left\langle \psi_{k+nQ} \psi_{k+mQ}^\dagger \right\rangle \\ F_m^{n\dagger} &\triangleq \left\langle \psi_{k+mQ}^\dagger \psi_{k+nQ} \right\rangle. \end{aligned}$$

For $NbSe_2$ we have an incommensurate system with three different ordering directions, requiring infinite-dimensional matrices. I will approximate these to have the structure of commensurate matrices by setting $\delta = 0$ in the CDW vectors $\mathbf{Q} = (1-\delta)\frac{2}{3}\overrightarrow{\Gamma M}$. In fact δ is irrational and close to 0.104. Writing the rows and columns of the gap matrix as $\mathbf{k} \pm \mathbf{Q}_i$, $i \in [1,3]$ would give a 7×7 matrix. Naïvely the gap matrix would be:

$$\underline{\underline{\Sigma}} \triangleq \begin{pmatrix} \Sigma_k & \Delta_{k+Q_1}^k & \Delta_{k-Q_1}^k & \Delta_{k+Q_2}^k & \Delta_{k-Q_2}^k & \Delta_{k+Q_3}^k & \Delta_{k-Q_3}^k \\ \Delta_k^{k+Q_1} & \Sigma_{k+Q_1} & 0 & 0 & 0 & 0 & 0 \\ \Delta_k^{k-Q_1} & 0 & \Sigma_{k-Q_1} & 0 & 0 & 0 & 0 \\ \Delta_k^{k+Q_2} & 0 & 0 & \Sigma_{k+Q_2} & 0 & 0 & 0 \\ \Delta_k^{k-Q_2} & 0 & 0 & 0 & \Sigma_{k-Q_2} & 0 & 0 \\ \Delta_k^{k+Q_3} & 0 & 0 & 0 & 0 & \Sigma_{k+Q_3} & 0 \\ \Delta_k^{k-Q_3} & 0 & 0 & 0 & 0 & 0 & \Sigma_{k-Q_3} \end{pmatrix}$$

that is, the state $\mathbf{k}$ couples to each of $\mathbf{k} \pm \mathbf{Q}_i$ and no other couplings occur. Note that $\Delta^{k+Q_n}_{k+Q_m} = \left(\Delta^{k+Q_m}_{k+Q_n}\right)^*$. There are two errors with the matrix, both stemming from symmetry constraints. With a CDW wavevector of $\mathbf{Q}_{CDW} = \frac{2}{3}\overrightarrow{\Gamma M}$ it follows that

$$\begin{aligned} \mathbf{Q}_1 + \mathbf{Q}_2 + \mathbf{Q}_3 &= \mathbf{0} \\ 2\mathbf{Q}_i &\equiv -\mathbf{Q}_i. \end{aligned} \tag{79}$$

This implies additional entries in the matrix: for example the diagonal entries Σ_{k-Q_1} and Σ_{k+Q_2} are coupled, since the difference in their wavevectors is $\mathbf{Q}_2 + \mathbf{Q}_1$ which is equivalent to $-\mathbf{Q}_3$. Employing a simplified notation $1 \triangleq k + Q_1$, $\overline{1} \triangleq k - Q_1$ *etc.* the matrix fills up like so:

$$\underline{\underline{\Sigma}} = \begin{pmatrix} \Sigma_k & \Delta^k_1 & \Delta^k_{\overline{1}} & \Delta^k_2 & \Delta^k_{\overline{2}} & \Delta^k_3 & \Delta^k_{\overline{3}} \\ \Delta^1_k & \Sigma_1 & \Delta^1_{\overline{1}} & 0 & \Delta^1_{\overline{2}} & 0 & \Delta^1_{\overline{3}} \\ \Delta^{\overline{1}}_k & \Delta^{\overline{1}}_1 & \Sigma_{\overline{1}} & \Delta^{\overline{1}}_2 & 0 & \Delta^{\overline{1}}_3 & 0 \\ \Delta^2_k & 0 & \Delta^2_{\overline{1}} & \Sigma_2 & \Delta^2_{\overline{2}} & 0 & \Delta^2_{\overline{3}} \\ \Delta^{\overline{2}}_k & \Delta^{\overline{2}}_1 & 0 & \Delta^{\overline{2}}_2 & \Sigma_{\overline{2}} & \Delta^{\overline{2}}_3 & 0 \\ \Delta^3_k & 0 & \Delta^3_{\overline{1}} & 0 & \Delta^3_{\overline{2}} & \Sigma_3 & \Delta^3_{\overline{3}} \\ \Delta^{\overline{3}}_k & \Delta^{\overline{3}}_1 & 0 & \Delta^{\overline{3}}_2 & 0 & \Delta^{\overline{3}}_3 & \Sigma_{\overline{3}} \end{pmatrix}$$

where the gap functions Δ are labelled with a top index according to their row (momentum scattered from) and a bottom index according to their column (momentum scattered to). Since states connected by $2\mathbf{Q}_1$ are in fact connected through the equivalence $2\mathbf{Q}_1 \equiv -\mathbf{Q}_1$ and so on, it also follows that we should include columns and rows corresponding to the momenta $\pm(\mathbf{Q}_1 - \mathbf{Q}_2)$, $\pm(\mathbf{Q}_1 - \mathbf{Q}_3)$, and $\pm(\mathbf{Q}_2 - \mathbf{Q}_3)$. In fact four of these six momenta turn out to be equivalent by the constraints of Equation 79, and we only need additional entries for (say) $\pm(\mathbf{Q}_1 - \mathbf{Q}_2)$. Including these the matrix takes the final form

$$\underline{\underline{\Sigma}} = \begin{pmatrix} \Sigma_k & \Delta & \Delta & \Delta & \Delta & \Delta & \Delta & 0 & 0 \\ \Delta & \Sigma_1 & \Delta & 0 & \Delta & 0 & \Delta & \Delta & \Delta \\ \Delta & \Delta & \Sigma_{\bar{1}} & \Delta & 0 & \Delta & 0 & \Delta & \Delta \\ \Delta & 0 & \Delta & \Sigma_2 & \Delta & 0 & \Delta & \Delta & \Delta \\ \Delta & \Delta & 0 & \Delta & \Sigma_{\bar{2}} & \Delta & 0 & \Delta & \Delta \\ \Delta & 0 & \Delta & 0 & \Delta & \Sigma_3 & \Delta & \Delta & \Delta \\ \Delta & \Delta & 0 & \Delta & 0 & \Delta & \Sigma_{\bar{3}} & \Delta & \Delta \\ 0 & \Delta & \Delta & \Delta & \Delta & \Delta & \Delta & \Sigma_{1\bar{2}} & 0 \\ 0 & \Delta & \Delta & \Delta & \Delta & \Delta & \Delta & 0 & \Sigma_{\bar{1}2} \end{pmatrix} \tag{80}$$

(the labels are this time omitted for clarity).

In Section 9.1 I calculated the self-energy of the electrons in the MMA. The result was Equation 75, and the effect on the generalized susceptibility D_2 was Equation 76. The proof relied on the Lehmann representation of the Green's function given in Equation 74 and proven in Appendix D.

Now that the Green's function has been generalized further to matrix form it is not clear that the method of Section 9.1 can be used. For example, in the MMA self-energy calculation the pole in the Green's function was shifted, but the general form of the Green's function was unaltered. In the present case the Green's function requires the numerical inversion of a 9×9 matrix, and the pole structure is completely undetermined.

In Appendix D I show that the Lehmann representation works for *any* function which is causal and normalizable. This includes the complicated form of the gapped G, and holds true at every level of the self-consistent calculation. We are therefore able to use Equations 75 and 76 even when G and Σ are promoted to matrices. Element (n, m) of the self-energy matrix is now

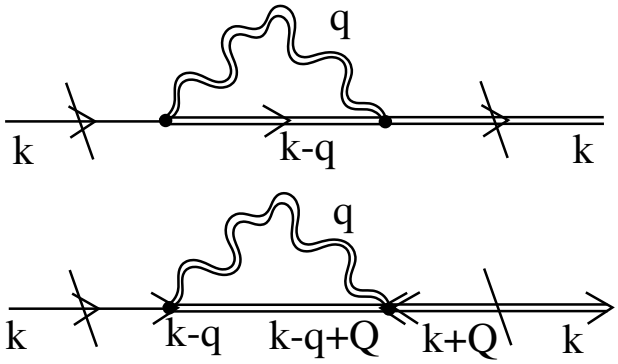


Figure 72: The Feynman diagrams for the self-energy in the Nambu-Gor'kov formalism. The top diagram is for the diagonal elements (self-energies). The slashes are used to indicate amputation, so this diagram is the same as Figure 66. The bottom diagram is for off-diagonal elements (gaps). The central fat electron line is the anomalous Green's function F^{k-q}_{k-q+Q} which does not conserve crystal momentum.

$$\begin{aligned}\Sigma^n_m\left(\mathbf{k},\epsilon+i\delta\right) &= -\frac{1}{\pi}\sum_{\mathbf{q}}\Bigg(g_{\mathbf{k}+\mathbf{Q}_n,\mathbf{k}-\mathbf{q}+\mathbf{Q}_n}g_{\mathbf{k}-\mathbf{q}+\mathbf{Q}_m,\mathbf{k}+\mathbf{Q}_m}\left(\frac{\Omega_{\mathbf{q}}}{\Omega_{RPA}}\right)\cdot \\ &\quad\cdot\int \mathrm{d}\epsilon'\mathfrak{Im}\left[G^n_m\left(\mathbf{k}-\mathbf{q},\epsilon'+i\delta\right)\right]\cdot \\ &\quad\cdot\left\{\frac{n_B\left(\Omega_{RPA}\right)+1-f\left(\epsilon'\right)}{\epsilon-\epsilon'-\Omega_{RPA}+i\delta}+\frac{n_B\left(\Omega_{RPA}\right)+f\left(\epsilon'\right)}{\epsilon-\epsilon'+\Omega_{RPA}+i\delta}\right\}\Bigg) \qquad (81)\end{aligned}$$

with the corresponding Feynman diagrams given in Figure 72. The generalized susceptibility is again given by

$$\left[D^{\Sigma}_2\left(\mathbf{q},i\Omega\right)\right]_{nm} = -\frac{1}{\pi}\sum_{\mathbf{k}}\left|g_{\mathbf{k},\mathbf{k}+\mathbf{q}}\right|^2\int \mathrm{d}\epsilon'\mathfrak{Im}\left(G_{nm}\left(\mathbf{k},\epsilon'+i\delta\right)\right)\mathfrak{Re}\left[\frac{f\left(\epsilon'\right)-f\left(\xi_{\mathbf{k}-\mathbf{q}}\right)}{\epsilon'-\xi_{\mathbf{k}-\mathbf{q}}-i\Omega}\right].$$

For the remainder of this chapter I will neglect the on-diagonal (self-energy) terms in the matrix of Equation 80. Neglecting self-energies means the calculation effectively takes place in the RPA rather than MMA, without internal phonon contributions to propagators.

10.1.3 Self-Consistent Gap Calculation

The numerical calculation of G and Δ is once again carried out self-consistently. The complex phase of the gap function plays no rôle in the physics of the system, and accordingly the phase of the seed survives as the phase of the self-consistent solution. This greatly

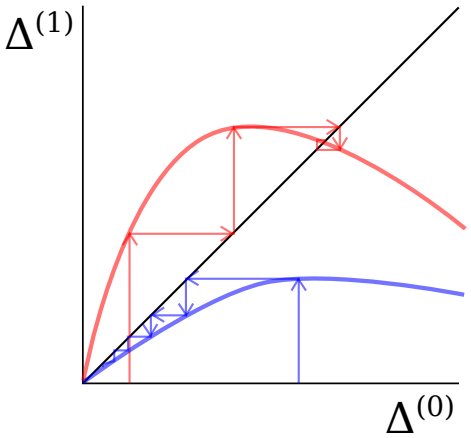


Figure 73: The intersection of the curve $\Delta^{(1)}\left(\Delta^{(0)}\right)$ with the line $\Delta^{(1)} = \Delta^{(0)}$ is a fixed point of the flow (the infinite-order solution). The solutions found in this study all take the form of either the red curve (a stable nonzero fixed point) or the blue curve (a stable fixed point at zero).

simplifies the search for a good seed value, and at a given $\mathbf{k}$ the problem can be solved exactly via an iterative method shown schematically in Figure 73. Inserting a seed $\Delta^{(0)}$, the first-order result $\Delta^{(1)}$ is calculated. In theory we could then feed this $\Delta^{(1)}$ back into the calculation to calculate $\Delta^{(2)}$, and so on. There is a neater method, though: finding $\Delta^{(1)}$ as a function of $\Delta^{(0)}$, the point at which the curve $\Delta^{(1)}\left(\Delta^{(0)}\right)$ intersects the straight line $\Delta^{(1)} = \Delta^{(0)}$ will either be a point of convergence of the series (stable fixed point) or a point of divergence (unstable fixed point).

The gap function is in general a rather complicated object: a complex-valued function of two 2D crystal momenta and two energies, $\Delta^{k+Q_n}_{k+Q_m}$. The complex phase is unimportant, and I assume the gap to be independent of energy, considering only the value at the Fermi level. This still leaves four degrees of freedom. If elements $(1, 2-7)$ of the self-energy matrix in Equation 80 turn out to be identical for the same seed this is sufficient to reduce the gap to a function of only one 2D momentum. The reason is as follows: if elements $(1, 2-7)$ are the same then $\Delta^{k}_{k+Q_1} = \Delta^{k}_{k-Q_1} = \Delta^{k}_{k+Q_2}$ *etc.*, and we can call this single function $\Delta(\mathbf{k})$; element $(2, 3)$ is $\Delta^{k+Q_1}_{k-Q_1}$, but from the two equivalence relations of Equation 79 this is equal to $\Delta^{k+Q_1}_{k+Q_1+Q_1}$ which is nothing other than $\Delta(\mathbf{k}+\mathbf{Q}_1)$. Similar reasoning holds for any off-diagonal element in the 9×9 matrix. Investigation of elements $(1, 2-7)$ reveals that they are indeed the same for all seeds at all points in the Brillouin zone which were tested, and the use of a real-valued function of one 2D momentum for Δ is justified. It would still be too computationally demanding to evaluate the gap function for all points

in the Brillouin zone. Instead I carried out a tight-binding fit to fifth-nearest neighbour, constraining the fit by finding the value of the gap function at a set of (high-symmetry) points. Defining

$$\begin{aligned} \xi &\triangleq \frac{1}{2}k_x \\ \eta &\triangleq \frac{\sqrt{3}}{2}k_y \end{aligned}$$

(units of 2π) the general tight-binding expression for an arbitrary real function to fifth-nearest neighbour is

$$\begin{aligned} \Delta_{\mathbf{k}} = \; & t_0 + t_1 \left(2\cos\left(\xi\right)\cos\left(\eta\right) + \cos\left(2\xi\right)\right) \\ & + t_2 \left(2\cos\left(3\xi\right)\cos\left(\eta\right) + \cos\left(2\eta\right)\right) \\ & + t_3 \left(2\cos\left(2\xi\right)\cos\left(2\eta\right) + \cos\left(4\xi\right)\right) \\ & + t_4 \left(\cos\left(\xi\right)\cos\left(3\eta\right) + \cos\left(5\xi\right)\cos\left(\eta\right) + \cos\left(4\xi\right)\cos\left(2\eta\right)\right) \\ & + t_5 \left(2\cos\left(3\xi\right)\cos\left(3\eta\right) + \cos\left(6\xi\right)\right) \end{aligned} \tag{82}$$

found by Fourier transforming the n^{th} nearest neighbour structure. I chose high symmetry points Γ, M, K, $\Lambda_1 \triangleq \frac{1}{2}\overrightarrow{\Gamma M}$, $\Lambda_2 \triangleq \frac{1}{3}\overrightarrow{\Gamma K}$, $\Lambda_3 \triangleq \frac{2}{3}\overrightarrow{\Gamma K}$ which focus somewhat on the K point around which the gap is seen to open in ARPES [105]. They are shown in Figure 74. Inserting them into Equation 82 gives six simultaneous equations which can be inverted analytically to give the set of t_n constrained by the set of Δ values. The result of the fit is also shown in Figure 74. In fact, of the six points tested the gap was non-zero only at the K point, which seems sensible with regard to the known experimental results.

10.2 The Spectral Function and Density of States

10.2.1 Analytic Expressions for Measurable Quantities

ARPES, as mentioned in Section 6.3, is a direct measure of the occupation of states at a given energy and wavevector. This suggests a link to the spectral function, introduced

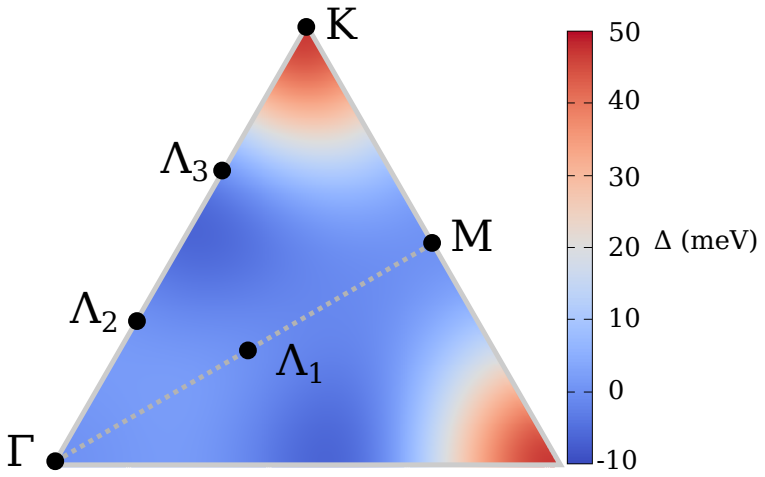


Figure 74: The tight-binding fit to the self-consistent gap function, with the high symmetry points (used to constrain the fit) marked.

in Equation 73 of Section 9.1 as giving the probability of a given electronic state $(\mathbf{k}, \epsilon)$ being occupied. Multiplying the spectral function by the Fermi-Dirac distribution of the electrons gives the ARPES intensity for a perfect, noise-free measurement:

$$I_{ARPES}(\mathbf{k}, \epsilon) \propto f(\epsilon) A(\mathbf{k}, \epsilon) + const. \tag{83}$$

It is relatively straightforward to convolve the expression with a noise function and experimental resolution [150], but in all cases relevant to this study the net result was simply a blurring of the image. I therefore consider the clean case only. There will be some background in any measurements corresponding to the constant offset in Equation 83, so I will consider both the overall scale and the zero of the intensity to be arbitrary.

The density of states can be measured in STS experiments, in which case only the energy can be varied (via the applied tunneling potential). Analytically the crystal momenta are summed over:

$$I_{DOS}(\epsilon) \propto \sum_{\mathbf{k}} I_{ARPES}(\mathbf{k}, \epsilon).$$

I will again neglect imperfections in the experiment. An important point to note is that while I_{DOS} contains an arbitrary scaling its zero is fixed.

10.2.2 Results for $NbSe_2$

Figures 75, 76, and 77 show the theoretical predictions for the ARPES intensity at E_F, the ARPES intensity across ϵ with $\mathbf{k}$ along $\overrightarrow{MK}$, and the DOS, respectively. The only free parameter in the theory remains the overall strength of the electron-phonon coupling. This was set to give a good match to both the experimental ARPES results of Borisenko *et al.* (Figs. 76 and 75) and the STS results of Soumayanaran *et al.* (Fig. 77). Of course it was not guaranteed that all could be satisfied simultaneously, but in fact the agreement is very good.

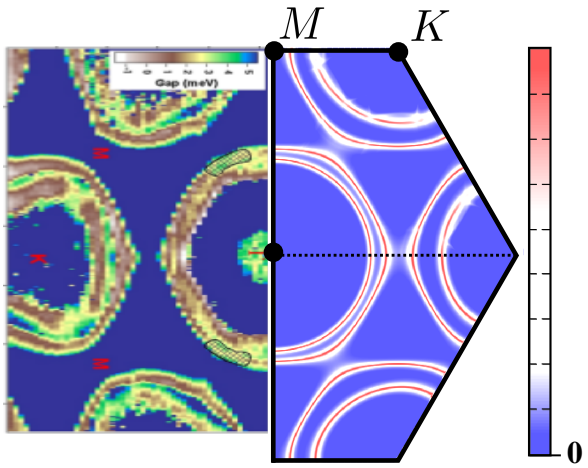


Figure 75: Left: experimental ARPES results of Borisenko *et al.* [105]. Right: the calculated ARPES intensity; upper half with the self-consistently calculated gap (evident along MK), lower half without the gap, for comparison. See also Figures 76 and 77.

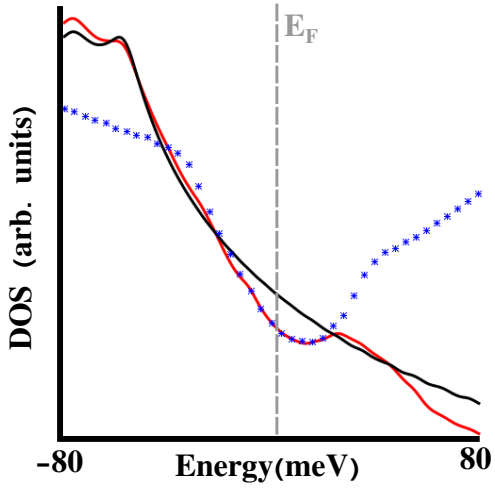


Figure 77: Blue points: the density of states as measured in STS (reproduced from [102]). Black: the theoretical prediction with no gap. Red: the theoretical prediction including the self-consistent gap. Note the centring of the gap above E_F. See also Figures 75 and 76.

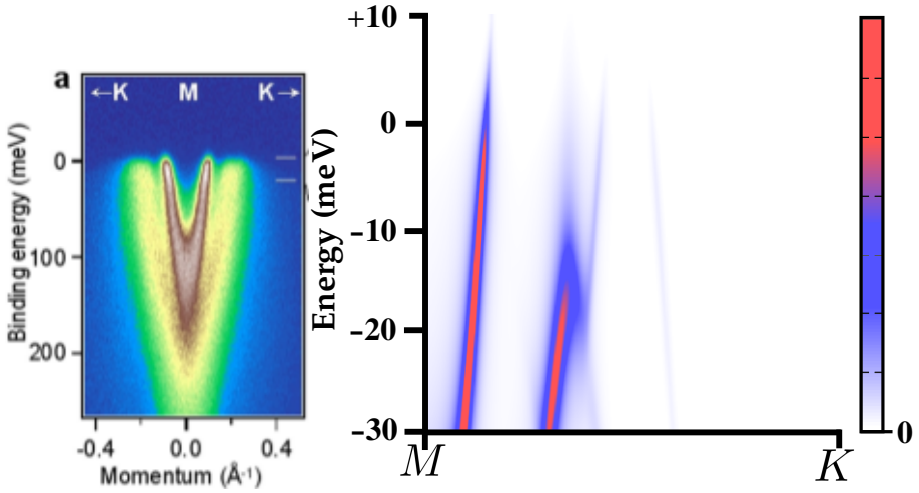


Figure 76: Left: the experimental ARPES intensity along the MK line (see Figure 75), reproduced from [105]. Right: theoretical result including the self-consistent CDW gap. Note the opening of the gap on the inner band only, and the corresponding backbending. The total gap is approximately twice the distance from the top of the band to E_F. See also Figure 75 and 77.

Figure 75 shows the experimentally observed ARPES intensity alongside the theoretical result both with and without the self-consistent gap. The key points to note are that the outer band does not develop a gap, and that the inner-band gap opens along the $\overrightarrow{MK}$ line. Figure 76 shows a different ARPES cut, this time through a range of energies along the $\overrightarrow{MK}$ line. The result is again in close agreement with the experimental results.

Finally, Figure 77 shows the density of states, both measured in STS and calculated theoretically. There are kinks in the experimental results around ± 35 meV. At this energy the tunneling electrons are able to excite phonons, so the density of states includes the many possible phonon states. The theory again matches the experiment, and both suggest a gap centred around 12 meV above E_F. I included a 4 meV chemical potential in the calculation for a closer match, which is well within the ± 16 meV uncertainty in the experiment leading to the phenomenological bandstructure fit employed throughout this work [106]. The gap is particle-hole asymmetric as the DOS is higher on the particle side than the hole side. This appears to derive predominantly from the bare bandstructure which is also shown.

11 Summary of Part II

In the second half of this thesis I have developed a model of the CDW phase transition in $NbSe_2$ deriving from electron-phonon coupling. The theory is 'strong coupling' in the sense that the details of the coupling are important to the development of the charge order; in particular I showed that it is necessary to include information about the orbital content of the bands, and the crystal momenta, scattered from and to. Having established the analytic form for the coupling as well as the bandstructure of the material in Chapter 7, I demonstrated that the model is capable of reproducing the full gamut of experimental results seen in $NbSe_2$, including the various 'anomalous' observations. These include an extended phonon softening rather than a sharp Kohn anomaly [103, 104], a particle/hole asymmetric gap which is offset from the Fermi level [102], the CDW gap opening in one band of the Fermi surface but not the other [105, 106], the incomplete gapping of the Fermi surface in the CDW state [105, 106], and the existence of a suppressed density of states over a range of temperatures above the transition [105, 108]. The fact that all of these results can be reproduced by a model containing only one free parameter suggests this description captures the essence of the physics at work in this system.

In quasi-1D materials, like those considered in Part I, the Peierls instability is undoubtedly the cause of the CDW transition. Nesting causes a divergence in the electronic susceptibility at $\mathbf{Q}_{CDW} = 2\mathbf{k}_F$, and an arbitrarily weak, shapeless coupling between electrons is sufficient to order the system. In Chapter 8 I definitively ruled out such weak-coupling theories as possible explanations for the CDW order in $NbSe_2$. Despite its anomalous behaviour, $NbSe_2$ is in most senses a typical quasi-2D CDW system. Presumably in time the results mentioned above will cease to be referred to as anomalous - when looked at from the correct, strong coupling, point of view, they are perfectly understandable.

In fact there is no reason to expect the Peierls story of 1D to apply to higher dimensions, and I would argue that it is time to retire it as a paradigm for charge order in two- and three-dimensional systems. Certainly nesting plays a significant rôle in some cases, such as the spin density wave in chromium [98, 97], but in general we should look instead to strong-coupling mechanisms like those in this thesis when we see charge order developing, and should consider ourselves lucky if the simpler nesting story can explain the order.

A discussion of charge order in layered materials will inevitably draw comparisons to the uncertain situation in the layered high-T_C cuprates and pnictides, and I should spend a few words discussing the similarities. The motivation is that, as noted at various points throughout this thesis, there is a close mathematical similarity between weak-coupling CDW ordering and the BCS theory of superconductivity; perhaps, then, the extension of the former to strong-coupling theories has some bearing on the extension of the latter to non-BCS superconductors[39]. At the same time, the high-T_C materials are believed to develop CDW order of their own simultaneously with their development of superconductivity, although whether the two orders help or hinder one another is at present unknown [96, 151, 5].

The model for the pseudogap I provide in Chapter 9, of CDW order suppressed by fluctuations of the phonon field, is close in spirit to the 'preformed pairs' theory of the high-T_C pseudogap [152] as discussed in that chapter. The fact that the CDW gap is highly anisotropic in k-space provides another point of contact. The phase diagrams established in Sections 8.5 and 9.3 as a function of strain suggest that the 1Q and 3Q CDW geometries are of a comparable stability in $NbSe_2$, with the 3Q forming naturally but very little uniaxial strain required to stabilize the 1Q state. In fact I find agreement with previous claims that the 1Q state seen on the surface of $NbSe_2$ in STM could be stabilized by local strain effects [102]. In the cuprates it is unclear whether the CDW state is of 1Q (stripe) type or 2Q (checkerboard) type, with theoretical work suggesting the materials sit on the boundary of stability between the two states [5, 4].

Of course, we cannot generalize the results presented here for $NbSe_2$ to the high-T_C materials. What is clear is that, as a result of the work presented in this thesis, $NbSe_2$ provides a well-understood example of how the aforementioned phenomena *can* arise in a layered material, which can therefore be used as a clear control case when plumbing the depths of the murkier waters of the cuprates and pnictides.

[39] Probably not.

Appendix A

Derivation of the non-interacting phonon action from first principles.

The Lagrangian for a 1D chain of N coupled classical harmonic oscillators, with generalized coördinates $q_j = (r_{j+1} - r_j)/a$ and natural frequency Ω, is

$$L = \sum_j^N \dot{q}_j \cdot p_j - \frac{p_j^2}{2m} - \frac{1}{2} m\Omega^2 q_j^2.$$

Take the continuum limit by defining the field $\varphi(r_j) = q_j/l$ and its conjugate momentum $\eta(r_j) = p_j l$, with $l = 1/\sqrt{m\Omega}$. This gives

$$L = \int \mathrm{d}r \dot{\varphi}(r)\,\eta(r) - \frac{\Omega}{2}\left(\eta(r)^2 + \varphi(r)^2\right).$$

When considering the interacting theory the electron field couples only to the phonon field φ and not its conjugate momentum η, so it would be nice to rewrite the action solely in terms of the former. This is done by completing the square:

$$\begin{aligned} S &= -\frac{\Omega}{2}\int \mathrm{d}t \int \mathrm{d}r \left\{\eta^2 + \varphi^2 - \frac{2}{\Omega}\dot{\varphi}\eta\right\} \\ &= -\frac{\Omega}{2}\int \mathrm{d}t \int \mathrm{d}r \left\{\varphi^2 + \left(\eta - \frac{1}{\Omega}\dot{\varphi}\right)^2 - \frac{1}{\Omega^2}\dot{\varphi}^2\right\} \end{aligned}$$

now when we consider the partition function $\int \frac{\mathscr{D}\varphi\mathscr{D}\eta}{\mathcal{N}} \exp(-S)$ the η integral cancels with its normalization factor. Thus we integrate out η exactly and find the 'effective' action

$$S = -\int \mathrm{d}t \int \mathrm{d}r \left\{\frac{\Omega}{2}\varphi^2 - \frac{1}{2\Omega}\dot{\varphi}^2\right\}$$

and integrating the second term by parts allows us to write

$$S = -\int \mathrm{d}t \int \mathrm{d}r \left\{\frac{1}{2}\varphi\left(\frac{\Omega^2 + \partial_t^2}{\Omega}\right)\varphi\right\}.$$

Carry out a spatial Fourier transform (crystal momentum q is not to be confused with generalized real space coördinate q_j) and include the wavevector dependence of the phonon frequency explicitly:

$$S = -\int \mathrm{d}t \sum_q \left\{ \frac{1}{2}\varphi_q \left(\frac{\Omega_q^2 + \partial_t^2}{\Omega_q} \right) \varphi_{-q} \right\}.$$

Wick rotate by writing $t = i\tau$

$$S = i\int_0^\beta \mathrm{d}\tau \sum_q \left\{ \frac{1}{2}\varphi_q \left(\frac{\Omega_q^2 - \partial_\tau^2}{\Omega_q} \right) \varphi_{-q} \right\}$$

where the cyclic nature of imaginary time has been invoked to write a finite cutoff on the integral. The cutoff is the inverse temperature of the system. Define the Euclidean action to be $S_E = iS$ so that

$$S_E = \int_0^\beta \mathrm{d}\tau \sum_q \left\{ \frac{1}{2}\varphi_q \left(\frac{\Omega_q^2 - \partial_\tau^2}{\Omega_q} \right) \varphi_{-q} \right\}.$$

Since imaginary time τ is cyclic we can decompose into Fourier modes, known as Matsubara frequencies:

$$S_E = \sum_{\Omega_n} \sum_q \left\{ \frac{1}{2}\varphi_{q,\Omega_n} \left(\frac{\Omega_q^2 + \Omega_n^2}{\Omega_q} \right) \varphi_{-q,-\Omega_n} \right\}$$

or

$$S_E = \sum_{\Omega_n} \sum_q \varphi_{q,\Omega_n} D^{-1}_{q,\Omega_n} \varphi_{-q,-\Omega_n}$$

defining the bare phonon propagator $D_{q,\Omega_n} = -2\Omega_q / \left((i\Omega_n)^2 - \Omega_q^2 \right)$. The same procedure holds for higher dimensional spaces, in which case the phonon field is a vector field with the dimension of the space. In this thesis I consider a 2D space, but focus on the longitudinal phonon modes, so often treat the field as a scalar. It should be understood that the vector field has been projected along the direction of the momentum transfer. The vector nature of the phonon field re-appears in the electron-phonon coupling, dealt with explicitly in the thesis.

Appendix B

Derivation of the 2-point correlator (propagator).

Define currents which couple to the fields like so:

$$\mathscr{Z}_{I,J} = \int \mathscr{D}\psi\mathscr{D}\varphi \exp(-S) \exp\left(\sum_k J_k^\dagger \varphi_k + \sum_k \varphi_k^\dagger J_k + \sum_k I_k^\dagger \psi_k + \sum_k \psi_k^\dagger I_k\right)$$

with S the non-interacting action for both the electron and phonon fields. The working can be done separately for each field, and I will do so as the prefactors are different in each case. First the bosonic phonon field:

$$\mathscr{Z}_J = \int \mathscr{D}\varphi \exp\left(-\left[\frac{1}{2}\sum_{kq} \varphi_k^\dagger A_{kq} \varphi_q + \sum_k J_k^\dagger \varphi_k + \sum_k \varphi_k^\dagger J_k\right]\right)$$

let

$$\varphi = \phi + \bar{\varphi}$$

with $\bar{\varphi}$ fixed (mean field). Expanding,

$$\begin{aligned}
\mathscr{Z}_J &= \int \mathscr{D}\phi \exp\left(-\left[\frac{1}{2}\sum_{kq} \phi_k^\dagger A_{kq} \phi_q + \frac{1}{2}\sum_{kq} \bar{\varphi}_k^\dagger A_{kq} \bar{\varphi}_q + \frac{1}{2}\sum_{kq} \phi_k^\dagger A_{kq} \bar{\varphi}_q \right.\right. \\
&\quad \left.\left. + \frac{1}{2}\sum_{kq} \bar{\varphi}_k^\dagger A_{kq} \phi_q + \sum_k J_k^\dagger \bar{\varphi}_k + \sum_k J_k^\dagger \phi_k + \sum_k \bar{\varphi}_k^\dagger J_k + \sum_k \phi_k^\dagger J_k\right]\right) \\
&= \int \mathscr{D}\phi \exp\left(-\left[\frac{1}{2}\sum_{kq} \phi_k^\dagger A_{kq} \phi_q + \frac{1}{2}\sum_{kq} \bar{\varphi}_k^\dagger A_{kq} \bar{\varphi}_q + \sum_k J_k^\dagger \bar{\varphi}_k + \sum_k \bar{\varphi}_k^\dagger J_k \right.\right. \\
&\quad \left.\left. + \sum_k \left(J_k^\dagger + \frac{1}{2}\sum_q \bar{\varphi}_q^\dagger A_{qk}\right)\phi_k + \sum_k \phi_k^\dagger \left(J_k + \frac{1}{2}\sum_q A_{kq}\bar{\varphi}_q\right)\right]\right)
\end{aligned}$$

so we remove the cross terms by picking

$$\bar{\varphi}_q = -2\sum_p A_{qp}^{-1} J_p$$

where the inverse is defined as

$$\sum_q A_{kq}\, A^{-1}_{qp} = \delta_{kp}.$$

Thus we have

$$\begin{aligned}
\mathscr{Z}_J &= \int \mathscr{D}\phi \exp\Bigg(-\Bigg[\frac{1}{2}\sum_{kq} \phi^\dagger_k A_{kq} \phi_q + 2\sum_{kqpm} J^\dagger_k A^{-1}_{pk} A_{kq} A^{-1}_{qm} J_m \\
&\quad -2\sum_{kp} J^\dagger_k A^{-1}_{kp} J_p - 2\sum_{kp} J^\dagger_p A^{-1}_{pk} J_k \Bigg]\Bigg) \\
&= \int \mathscr{D}\phi \exp\left(-\left[\frac{1}{2}\sum_{kq} \phi^\dagger_k A_{kq} \phi_q - 2\sum_{kp} J^\dagger_p A^{-1}_{pk} J_k \right]\right)
\end{aligned}$$

or

$$\frac{\mathscr{Z}_J}{\mathscr{Z}_0} = \exp\left(2\sum_{kq} J^\dagger_k A^{-1}_{kq} J_q \right).$$

Now we take two functional derivatives with respect to the currents on each side to get

$$\begin{aligned}
\delta_{J^\dagger_q}\delta_{J_k}\frac{\mathscr{Z}_J}{\mathscr{Z}_0} &= \mathscr{Z}_0^{-1}\int \mathscr{D}\varphi \left(-\varphi^\dagger_k\right)\left(-\varphi_q\right)\exp\left(-\left[\frac{1}{2}\sum_{kq}\varphi^\dagger_k A_{kq}\varphi_q + \sum_k J^\dagger_k\varphi_k + \sum_k \varphi^\dagger_k J_k \right]\right) \\
&= \delta_{J^\dagger_q}\delta_{J_k}\exp\left(2\sum_{kq} J^\dagger_k A^{-1}_{kq} J_q \right) \\
\left\langle \varphi^\dagger_k\varphi_q \right\rangle_J &= \delta_{J^\dagger_q}\left(2A^{-1}_{qk} J^\dagger_q \right)\exp\left(2\sum_{kq} J^\dagger_k A^{-1}_{kq} J_q \right) \\
\left\langle \varphi^\dagger_k\varphi_q \right\rangle_0 &= 2A^{-1}_{qk}.
\end{aligned}$$

For the action considered in this thesis we have

$$A_{kq} = \left(\frac{\Omega_n^2}{\Omega_{\mathbf{q}}} + \Omega_{\mathbf{q}} \right)\delta_{kq}$$

so

$$\boxed{\left\langle \varphi_k^\dagger \varphi_q \right\rangle_0 = 2\left(\frac{\Omega_n^2}{\Omega_{\mathbf{q}}} + \Omega_{\mathbf{q}}\right)^{-1} \delta_{kq} \triangleq D_q \delta_{kq}}.$$

Now the Grassman electron field:

$$\mathscr{Z}_I = \int \mathscr{D}\psi \exp\left(-\left[\sum_{kq} \psi_k^\dagger B_{kq} \psi_q + \sum_k I_k^\dagger \psi_k + \sum_k \psi_k^\dagger I_k\right]\right)$$

let

$$\psi_k = \Psi_k + \bar{\psi}_k$$

with $\overline{\psi}_k$ fixed (mean field) and as before

$$\begin{aligned}\mathscr{Z}_I \;=\; & \int \mathscr{D}\Psi \exp\left(-\left[\sum_{kq} \Psi_k^\dagger B_{kq} \Psi_q + \sum_{kq} \bar{\psi}_k^\dagger B_{kq} \bar{\psi}_q + \sum_k I_k^\dagger \bar{\psi}_k + \sum_k \bar{\psi}_k^\dagger I_k\right.\right. \\ & \left.\left. + \sum_q \left(I_q^\dagger + \sum_k \bar{\psi}_k^\dagger B_{kq}\right) \Psi_q + \sum_k \Psi_k^\dagger \left(I_k + \sum_q B_{kq} \bar{\psi}_q\right)\right]\right)\end{aligned}$$

choose

$$\bar{\psi}_q = -\sum_p B_{qp}^{-1} I_p$$

with

$$\sum_p B_{kp}\, B_{pq}^{-1} = \delta_{kq}$$

to give

$$\mathscr{Z}_I = \int \mathscr{D}\Psi \exp\left(-\left[\sum_{kq} \Psi_k^\dagger B_{kq} \Psi_q - \sum_{kp} I_k^\dagger B_{kp}^{-1} I_p\right]\right)$$

and again

$$\frac{\mathscr{Z}_I}{\mathscr{Z}_0} = \exp\left(\sum_{kq} I_k^\dagger B_{kq}^{-1} I_q\right).$$

Taking functional derivatives gives

$$\delta_{I_q^\dagger}\delta_{I_k}\frac{\mathscr{Z}_I}{\mathscr{Z}_0}\bigg|_{I=0} = \left\langle \psi_k^\dagger \psi_q \right\rangle_0 = B_{qk}^{-1}$$

or comparing to the original action,

$$\boxed{\left\langle \psi_k^\dagger \psi_q \right\rangle_0 = G_k \delta_{kq}}.$$

Appendix C

Matsubara sums.

Matsubara's trick is to replace a sum over discrete frequencies by a contour integral with a weight function containing poles sat at those frequencies. The two are equivalent through Cauchy's theorem. For fermions the frequencies are at $i\omega_n = (2n+1)\,\pi/\beta$, and the function with poles at the required places in the complex plane is the Fermi-Dirac distribution function

$$f(z) = \frac{1}{\exp(\beta z)+1}.$$

For bosons the frequencies are at $i\Omega_n = 2\pi n/\beta$, and the function is the Bose-Einstein distribution function

$$n_B(z) = \frac{1}{\exp(\beta z)-1}.$$

Thus for sums over fermionic frequencies we employ the relation

$$\sum_{\omega_n} h(i\omega_n) = \frac{1}{2\pi i}\oint_{\mathcal{C}} \mathrm{d}z h(z) f(z)$$

and for bosonic frequencies

$$\sum_{\Omega_n} h(i\Omega_n) = \frac{1}{2\pi i}\oint_{\mathcal{C}} \mathrm{d}z h(z) n_B(z)$$

where in both cases the contour $\mathcal{C}$ encircles the infinite set of poles running up the imaginary axis. This well-established technique is clearly explained in references [127, 140, 128, 10] to name a few.

The susceptibility loop for χ and D_2

Both the electronic susceptibility

$$\chi = -\sum_{\mathbf{k}}\sum_{\omega_n} G_k G_{k+q}$$

and the generalized susceptibility

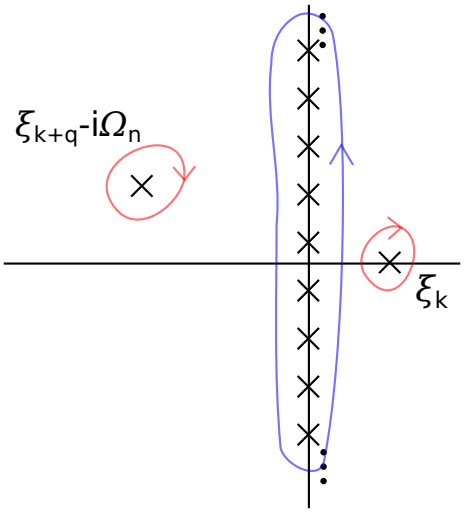


Figure 78: The contour $\mathcal{C}_1$ used in Equation 84. The blue contour, stretching to $\pm i\infty$, is continuously deformable into the contour defined by the two red lines. The latter is employed in the Matsubara sum.

$$D_2 = -\sum_{\mathbf{k}} \sum_{\omega_n} |g_{\mathbf{k},\mathbf{k+q}}|^2 G_k G_{k+q}$$

require evaluation of the fermionic Matsubara sum

$$\sum_{\omega_n} G_k G_{k+q} = \sum_{\omega_n} \frac{1}{i\omega_n - \xi_{\mathbf{k}}} \frac{1}{i\omega_n + i\Omega_n - \xi_{\mathbf{k+q}}}$$

where I have absorbed the chemical potential into the definition of the energy ξ for convenience. Using the Cauchy trick this can be re-written

$$\sum_{\omega_n} G_k G_{k+q} = \frac{1}{2\pi i} \oint_{\mathcal{C}_1} \mathrm{d}z \frac{f(z)}{z - \xi_{\mathbf{k}}} \frac{1}{z + i\Omega_n - \xi_{\mathbf{k+q}}}. \tag{84}$$

The contour $\mathcal{C}_1$ can be deformed into lines encircling the two poles, as shown in Figure 78. Using the residue theorem the result is

$$\sum_{\omega_n} G_k G_{k+q} = \frac{f(\xi_{\mathbf{k}})}{\xi_{\mathbf{k}} + i\Omega_n - \xi_{\mathbf{k+q}}} + \frac{f(\xi_{\mathbf{k+q}} - i\Omega_n)}{\xi_{\mathbf{k+q}} - i\Omega_n - \xi_{\mathbf{k}}}.$$

Noting that $f(z - i\Omega_n) = f(z - 2\pi i n/\beta) \equiv f(z)$ the result simplifies to

$$\sum_{\omega_n} G_k G_{k+q} = \frac{f(\xi_{\mathbf{k}}) - f(\xi_{\mathbf{k+q}})}{\xi_{\mathbf{k}} - \xi_{\mathbf{k+q}} + i\Omega_n}.$$

The function D_2 including the electron self-energy

In this case we have

$$D_2^{\Sigma} = -\sum_{\mathbf{k}} \sum_{\omega_n} \left|g_{\mathbf{k},\mathbf{k+q}}\right|^2 G_k^{\Sigma} G_{k+q}$$

requiring evaluation of the fermionic Matsubara sum

$$\sum_{\omega_n} G_k^{\Sigma} G_{k+q} = \sum_{\omega_n} \frac{1}{i\omega_n - \xi_{\mathbf{k}} - \Sigma_k} \frac{1}{i\omega_n + i\Omega_n - \xi_{\mathbf{k+q}}}.$$

The problem is that the pole in the first propagator is now in an unknown location. In fact it has become a branch cut. The simplest way to deal with this is to use the Lehmann representation, which is a general identity for complex functions following from the Kramers-Kronig relation. The identity states that

$$G^{\Sigma}\left(\mathbf{k}, i\omega_n\right) \equiv -\frac{1}{\pi} \int \mathrm{d}\epsilon \frac{\Im\mathfrak{m}\left(G^{\Sigma}\left(\mathbf{k}, \epsilon\right)\right)}{i\omega_n - \epsilon}$$

where the branch cut along the line $i\omega_n = \epsilon$ is apparent. As noted in the main text this representation includes the spectral function

$$A\left(\mathbf{k}, \epsilon\right) \triangleq -\frac{1}{\pi} \Im\mathfrak{m} G\left(\mathbf{k}, \epsilon\right)$$

about which I will shortly prove two useful facts.

The integral to evaluate is

$$\sum_{\omega_n} G_k^{\Sigma} G_{k+q} = \int \mathrm{d}\epsilon A\left(\mathbf{k}, \epsilon\right) \frac{1}{2\pi i} \oint_{\mathcal{C}_2} \mathrm{d}z \frac{f\left(z\right)}{z - \epsilon} \frac{1}{z + i\Omega_n - \xi_{\mathbf{k+q}}} \tag{85}$$

with the contour $\mathcal{C}_2$ given in Figure 79. The result is

$$\sum_{\omega_n} G_k^{\Sigma} G_{k+q} = \int \mathrm{d}\epsilon A\left(\mathbf{k}, \epsilon\right) \left[\frac{f\left(\epsilon\right)}{\epsilon + i\Omega_n - \xi_{\mathbf{k+q}}} + \frac{f\left(\xi_{\mathbf{k+q}} - i\Omega_n\right)}{\xi_{\mathbf{k+q}} - i\Omega_n - \epsilon}\right]$$

simplifying as in the previous case to

$$\sum_{\omega_n} G_k G_{k+q} = \int \mathrm{d}\epsilon A\left(\mathbf{k}, \epsilon\right) \left[\frac{f\left(\epsilon\right) - f\left(\xi_{\mathbf{k+q}}\right)}{\epsilon + i\Omega_n - \xi_{\mathbf{k+q}}}\right].$$

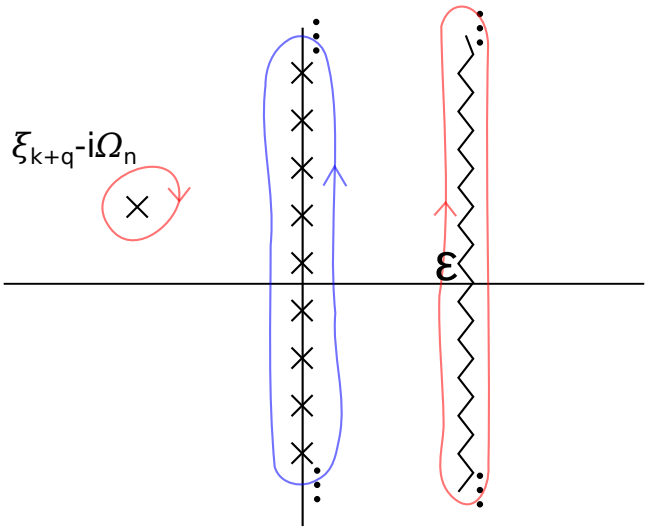


Figure 79: The contour $\mathcal{C}_2$ employed in Equation 85. A branch cut is present along $z = \epsilon$, but takes the form of an integral over simple poles.

Note that

$$A\left(\mathbf{k}, \epsilon + i\delta\right) = -\frac{1}{\pi}\mathfrak{Im}\left(\frac{1}{\epsilon + i\delta - \xi_{\mathbf{k}} - \Sigma_k}\right)$$

so if $\Sigma = 0$, $\delta \to 0^+$ it becomes a delta function:

$$\begin{aligned}
& -\frac{1}{\pi}\int_{-\Lambda}^{\Lambda} \mathrm{d}\epsilon \mathfrak{Im} \frac{1}{\epsilon - \xi_{\mathbf{k}} + i\delta} \\
= \; & \frac{1}{\pi}\int_{-\Lambda}^{\Lambda} \mathrm{d}\epsilon \frac{\delta}{\left(\epsilon - \xi_{\mathbf{k}}\right)^2 + \delta^2} \\
\epsilon - \xi_{\mathbf{k}} = \delta\tan\left(\theta\right) \; & \downarrow \\
= \; & \frac{1}{\pi}\left[\operatorname{atan}\left(\frac{\Lambda - \xi_{\mathbf{k}}}{\delta}\right) - \operatorname{atan}\left(\frac{-\Lambda - \xi_{\mathbf{k}}}{\delta}\right)\right] \\
\lim_{\delta \to 0^+} \; & = 1 \\
\therefore \lim_{\delta \to 0^+} A\left(\mathbf{k}, \epsilon + i\delta\right)|_{\Sigma=0} \; & = \delta\left(\epsilon - \xi_{\mathbf{k}}\right).
\end{aligned}$$

Therefore the expression for D_2^{Σ} reduces smoothly to the non-interacting case D_2 as the self-energy is turned off.

The second useful fact amounts to the observation that A is Lorentzian, so the integral over all energy should be independent of Σ, δ, and $\mathbf{k}$, up to sign. Defining $\Sigma' \triangleq \mathfrak{Re}\left[\Sigma\right]$,

$\Sigma'' \triangleq \mathfrak{Im}\left[\Sigma\right]$,

$$\begin{array}{rcl} -\frac{1}{\pi}\int_{-\infty}^{\infty} \mathrm{d}\epsilon \mathfrak{Im} G^{\Sigma}\left(\mathbf{k}, \epsilon + i\delta\right) & = & -\frac{1}{\pi}\int_{-\infty}^{\infty} \mathrm{d}\epsilon \frac{\Sigma'' - \delta}{\left(\epsilon - \xi_{\mathbf{k}} - \Sigma'\right)^2 + \left(\Sigma'' - \delta\right)^2} \\ \epsilon - \xi_{\mathbf{k}} - \Sigma' \triangleq \left(\Sigma'' - \delta\right)\tan\left(\theta\right) & \downarrow & \\ & = & -\frac{1}{\pi}\left[\mathrm{atan}\left(\frac{\infty}{\Sigma'' - \delta}\right) - \mathrm{atan}\left(-\frac{\infty}{\Sigma'' - \delta}\right)\right] \\ & = & \begin{cases} -1, & \Sigma'' > \delta \\ 1, & \delta > \Sigma'' \end{cases} \end{array}$$

constituting a 'sum rule'.

The functions D_3 and D_4

Neglecting band indices, which simply follow their respective momentum label, the non-linear response terms are

$$\begin{array}{rcl} D_3\left(\mathbf{q}, \mathbf{p}\right) & \triangleq & \sum_k g_{\mathbf{k},\mathbf{k+q}} g_{\mathbf{k+q},\mathbf{k+q+p}} g_{\mathbf{k+q+p},\mathbf{k}} G_k G_{k+q} G_{k+p+q} \\ D_4\left(\mathbf{q}, \mathbf{p}, \mathbf{l}\right) & \triangleq & \sum_k g_{\mathbf{k},\mathbf{k+q}} g_{\mathbf{k+q},\mathbf{k+q+p}} g_{\mathbf{k+q+p},\mathbf{k+q+p+l}} g_{\mathbf{k+q+p+l},\mathbf{k}} G_k G_{k+q} G_{k+p+q} G_{k+p+q+l}. \end{array}$$

The method is the same for both so I will treat only the simpler case of D_3 here. The fermion frequencies corresponding to the p and q labels are set to zero. Keeping them finite could act as a regularization in numerical calculations, but the physical motivation would not be clear. Instead I will employ an explicit regularization shortly. The sum is

$$\sum_{\omega_n} G_k G_{k+q} G_{k+p+q} = \frac{1}{2\pi i}\oint \mathrm{d}z \frac{f\left(z\right)}{z - \xi_{\mathbf{k}}} \frac{1}{z - \xi_{\mathbf{k+q}}} \frac{1}{z - \xi_{\mathbf{k+p+q}}}$$

and evaluates to

$$\frac{f\left(A\right)}{\left(A - B\right)\left(A - C\right)} + \frac{f\left(B\right)}{\left(B - A\right)\left(B - C\right)} + \frac{f\left(C\right)}{\left(C - A\right)\left(C - B\right)}$$

with

$$\begin{aligned} A &\triangleq \xi_{\mathbf{k}} \\ B &\triangleq \xi_{\mathbf{k+q}} \\ C &\triangleq \xi_{\mathbf{k+p+q}}. \end{aligned}$$

Numerically there is an issue whenever any of the terms A, B, or C are coincident. As noted in [5, 4] the divergences are in fact cancelled by contributions from the different terms. The correct time to deal with them is before carrying out the frequency sum.

If $A = B$, the expression

$$\frac{1}{2\pi i}\oint \mathrm{d}z \frac{f(z)}{z-A}\frac{1}{z-B}\frac{1}{z-C}$$

reduces to

$$\frac{1}{2\pi i}\oint \mathrm{d}z \frac{f(z)}{(z-B)^2}\frac{1}{z-C}.$$

The residue of the second-order pole can be evaluated using the relation

$$\mathrm{Res}\left(\frac{g(z)}{(z-a)^n}\right) = \frac{1}{(n-1)!}\left.\frac{\partial^{n-1} g(z)}{\partial z^{n-1}}\right|_{z=a}.$$

Thus

$$\frac{1}{2\pi i}\oint \mathrm{d}z \frac{f(z)}{(z-B)^2}\frac{1}{z-C} = \frac{f(C)-f(B)}{(C-B)^2} + \frac{f'(B)}{B-C}.$$

For $A = B = C$ we have

$$\frac{1}{2\pi i}\oint \mathrm{d}z \frac{f(z)}{(z-C)^3} = \frac{1}{2}f''(C).$$

For D_4 the corresponding expression is

$$\frac{f(A)}{(A-B)(A-C)(A-D)}+\frac{f(B)}{(B-A)(B-C)(B-D)}$$
$$+\frac{f(C)}{(C-A)(C-B)(C-D)}+\frac{f(D)}{(D-A)(D-B)(D-C)}$$

and the regularized cases are

$A=B$:

$$\frac{f(C)}{(C-B)^2(C-D)}+\frac{f(D)}{(D-B)^2(D-C)}+\frac{f'(B)}{(B-C)(B-D)}$$
$$-f(B)\left[\frac{1}{(B-C)^2(B-D)}+\frac{1}{(B-C)(B-D)^2}\right]$$

$A=B=C$:

$$\frac{f''(C)}{C-D}-2\frac{f'(C)}{(C-D)^2}+2\frac{f(C)-f(D)}{(C-D)^3}$$

$A=B$, $C=D$:

$$\frac{f'(B)+f'(D)}{(B-D)^2}-2\frac{f(B)-f(D)}{(B-D)^3}$$

$A=B=C=D$:

$$\frac{1}{6}f'''(D).$$

Numerically a cutoff distance is implemented, such that when two terms approach within the cutoff the regularized form is used.

The electron self-energy Σ

The expression for the self-energy is

$$\Sigma(\mathbf{k},i\omega_n)=-\sum_{\mathbf{q}}\sum_{\Omega_n}|g_{\mathbf{k},\mathbf{k}-\mathbf{q}}|^2\,G(\mathbf{k}-\mathbf{q},i\omega_n-i\Omega_n)\,D(\mathbf{q},i\Omega_n)$$

with D the RPA renormalized phonon propagator and G containing the self-energy (the

calculation therefore being self-consistent). Expanding, using the Lehmann representation for the electron propagator,

$$\begin{aligned}\Sigma\left(\mathbf{k},i\omega_n\right) &= -\sum_{\mathbf{q}}\sum_{i\Omega_n}\left(\left|g_{\mathbf{k},\mathbf{k}-\mathbf{q}}\right|^2\frac{-2\Omega_0\left(\mathbf{q}\right)}{\left(i\Omega_n+\Omega_{RPA}\left(\mathbf{q}\right)\right)\left(i\Omega_n-\Omega_{RPA}\left(\mathbf{q}\right)\right)}\cdot\right.\\ &\left.\cdot\left(-\frac{1}{\pi}\right)\int \mathrm{d}\epsilon'\frac{\Im\mathfrak{m}\left[G\left(\mathbf{k}-\mathbf{q},\epsilon'+i\delta\right)\right]}{i\omega_n-i\Omega_n-\epsilon'}\right).\end{aligned}$$

Carrying out the Matsubara sum

$$\begin{aligned}\Sigma\left(\mathbf{k},i\omega_n\right) &= \frac{1}{\pi}\sum_{\mathbf{q}}\left|g_{\mathbf{k},\mathbf{k}-\mathbf{q}}\right|^2 2\Omega_0\int \mathrm{d}\epsilon'\Im\mathfrak{m}\left[G\left(\mathbf{k}-\mathbf{q},\epsilon'+i\delta\right)\right]\cdot\\ &\cdot\oint \mathrm{d}z\frac{n_B\left(z\right)}{\left(z+\Omega_{RPA}\right)\left(z-\Omega_{RPA}\right)}\frac{1}{z+\epsilon'-i\omega_n}\\ &= -\frac{1}{\pi}\sum_{\mathbf{q}}\left|g_{\mathbf{k},\mathbf{k}-\mathbf{q}}\right|^2\left(\frac{\Omega_0}{\Omega_{RPA}}\right)\int \mathrm{d}\epsilon'\Im\mathfrak{m}\left[G\left(\mathbf{k}-\mathbf{q},\epsilon'+i\delta\right)\right]\cdot\\ &\cdot\left\{\frac{n_B\left(\Omega_{RPA}\right)+1-f\left(\epsilon'\right)}{i\omega_n-\epsilon'-\Omega_{RPA}}+\frac{n_B\left(\Omega_{RPA}\right)+f\left(\epsilon'\right)}{i\omega_n-\epsilon'+\Omega_{RPA}}\right\}\end{aligned}$$

where the relations

$$\begin{aligned}n_B\left(-\Omega_{RPA}\right) &= -1-n_B\left(\Omega_{RPA}\right)\\ n_B\left(i\omega_n-\epsilon'\right) &= -f\left(-\epsilon'\right)=f\left(\epsilon'\right)-1\end{aligned}$$

have been employed.

The Wick rotated version $i\omega_n\rightarrow\epsilon+i\delta$ is

$$\begin{aligned}\Sigma\left(\mathbf{k},\epsilon\right) &= -\frac{1}{\pi}\sum_{\mathbf{q}}\left|g_{\mathbf{k},\mathbf{k}-\mathbf{q}}\right|^2\left(\frac{\Omega_0}{\Omega_{RPA}}\right)\int \mathrm{d}\epsilon'\Im\mathfrak{m}\left[G\left(\mathbf{k}-\mathbf{q},\epsilon'+i\delta\right)\right]\cdot\\ &\cdot\left\{\frac{n_B\left(\Omega_{RPA}\right)+1-f\left(\epsilon'\right)}{\epsilon-\epsilon'-\Omega_{RPA}+i\delta}+\frac{n_B\left(\Omega_{RPA}\right)+f\left(\epsilon'\right)}{\epsilon-\epsilon'+\Omega_{RPA}+i\delta}\right\}.\end{aligned}$$

The calculation starts with a seed $\Sigma^{(0)}$; substituting this into G gives the first order ex-

pression

$$\begin{aligned}
\Sigma^{(1)}(\mathbf{k},\epsilon) &= -\frac{1}{\pi}\sum_{\mathbf{q}}\left|g_{\mathbf{k},\mathbf{k}-\mathbf{q}}\right|^2\left(\frac{\Omega_0}{\Omega_{RPA}}\right)\int \mathrm{d}\epsilon' \mathfrak{Im}\left[\frac{1}{\epsilon'-\xi_{\mathbf{k}-\mathbf{q}}-\Sigma^{(0)}+i\delta}\right]\cdot \\
&\quad \cdot\left\{\frac{n_B(\Omega_{RPA})+1-f(\epsilon')}{\epsilon-\epsilon'-\Omega_{RPA}+i\delta}+\frac{n_B(\Omega_{RPA})+f(\epsilon')}{\epsilon-\epsilon'+\Omega_{RPA}+i\delta}\right\} \\
&= -\frac{1}{\pi}\sum_{\mathbf{q}}\int \mathrm{d}\epsilon'\left|g_{\mathbf{k},\mathbf{k}-\mathbf{q}}\right|^2\frac{\Sigma^{(0)''}}{\left(\epsilon'-\xi_{\mathbf{k}-\mathbf{q}}-\Sigma^{(0)'}\right)^2+\left(\Sigma^{(0)''}-\delta\right)^2}\frac{\Omega_0}{\Omega_{RPA}}\cdot \\
&\quad \cdot\left\{\frac{n_B(\Omega_{RPA})+1-f(\epsilon')}{\epsilon-\epsilon'-\Omega_{RPA}+i\delta}+\frac{n_B(\Omega_{RPA})+f(\epsilon')}{\epsilon-\epsilon'+\Omega_{RPA}+i\delta}\right\}
\end{aligned}$$

with $\Sigma' = \mathfrak{Re}[\Sigma]$, $\Sigma'' = \mathfrak{Im}[\Sigma]$.

Appendix D

Proof of the Lehmann representation for arbitrary causal normalizable functions.

Consider a complex function $G(z)$. I will refer to G as a Green's function, but it can be any function obeying the relations I am about to define. Let

$$G^{+}(z) \triangleq G\left(z^{+}\right) \triangleq G\left(z+i0^{+}\right).$$

If G is a Green's function then G^{+} is the corresponding retarded Green's function. Define this to have the following properties:

1. $G^{+}(z)$ has no poles in the upper half of the complex plane (is causal)
2. $G(z)$ drops off at least as fast as $|z|^{-1}$ at large $|z|$ (is normalizable).

Point (2) implies the same for G^{+}. It follows from the definition of zero[40] that

$$G^{+}(z) = G^{+}\left(z^{+}\right).$$

Note that one cannot simply subtract $i0$ from each side to obtain the same relation for G:

$$G(z) \neq G\left(z^{+}\right). \tag{86}$$

This is clear physically. From points (1) and (2) it follows that a contour integral of G^{+} around the upper half plane is zero. Thus from a trivial application of the residue theorem it follows that

$$\begin{aligned} G^{+}\left(z^{+}\right) &= \frac{1}{2\pi i}\oint_{C} \mathrm{d}\omega' \frac{G^{+}(\omega')}{\omega' - z^{+}} \\ G\left(z^{+}\right) &= \frac{1}{2\pi i}\oint_{C} \mathrm{d}\omega' \frac{G^{+}(\omega')}{\omega' - z^{+}} \end{aligned}$$

where C is the red contour shown in Figure 80. We are free to deform that contour, at zero cost, by introducing an infinitesimal nubbin directly below the pole at $\omega' = z^{+}$(shown in blue in the same Figure). Call the new contour C'.

[40] *i.e.* $0x = 0 \forall x$

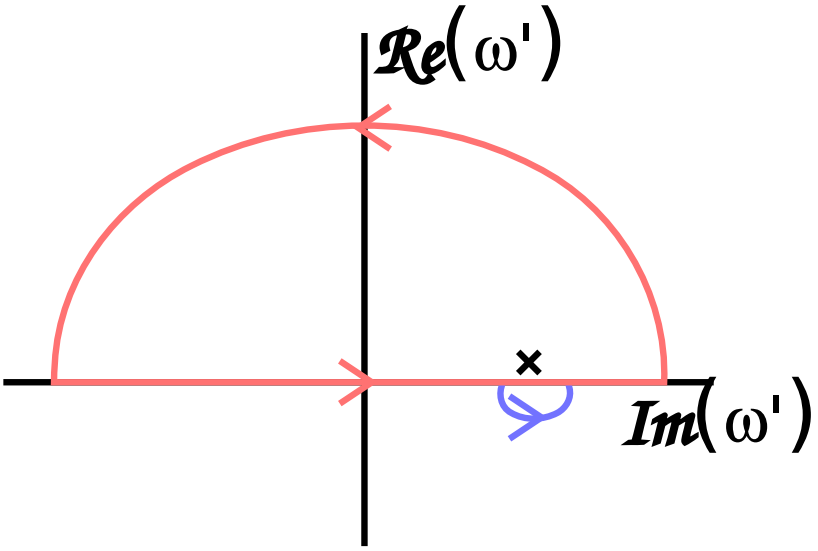


Figure 80: Two equivalent contours used in the text. The radius of the arc is taken to infinity.

Now it is clear that, without invoking the forbidden relation of Equation 86, for this specific contour we *can* shift the pole down into the nubbin. Thus it follows that

$$G(z) = \frac{1}{2\pi i} \oint_{C'} \mathrm{d}\omega' \frac{G^{+}(\omega')}{\omega' - z}. \tag{87}$$

As a point of interest, the arc in C contributes nothing to the integral, and so it follows straightforwardly that

$$G(z) = \frac{1}{\pi i} P \int \mathrm{d}\omega' \frac{G(\omega')}{\omega' - z}$$

with P the principal part (the average of avoiding the pole above and below the axis). This is the Kramers-Kronig relation.

Consider the spectral function, defined to be

$$A(z) \triangleq -\frac{1}{\pi} \mathfrak{Im}\left(G^{+}(z)\right).$$

This can be rewritten

$$\begin{aligned} A(z) &= -\frac{1}{2\pi i}\left(G^{+}(z) - \left(G^{+}(z)\right)^{*}\right) \\ &\equiv -\frac{1}{2\pi i}\left(G^{+}(z) - G^{-}(z)\right). \end{aligned}$$

Introducing a pole and integrating both sides we have

$$\int_{-\infty}^{\infty} \mathrm{d}\omega' \frac{A(\omega')}{\omega' - z} = -\frac{1}{2\pi i}\int_{-\infty}^{\infty} \mathrm{d}\omega' \frac{G^{+}(\omega') - G^{-}(\omega')}{\omega' - z}.$$

In the following I will only consider $\mathfrak{Im}(z) > 0$, but the argument can be extended to $\mathfrak{Im}(z) < 0$ by mirroring in the real axis. From property (2) we can introduce infinite arcs to close the contours at no cost. Choose the contour C' for the G^{+} integral. We can choose the G^{-} contour independently, and in fact I will take the same line (with nubbin) along the real axis but close this contour in the lower half plane instead. This is because G^{-}, the advanced Green's function, has no poles in this region (clear from the definition $(G^{+})^{*} \triangleq G^{-}$). Since there are no poles of either G^{-} or $\omega' - z$ in this contour it evaluates to zero. The result is therefore

$$\int_{-\infty}^{\infty} \mathrm{d}\omega' \frac{A(\omega')}{\omega' - z} = -\frac{1}{2\pi i}\oint_{C'} \mathrm{d}\omega' \frac{G^{+}(\omega')}{\omega' - z}.$$

From Equation 87 and the definition of A it follows that

$$\frac{1}{\pi}\int_{-\infty}^{\infty} \mathrm{d}\omega' \frac{\mathfrak{Im}\left(G^{+}(\omega')\right)}{\omega' - z} = G(z) \tag{88}$$

for any function obeying properties (1) and (2). This is the Lehmann representation.
It is important in this thesis that even though G may only be known numerically, and its pole structure is undetermined, G^{+} will never have poles in the upper half plane. This is simply because G is the propagator for quasiparticles we can observe in the lab, and such things are always causal.

Appendix E

The weak-coupling perturbation theory expansion of the action to fourth order in the phonon field.

The partition function is

$$\mathscr{Z} = \int \mathscr{D}\psi \int \mathscr{D}\varphi \exp\left(-S_\psi\right) \exp\left(-S_\varphi\right) \left[1 - S_{int} + \frac{1}{2}S_{int}^2 - \frac{1}{3!}S_{int}^3 + \frac{1}{4!}S_{int}^4\right]$$

with

$$S_{int} = \sum_{kq} g_{\mathbf{k},\mathbf{k}'} \varphi_{k-k'} \psi_{k'}^\dagger \psi_k.$$

The following definitions are employed:

$$\begin{aligned} \left\langle \mathscr{O}\left[\psi\right]\right\rangle_\psi &\triangleq \int \mathscr{D}\psi \exp\left(-S_\psi\right) \mathscr{O}\left[\psi\right] \\ \left\{\psi_k^\dagger, \psi_q\right\} &\triangleq 0 \end{aligned}$$

where S_ψ indicates the free electron action. The rule for field averages is

$$\left\langle \psi_k^\dagger \psi_q \right\rangle_\psi = G_k \delta_{kq}$$

with working provided in Appendix B.

Order 1

$$-\left\langle S_{int}\right\rangle = -\sum_{11'} g_{11'}\varphi_{1-1'} \left\langle \psi_1^\dagger \psi_{1'}\right\rangle = -\sum_{11'} g_{11'}\varphi_{1-1'}\delta_{11'} = -\sum_{1} g_{11}\varphi_0 = 0.$$

The notation $\psi_1 \triangleq \psi_{k_1}$ and $1 \triangleq \psi_1^\dagger$ is assumed from now on.

Order 2

$$\frac{1}{2}\left\langle S_{int}^2\right\rangle = \frac{1}{2}\sum_{11'22'} g_{11'}g_{22'}\varphi_{1-1'}\varphi_{2-2'} \left\langle 1^\dagger 1' 2^\dagger 2'\right\rangle$$

Wick's theorem states that functional average of (time-ordered) products of states can be decomposed into products of two-point correlators [127]. The only contraction which is nonzero, connected (in the sense of all external legs being joined) and conserves momentum is $\left\langle 1^{\dagger}2'\right\rangle \left\langle 2^{\dagger}1'\right\rangle$. This requires three interchanges, giving an overall negative sign. Therefore

$$\left\langle 1^{\dagger}1'2^{\dagger}2'\right\rangle_{connected} = -G_1 G_2 \delta_{12'}\delta_{21'}$$

and we get

$$\begin{aligned}\frac{1}{2}S_{int}^2 &= -\frac{1}{2}\sum_{12} g_{12}g_{21}\varphi_{1-2}\varphi_{2-1}G_1G_2 \\ &= -\frac{1}{2}\sum_{12} \left|g_{12}\right|^2 \left|\varphi_{1-2}\right|^2 G_1G_2.\end{aligned}$$

Disconnected terms disappear under re-exponentiation, hence their neglection.

Order 3

The Wick product is this time

$$\left\langle 1^{\dagger}1'2^{\dagger}2'3^{\dagger}3'\right\rangle$$

and the two surviving contractions are

$$\begin{aligned}a: &\quad \left\langle 1^{\dagger}2'\right\rangle \left\langle 2^{\dagger}3'\right\rangle \left\langle 3^{\dagger}1'\right\rangle \\ b: &\quad \left\langle 1^{\dagger}3'\right\rangle \left\langle 2^{\dagger}1'\right\rangle \left\langle 3^{\dagger}2'\right\rangle\end{aligned}$$

a quick check (counting how many times operators have passed each other) shows that both terms have a positive sign. So the highest order in G contribution is

$$-\frac{1}{3!}S^3_{int} = -\frac{1}{3!}\sum_{\substack{11'\\22'\\33'}} g_{11'}g_{22'}g_{33'}\varphi_{1-1'}\varphi_{2-2'}\varphi_{3-3'}G_1G_2G_3\left(\delta_{12'}\delta_{23'}\delta_{31'}+\delta_{13'}\delta_{21'}\delta_{32'}\right)$$

$$\begin{aligned} &= -\frac{1}{3!}\sum_{123}\left(g_{13}g_{21}g_{32}\varphi_{1-3}\varphi_{2-1}\varphi_{3-2}+g_{12}g_{23}g_{31}\varphi_{1-2}\varphi_{2-3}\varphi_{3-1}\right)G_1G_2G_3 \\ &= \frac{-2}{3!}\sum_{123}\mathfrak{Re}\left(g_{12}g_{23}g_{31}\varphi_{1-2}\varphi_{2-3}\varphi_{3-1}\right)G_1G_2G_3 \end{aligned}$$

using in the last line the reality of the interaction Lagrangian, that is, $a = b^*$.

Order 4

The Wick product

$$\left\langle 1^\dagger 1' 2^\dagger 2' 3^\dagger 3' 4^\dagger 4' \right\rangle$$

has 9 physical contractions. They fall into four classes:

a	$\delta_{12'}\delta_{23'}\delta_{34'}\delta_{41'}$
a^*	$\delta_{21'}\delta_{32'}\delta_{43'}\delta_{14'}$
b	$\delta_{12'}\delta_{24'}\delta_{31'}\delta_{43'}$
b^*	$\delta_{21'}\delta_{42'}\delta_{13'}\delta_{34'}$
c	$\delta_{13'}\delta_{24'}\delta_{32'}\delta_{41'}$
c^*	$\delta_{31'}\delta_{42'}\delta_{23'}\delta_{14'}$
d_1	$\delta_{12'}\delta_{21'}\delta_{34'}\delta_{43'}$
d_2	$\delta_{13'}\delta_{31'}\delta_{24'}\delta_{42'}$
d_3	$\delta_{14'}\delta_{41'}\delta_{23'}\delta_{32'}$

in each case the conjugate has the primes swapped on each δ, which results in a complex conjugation of the total result as in the cubic case. Examining the d terms:

$$\begin{aligned}
& \frac{1}{4!}\sum_{1234}|g_{12}|^2\,|\varphi_{1-2}|^2\,G_1G_2\,|g_{34}|^2\,|\varphi_{3-4}|^2\,G_3G_4 \\
& +\frac{1}{4!}\sum_{1234}|g_{13}|^2\,|\varphi_{1-3}|^2\,G_1G_3\,|g_{24}|^2\,|\varphi_{2-4}|^2\,G_2G_4 \\
& +\frac{1}{4!}\sum_{1234}|g_{14}|^2\,|\varphi_{1-4}|^2\,G_1G_4\,|g_{32}|^2\,|\varphi_{3-2}|^2\,G_3G_2 \\
= \quad & \frac{3}{4!}\left[\sum_{12}|g_{12}|^2\,|\varphi_{1-2}|^2\,G_1G_2\right]^2 \\
= \quad & \frac{1}{2}\left[-\frac{1}{2}\sum_{12}|g_{12}|^2\,|\varphi_{1-2}|^2\,G_1G_2\right]^2
\end{aligned}$$

so this provides the next term in the reëxponentiation of the Order 2 (quadratic) term.

The other terms add to give

$$\begin{aligned}
\frac{1}{4!}S^4_{int} \quad = \quad & \frac{-2}{4!}\sum_{1234}G_1G_2G_3G_4\mathfrak{Re}\left[g_{12}g_{23}g_{34}g_{41}\varphi_{1-2}\varphi_{2-3}\varphi_{3-4}\varphi_{4-1}\right. \\
& +g_{12}g_{24}g_{31}g_{43}\varphi_{1-2}\varphi_{2-4}\varphi_{3-1}\varphi_{4-3} \\
& \left.+g_{14}g_{23}g_{31}g_{42}\varphi_{1-4}\varphi_{2-3}\varphi_{3-1}\varphi_{4-2}\right] \\
= \quad & \frac{-3\times 2}{4!}\sum_{1234}G_1G_2G_3G_4\mathfrak{Re}\left[g_{12}g_{23}g_{34}g_{41}\varphi_{1-2}\varphi_{2-3}\varphi_{3-4}\varphi_{4-1}\right].
\end{aligned}$$

Effective action

Combining the results we have that

$$
\begin{aligned}
\mathscr{Z} &= \int \mathscr{D}\psi\mathscr{D}\varphi \exp\left(-S_\psi - S_\varphi\right) \Bigg[1 - \frac{1}{2}\sum_{12} \left|g_{12}\right|^2 \left|\varphi_{1-2}\right|^2 G_1 G_2 \\
&+ \frac{1}{2}\left(-\frac{1}{2}\sum_{12} \left|g_{12}\right|^2 \left|\varphi_{1-2}\right|^2 G_1 G_2 \right)^2 \\
&- \frac{1}{3}\sum_{123} \mathfrak{Re}\left(g_{12} g_{23} g_{31} \varphi_{1-2} \varphi_{2-3} \varphi_{3-1} \right) G_1 G_2 G_3 \\
&- \frac{1}{4}\sum_{1234} \mathfrak{Re}\left(g_{12} g_{23} g_{34} g_{41} \varphi_{1-2} \varphi_{2-3} \varphi_{3-4} \varphi_{4-1} \right) G_1 G_2 G_3 G_4 + \mathcal{O}\left(g^5\right) \Bigg] .
\end{aligned}
$$

To this order this is equal to

$$
\mathscr{Z} = \int \mathscr{D}\psi\mathscr{D}\varphi \exp\left(-S_\psi\right) \exp\left(-S_\varphi^{eff}\right)
$$

with

$$
\begin{aligned}
S_\varphi^{eff} &= \frac{1}{2}\sum_q \varphi_q \left(\frac{\omega_n^2}{\Omega} + \Omega \right) \varphi_{-q} + \frac{1}{2}\sum_{12} \left|g_{12}\right|^2 \left|\varphi_{1-2}\right|^2 G_1 G_2 \\
&+ \frac{1}{3}\sum_{123} \mathfrak{Re}\left(g_{12} g_{23} g_{31} \varphi_{1-2} \varphi_{2-3} \varphi_{3-1} \right) G_1 G_2 G_3 \\
&+ \frac{1}{4}\sum_{1234} \mathfrak{Re}\left(g_{12} g_{23} g_{34} g_{41} \varphi_{1-2} \varphi_{2-3} \varphi_{3-4} \varphi_{4-1} \right) G_1 G_2 G_3 G_4
\end{aligned}
$$

or

$$
\begin{aligned}
S_\varphi^{eff} &= \frac{1}{2}\sum_q \varphi_q \left(\frac{\omega_n^2}{\Omega} + \Omega + \sum_k \left|g_{k,k+q}\right|^2 G_k G_{k+q} \right) \varphi_{-q} \\
&+ \frac{1}{3}\sum_{kqp} \mathfrak{Re}\left(g_{k+p+q,k+p} g_{k+p,k} g_{k,k+p+q} \varphi_q \varphi_p \varphi_{-p-q} \right) G_{k+p+q} G_{k+p} G_k \\
&+ \frac{1}{4}\sum_{kpql} \Bigg(\mathfrak{Re}\left(g_{k+l+p+q,k+l+p} g_{k+l+p,k+l} g_{k+l,k} g_{k,k+l+p+q} \varphi_q \varphi_p \varphi_l \varphi_{-l-p-q} \right) \cdot \\
&\cdot G_{k+l+p+q} G_{k+l+p} G_{k+l} G_k \Bigg) .
\end{aligned}
$$

The complex phase of the combination $g\varphi$ can be shifted onto either g or φ individually.

Choosing the former, *i.e.* real phonon displacement fields, Hermiticity of the Hamiltonian requires

$$g_{k,k'} = g^*_{k',k}$$

and noting that the labels on the strings of gs form closed loops we see that taking the real part is gratuitous. The reality of these expressions follows simply from gauge invariance of the action.

References

[1] R. Peierls. *More Surprises in Theoretical Physics.* Princeton University Press, 1991.

[2] G. Grüner. *Rev. Mod. Phys.*, 60(4), 1988.

[3] W. Kohn. *Phys. Rev. Lett.*, 2:393–394, 1959.

[4] A. Melikyan and M. R. Norman. Symmetry of the charge density wave in cuprates (supplementary material). *Phys. Rev. B*, 89:024507, 2014.

[5] A. Melikyan and M. R. Norman. *Phys. Rev. B*, 89:024507, 2014.

[6] G. Grüner. *Density Waves in Solids.* Perseus (Cambridge Mass.), 1994.

[7] R. Shankar. *Rev. Mod. Phys.*, 66:129–192, 1994.

[8] A. M. Gabovich, A. I. Voitenko, J. F. Annett, and M. Ausloos. *Supercond. Sci. Technol.*, 14:R1–R27, 2001.

[9] J. Solyom. *Adv. Phys.*, 28(2):201–303, 1979.

[10] J. R. Schrieffer. *Theory of Superconductivity.* Westview Press, 1964.

[11] J. A. Wilson, F. J. Di Salvo, and S. Mahajan. *Adv. Phys.*, 24:117, 1975.

[12] J. L. Hodeau, M. Marezio, C. Roucau, R. Ayroles, A. Meerschaut, J. Rouxel, and P. Monceau. *J. Phys. C*, 11:4117, 1978.

[13] C. F. Eagen, S. A. Werner, and R. B. Saillant. *Phys. Rev. B*, 12:2036, 1975.

[14] F. Denoyer, F. Comès, A. F. Garito, and A. J. Heeger. *Phys. Rev. Lett.*, 35:445, 1975.

[15] D. Jérome. *Chem. Rev.*, 104:5565, 2004.

[16] D. R. Hofstadter. *Phys. Rev. B*, 14:2239–2249, 1976.

[17] C. Nayak. *Phys. Rev. B*, 62:4880–4889, 2000.

[18] J. B. Marston and I. Affleck. *Phys. Rev. B*, 39:11538–11558, 1989.

[19] M.G.V. Bogle, J.E. Hearst, V.F.R. Jones, and L. Stoilov. *Journal of Knot Theory and Its Ramifications*, 3(2):121–140, 1994.

[20] A. Sossinsky. *Knots: Mathematics With a Twist*. Harvard University Press, 2002.

[21] D. S. Richeson. *Euler's Gem: The Polyhedron Formula and the Birth of Topology*. Princeton University Press, 2008.

[22] C. R. Dean, L. Wang, P. Maher, C. Forsythe, F. Ghahari, Y. Gao, J. Katoch, M. Ishigami, P. Moon, M. Koshino, T. Taniguchi, K. Watanabe, K. L. Shepard, J. Hone, and P. Kim. *Nature*, 497:598–602, 2013.

[23] Y. E. Kraus, Y. Lahini, Z. Ringel, M. Verbin, and O. Zilberberg. *Phys. Rev. Lett.*, 109:106402, 2012.

[24] E. H. Hall. *American Journal of Mathematics*, 2(3):287–292, 1879.

[25] J. (Ed.) Hajdu. *Introduction to the theory of the integer quantum hall effect*. VCH, Weinheim, 1994.

[26] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs. *Phys. Rev. Lett.*, 49:405–408, 1982.

[27] X.-G. Wen. *Quantum Field Theory of Many-Body Systems*. Oxford University Press, 2004.

[28] M. V. Berry. *Proc. R. Soc. A*, 392:45–57, 1984.

[29] Y. Aharonov and D. Bohm. *Phys. Rev.*, 115:485–491, 1959.

[30] J. Zak. *Phys. Rev. Lett.*, 62:2747–2750, 1989.

[31] D. J. Thouless. *Topological Quantum Numbers in Nonrelativistic Physics*. World Scientific, 1997.

[32] D. Calvani, A. Cuccoli, N. I. Gidopoulos, and P. Verrucchi. *Proceedings of the National Academy of Sciences*, 110(17):6748–6753, 2013.

[33] M.-C. Chang and Q. Niu. *Phys. Rev. Lett.*, 75:1348–1351, 1995.

[34] M.-C. Chang and Q. Niu. *Phys. Rev. B*, 53:7010–7023, 1996.

[35] M. Kohmoto. *Ann. Phys.*, 160:343–354, 1985.

[36] R. D. King-Smith and D. Vanderbilt. *Phys. Rev. B*, 47:1651–1654, 1993.

[37] D. J. Thouless. *Surface Science*, 142:147–154, 1984.

[38] C. M. Bender and S. A. Orszag. *Advanced Mathematical Methods for Scientists and Engineers: Asymptotic Methods and Perturbation Theory.* Springer, 1999.

[39] M. Senechal. *Quasicrystals and Geometry.* Cambridge University Press, 1995.

[40] J. E. S. Socolar and P. J. Steinhardt. *Phys. Rev. B*, 34:617–647, 1986.

[41] C. Janot. *Quasicrystals: A Primer.* Clarendon Press (Oxford), 1994.

[42] N. W. Ashcroft and N. D. Mermin. *Solid State Physics.* Holt, Rinehart and Winston, 1976.

[43] D. Shechtman, I. Blech, D. Gratias, and J. W. Cahn. *Phys. Rev. Lett.*, 53:1951–1953, 1984.

[44] V. Bargmann. *Journal of Mathematical Physics*, 5(7), 1964.

[45] M. Z. Hasan and C. L. Kane. *Rev. Mod. Phys.*, 82:3045–3067, 2010.

[46] F. J. Dyson. *Journal of Mathematical Physics*, 3(6), 1962.

[47] M. Zirnbauer. Symmetry classes of disordered fermions and topological insulators. *Sommerfeld Theory Colloquium at LMU Munich.*

[48] M. E. Peskin and D. V. Schroeder. *An Introduction to Quantum Field Theory.* Westview Press, 1995.

[49] A. Altland and M. R. Zirnbauer. *Phys. Rev. B*, 55:1142–1161, 1997.

[50] S. Ryu, A. P. Schnyder, A. Furusaki, and A. W. W. Ludwig. *New Journal of Physics*, 12(6):065010, 2010.

[51] J. S. Baez. The ten-fold way. *math.ucr.edu/home/baez/tenfold.html, retrieved 14/5/15.*

[52] G. W. Moore. Quantum symmetries and compatible hamiltonians. *Notes for Physics 695, Rutgers University.*

[53] C. L. Kane and E. J. Mele. *Phys. Rev. Lett.*, 95:146802, 2005.

[54] L. Fu and C. L. Kane. *Phys. Rev. B*, 76:045302, 2007.

[55] S. Ryu. Classification of topological insulators and superconductors and d-branes. *Lecture Notes.*

[56] R.-J. Slager, A. Mesaros, V. Juricic, and J. Zaanen. *Nature Phys.*, 9:98, 2013.

[57] C. Fang, M. J. Gilbert, and B. A. Bernevig. *Phys. Rev. B*, 86:115112, 2012.

[58] T. Morimoto and A. Furusaki. *Phys. Rev. B*, 88:125129, 2013.

[59] C.-K. Chiu, H. Yao, and S. Ryu. *Phys. Rev. B*, 88:075142, 2013.

[60] Y. E. Kraus and O. Zilberberg. *Phys. Rev. Lett.*, 109:116404, 2012.

[61] M. Verbin, O. Zilberberg, Y. Lahini, Y. E. Kraus, and Y. Silberberg. *Phys. Rev. B*, 91:064201, 2015.

[62] Y. E. Kraus, Z. Ringel, and O. Zilberberg. *Phys. Rev. Lett.*, 111:226401, 2013.

[63] A. Quandt. *Physics*, 5:99, 2012.

[64] K. A. Madsen, E. J. Bergholtz, and P. W. Brouwer. *Phys. Rev. B*, 88:125118, 2013.

[65] A. Avila and S. Jitomirskaya. *Annals of Mathematics*, 170:303–342, 2009.

[66] B. M. Tahar and H. Oyono-Oyono. *Gap-labelling for quasicrystals (proving a conjecture by J. Bellissard) in proc. of Operator Algebras and Mathematical Physics (Constanta 2001).*

[67] J. Kaminker and I. Putnam. *Michigan Mathematical Journal*, 51(3), 2003.

[68] W. L. McMillan. *Phys. Rev. B*, 12:1187–1196, 1975.

[69] W. L. McMillan. *Phys. Rev. B*, 14:1496–1502, 1976.

[70] W. L. McMillan. *Phys. Rev. B*, 16(2):643–650, 1977.

[71] D. E. Moncton, J. D. Axe, and F. J. DiSalvo. *Phys. Rev. Lett.*, 34:734–737, 1975.

[72] S. Kagoshima, T. Ishiguro, and H. Anzai. *J. Phys. Soc. Jap.*, 41:2061–2071, 1976.

[73] H. Fukuyama and P. A. Lee. *Phys. Rev. B*, 17:535–541, 1978.

[74] J. Bardeen. *Physica Scripta*, T27:136–143, 1989.

[75] H. Fukuyama. *J. Phys. Soc. Jpn.*, 41(2), 1976.

[76] H. Fukuyama. *J. Phys. Soc. Jpn.*, 45(5), 1978.

[77] C.-Y. Lai, C. Shi, and S.-W. Tsai. *Phys. Rev. B*, 87:075134, 2013.

[78] M. Aidelsburger, M. Lohse, C. Schweizer, M. Atala, J. T. Barreiro, S. Nascimbene, N. R. Cooper, I. Bloch, and N. Goldman. Revealing the topology of hofstadter bands with ultracold bosonic atoms. *arxiv:1407.4205 [cond-mat.quant-gas]*.

[79] K. Singh and D. M. Weld. Fibonacci optical lattices for tunable quantum quasicrystals. *arXiv:1504.06769 [cond-mat.quant-gas]*.

[80] J. Bellissard and B. Simon. *J. Funct. Anal.*, 48:408–419, 1982.

[81] J. F. Lindner, V. Kohar, B. Kia, M. Hippke, J. G. Learned, and W. L. Ditto. *Phys. Rev. Lett.*, 114:054101, 2015.

[82] E. J. Bergholtz and A. Karlhede. *Phys. Rev. B*, 77:155308, 2008.

[83] J. Hubbard. *Phys. Rev. B*, 17:494–505, 1978.

[84] L. Bindi, P. J. Steinhardt, N. Yao, and P. J. Lu. *Science*, 324(5932):1306–1309, 2009.

[85] L. Bindi, N. Yao, C. Lin, L. S. Hollister, C. L. Andronicos, V. V. Distler, M. P. Eddy, A. Kostin, V. Kryachko, G. J. MacPherson, W. M. Steinhardt, M. Yudovskaya, and P. J. Steinhardt. *Sci. Rep.*

[86] G. Wexler and A. M. Woolley. *J. Phys. C: Cond. Mat. Phys.*, 9:1189–1200, 1976.

[87] N. J. Doran, D. J. Titterington, B. Ricco, and G. Wexler. *J. Phys. C: Solid State Phys.*, 11(4):685, 1978.

[88] N. J. Doran, D. Titterington, B. Ricco, M. Schreiber, and G. Wexler. *Journal of Physics C: Solid State Physics*, 11(4):699, 1978.

[89] J. van Wezel, P. Nahai-Williamson, and S. S. Saxena. *Europhys. Lett.*, 89:47004, 2010.

[90] J. van Wezel. *Europhys. Lett.*, 96:67011, 2011.

[91] J. E. Hoffman, E. W. Hudson, K. M. Lang, V. Madhavan, H. Eisaki, S. Uchida, and J. C. Davis. *Science*, 295(5554):466–469, 2002.

[92] J. E. Hoffman, K. McElroy, D.-H. Lee, K. M Lang, H. Eisaki, S. Uchida, and J. C. Davis. *Science*, 297(5584):1148–1151, 2002.

[93] C. V. Parker, P. Aynajian, E. H. da Silva Neto, A. Pushp, S. Ono, J. Wen, Z. Xu, G. Gu, and A. Yazdani. *Nature*, 468:677–680, 2010.

[94] P. Doan, M. Gooch, Z. Tang, B. Lorenz, A. Möller, J. Tapp, P. C. W. Chu, and A. M. Guloy. *J. Am. Chem. Soc.*, 134:16520, 2012.

[95] K. Rossnagel, O. Seifarth, L. Kipp, M. Skibowski, D. Voß, P. Krüger, A. Mazur, and J. Pollmann. *Phys. Rev. B*, 64:235119, 2001.

[96] J. Chang, E. Blackburn, A. T. Holmes, N. B. Christensen, J. Larsen, J. Mesot, Ruixing Liang, D. A. Bonn, W. N. Hardy, A. Watenphul, M. v. Zimmermann, E. M. Forgan, and S. M. Hayden. *Nature Physics*, 8:871–876, 2012.

[97] E. Fawcett. *Rev. Mod. Phys.*, 60:209–283, 1988.

[98] D. G. Laurent, J. Callaway, J. L. Fry, and N. E. Brener. *Phys. Rev. B*, 23:4977–4987, 1981.

[99] N. J. Doran. *J. Phys. C: Solid State Phys.*, 11, 1978.

[100] D. E. Moncton, J. D. Axe, and F. J. DiSalvo. *Phys. Rev. B*, 16:801–819, 1977.

[101] T. Kiss, T. Yokoya, A. Chainani, S. Shin, T. Hanaguri, M. Nohara, and H. Takagi. *Nature Physics*, 3(10):720–725, 2007.

[102] A. Soumyanarayanan, Michael M. Yee, Y. He, J. van Wezel, D. J. Rahn, K. Rossnagel, E. W. Hudson, M. R. Norman, and J. E. Hoffman. *Proceedings of the National Academy of Sciences*, 110(5):1623–1627, 2013.

[103] F. Weber, S. Rosenkranz, J.-P. Castellan, R. Osborn, R. Hott, R. Heid, K.-P. Bohnen, T. Egami, A. H. Said, and D. Reznik. *Phys. Rev. Lett.*, 107:107403, 2011.

[104] F. Weber, R. Hott, R. Heid, K.-P. Bohnen, S. Rosenkranz, J.-P. Castellan, R. Osborn, A. H. Said, B. M. Leu, and D. Reznik. *Phys. Rev. B*, 87:245111, 2013.

[105] S. V. Borisenko, A. A. Kordyuk, V. B. Zabolotnyy, D. S. Inosov, D. Evtushinsky, B. Büchner, A. N. Yaresko, A. Varykhalov, R. Follath, W. Eberhardt, L. Patthey, and H. Berger. *Phys. Rev. Lett.*, 102:166402, 2009.

[106] D. J. Rahn, S. Hellmann, M. Kalläne, C. Sohrt, T. K. Kim, L. Kipp, and K. Rossnagel. *Phys. Rev. B*, 85:224532, 2012.

[107] H. F. Hess, R. B. Robinson, and J. V. Waszczak. *Physica B: Condensed Matter*, 169(1-4):422–431, 1991.

[108] C. J. Arguello, S. P. Chockalingam, E. P. Rosenthal, L. Zhao, C. Gutiérrez, J. H. Kang, W. C. Chung, R. M. Fernandes, S. Jia, A. J. Millis, R. J. Cava, and A. N. Pasupathy. *Phys. Rev. B*, 89:235115, 2014.

[109] U. Chatterjee, J. Zhao, M. Iavarone, R. Di Capua, J. P. Castellan, G. Karapetrov, C. D. Malliakas, M. G. Kanatzidis, H. Claus, J. P. C. Ruff, F. Weber, J. van Wezel, J. C. Campuzano, R. Osborn, M. Randeria, N. Trivedi, M. R. Norman, and S. Rosenkranz. *Nature Communications*, 6:6313, 2015.

[110] M. R. Norman, D. Pines, and C. Kallin. *Advances in Physics*, 54(8):715–733, 2005.

[111] T. M. Rice and G. K. Scott. *Phys. Rev. Lett.*, 35:120–123, 1975.

[112] M. D. Johannes, I. I. Mazin, and C. A. Howells. *Phys. Rev. B*, 73:205102, 2006.

[113] M. Iavarone, R. Di Capua, X. Zhang, M. Golalikhani, S. A. Moore, and G. Karapetrov. *Phys. Rev. B*, 85:155103, 2012.

[114] O. Anderson, R. Manzke, and M. Skibowski. *Phys. Rev. Lett.*, 55:2188–2191, 1985.

[115] T. Pillo, J. Hayoz, H. Berger, F. Lévy, L. Schlapbach, and P. Aebi. *Phys. Rev. B*, 61:16213–16222, 2000.

[116] C. H. Chen, W. Fabian, F. C. Brown, K. C. Woo, B. Davies, B. DeLong, and A. H. Thompson. *Phys. Rev. B*, 21:615–624, 1980.

[117] T. E. Kidd, T. Miller, M. Y. Chou, and T.-C. Chiang. *Phys. Rev. Lett.*, 88:226402, 2002.

[118] B. Zenker, H. Fehske, H. Beck, C. Monney, and A. R. Bishop. *arXiv:1304.1965 [cond-mat.str-el].*

[119] A. A. Kordyuk. *Low Temp. Phys.*, 40:286, 2014.

[120] A. Damascelli, Z. Hussain, and Z.-X. Shen. *Rev. Mod. Phys.*, 75:473–541, 2003.

[121] H. J. W. Zandvliet and A. van Houselt. *Annual Review of Analytical Chemistry*, 2:37–55, 2009.

[122] E. Blackburn, J. Chang, A. H. Said, B. M. Leu, Ruixing Liang, D. A. Bonn, W. N. Hardy, E. M. Forgan, and S. M. Hayden. *Phys. Rev. B*, 88:054506, 2013.

[123] G. W. F. (Ed.) Drake. *Atomic, Molecular, and Optical Physics Handbook.* AIP, Woodbury, NY, 1996.

[124] J. C. Slater and G. F. Koster. *Phys. Rev.*, 94:1498–1524, 1954.

[125] C. M. Varma, E. I. Blount, P. Vashishta, and W. Weber. *Phys. Rev. B*, 19:6130–6141, 1979.

[126] G. D. Gaspari and B. L. Györffy. *Phys. Rev. Lett.*, 28:801–805, 1972.

[127] A. Altland and B. D. Simons. *Condensed Matter Field Theory.* Cambridge University Press, 2010.

[128] G. D. Mahan. *Many Particle Physics.* Plenum Press, New York, 1990.

[129] Y. Feng, J. van Wezel, J. Wang, F. Flicker, D. M. Silevitch, P. B. Littlewood, and T. F. Rosenbaum. *in preparation.*

[130] C. M. Varma and A. L. Simons. *Phys. Rev. Lett.*, 51:138–141, 1983.

[131] H. Yoshiyama, Y. Takaoka, N. Suzuki, and K. Motizuki. *J. Phys. C: Solid State Phys.*, 19:5591, 1986.

[132] J. M. E. Harper, T. H. Geballe, and F. J. Di Salvo. *Phys. Lett. A*, 54(1):27–28, 1975.

[133] T. M. Rice. *Festkörperprobleme*, XX:393–411, 1980.

[134] H. J. Conley, B. Wang, J. I. Ziegler, R. F. Haglund, S. T. Pantelides, and K. I. Bolotin. *Nano Letters*, 13(8):3626–3630, 2013.

[135] N. Mounet and N. Marzari. *Phys. Rev. B*, 71:205214, 2005.

[136] J. E. Inglesfield. *Journal of Physics C: Solid State Physics*, 13(1):17, 1980.

[137] L. D. Landau, E. M. Lifshitz, and L. P. Pitaevskii. *Electrodynamics of Continuous Media.* Butterworth Heinemann, 1981.

[138] A. V. Chubukov, D. L. Maslov, and V. I. Yudson. *arXiv:1401.1461 [cond-mat.str-el].*

[139] A. B. Migdal. *Soviet Phys. JETP*, 7, 1958.

[140] M. V. Sadovskii. *Diagrammatics.* World Scientific, 2006.

[141] P. A. Lee, T. M. Rice, and P. W. Anderson. *Phys. Rev. Lett.*, 31(7):462–465, 1973.

[142] M. J. Lawler, K. Fujita, J. Lee, A. R. Schmidt, Y. Kohsaka, C. K. Kim, H. Eisaki, S. Uchida, J. C. Davis, J. P. Sethna, and E.-A. Kim. *Nature*, 466:347–351, 2010.

[143] A. Hackl, M. Vojta, and S. Sachdev. *Phys. Rev. B*, 81:045182, 2010.

[144] R. Daou, J. Chang, D. LeBoeuf, O. Cyr-Choiniere, F. Laliberte, N. Doiron-Leyraud, B. J. Ramshaw, R. Liang, D. A. Bonn, W. N. Hardy, and L. Taillefer. *Nature*, 463:519–522, 2010.

[145] M. Hashimoto, R.-H. He, K. Tanaka, J.-P. Testaud, W. Meevasana, R. G. Moore, D. Lu, H. Yao, Y. Yoshida, H. Eisaki, T. P. Devereaux, Z. Hussain, and Z.-X. Shen. *Nat. Phys.*, 6:414–418, 2010.

[146] Y. Nambu. *Phys. Rev.*, 117:648–663, 1960.

[147] L. P. Gor'kov. *Soviet. Phys. JETP*, 9:1364, 1959.

[148] N. N. Bogoliubov. *Soviet Phys.*, 34, 1958.

[149] J. Bardeen, L. N. Cooper, and J. R. Schrieffer. *Phys. Rev.*, 106:162–164, 1957.

[150] M. R. Norman, H. Ding, H. Fretwell, M. Randeria, and J. C. Campuzano. *Phys. Rev. B*, 60:7585–7590, 1999.

[151] G. Ghiringhelli, M. Le Tacon, M. Minola, S. Blanco-Canosa, C. Mazzoli, N. B. Brookes, G. M. De Luca, A. Frano, D. G. Hawthorn, F. He, T. Loew, M. Moretti Sala, D. C. Peets, M. Salluzzo, E. Schierle, R. Sutarto, G. A. Sawatzky, E. Weschke, B. Keimer, and L. Braicovich. *Science*, 337:821, 2012.

[152] T. Timusk and B. Statt. *Reports on Progress in Physics*, 62:61–122, 1999.