4 The Variational Cluster Approximation

Robert Eder
Institut für Festkörperphysik
Karlsruhe Institute of Technology

Contents

E. Pavarini, E. Koch, and U. Schollwöck
Emergent Phenomena in Correlated Matter
Modeling and Simulation Vol. 3
Forschungszentrum Jülich, 2013, ISBN 978-3-89336-884-6
http://www.cond-mat.de/events/correl13

1 Introduction

A quantity of central importance for the description of correlated electron systems is the electronic self-energy $\boldsymbol{\Sigma}(\boldsymbol{k},\omega)$. It may be viewed as a momentum- and energy-dependent correction to the energy of an electron that describes the effects of its interaction with the other electrons. Here the word 'correction' is by no means supposed to imply that $\boldsymbol{\Sigma}(\boldsymbol{k},\omega)$ is small. Quite the contrary, for example, in a Mott-insulator $\boldsymbol{\Sigma}(\boldsymbol{k},\omega)$ contains a term of the form $U/\hbar(\omega-\omega_0)$, with U the intra-atomic Coulomb repulsion, and this term is both large and strongly dependent on the energy ω. In fact, the very reason why density functional calculations do not reproduce the single-particle excitation spectrum – or 'band structure' – of Mott-insulators is that they implicitly assume an ω-independent self-energy and thus miss this 'correction' of order U.
It would therefore seem desirable to have a theoretical principle that allows us to actually compute the self-energy of a correlated electron system, and in fact it can be shown that $\boldsymbol{\Sigma}(\boldsymbol{k},\omega)$ obeys a stationarity condition that can be used for that purpose. More precisely, Luttinger and Ward have shown in a seminal paper [1] that the grand canonical potential Ω of an interacting fermion system can be expressed as a functional of its self-energy, $\Omega = \Omega[\boldsymbol{\Sigma}]$, and that this functional is stationary with respect to variations of $\boldsymbol{\Sigma}$:

$$\frac{\delta\Omega}{\delta\boldsymbol{\Sigma}(\boldsymbol{k},\omega)} = 0\ .$$

Unfortunately a straightforward application of the stationarity property – e.g. by introducing 'trial self-energies' that depend on a number of variational parameters – is not possible because $\Omega[\boldsymbol{\Sigma}]$ involves the so-called Luttinger-Ward functional $F[\boldsymbol{\Sigma}]$, which is defined as a sum over infinitely many Feynman diagrams and thus cannot be evaluated for a given trial self-energy. A possible approximation would be to truncate the Luttinger-Ward functional, thereby keeping only a selected class of Feynman diagrams, typically 'bubbles' or 'ladders'. These are the famous conserving approximations of Baym and Kadanoff [2] and one example for such an approximation is the very successful GW-approximation proposed by Hedin [3]. On the other hand, the truncation of the Luttinger-Ward functional ultimately is a poorly controlled approximation that may be less suitable for strongly correlated electrons.
In 2003, however, an entirely new idea on how to apply the stationarity principle for $\boldsymbol{\Sigma}$ in strongly correlated electron systems was put forward by Potthoff, the so-called variational cluster approximation (VCA) [4–6]. The basic idea of the VCA is to generate trial self-energies $\boldsymbol{\Sigma}$ for an infinite lattice by exact diagonalization of finite clusters and, in the course of doing so, to evaluate the exact value of the Luttinger-Ward functional $F[\boldsymbol{\Sigma}]$ numerically. Variation of $\boldsymbol{\Sigma}$ is performed by varying the single-particle terms of a cluster that serves as the 'self-energy preparation-lab'. Put another way, the VCA seeks the best approximation to the self-energy of an infinite lattice amongst 'cluster-representable' ones, i.e. functions $\boldsymbol{\Sigma}(\boldsymbol{k},\omega)$ which can be generated as the exact self-energies of finite clusters. This is a new way for generating approximations in strongly correlated electron systems and in the following the variational principle itself, the basic idea of the VCA, and some selected applications will be presented.

2 Notation and brief review of field theory

First we define the notation and give a brief review of some concepts from field theory. While this will be rather cursory, introductions to the use of field theory in statistical physics can be found in many textbooks [7–9], in the present notes we try to be consistent with Fetter and Walecka (FW) [8].

We assume that the solid in question can be described as a periodic array of atomic orbitals centered on the nuclei of the atoms that form the basis of the lattice and we assume periodic boundary conditions. We choose the unit of length such that the unit cell has volume 1. All orbitals are taken as mutualy orthogonal. The number of unit cells in the crystal is N and the number of atoms in the basis n_{Atom}. The orbitals can be labeled by a triple of indices (i, n, ν) where $i \in \{1, \ldots, N\}$ denotes the unit cell, $n \in \{1, \ldots, n_{Atom}\}$ the basis atom and $\nu \in \{s, p_x, p_y, p_z, d_{xy} \ldots\}$ the type of orbital. The number of orbitals per unit cell is n_{orb} [10]. We introduce fermionic creation and annihilation operators, $c^\dagger_{i,n,\nu,\sigma}$ and $c_{i,n,\nu,\sigma}$, for electrons in these orbitals, where σ denotes the z-component of spin. It will often be convenient to contract (i, n, ν, σ) into a single 'compound index' α, so that the Hamiltonian – assumed to be time-independent – can be written as $H = H_0 + H_1$, with

$$H_0 = \sum_{\alpha,\beta} t_{\alpha,\beta}\, c^\dagger_\alpha c_\beta, \tag{1}$$

$$H_1 = \frac{1}{2} \sum_{\alpha,\beta,\gamma,\delta} V_{\alpha,\beta,\delta,\gamma}\, c^\dagger_\alpha c^\dagger_\beta c_\gamma c_\delta\ . \tag{2}$$

Note the factor of $1/2$ and the 'inverted' order of indices on the interaction matrix element V in (2) which follows from the prescription of second quantization [7–9], see e.g. the lecture of E. Koch. The Fourier transform of the Fermion operators reads

$$c^\dagger_{\mathbf{k},\beta} = \frac{1}{\sqrt{N}} \sum_i e^{i\mathbf{k}\cdot(\mathbf{R}_i + \mathbf{r}_n)}\, c^\dagger_{i,n,\nu,\sigma}\ ,$$

where we have introduced the orbital index $\beta = (n, \nu, \sigma)$. Since this second 'compound index' always comes together with either a momentum $\mathbf{k}$ or a cell index i no misunderstanding is possible. The Hamiltonian now can be written as

$$H_0 = \sum_{\mathbf{k}} \sum_{\alpha,\beta} \mathbf{t}_{\alpha,\beta}(\mathbf{k})\, c^\dagger_{\mathbf{k},\alpha}\, c_{\mathbf{k},\beta}, \tag{3}$$

$$H_1 = \frac{1}{2N} \sum_{\mathbf{k},\mathbf{k}',\mathbf{q}} \sum_{\alpha,\beta,\gamma,\delta} V_{\alpha,\beta,\delta,\gamma}(\mathbf{k}, \mathbf{k}', \mathbf{q})\, c^\dagger_{\mathbf{k}+\mathbf{q},\alpha}\, c^\dagger_{\mathbf{k}'-\mathbf{q},\beta}\, c_{\mathbf{k}',\gamma}\, c_{\mathbf{k},\delta}\ . \tag{4}$$

Equation (3) defines the $2n_{orb} \times 2n_{orb}$ matrix $\mathbf{t}(\mathbf{k})$, whose eigenvalues $E_n(\mathbf{k})$ give the noninteracting band structure. This formulation allows H_0 to describe magnetic systems or include spin-orbit coupling. With the explicit prefactor of $1/N$ in (4) the matrix elements V in (4) are of order 1.

In all that follows we consider a grand canonical ensemble with inverse temperature $\beta = 1/k_B T$ and chemical potential μ. The thermal average of any operator $\hat{O}$ is given by

$$\langle \hat{O} \rangle_{th} = \frac{1}{Z} \mathrm{Tr} \left(e^{-\beta(H-\mu N)} \hat{O} \right) \tag{5}$$

with the grand partition function

$$Z = \mathrm{Tr} \left(e^{-\beta(H-\mu N)} \right). \tag{6}$$

Introducing the imaginary-time Heisenberg operator

$$c_\alpha(\tau) = e^{\tau(H-\mu N)/\hbar} \, c_\alpha \, e^{-\tau(H-\mu N)/\hbar} \, ,$$

the *imaginary time Green's function* is defined as

$$G_{\alpha,\beta}(\tau) = -\Theta(\tau) \left\langle c_\alpha(\tau) \, c_\beta^\dagger \right\rangle_{th} + \Theta(-\tau) \left\langle c_\beta^\dagger \, c_\alpha(\tau) \right\rangle_{th} \tag{7}$$

$$\begin{aligned} = \frac{1}{Z} \Bigg(& - \Theta(+\tau) \sum_{i,j} e^{-\beta(E_i - \mu N_i)} \, e^{\tau(E_i - E_j + \mu)/\hbar} \, \langle i | c_\alpha | j \rangle \langle j | c_\beta^\dagger | i \rangle \\ & + \Theta(-\tau) \sum_{i,j} e^{-\beta(E_i - \mu N_i)} \, e^{\tau(E_j - E_i + \mu)/\hbar} \, \langle i | c_\beta^\dagger \, | j \rangle \langle j | c_\alpha | i \rangle \Bigg) , \end{aligned} \tag{8}$$

where $|i\rangle$ are the exact eigenstates of the system with energies E_i and particle number N_i and $\Theta(\tau)$ is the Heaviside step function. $\mathbf{G}(\tau)$ is a matrix with row dimension $2N n_{orb}$, which can be made block-diagonal by introducing the spatial Fourier transform

$$G_{(n,\nu,\sigma),(n',\nu',\sigma')}(\mathbf{k}, \tau) = \frac{1}{N} \sum_{i,j} e^{i\mathbf{k} \cdot (\mathbf{R}_i - \mathbf{R}_j + \mathbf{r}_n - \mathbf{r}_{n'})} \, G_{(i,n,\nu,\sigma),(j,n',\nu',\sigma')}(\tau) \, ,$$

where $\mathbf{G}(\mathbf{k}, \tau)$ is a $2n_{orb} \times 2n_{orb}$ matrix.
From (8) it is easy to see that $\mathbf{G}$ is well-defined only if $\tau \in [-\beta\hbar, \beta\hbar]$ when E_i are unbounded from above [11], and that for $\tau \in [-\beta\hbar, 0]$ one has $\mathbf{G}(\tau + \beta\hbar) = -\mathbf{G}(\tau)$. It follows that $\mathbf{G}(\tau)$ has the Fourier transform (see equation (25.10) in FW)

$$\begin{aligned} \mathbf{G}(\tau) &= \frac{1}{\beta\hbar} \sum_{\nu=-\infty}^{\infty} e^{-i\omega_\nu \tau} \, \mathbf{G}(i\omega_\nu), \\ \mathbf{G}(i\omega_\nu) &= \int_0^{\beta\hbar} d\tau \, e^{i\omega_\nu \tau} \, \mathbf{G}(\tau) \, , \end{aligned}$$

The $\omega_\nu = (2\nu + 1)\pi/\beta\hbar$ are called the (Fermionic) Matsubara frequencies. From (8) we obtain

$$\begin{aligned} G_{\alpha\beta}(i\omega_\nu) &= \frac{1}{Z} \sum_{i,j} \frac{e^{-\beta(E_i - \mu N_i)} + e^{-\beta(E_j - \mu N_j)}}{i\omega_\nu + \frac{1}{\hbar}\mu - \frac{1}{\hbar}(E_j - E_i)} \, \langle i | c_\alpha | j \rangle \langle j | \, c_\beta^\dagger | i \rangle \\ &= \left\langle c_\alpha \, \frac{1}{i\omega_\nu + \frac{1}{\hbar}\mu - \frac{1}{\hbar} L} \, c_\beta^\dagger \right\rangle_{th} - \left\langle c_\beta^\dagger \, \frac{1}{-i\omega_\nu - \frac{1}{\hbar}\mu - \frac{1}{\hbar} L} \, c_\alpha \right\rangle_{th} \end{aligned} \tag{9}$$

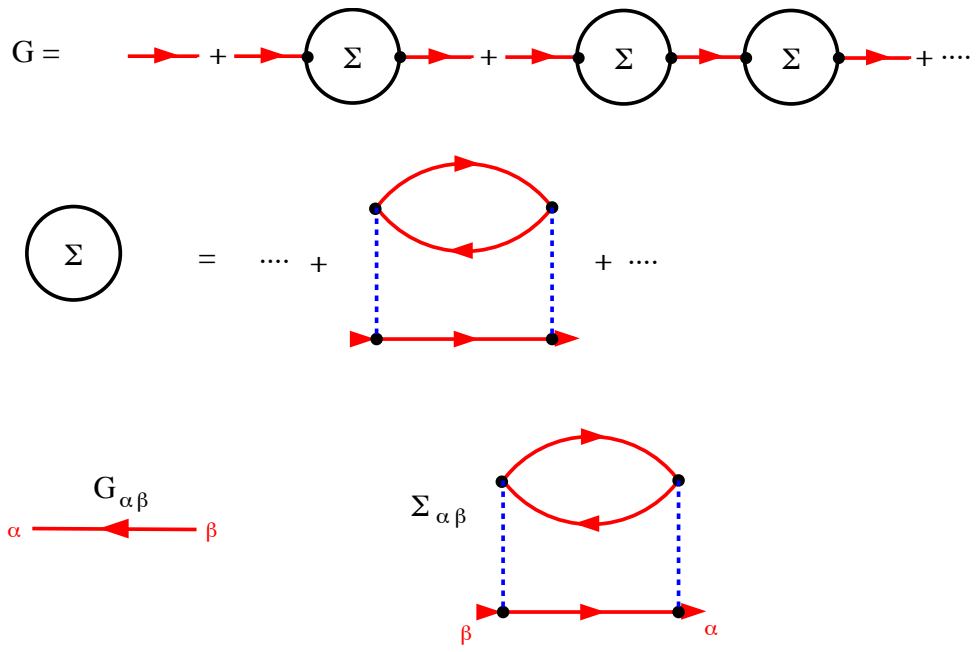

Fig. 1: *Top: Graphical representation of the Dyson equation. Middle: Self-energy diagrams have two open ends. Bottom: The convention for the representation of the Green's function implies the labeling of the entry points of the self-energy.*

where the Liouville operator L is defined by $LX = [H, X]$. When viewed as function of a complex variable z, all elements of $\mathbf{G}(z)$ are analytic in the complex z-plane except for the real axis, where there are poles at $z = (E_j - E_i - \mu)/\hbar$. It is this property on which the usefulness of the imaginary-time Green's function is based: its Fourier transform $\mathbf{G}(z)$ can be analytically continued to a line infinitesimally above the real axis and then gives the Fourier transform of the retarded real-time Green's function – from which single-particle spectral function i.e. the photoemission and inverse photoemission spectrum of a system can be obtained. For this reason the Fourier transform (9) is often called 'the' Green's function and when we speak of the Green's function in the following we always mean (9). The first line in (9) is the *Lehmann representation* of the Green's function.

It is shown in textbooks of field theory [7–9] that the imaginary-time Green's function can be expanded in Feynman diagrams and the self-energy $\mathbf{\Sigma}(\mathbf{k}, i\omega_\nu)$ can be introduced in the standard way, see Figure 1. The self-energy can be expanded in diagrams which have two 'entry points' an incoming and an outgoing one, see Figure 1. Following FW [8], we represent the Green's function $G_{\alpha\beta}$ by a directed line with an arrow running $\beta \to \alpha$ (the reason is that it is the *creation operator* that has the index β, see (7)). In the Dyson equation the orbital indices of the Green's function and the self-energy must take the form of consecutive matrix products, e.g. $G^0_{\delta\alpha}\Sigma_{\alpha\beta}G^0_{\beta\gamma}$ - otherwise the summation of the geometric series would not be possible. It follows that the element $\Sigma_{\alpha\beta}$ must have the label α on the outgoing entry and the label β on the incoming one, see Figure 1. This will be of some importance later on.

Note that the real time Green's function at finite temperature does not allow for a Feynman diagram expansion – this is why the digression of calculating the imaginary time Green's function

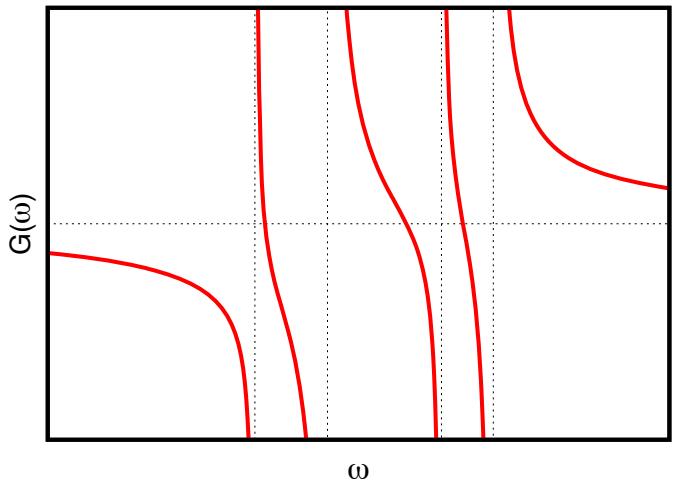

Fig. 2: *Real part of the Green's function $G(\omega)$ for real ω. The dashed vertical lines give the position of the poles, ω_i.*

and analytically continuing its Fourier transform is necessary. It follows from the diagrammatic expansion that the Green's-function obeys the Dyson equation (see (26.5) of FW)

$$\Big(i\omega_\nu + \mu/\hbar - \mathbf{t}(\mathbf{k})/\hbar - \mathbf{\Sigma}(\mathbf{k}, i\omega_\nu)\Big)\mathbf{G}(\mathbf{k}, i\omega_\nu) = 1$$

$$\big(-\partial_\tau + \mu/\hbar - \mathbf{t}(\mathbf{k})/\hbar\big)\mathbf{G}(\mathbf{k},\tau) - \int_0^{\beta\hbar} \mathbf{\Sigma}(\mathbf{k}, \tau-\tau')\ \mathbf{G}(\mathbf{k},\tau')d\tau' = \delta(\tau). \qquad (10)$$

where the second equation is the Fourier-transform of the first and FW (25.21) was used.
Let us finally briefly discuss the analytic structure of the Green's function and self-energy. For simplicity we specialize to a single band and assume that the z-component of spin is a good quantum number so that the Green's function is a scalar. It can be seen from (9) that the Fourier transform of the Green's function has the form

$$G(\omega) = \sum_i \frac{\alpha_i^2}{\omega - \omega_i}$$

where α_i and ω_i are real numbers. It has poles on the real axis and the real part of $G(\omega)$ looks like in Figure 2. This shows that in between any two successive poles ω_i and ω_{i+1} the Green's function crosses zero with a negative slope

$$G(\omega) \approx -\beta_i^2(\omega - \zeta_i).$$

Near the crossing point ζ_i we thus have

$$\begin{aligned}\Sigma(\omega) &= -G^{-1}(\omega) + \omega + \mu - t_{\mathbf{k}} \\ &= \frac{\sigma_i}{\omega - \zeta_i} + \dots\end{aligned}$$

where $\sigma_i = 1/\beta_i^2$. The self-energy thus has poles on the real axis as well, and these poles are 'sandwiched' in between the poles of the Green's function. Luttinger has shown [12] that $\Sigma(\omega)$ is essentially determined by these poles and their residua in that it can be written as

$$\Sigma(\omega) = \eta + \sum_i \frac{\sigma_i}{\omega - \zeta_i} \qquad (11)$$

with a real constant η.

3 Proof of the theorem by Luttinger and Ward

3.1 Statement of the theorem

The grand canonical potential $\Omega(T,\mu)$ contains all thermodynamical information about a system at fixed temperature T and chemical potential μ. It is defined as the logarithm of the grand partition function

$$\begin{aligned} \Omega &= -\frac{1}{\beta}\ln(Z). \\ Z &= \sum_i e^{-\beta(E_i-\mu N_i)}, \end{aligned}$$

where the sum is over all eigenstates of the system with energy E_i and particle number N_i. The latter can indeed be evaluated for noninteracting particles and in this way one obtains for example the grand canonical potential of noninteracting Bloch electrons

$$\Omega_{\mathrm{non-int}} = -\frac{1}{\beta}\sum_{n=1}^{2n_{orb}}\sum_{\mathbf{k}}\ln\left(1+e^{-\beta(E_n(\mathbf{k})-\mu)}\right). \tag{12}$$

As shown in textbooks of statistical mechanics, expression (12) allows one to derive the complete thermodynamics of metals. However, it is in general not possible to evaluate the grand partition function for a system of *interacting particles* of macroscopic size.
Luttinger and Ward, however, derived a relation for for the grand canonical potential of interacting fermions [1]. More precisely, they considered the following quantity

$$\tilde{\Omega} = -\lim_{\eta\to 0^+}\frac{1}{\beta}\sum_{\mathbf{k},\nu} e^{i\omega_\nu\eta}\left(\ln\,\det\Big(-\mathbf{G}^{-1}(\mathbf{k},i\omega_\nu)\Big)+\mathrm{Tr}\Big(\mathbf{G}(\mathbf{k},i\omega_\nu)\mathbf{\Sigma}(\mathbf{k},i\omega_\nu)\Big)\right)+\Phi[\mathbf{G}]\,. \tag{13}$$

Here $\sum_\nu$ denotes summation over the Fermionic Matsubara frequencies and $\Phi[\mathbf{G}]$ is the so-called Luttinger-Ward functional which is defined as a sum over closed, linked Feynman-diagrams (the precise definition will be discussed below). The important point here is that a closed Feynman diagram is simply a number, so that $\Phi[\mathbf{G}]$ indeed assigns a (real) number to each possible Green's function $\mathbf{G}$. Regarding the logarithm of the determinant in (13), we recall that the determinant of a matrix is given by the product of its eigenvalues (the matrix need not be Hermitian for this to be true), so the logarithm of the determinant is the sum of the logarithms of the (complex) eigenvalues of $-\mathbf{G}^{-1}$.

In the following, we want to show that in fact $\tilde{\Omega} = \Omega$, the true grand canonical potential and thereby follow the original proof by Luttinger and Ward. The basic idea is to multiply the interaction part of the Hamiltonian, (2), by a scale factor, $H_1 \to \lambda H_1$, then show $\tilde{\Omega} = \Omega$ for $\lambda = 0$ – i.e. the noninteracting limit – and next show that $\partial_\lambda\tilde{\Omega} = \partial_\lambda\Omega$. Obviously, this proves the identity of the two expressions for any λ.

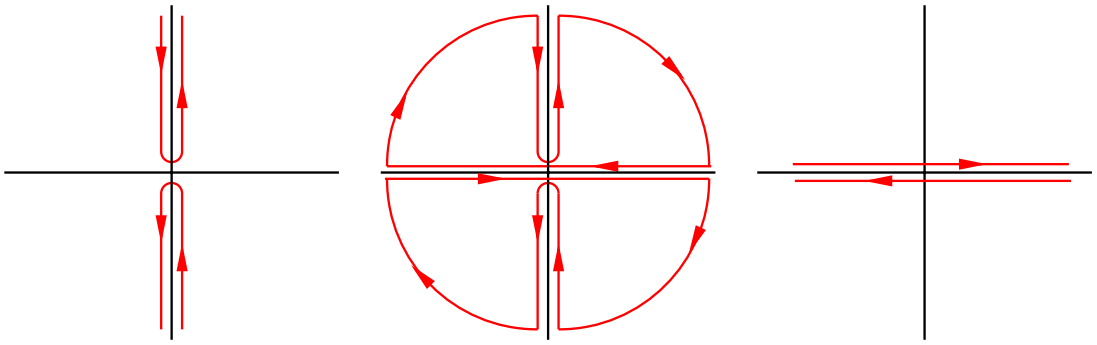

Fig. 3: *(a) Integration contour $\mathcal{C}$ used in (15). Since the integrals along the two contours in (b) are zero and the contributions from the circular arcs vanish, the integral along the contour in (a) is equal to that over the contour $\mathcal{C}'$ in (c).*

3.2 The case $\lambda = 0$

In this limit, $\mathbf{\Sigma} = 0$ and $\Phi[\mathbf{G}] = 0$ (the latter property follows because all interaction lines in all diagrams are zero) so that only the first term in (13) remains and

$$\begin{aligned} \mathbf{G}^{-1}(\mathbf{k},\omega) &= \omega + (\mu - \mathbf{t}(\mathbf{k}))/\hbar, \\ \ln\,\det\big(-\mathbf{G}^{-1}(\mathbf{k},\omega)\big) &= \sum_{n=1}^{2n_{orb}} \ln\Big(-\omega - \big(\mu - E_n(\mathbf{k})\big)/\hbar\Big)\,. \end{aligned} \tag{14}$$

We now replace the sum over Matsubara frequencies by a contour integration, a standard trick used in field theory (see e.g. section 25 of FW) and obtain

$$-\frac{1}{\beta}\sum_{\nu} e^{i\omega_\nu \eta} \ln\,\det\big(-\mathbf{G}^{-1}(\mathbf{k},i\omega_\nu)\big) = \frac{\hbar}{2\pi i}\oint_{\mathcal{C}} d\omega\, f(\omega)\, e^{\omega\eta} \ln\,\mathbf{det}\big(-\mathbf{G}^{-1}(\mathbf{k},\omega)\big)\,, \tag{15}$$

where

$$f(\omega) = \frac{1}{e^{\beta\hbar\omega}+1},$$

is the Fermi function and the contour $\mathcal{C}$ encircles the imaginary axis in counterclockwise fashion, see Figure 3a. Next we note that the integrals along the two clover-shaped contours in Figure 3b are zero, *provided* the integrand is analytic in the interior of the two curves. Since the Fermi function has all of its poles along the imaginary axis, which is outside of the curves in 3b, we only need to consider possible singularities of $\ln\,\det(-\mathbf{G}^{-1}(\mathbf{k},\omega))$. In principle, the complex logarithm has a branch-cut along the negative real axis which could be problematic. However, a quick glance at (14) shows that as long as ω has a nonvanishing imaginary part, the argument of the logarithm can never be purely real. Singularities of the logarithm thus occur only on the real axis, which also is exterior to the contours 3b. The integral along these contours therefore is indeed zero. Next, Jordan's lemma can be invoked to establish that the integral along the large semicircles vanishes. Here, the Fermi function $f(\omega)$ guarantees that the

contribution from the semicircle with $\Re(\omega) > 0$ vanishes, whereas the factor $e^{\omega\eta}$ does the same for the semicircle with $\Re(\omega) < 0$. It follows that the integral along the contour $\mathcal{C}$ in Figure 3a is equal to that along the contour $\mathcal{C}'$ in 3c (note the inverted direction of the curves in 3c as compared to 3b). Next, we insert

$$f(\omega) = -\frac{1}{\beta\hbar}\,\frac{d}{d\omega}\ln\left(1+e^{-\beta\hbar\omega}\right) \tag{16}$$

and integrate by parts. Thereby the Fermi function and the factor $e^{\eta\omega}$ again make sure that the contributions from $\Re(\omega) = \pm\infty$ vanish and we obtain

$$\begin{aligned}
&\frac{1}{\beta}\frac{1}{2\pi i}\oint_{\mathcal{C}'} d\omega\,\ln\left(1+e^{-\beta\hbar\omega}\right)\frac{d}{d\omega}\left(e^{\eta\omega}\sum_{n=1}^{2n_{orb}}\ln\left(-\omega+(\mu-E_n(\mathbf{k}))/\hbar\right)\right)\\
&= \frac{1}{\beta}\frac{1}{2\pi i}\oint_{\mathcal{C}'} d\omega\,\ln\left(1+e^{-\beta\hbar\omega}\right)\,e^{\eta\omega}\sum_{n=1}^{2n_{orb}}\frac{1}{\omega+(\mu-E_n(\mathbf{k}))/\hbar}+\mathcal{O}(\eta)\,.
\end{aligned}$$

Now we substitute $\hbar\omega \to z$ and use the theorem of residues (thereby remembering that $\mathcal{C}'$ encircles the poles of the Green's function on the real axis in *clockwise* fashion) and after taking the limit $\eta \to 0$ obtain the expression (12), which completes the first step of the proof.

3.3 Calculation of $\partial\Omega/\partial\lambda$

To obtain the derivative of the true grand potential Ω with respect to λ we start from the formula

$$\begin{aligned}
\lambda\frac{\partial}{\partial\lambda}\Omega(\lambda) &= -\frac{\lambda}{\beta}\frac{\partial}{\partial\lambda}\ln\left(\mathrm{Tr}\left(e^{-\beta(H_0+\lambda H_1)-\mu N}\right)\right)\\
&= \frac{1}{Z}\,\mathrm{Tr}\left(\lambda H_1\,e^{-\beta(H_0+\lambda H_1)-\mu N}\right)\\
&= \langle\lambda H_1\rangle_\lambda
\end{aligned}$$

where $\langle ...\rangle_\lambda$ denotes the thermal average *calculated at interaction strength* λ. The last quantity thus is the expectation value of the interaction Hamiltonian for interaction strength λ. This can be computed by making use of the equation of motion of the Green's function, a procedure found in many textbooks, see e.g. Equation (23.14) of FW. One has

$$\langle\lambda H_1\rangle_\lambda = -\frac{1}{2}\lim_{\tau\to 0^-}\sum_{\mathbf{k}}\mathrm{Tr}\left(\hbar\frac{\partial}{\partial\tau}-\mu+\boldsymbol{t}(\mathbf{k})\right)\mathbf{G}_\lambda(\boldsymbol{k},\tau)\,,$$

where the subscript λ on the Green's function implies that this is the exact Green's function for interaction strength λ. Next we recall the Dyson equation (10), which holds for any λ

$$\left(-\partial_\tau+(\mu-\mathbf{t}(\mathbf{k}))/\hbar\right)\mathbf{G}(\mathbf{k},\tau)-\int_0^{\beta\hbar}d\tau'\,\boldsymbol{\Sigma}(\mathbf{k},\tau-\tau')\,\mathbf{G}(\mathbf{k},\tau') = \delta(\tau).$$

Since $\delta(\tau<0)=0$ we have $\lim_{\tau\to 0^-}\delta(\tau)=0$ and obtain

$$\begin{aligned}
\lambda\frac{\partial}{\partial\lambda}\Omega(\lambda) &= \frac{\hbar}{2}\lim_{\tau\to 0^-}\sum_{\mathbf{k}}\int_0^{\beta\hbar}d\tau'\,\mathrm{Tr}\big(\boldsymbol{\Sigma}_\lambda(\boldsymbol{k},\tau-\tau')\,\boldsymbol{G}_\lambda(\boldsymbol{k},\tau')\big)\\
&= \frac{1}{2\beta}\sum_{\mathbf{k},\nu}\mathrm{Tr}\,\boldsymbol{\Sigma}_\lambda(\boldsymbol{k},i\omega_\nu)\,\boldsymbol{G}_\lambda(\boldsymbol{k},i\omega_\nu)\,.
\end{aligned} \tag{17}$$

3.4 Definition and properties of the Luttinger-Ward functional

As already mentioned, the Luttinger-Ward functional $\Phi[\mathbf{G}]$ is defined as a sum over infinitely many Feynman diagrams with certain properties. The diagrams which contribute are *closed*, which means they have no external lines. They are moreover *connected*, which means that they cannot be decomposed into sub-diagrams that are not connected by lines. And finally, only *skeleton diagrams* are taken into account in the Luttinger-Ward functional. A skeleton diagram is a diagram where no Green's function line contains a self-energy insertion. In other words, it is impossible to draw a box around any part of the diagram so that only two Green's function lines cross the box.

At this point we need to discuss an important property of the skeleton diagrams. Let us consider a self-energy diagram. It contains one Green's function line from the entry-point to the exit-point, and a number of Green's function loops. Starting from the entry-point we may follow the Green's function line and draw a circle around each self-energy insertion that we encounter until we reach the exit point. This procedure will eliminate a number of loops, that means enclose them in a self-energy insertion. Then, we continue along the first interaction line which is not eliminated until we reach a Fermion loop that is not yet eliminated. We follow the Green's function line along this loop and again draw a circle around each self-energy insertion. We proceed to the next interaction line that has not yet been eliminated and so on. We end up with a diagram in which all self-energy insertions are inside circles. Replacing the circles by straight lines we obviously obtain a skeleton-diagram for the self-energy. It is easy to see that the skeleton diagram to which a given self-energy diagram is reduced is unique. All self-energy diagrams thus can be grouped into classes such that all members of one class can be reduced to the same skeleton diagram. Conversely, all members of one class can be obtained by starting out from the skeleton-diagram and inserting the full Green's function for each Green's function line in the diagram, which we write as

$$\mathbf{\Sigma}(\mathbf{k},\omega) \;=\; \sum_n \mathbf{\Sigma}^{(s,n)}(\mathbf{k},\omega). \tag{18}$$

Here $\mathbf{\Sigma}^{(s,n)}$ denotes the set of all n^{th} order skeleton diagrams (i.e. diagrams with n interaction lines) with the Green's function lines replaced by the full Green's function.

Having defined the diagrams contributing to $\Phi[\mathbf{G}]$ each diagram is now translated into a multiple sum according to the standard Feynman rules for the imaginary-time Green's function in momentum space (see section 25 of FW). However, there is one crucial difference: whereas in a standard Feynman diagram a Green's function line corresponds to a factor $\mathbf{G}^0(\mathbf{k},\omega)$ (the noninteracting Green's function), in the Luttinger-Ward functional we replace $\mathbf{G}^0(\mathbf{k},\omega) \to \mathbf{G}(\mathbf{k},\omega)$ where $\mathbf{G}(\mathbf{k},\omega)$ is the argument of the functional $\Phi[\mathbf{G}]$. As an example, the expression corresponding to the diagram in Figure 4 is

$$\left(\frac{-1}{\beta\hbar^2 N}\right)^2 (-1)^2 \sum_{\mathbf{k},\mathbf{k}',\mathbf{q}} \sum_{\alpha,\beta,\gamma,\delta} \sum_{\alpha_1,\beta_1,\gamma_1,\delta_1} \sum_{\nu,\nu',\mu} V_{\alpha,\beta,\delta,\gamma}(\mathbf{k},\mathbf{k}',\mathbf{q})\, V_{\delta_1,\gamma_1,\alpha_1,\beta_1}(\mathbf{k}+\mathbf{q},\mathbf{k}'-\mathbf{q},-\mathbf{q})$$
$$\times\, G_{\alpha_1,\alpha}(\mathbf{k}+\mathbf{q}, i\omega_\nu+\omega_\mu)\, G_{\delta,\delta_1}(\mathbf{k}, i\omega_\nu)\, G_{\beta_1,\beta}(\mathbf{k}'-\mathbf{q}, \omega_{\nu'}-\omega_\mu)\, G_{\gamma,\gamma_1}(\mathbf{k}', \omega_{\nu'})\ . \tag{19}$$

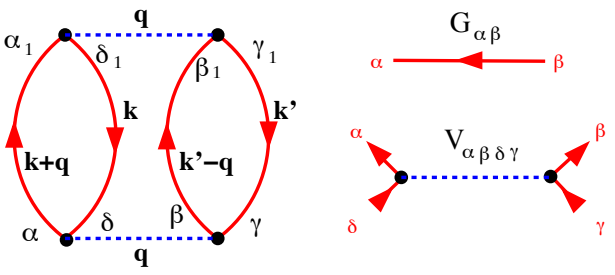

Fig. 4: *Left: A diagram contributing to the Luttinger-Ward functional. Right: the elements of the diagram.*

The Luttinger-Ward functional $\Phi[\mathbf{G}]$ thus consists of an infinite sum of multiple sums which involve the interaction matrix elements V of the Hamiltonian (4) and the function $\mathbf{G}$ for which the functional is to be evaluated.

Let us briefly discuss the scaling with system size, N. By the Feynman rules an n^{th} order diagram has the prefactor $(1/N)^n$. On the other hand, there are n interaction lines, and $2n$ Green's function lines, so that there are $3n$ momenta. The n interaction lines give rise to $2n$ momentum conservation conditions, one for each end of a line. However, in a *closed* diagram one of these momentum conservation conditions is fulfilled trivially so that there remain $n+1$ momenta to be summed over (see the above example). Each sum runs over N momenta so that the total diagram is of order N – as it has to be because Ω is an extensive quantity.

In addition to the factors originating from the Feynman rules, each diagram is multiplied by $-1/(\beta S)$ where the positive integer S is the *symmetry factor* of the diagram. A very detailed discussion of these symmetry factors is given in section 2.3 of Negele-Orland [9]. The definition is as follows: first, the diagram is drawn such that all interaction lines are in x-direction. The n interaction lines of a diagram are labeled by integers $i \in \{1 \ldots n\}$ and the ends of each interaction line are labeled by R and L (for 'right end' and 'left end'), see Figure 5a. Any Green's function line in the diagram now can be labeled by the ends of the interaction lines where it departs and where it ends: $(i, S_1) \rightarrow (j, S_2)$ with $i, j \in \{1 \ldots n\}$ and $S_1, S_2 \in \{R, L\}$. Obviously, the diagram is characterized completely by the $2n$ 'directed quadruples' $(i, S_1) \rightarrow (j, S_2)$. Then, we consider the following operations on the diagrams: a) any permutation of the indices i, b) exchange of the labels R and L on an arbitrary number of interaction lines, c) any combination of a permutation followed by label exchanges. Such an operation obviously changes the quadruples which characterize the connectivity of the diagram: $[(i, S_1) \rightarrow (j, S_2)] \rightarrow [(i', S_1') \rightarrow (j', S_2')]$. The symmetry factor of a diagram then is the number of symmetry operations – including identity – where the new labels $(i', S_1') \rightarrow (j', S_2')$ are a permutation of the old ones, $(i, S_1) \rightarrow (j, S_2)$ (Negele-Orland then call the transformed diagram a deformation of the first one). As an example, consider the diagram in Figure 5a. Label exchange on, say, the interaction line 2 leads to the diagram shown in 5b, which however is not a deformation of the original diagram. This can be seen by considering e.g. the line connecting the R-end of 1 and the R-end of 2. In 5a this line would have the label $(2, R) \rightarrow (1, R)$, whereas it would be $(1, R) \rightarrow (2, R)$ in 5b. This means that the direction of momentum flow along the

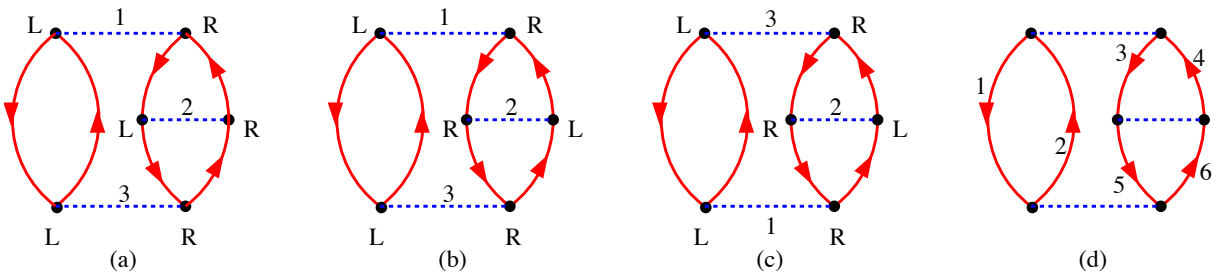

Fig. 5: *Determination of the symmetry factor S for a diagram.*

line would be reversed. On the other hand, the permutation of the labels 1 and 3 followed by label exchange on interaction line 2 leads to the diagram 5c which indeed is a deformation of the original diagram. In Figure 5d the Green's function lines are numbered by $1 \to 6$ and Table 1 gives the quadruples corresponding to these lines in Figures 5a and 5c. Obviously the two sets of quadruples are a permutation of each other. It turns out that this is the only symmetry operation which leaves the diagram invariant, so that, taking into account the identity operation, the diagram has $S = 2$. Since a symmetry operation corresponds to a permutation of the quadruples $(i, S_1) \to (j, S_2)$ that characterize the individual Green's function lines in a diagram, it defines a mapping between these lines whereby each line is mapped onto the one that gets its label. For example, from Table 1 one reads off the corresponding mapping for the operation connecting 5a and 5c:

$$\begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 1 & 6 & 5 & 4 & 3 \end{array}$$

If two Green's function lines i and j are mapped onto each other the lines are equivalent in the sense that the diagram could be deformed such that the deformed diagram is precisely the same as the original one but with line j now taking the place of line i and vice versa.
Let us now assume that a diagram has the symmetry factor S. This means that all Green's function lines can be grouped into disjunct classes such that the lines belonging to one class are mapped onto each other by one of the S symmetry operations. For example, the diagram in 5 has the classes $(1, 2)$, $(3, 6)$ and $(4, 5)$. Since a diagram with n interaction lines has $2n$ Green's function lines the number of classes is $2n/S$ which will be of importance later on.

Line	5(a)	5(c)
1	(1,L)→(3,L)	(3,L)→(1,L)
2	(3,L)→(1,L)	(1,L)→(3,L)
3	(1,R)→(2,L)	(3,R)→(2,R)
4	(2,R)→(1,R)	(2,L)→(3,R)
5	(2,L)→(3,R)	(2,R)→(1,R)
6	(3,R)→(2,R)	(1,R)→(2,L)

Table 1: *Quadruples describing the connectivity of the diagrams Figure 5a and Figure 5c. The numbers of the Green's function lines are given in Figure 5d.*

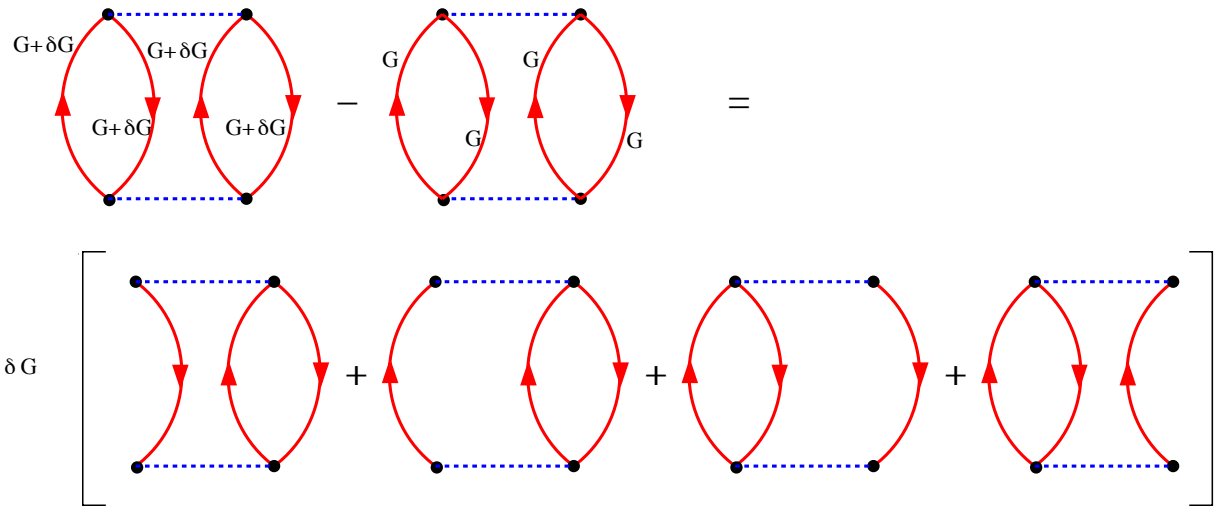

Fig. 6: *Variation of* $\mathbf{G}$ *implies opening the lines of a Feynman diagram.*

Next, we want to see the meaning of this definition. In fact, the Luttinger-Ward functional is the *generating functional* of the self-energy, or, more precisely:

$$\frac{\partial \Phi}{\partial G_{\alpha,\beta}(\mathbf{k}, i\omega_\nu)} = \frac{1}{\beta}\, \Sigma_{\beta,\alpha}(\mathbf{k}, i\omega_\nu). \tag{20}$$

To see this, consider an infinitesimal change $G_{\alpha\beta}(\mathbf{k}, i\omega_\nu) \rightarrow G_{\alpha\beta}(\mathbf{k}, i\omega_\nu) + \delta G_{\alpha\beta}(\mathbf{k}, i\omega_\nu)$ as in Figure 6. The initial diagrams correspond to multiple sums over products of Green's functions where all internal frequencies, momenta, and orbital indices are summed over, subject to the condition of energy/momentum conservation at each interaction vertex, see (19). The first-order change then also can be viewed as a sum of diagrams but with a single missing line, which corresponds to the variation δG that has been factored out. Another way to state this is to say that differentiating with respect to an element of $\mathbf{G}$ amounts to successively opening each of the lines in the initial closed diagram and summing the remaining diagrams. These remaining diagrams obviously look like self-energy diagrams in that they have two entry points. We now need to show, however, that the diagrams not only look like possible contributions to the skeleton diagram expansion of the self-energy, but that they come with exactly the right numerical prefactors. At this point, the additional prefactors of $-1/\beta S$ turn out to be crucial.

We first note that the momentum and frequency which flow into/out-of the diagram are fixed by the momentum and frequency of δG. Regarding the orbital indices, we recall that $G_{\alpha\beta}$ corresponds to a directed line $\beta \rightarrow \alpha$. The resulting self-energy-like diagrams therefore all have the matrix index α on their incoming entry and β on their outgoing entry and, comparing with Fig. 4, we see that this assignment of indices corresponds to $\Sigma_{\beta\alpha}$. Moreover, all internal momenta, frequencies and matrix indices in the remaining diagrams are summed over – subject to the condition of frequency and momentum conservation at the interaction lines – as would be the case in the true self-energy diagrams. Second, the order n of a diagram, i.e. the number of interaction lines, is not changed by opening a Green's function line, so that the prefactor $(-1/\beta\hbar^2 N)^n$ of the closed diagram is also the correct prefactor for the resulting self-energy diagram. Third, opening a Green's function line reduces the number of closed fermion loops by 1

and the factor (-1) in $-1/\beta S$ takes care of this. Lastly, we need to discuss the symmetry factor S. Let us consider a diagram with n interaction lines, which accordingly has $2n$ Green's function lines and moreover assume that the diagram has the symmetry factor S. As we saw above, the $2n$ Green's function lines can be divided into classes of S members which are mapped into each other by the symmetry operations, and the number of these classes is $2n/S$. A symmetry operation maps a Green's function line i onto an equivalent one j, so it is possible to deform the diagram such that it looks exactly the same as the original one but with line j in place of line i. This means, however, that 'opening' the line i also gives *exactly* the same self-energy diagram as opening line j. Accordingly, from the single closed diagram of degree n with symmetry factor S we obtain $2n/S$ different skeleton diagrams for the self-energy, and each is produced S times, see also Figure 7. This factor of S, however, precisely cancels the prefactor $1/S$. It follows that each skeleton-diagram for the self-energy is produced with the same prefactor $1/\beta$. Differentiating $\Phi[\mathbf{G}]$ with respect to $G_{\alpha\beta}(\mathbf{k}, i\omega_\nu)$ thus gives $1/\beta$ times the sum of all skeleton diagrams for $\Sigma_{\beta\alpha}(\mathbf{k}, i\omega_\nu)$, with the noninteracting Green's function replaced by the full one, and this is exactly $\Sigma_{\beta\alpha}(\mathbf{k}, i\omega_\nu)$ itself, see (18), proving (20).

We have just seen that all skeleton-diagrams for the self-energy can be obtained by differentiating the Luttinger-Ward functional with respect to $\mathbf{G}$, whereby the differentiation corresponds to opening one line in a closed diagram. We then may ask if this operation can be reversed, namely if the Luttinger-Ward functional can be obtained by starting from the skeleton-diagram expansion of the self-energy and closing the diagrams by reconnecting the entry-points of the self-energy by a Green's function. More precisely, we consider an expression of the form

$$\frac{1}{\beta} \sum_{\nu,\mathbf{k}} \sum_{\alpha,\beta} \mathbf{G}_{\alpha,\beta}(\mathbf{k}, i\omega_\nu)\, \mathbf{\Sigma}^{(s,n)}_{\beta,\alpha}(\mathbf{k}, i\omega_\nu)\ . \tag{21}$$

We have seen that an n^{th} order diagram contributing to $\Phi[\mathbf{G}]$ with symmetry factor S produces $2n/S$ different skeleton-self-energy diagrams, and each of them S times and with a factor of (-1), so that the remaining prefactor was $1/\beta$. Upon closing the fermion line again, according to (21), each of these diagrams produces the original closed diagram (it is easy to see that for each self-energy diagram there is exactly one closed diagram from which it can be obtained). Since there are $2n/S$ self-energy diagrams originating from the original closed diagram the latter is produced $2n/S$ times and thus has the additional prefactor $-2n/S\beta$, where the factor of (-1) is due to the additional fermion loop in the closed diagram. In the expansion of $\Phi[\mathbf{G}]$, however, the diagram would have had the prefactor $-1/S\beta$, or, put another way, closing the sum of all n^{th} order skeleton diagrams for $\mathbf{\Sigma}$ according to (21) produces the n^{th} order contribution to $\Phi[\mathbf{G}]$ with an additional prefactor of $2n$ so that

$$\Phi^{(n)} = \frac{1}{2n\beta} \sum_{\nu,\mathbf{k}} \mathrm{Tr}\, \mathbf{G}(\mathbf{k}, i\omega_\nu)\, \mathbf{\Sigma}^{(s,n)}(\mathbf{k}, i\omega_\nu). \tag{22}$$

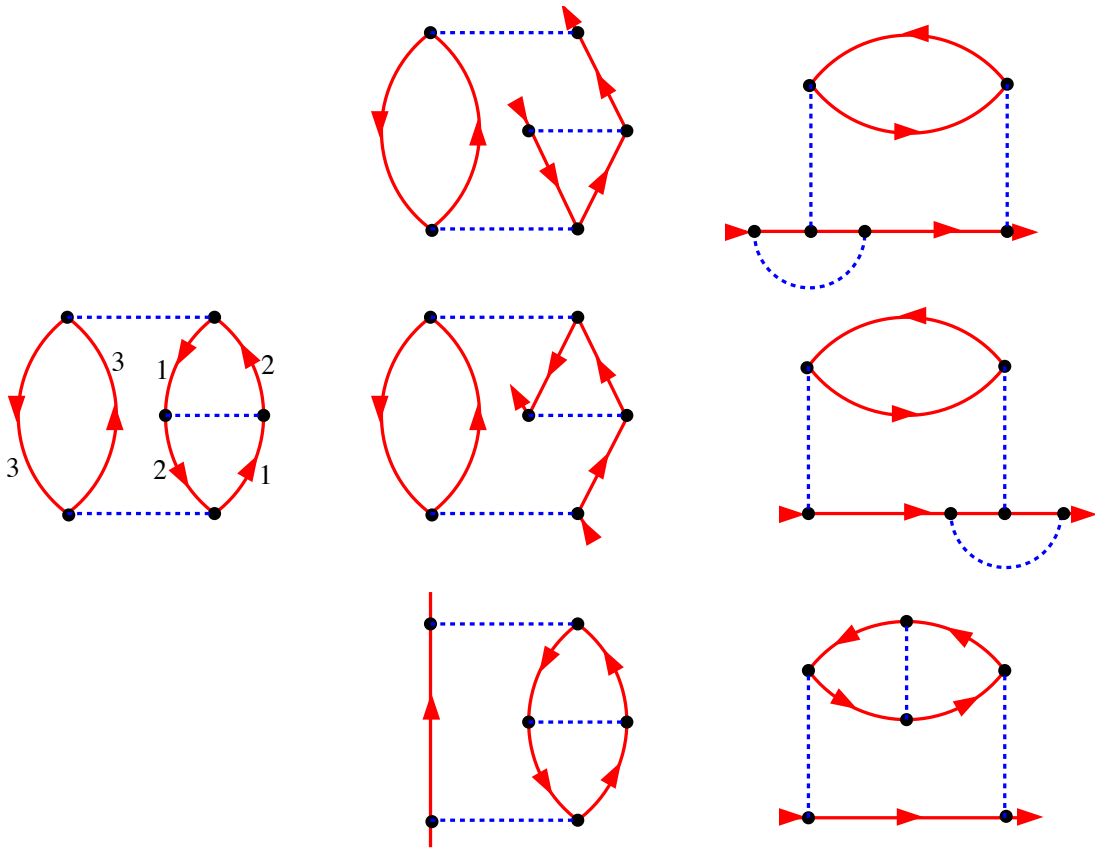

Fig. 7: *The diagram on the left has $n = 3$ and $S = 2$ and accordingly 3 classes of symmetry-equivalent Green's function lines. The lines are labeled by the number of the classes, compare Figure (5) and Table 1. Successively opening the lines of the diagram produces the three different self-energy diagrams in the center column and each of them is produced $S = 2$ times. The right column shows the diagrams redrawn to more look like self-energy diagrams.*

3.5 Calculation of $\partial\tilde{\Omega}/\partial\lambda$

We proceed to the the final step of the proof and compute $\partial\tilde{\Omega}/\partial\lambda$. If we vary the interaction strength λ there are two places in the expression $\tilde{\Omega}$ in (13) where this makes itself felt. Namely, the self-energy $\mathbf{\Sigma}$ will change, and moreover the interaction matrix elements V in the Luttinger-Ward functional (compare 19) that have a prefactor of λ will also contribute to the variation. Let us first consider the variation of $\mathbf{\Sigma}$ and compute

$$\frac{\partial\tilde{\Omega}}{\partial\Sigma_{\alpha,\beta}(\mathbf{k}, i\omega_\nu)}.$$

There are three terms in (13) and we consider them one after the other. The first two terms involve a sum over momentum and frequency and obviously only those terms with momentum $\mathbf{k}$ and frequency ω_ν will contribute. Accordingly, in the following equations we omit the

arguments $(\mathbf{k}, i\omega_\nu)$ for brevity. Then we find by using the chain rule for differentiation

$$\begin{aligned}\frac{\partial}{\partial \Sigma_{\alpha,\beta}}\left(-\frac{1}{\beta}\ln\,\det\left(-\mathbf{G}^{-1}\right)\right) &= -\frac{1}{\beta}\sum_{\mu,\nu}\left(\frac{\partial}{\partial(-G^{-1}_{\mu,\nu})}\,\ln\,\det\left(-\mathbf{G}^{-1}\right)\right)\frac{\partial(-G^{-1}_{\mu,\nu})}{\partial\Sigma_{\alpha,\beta}}\\ &= -\frac{1}{\beta}\sum_{\mu,\nu}(-G_{\nu,\mu})\;\delta_{\mu,\alpha}\;\delta_{\nu,\beta}\\ &= \frac{1}{\beta}\,G_{\beta,\alpha}\;.\end{aligned}$$

In going to the 2^{nd} line we used the identity from the Appendix and the Dyson equation

$$-\mathbf{G}^{-1} = -\omega - \mu/\hbar + \mathbf{\Sigma}$$

from which it follows that

$$\frac{\partial(-G^{-1}_{\mu,\nu})}{\partial\Sigma_{\alpha,\beta}} = \delta_{\mu,\alpha}\;\delta_{\nu,\beta}.$$

We proceed to the second term:

$$\frac{\partial}{\partial\Sigma_{\alpha,\beta}}\left(-\frac{1}{\beta}\mathrm{Tr}\,\mathbf{\Sigma}\,\mathbf{G}\right) = \frac{\partial}{\partial\Sigma_{\alpha,\beta}}\left(-\frac{1}{\beta}\sum_{\mu,\nu}\Sigma_{\nu,\mu}\;G_{\mu,\nu}\right) = -\frac{1}{\beta}\left(G_{\beta,\alpha} + \sum_{\mu,\nu}\Sigma_{\nu,\mu}\frac{\partial G_{\mu,\nu}}{\partial\Sigma_{\alpha,\beta}}\right).$$

Lastly we consider the Luttinger-Ward functional. Using again the chain rule we find

$$\frac{\partial\Phi[G]}{\partial\Sigma_{\alpha,\beta}} = \sum_{\mu,\nu}\frac{\partial\Phi[G]}{\partial G_{\mu,\nu}}\frac{\partial G_{\mu,\nu}}{\partial\Sigma_{\alpha,\beta}} = \frac{1}{\beta}\sum_{\mu,\nu}\Sigma_{\nu,\mu}\frac{\partial G_{\mu,\nu}}{\partial\Sigma_{\alpha,\beta}}\;.$$

Adding up the three terms we thus obtain the important result

$$\frac{\partial\tilde{\Omega}}{\partial\Sigma_{\alpha,\beta}(\mathbf{k},i\omega_\nu)} = 0. \tag{23}$$

In other words: the expression $\tilde{\Omega}$, which will be seen to be equal to the grand potential Ω in a moment, is *stationary with respect to variations of the self-energy.* This is the stationarity condition for $\mathbf{\Sigma}$ which is the basis of the VCA.

First, however, we have to complete the proof and evaluate $\lambda\frac{\partial}{\partial\lambda}\tilde{\Omega}(\lambda)$. Since there is no variation of $\tilde{\Omega}$ due to a variation of $\mathbf{\Sigma}$, the only remaining source of variation are the interaction lines in the Luttinger-Ward functional. Namely any n^{th} order diagram has the prefactor of λ^n so that

$$\lambda\,\frac{\partial}{\partial\lambda}\,\Phi^{(n)} = n\,\Phi^{(n)}$$

Using (22) we thus obtain

$$\begin{aligned}\lambda\frac{d\tilde{\Omega}}{d\lambda} = \sum_n n\,\Phi^{(n)} &= \sum_n\frac{1}{2\beta}\sum_{\nu,\mathbf{k}}\mathrm{Tr}\,\mathbf{G}_\lambda(\mathbf{k},i\omega_\nu)\,\mathbf{\Sigma}^{(s,n)}_\lambda(\mathbf{k},i\omega_\nu)\\ &= \frac{1}{2\beta}\sum_{\nu,\mathbf{k}}\mathrm{Tr}\,\mathbf{G}_\lambda(\mathbf{k},i\omega_\nu)\left(\sum_n\mathbf{\Sigma}^{(s,n)}_\lambda(\mathbf{k},i\omega_\nu)\right)\\ &= \frac{1}{2\beta}\sum_{\nu,\mathbf{k}}\mathrm{Tr}\,\mathbf{G}_\lambda(\mathbf{k},i\omega_\nu)\,\mathbf{\Sigma}_\lambda(\mathbf{k},i\omega_\nu)\;.\end{aligned}$$

Comparing with (17), we see that this is equal to $\lambda\frac{\partial}{\partial\lambda}\Omega(\lambda)$ which completes the proof.
Let us summarize the results that we have obtained:

1. The grand canonical potential Ω of an interacting Fermi system is given by eqn. (13).

2. The Luttinger-Ward functional is the generating functional of $\mathbf{\Sigma}(\mathbf{k}, i\omega_\nu)$, see eqn. (20).

3. $\Phi[\mathbf{G}]$ depends only on the interaction matrix elements $V_{\alpha\beta\delta\gamma}$ in the Hamiltonian and the Green's function $\mathbf{G}$ which is the argument of the functional.

4. Ω is stationary under variations of $\mathbf{\Sigma}(\mathbf{k}, i\omega_\nu)$ see (23).

Looking at the above proof one might worry about the fact that it assumes a continuous evolution of the system with increasing interaction strength λ - whereas we are interested e.g. in Mott-insulators where we have reason to believe that a phase transition occurs as a function of λ. However, Potthoff has recently given a nonperturbative proof of the theorem [6, 13] that means all of the above properties of the grand potential, the Luttinger-Ward functional and the self-energy remain valid in a strongly correlated electron system where a Feynman-diagram expansion of the Green's function and the adiabatic continuity with the noninteracting system can no longer be assumed valid.

4 The variational cluster approximation

In the preceding section we have seen that the grand canonical ensemble of a system is stationary with respect to variations of the self-energy. In order to rewrite Ω as a functional of the self-energy we need to change the argument of the Luttinger-Ward functional from $\mathbf{G}$ to $\mathbf{\Sigma}$. Since $\mathbf{\Sigma}$ is the derivative of Φ with respect to $\mathbf{G}$ this can be achieved, following [4], by introducing the Legendre-transform of the Luttinger-Ward functional:

$$\begin{aligned} F[\mathbf{\Sigma}] &= \Phi[\mathbf{G}[\mathbf{\Sigma}]] - \sum_{\mathbf{k},\nu}\sum_{\alpha,\beta}\frac{\partial\Phi}{\partial G_{\alpha\beta}(\mathbf{k}, i\omega_\nu)}\,G_{\alpha\beta}(\mathbf{k}, i\omega_\nu) \\ &= \Phi[\mathbf{G}[\mathbf{\Sigma}]] - \frac{1}{\beta}\sum_{\mathbf{k},\nu}\sum_{\alpha,\beta}G_{\alpha\beta}(\mathbf{k}, i\omega_\nu)\,\Sigma_{\beta\alpha}(\mathbf{k}, i\omega_\nu) \\ &= \Phi[\mathbf{G}[\mathbf{\Sigma}]] - \frac{1}{\beta}\sum_{\mathbf{k},\nu}\quad \mathrm{Tr}\,\mathbf{G}(\mathbf{k}, i\omega_\nu)\,\mathbf{\Sigma}(\mathbf{k}, i\omega_\nu)\ . \end{aligned}$$

By virtue of being a Legendre transform this new functional obviously satifies

$$\frac{\partial F}{\partial\Sigma_{\alpha\beta}(\mathbf{k}, i\omega_\nu)} = -\frac{1}{\beta}G_{\beta\alpha}(\mathbf{k}, i\omega_\nu)\ . \tag{24}$$

Moreover, the second and third term in (13) together are nothing but $F[\mathbf{\Sigma}]$, whence

$$\Omega = -\lim_{\eta\to 0^+}\frac{1}{\beta}\sum_{\mathbf{k},\nu}e^{i\omega_\nu\eta}\left[\ln\,\det\left(-i\omega_\nu + (\mathbf{t}(\mathbf{k}) - \mu)/\hbar + \mathbf{\Sigma}(\mathbf{k}, i\omega_\nu)\right)\right] + F[\mathbf{\Sigma}]\ . \tag{25}$$

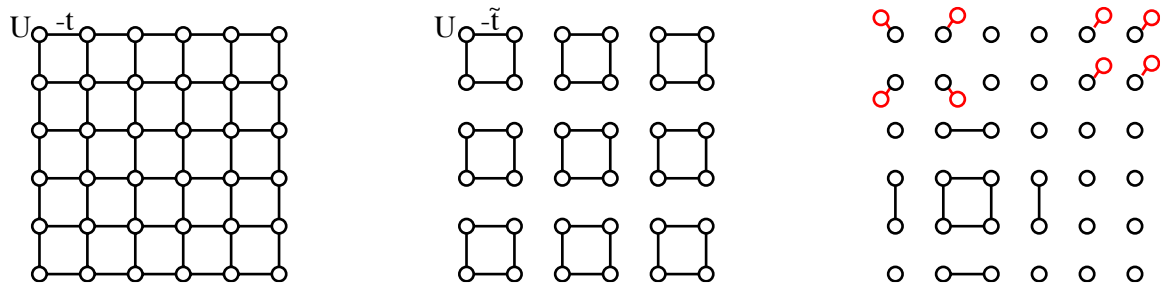

Fig. 8: *Left: The 2D Hubbard model. Center: The 2D Hubbard model partitioned into* 2×2 *clusters. Right: Partitioning into larger/smaller clusters, with or without additional bath sites and additional hopping integrals. All systems have a different kinetic energy* H_0 *but exactly the same interaction part* H_1*; accordingly, they all have the same Luttinger Ward functional.*

Here we have used the Dyson equation to replace $-\mathbf{G}^{-1}$ in the first term. This expresses Ω as a functional of $\mathbf{\Sigma}$, and this functional is known to be stationary with respect to variations of its argument at the exact $\mathbf{\Sigma}$ (this is also easily verified using the identity from the Appendix together with (24)). One might now try and either derive Euler-Lagrange equations or introduce a trial self-energy containing some variational parameters, e.g., of the form (11) with only a certain number of poles (σ_i, ζ_i) retained, and perform the variation with respect to these parameters. Unfortunately this procedure does not work, because the functional $F[\mathbf{\Sigma}]$ was defined as the Legendre transform of the Luttinger-Ward functional $\Phi[\mathbf{G}]$, which in turn was defined as a sum over an infinite number of Feynman diagrams and thus is completely impossible to evaluate for a given trial self-energy.

At this point Potthoff's new idea comes into play. For definiteness let us assume that we are interested in a 2D Hubbard model, shown schematically in Figure 8a. Then, we might partition the plane into finite clusters and set the hopping between the clusters to zero, so that they become disconnected, see Fig. 8b. The resulting array of clusters has been termed the *reference system*. The finite clusters also can be decorated in various ways by noninteracting bath orbitals, they can be larger than just 2×2 or contain hopping terms not included in the original Hubbard Hamiltonian, see Fig. 8c. As long as the resulting clusters are not too big, the Fock space of a single cluster has a manageable size and the clusters can be treated by exact diagonalization. This gives us all eigenstates $|i\rangle$ together with their energies E_i and particle numbers N_i. Using these, we may numerically evaluate the grand partition function $\breve{Z}$ and obtain the potential $\breve{\Omega}$ for a single cluster. In addition, we can calculate the Green's function matrix $\breve{\mathbf{G}}(\omega)$ using the Lehmann representation or the Lanczos algorithm [14], invert it and extract the self-energy $\breve{\mathbf{\Sigma}}(\omega)$. For all of this it is actually sufficient to know all eigenstates with $E_i - \mu N_i$ within a window of $\approx 10\ k_B T$ above the minimum value, which can be obtained by the Lanczos algorithm even for clusters of size $\tilde{N} \approx 10 - 20$. Next, we revert to expression (25) and obtain the numerical value of $F[\tilde{\mathbf{\Sigma}}]$:

$$F[\breve{\mathbf{\Sigma}}] = \breve{\Omega} + \lim_{\eta \to 0^+} \frac{1}{\beta} \sum_{\nu} e^{i\omega_\nu \eta} \left[\ln \det \left(-i\omega_\nu + (\breve{\mathbf{t}} - \mu)/\hbar + \breve{\mathbf{\Sigma}}(i\omega_\nu) \right) \right] , \tag{26}$$

where $\breve{\mathbf{t}}$ is the kinetic energy of the cluster, i.e., the matrix $t_{\alpha\beta}$ in (1). This procedure gives

us the *exact* self-energy $\breve{\mathbf{\Sigma}}(\omega)$ of the cluster together with the *exact* numerical value of the corresponding Luttinger-Ward functional. An important point is that all matrix elements $\breve{\Sigma}_{\alpha\beta}$ that have one index α or β on a bath site are zero. This can be seen, e.g., from the diagrammatic expansion of the self-energy. The cluster self-energy $\breve{\mathbf{\Sigma}}$ therefore has non-vanishing entries only for the correlated sites of the original lattice problem.

At this point the crucial observation by Potthoff comes into play: we have seen above that the Luttinger-Ward functional $\Phi[\mathbf{G}]$ was a sum over Feynman diagrams into which, apart from numerical factors, only two quantities do enter: the interaction matrix elements V of the Hamiltonian and the Green's function $\mathbf{G}$ which is the argument of the functional, see e.g. (19). In our example with the 2D Hubbard model, however, the full 2D Hubbard model and the array of clusters, which may include non-interacting bath sites, differ only in their single-particle terms H_0 but do have exactly the same interaction part H_1. It follows that the functional $\Phi[\mathbf{G}]$ and hence its Legendre transform $F[\mathbf{\Sigma}]$ are identical for the two systems. Since, however, we are able to calculate the self-energy of the cluster and the corresponding value of the Luttinger-Ward functional exactly, we may use these as trial self-energies for the lattice system. In other words, we make the ansatz for the lattice system

$$\begin{aligned} \Omega_{latt} &= -\lim_{\eta\to 0^+}\frac{1}{\beta}\sum_{\mathbf{k},\nu} e^{i\omega_\nu\eta}\left[\ln\det\left(-\mathbf{G}'(\mathbf{k},i\omega_\nu)^{-1}\right)\right] + F[\breve{\mathbf{\Sigma}}], \\ \mathbf{G}'(\mathbf{k},\omega) &= \Big(\omega + (\mu - \mathbf{t}(\mathbf{k}))/\hbar - \tilde{\mathbf{\Sigma}}(\mathbf{k},\omega)\Big)^{-1}. \end{aligned} \tag{27}$$

Here $\mathbf{t}(\mathbf{k})$ is now the kinetic energy matrix of the *lattice system* whereas $\breve{\mathbf{\Sigma}}(\mathbf{k},\omega)$ is the spatial Fourier transform of the *cluster self-energy* (which may have no $\mathbf{k}$-dependence at all, depending on the geometry of the reference system). Accordingly, $\mathbf{G}'$ is the approximate Green's function of the lattice system.

Then, how do we perform the variation of the self-energy? The answer is that the single-particle Hamiltonian $\breve{H}_0$ of the cluster used to compute the trial self-energies $\breve{\mathbf{\Sigma}}$ is completely arbitrary, because the only requirement for the equality of the Luttinger-Ward functionals was the equality of the interaction part H_1. If we change the single-particle terms of the reference system, i.e. the hopping integrals or site-energies, the self-energy of the cluster will change. The self-energy and its Luttinger-Ward functional thus become functions of the single-particle terms $\breve{t}_{\alpha\beta}$ of the reference system: $\Omega_{latt} = \Omega_{latt}(\breve{t}_{\alpha\beta})$. Then, we demand that

$$\frac{\partial\Omega_{latt}}{\partial\breve{t}_{\alpha\beta}} = 0, \tag{28}$$

which is a condition on the parameters of the reference system, $\breve{t}$, and we denote the solution of (28) by $\breve{t}^*$. The physical interpretation is that the VCA is seeking the best approximation to the self-energy of the lattice-system amongst those functions $\breve{\mathbf{\Sigma}}(\mathbf{k},i\omega_\nu)$ that can be represented as the exact self-energies of the reference system for some values of the single-particle parameters $\breve{t}_{\alpha\beta}$. After solving (28) we obtain an approximate self-energy $\breve{\mathbf{\Sigma}}(\mathbf{k},i\omega_\nu)$ and an approximate value of the grand canonical potential Ω_{latt}. Since Ω_{latt} can be obtained for arbitrary values

of T and μ or other external parameters, thermodynamical quantities such as particle number, entropy or specific heat can be obtained by doing the procedure for different T and μ and differentiating.
As an example we address an interesting property of the VCA The particle number N_e of any system can be obtained in two different ways (the second is a combination of FW (23.9) and (25.10)):

$$N_e = -\frac{\partial \Omega}{\partial \mu} = \frac{1}{\beta\hbar} \lim_{\eta\to 0^+} \sum_{\mathbf{k},\nu} e^{i\omega_\nu \eta} \operatorname{Tr} \mathbf{G}'(\mathbf{k}, i\omega_\nu).$$

Since the VCA gives both Ω, and $\mathbf{G}'(\omega)$, it is natural to ask if the two ways of calculating N_e give the same result, and this question has been addressed by Aichhorn *et al.* [15]. We first note that the chemical potential of the reference system has to be the same as that of the physical system. Next, let us assume that we regroup the orbital energies of the cluster, $\tilde{t}_{\alpha\alpha}$, by separating the center of gravity

$$\epsilon = \frac{1}{2n_{orb}} \operatorname{Tr} \tilde{t}$$

and introducing $2n_{orb} - 1$ relative energies $\tilde{t}'_{\alpha\alpha}$ so that $\tilde{t}_{\alpha\alpha} = \tilde{t}'_{\alpha\alpha} + \epsilon$. Since in all calculations for the reference system the chemical potential μ and ϵ only appear in the combination $\mu - \epsilon$, the derivative of any cluster quantity $\breve{A}$ with respect to the chemical potential μ obeys

$$\left.\frac{\partial \breve{A}}{\partial \epsilon}\right|_{\breve{t}',\mu} = -\left.\frac{\partial \breve{A}}{\partial \mu}\right|_{\breve{t}',\epsilon}.$$

Next we consider the change of the approximate Ω_{latt} induced by a change of μ. A variation of μ will make itself felt at a variety of places. Looking at (26) and (27) we see that μ appears explicitely in these. Moreover, μ appears in the grand partition function $\breve{Z}$ and the Green's function $\breve{\mathbf{G}}$ of the reference system, so that the cluster self-energy itself will change with μ. As a consequence of these changes, we have to take into account that $\breve{t}^*$, the solution of (28), will change as well: $\breve{t}^* \to \breve{t}^* + \delta\breve{t}^*$, so that the situation becomes somewhat complicated. Fortunately enough, the first-order change of Ω_{latt} due to a variation of $\breve{t}^*$ is zero; this is exactly the stationarity condition (28). We thus need to consider only the change of Ω_{latt} for fixed parameters $\breve{t}$. Using the last identity in the Appendix we obtain

$$\begin{aligned}
-\left.\frac{\partial \Omega_{latt}}{\partial \mu}\right|_{\breve{t}',\epsilon} &= \lim_{\eta\to 0^+} \frac{1}{\beta} \sum_{\mathbf{k},\nu} e^{i\omega_\nu \eta} \operatorname{Tr} \left(\mathbf{G}'(\mathbf{k}, i\omega_\nu) \left(\frac{1}{\hbar} - \left.\frac{\partial \breve{\mathbf{\Sigma}}(\mathbf{k}, i\omega_\nu)}{\partial \mu}\right|_{\breve{t}',\epsilon} \right) \right) - \frac{\partial F[\tilde{\mathbf{\Sigma}}]}{\partial \mu}|_{\tilde{t}',\epsilon} \\
&= \lim_{\eta\to 0^+} \frac{1}{\beta} \sum_{\mathbf{k},\nu} e^{i\omega_\nu \eta} \operatorname{Tr} \left(\mathbf{G}'(\mathbf{k}, i\omega_\nu) \left(\frac{1}{\hbar} + \left.\frac{\partial \breve{\mathbf{\Sigma}}(\mathbf{k}, i\omega_\nu)}{\partial \epsilon}\right|_{\breve{t}',\mu} \right) \right) + \frac{\partial F[\tilde{\mathbf{\Sigma}}]}{\partial \epsilon}|_{\tilde{t}',\mu} \\
&= \lim_{\eta\to 0^+} \frac{1}{\beta\hbar} \sum_{\mathbf{k},\nu} e^{i\omega_\nu \eta} \operatorname{Tr} \mathbf{G}'(\mathbf{k}, i\omega_\nu) - \left.\frac{\partial \Omega_{latt}}{\partial \epsilon}\right|_{\breve{t}',\mu}.
\end{aligned}$$

The presence of the first term in the last line can be be understood by noting that μ appears *explictly* in the approximate cluster Green's function $\mathbf{G}'$ (see (27)) whereas ϵ does not. At this

point we note that if ϵ has been included into the set of cluster parameters which are subject to variation, the last term vanishes (because this is exactly equation (28) for $\breve{t}_{\alpha\beta} = \epsilon$) and the two expressions for the particle number indeed give the same result. The VCA thus gives a thermodynamically consistent particle number if and only if the center of gravity of the orbital energies in the reference system is included into the set of parameters to be varied.
To conclude this section we briefly comment on the evaluation of terms like

$$S = -\frac{1}{\beta}\sum_{\nu} e^{i\omega_\nu 0^+} \ln \det(-\mathbf{G}^{-1}(i\omega_\nu))\,,$$

where $\mathbf{G}$ may be either the Green's function of the reference system as in (26) or the approximate Green's function of the lattice system as in (27). The form of this term suggests that we proceed exactly as in the first step of the proof of the Luttinger-Ward theorem, namely to convert the sum over Matsubara-frequencies into a contour-integral, deform the contour using Jordan's lemma as in Figure 3, replace the Fermi function according to (16) and integrate by parts. One obtains

$$S = -\frac{1}{2\pi\beta i}\oint_{\mathcal{C}'} d\omega \log(1+e^{-\beta\omega}) \sum_n \frac{1}{\lambda_n(\omega)}\frac{\partial\lambda_n(\omega)}{\partial\omega}$$

where $\lambda_n(\omega)$ are the eigenvalues of $\boldsymbol{G}(\omega)$. There are two types of singularities of the integrand in this expression:

1. zeros of an eigenvalue (which corresponds to a singularity of an eigenvalue of $\boldsymbol{\Sigma}$) i.e.

$$\lambda(\omega) \approx a_\nu(\omega-\zeta_\nu) \;\rightarrow\; \frac{1}{\lambda(\omega)}\frac{\partial\lambda(\omega)}{\partial\omega} = \frac{1}{\omega-\zeta_\nu}$$

2. singularities of an eigenvalue, i.e.

$$\lambda(\omega) \approx \frac{b_\mu}{\omega-\eta_\mu} \;\rightarrow\; \frac{1}{\lambda(\omega)}\frac{\partial\lambda(\omega)}{\partial\omega} = -\frac{1}{\omega-\eta_\mu}\,.$$

In this way we obtain the expression derived by Potthoff [4]:

$$S = -\frac{1}{\beta}\left(\sum_\mu \log(1+e^{-\beta\eta_\mu}) - \sum_\nu \log(1+e^{-\beta\zeta_\nu})\right),$$

An alternative is to simply evaluate the contour integral numerically.

5 Applications of the VCA

5.1 Metal-insulator transition in a dimer

As a simple illustration of the procedure we study Potthoff's re-derivation of the phase diagram for the metal-insulator transition in the Hubbard model [16]. We consider a half-filled single-

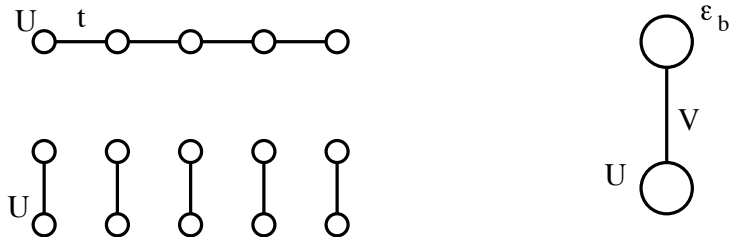

Fig. 9: *Left: The physical Hubbard model (top) which is a true infinite lattice system and the reference system which is an array of identical dimers. Right: Schematic representation of the parameters of a single dimer (see the Hamiltonian (29)).*

band Hubbard model on a bipartite N-site lattice

$$\begin{aligned} H &= \sum_{\mathbf{k},\sigma} t(\mathbf{k})\, c^\dagger_{\mathbf{k},\sigma} c_{\mathbf{k},\sigma} + \frac{U}{2} \sum_{i=1}^{N} (n_i - 1)(n_i - 1) - N\frac{U}{2} \\ &= \sum_{\mathbf{k},\sigma} t(\mathbf{k})\, c^\dagger_{\mathbf{k},\sigma} c_{\mathbf{k},\sigma} + U \sum_{i=1}^{N} n_{i,\uparrow}\, n_{i,\downarrow} - \frac{U}{2} \sum_{i=1}^{N} n_i \end{aligned}$$

where $n_i = c^\dagger_{i,\uparrow} c_{i,\uparrow} + c^\dagger_{i,\downarrow} c_{i,\downarrow}$. For simplicity we assume this Hamiltonian has particle-hole symmetry. More precisely, under the transformation $c^\dagger \leftrightarrow c$ we have $n_i - 1 \to 1 - n_i$ so that the interaction part is invariant, whereas the first term changes sign. If the hopping term connects only sites on different sublattices, which is what we assume, this sign change can be compensated by the gauge transformation $c^\dagger_{i,\sigma} \to -c^\dagger_{i,\sigma}$ on the sites i of one sublattice. This transformation exchanges photoemission and inverse photoemission spectrum and implies $\mu = U/2$.

For the reference system, Potthoff chose N dimers with one 'Hubbard-site' hybridizing with one bath-site, see Figure 9, whereby the Hamiltonian for one dimer reads

$$H - \mu N = -V \sum_{\sigma} (c^\dagger_\sigma b_\sigma + b^\dagger_\sigma c_\sigma) + \left(\epsilon_b - \frac{U}{2}\right) \sum_{\sigma} b^\dagger_\sigma b_\sigma + \frac{U}{2}(n_c - 1)(n_c - 1) - \frac{U}{2}\,. \quad (29)$$

Here $b^\dagger_\sigma$ creates an electron in a bath site and $n_c = c^\dagger_\uparrow c_\uparrow + c^\dagger_\downarrow c_\downarrow$. We have to write $\epsilon_b - U/2$ because $\mu = U/2$. Since we want to generate particle-hole symmetric self-energies we have to impose particle-hole symmetry also in the reference system. The transformation $c^\dagger \leftrightarrow c$, $b^\dagger \leftrightarrow -b$ indeed converts the Hamiltonian into itself *except* for the second term. Setting $\epsilon_b = U/2$, however, eliminates this term and particle-hole symmetry is restored. The only remaining parameter to be varied therefore is V.

The Fock space of the dimer has a dimension of $4 \times 4 = 16$, so all eigenstates can be readily obtained. If we construct basis functions with fixed particle number, spin, and z-component of the spin, the problem in fact can be broken down to diagonalizing 2×2 matrices, i.e. the reference system can be solved analytically. To further simplify the calculations, Potthoff used a semielliptical density of states of width $W = 4$ for the conduction band

$$\rho_0(\epsilon) = \frac{1}{2\pi}\sqrt{4 - \epsilon^2}.$$

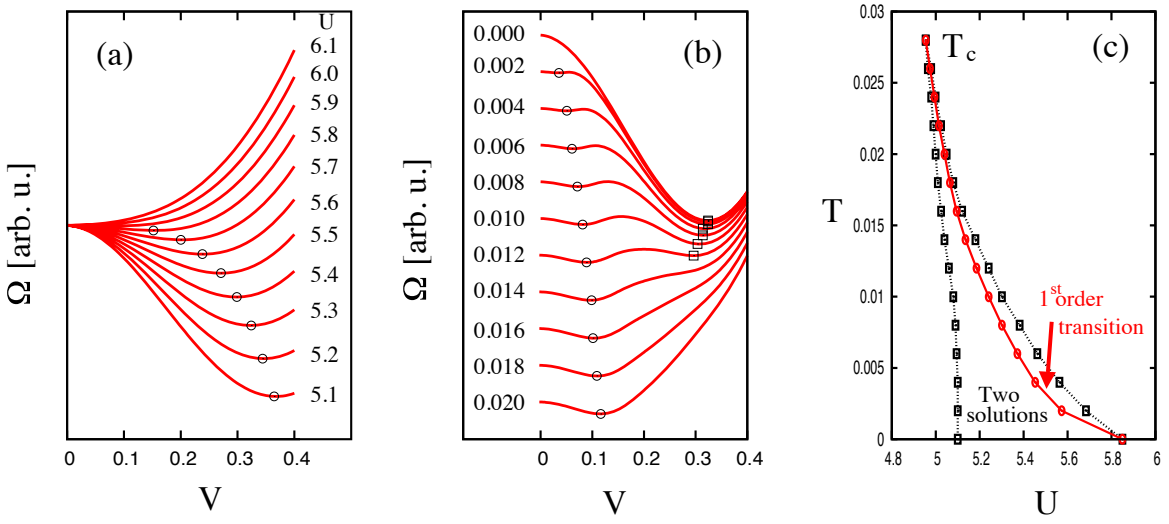

Fig. 10: *(a): Ω versus V at $T = 0$, variation of U, (b): Ω versus V at $U = 5.2$, variation with T, (c): the resulting phase diagram.*

Figure 10a then shows $\Omega(V)$ at $T = 0$ for different values of U. For smaller U there are two stationary points: a maximum at $V = 0$ and a minimum at finite V, which is the physical solution. At $U_c \approx 5.85$ the two extrema coalesce into a single minimum at $V = 0$, which is the only stationary point for larger U. This change from finite V to $V = 0$ precisely corresponds to the metal-insulator transition. To see this we note that in the special case of $T = 0$ and $\mu = U/2$ the self-energy of the dimer can be evaluated exactly [17]:

$$\Sigma(\omega) \;=\; \frac{U}{2} + \frac{U^2}{8}\left(\frac{1}{\omega + 3V} + \frac{1}{\omega - 3V}\right). \tag{30}$$

Note that this has exactly the form (11) derived by Luttinger [12]. The $\mathbf{k}$ integrated Green's function then is (note that $\mu = U/2$)

$$G(\omega) \;=\; \int_{-2}^{2} d\epsilon \, \frac{\rho_0(\epsilon)}{\omega + U/2 - \epsilon - \Sigma(\omega)} \,. \tag{31}$$

For real ω the single particle spectral density, i.e. the combined photoemission and inverse photoemission spectrum, is given by

$$A(\omega) \;=\; -\frac{1}{\pi} \lim_{\delta \to 0} \Im\, G(\omega + i\delta) \,.$$

This is shown in Figure 11 for different V, together with the imaginary part of the self-energy. Since we only want to see qualitatively the effect of vanishing V, $U = 5$ was kept throughout. Then $\Im\Sigma(\omega)$ shows two Lorentzian peaks located at $\pm 3V$ as expected from (30). Each of these peaks creates a gap in the density of states, so that there are three regions with nonvanishing spectral density. As $V \to 0$ the two poles of $\Sigma(\omega)$ approach each other and the spectral weight in the inner region around $\omega = 0$ which corresponds to the Fermi energy, becomes smaller and smaller. Eventually, at $V = 0$ the two peaks merge and there is no more spectral weight

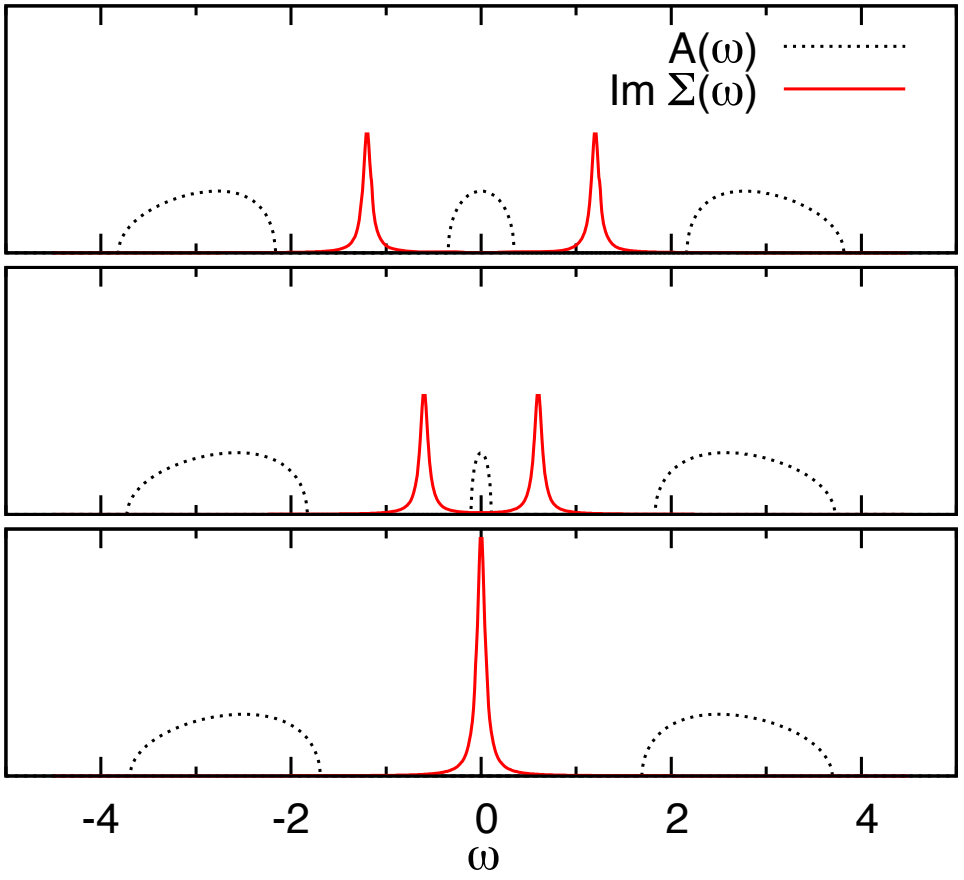

Fig. 11: *Single particle spectral function and imaginary part of the self-energy (calculated with an imaginary part of* 0.05 *for the frequency) obtained from the angle integrated Green's function (31) and the self-energy (30). Parameter values are* $U = 5$ *and* $V = 0.4$ *(top),* $V = 0.2$ *(center) and* $V = 0$ *(bottom).*

at the chemical potential; the system is an insulator now, which is the scenario predicted by dynamical mean-field theory (DMFT) [18]. It remains to be mentioned that DMFT calculations find $U_c \approx 5.84$ [19].

Next, Figure 10b shows $\Omega(V)$ for the fixed value of $U = 5.2$ and different temperatures T. For most temperatures there are three stationary points whereby the local maximum can be discarded. It follows that there are actually two possible solutions for each temperature between $T = 0.10$ and $T = 0.12$. This implies that there is a 1^{st} order phase transition between these two temperatures. Repeating the procedure for various U gives the phase diagram in Figure 10c. There is only a metallic solution for small U, at a first U_{c1} a second insulating solution starts to appear, at U_c there is a first order metal-insulator transition and on from U_{c2} there is only an insulating solution. The results obtained in this way by the, essentially analytical, solution of a dimer are qualitatively very similar to those obtained by extensive numerical renormalization group [20] and quantum Monte Carlo [21] calculations in the framework of DMFT. The main deficiency of the dimer calculation is the underestimation of the critical temperature T_c in Figure 10c by about a factor of two.

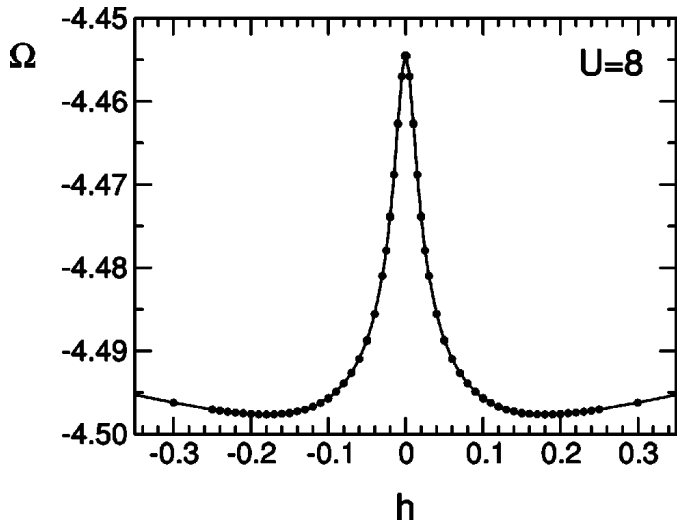

Fig. 12: *Variation of Ω with h in (32) for the half-filled 2D Hubbard model. The reference system is an array of* 10*-site clusters. Reprinted with permission from [22], Copyright 2004 by the American Physical Society.*

5.2 Discussion of spontaneous symmetry breaking

As already mentioned, the VCA gives an estimate of the grand potential Ω. This property makes the VCA of particular usefulness for the discussion of ordering transitions. For definiteness, let us assume we want to discuss antiferromagnetism in the 2D Hubbard model and let us assume that we partition the planar model into an array of finite clusters as in Figure 8. Then, since the single-particle terms of the reference system are completely arbitrary, we may include a term

$$\tilde{H}_S \;=\; h \sum_j e^{i\boldsymbol{Q}\cdot\boldsymbol{R}_j} \left(n_{j,\uparrow} - n_{j,\downarrow}\right) \tag{32}$$

with $\boldsymbol{Q} = (\pi, \pi)$ into $\breve{H}_0$. This term represents a staggered magnetic field which breaks the spin-rotation symmetry of the Hamiltonian. It has to be stressed, however, that *no* magnetic field whatsoever is added to the Hamiltonian of the *lattice system*. The self-energy $\breve{\boldsymbol{\Sigma}}(\omega)$ computed in the reference system with $h \neq 0$, however, incorporates this broken symmetry in various ways; for example the self-energy for the two spin-directions will be different and the sites of the cluster are divided into inequivalent sublattices. If we now determine the optimum value of the parameter h from the standard requirement

$$\frac{\partial \Omega_{latt}}{\partial h} = 0$$

there are two possible outcomes: we will usually always find a solution with $h^* = 0$, which corresponds to the paramagnetic state. It may happen, however, that there is a second solution with $h^* \neq 0$, see Figure 12 for an example, and if this gives a lower Ω it follows that even in the complete absence of any magnetic field the grand potential of the lattice system can be lowered by a self-energy which incorporates broken symmetry. The lattice system thus undergoes a transition to a state of spontaneously broken symmetry. In this way, not only various kinds of magnetic ordering but also superconductivity, charge, or orbital ordering can be discussed, see the review by Potthoff [6].

5.3 Photoemission spectra of NiO, CoO and MnO

Lastly, we consider beyond-band-structure calculations for realistic models of $3d$ transition metal compounds. Very often these have a rock-salt structure, such as NiO, CoO and MnO, or a Perovskite structure, such as $LaCoO_3$. In both cases the transition metal ion is surrounded by an octahedron of oxygen ions.
It is well-known that these materials are often not well described by LDA calculations and it is widely accepted that the reason is the strong Coulomb-repulsion between electrons in the $3d$ shells of the transition metal ions. The description of this Coulomb interaction is the subject of *multiplet theory*, which was initiated in the 1920's to explain the optical spectra of atoms and ions in the gas phase. Multiplet theory is discussed in many textbooks of atomic physics, in particular the books by Slater [23] and Griffith [24] should be mentioned, as well as [25].
We assume that the orbitals which describe the $3d$ electrons in the Hamiltonian (2) are analogous to atomic wave functions in that they can be labeled by the set of quantum numbers $\nu = (n, l, m, \sigma)$ where $n = 3$ is the principal quantum number, $l = 2$ the total orbital angular momentum quantum number, $m \in \{-l, \dots l\}$ the z-component of orbital angular momentum, and $\sigma = \pm 1/2$ the z-component of spin. n and l could be omitted because they are identical for all $3d$ orbitals, but we keep them to stay consistent with Slater and Griffith. We introduce creation and annihilation operators $d^\dagger_{i,\nu}$ and $d_{i,\nu}$ for electrons in the $3d$ shell of the transition metal ion i. The Coulomb interaction between the $3d$ electrons then can be written as

$$\begin{aligned} H_1 &= \frac{1}{2}\sum_i \sum_{\nu_1,\nu_2,\nu_3,\nu_4} V(\nu_1,\nu_2,\nu_3,\nu_4)\; d^\dagger_{i,\nu_1} d^\dagger_{i,\nu_2} d_{i,\nu_3} d_{i,\nu_4} \\ V(\nu_1,\nu_2,\nu_3,\nu_4) &= \delta_{\sigma_1,\sigma_4}\,\delta_{\sigma_2,\sigma_3}\,\delta_{m_1+m_2,m_3+m_4} \\ &\quad \times \sum_{k=0}^{\infty} c^k(l_1 m_1; l_4 m_4)\; c^k(l_3 m_3; l_2 m_2)\; R^k(n_1 l_1, n_2 l_2, n_3 l_3, n_4 l_4)\,. \end{aligned} \tag{33}$$

Here the *Gaunt coefficients* $c^k(lm; l'm')$ are given by

$$c^k(lm; l'm') = \sqrt{\frac{4\pi}{2k+1}} \int_0^{2\pi} d\phi \int_{-1}^{1} dcos(\Theta)\, Y^*_{lm,}(\Theta,\phi)\, Y_{k,m-m'}(\Theta,\phi)\, Y_{l',m'}(\Theta,\phi) \tag{34}$$

and the *Slater integrals* R^k by

$$R^k(n_1 l_1, n_2 l_2, n_3 l_3, n_4 l_4) = e^2 \int_0^\infty dr\, r^2 \int_0^\infty dr'\, r'^2 R_{n_1 l_1}(r) R_{n_2 l_2}(r') \frac{r^k_<}{r^{k+1}_>} R_{n_4 l_4}(r) R^l_{n_3 l_3}(r')\,. \tag{35}$$

The Gaunt coefficients are pure numbers, which do not depend on the specific ion and are tabulated in textbooks [23–25]. The calculation of the Slater integrals requires knowledge of the radial wave function $R_{3,2}(r)$ of the $3d$ shell, which is often obtained from Hartree-Fock wave functions for the free transition metal ion in question. In any case, the parameters c^k and R^k can be assumed to be known. More detailed analysis shows, moreover, that for a d-shell only the terms with $k = 0, 2, 4$ in the sum in (33) differ from zero; the sum thus is finite and the Coulomb matrix elements can be calculated without problems. The noninteracting part of the

Hamiltonian can be written as

$$H_0 = \sum_{i,j}\sum_{\nu,\lambda}\Big(t_{(i,\nu),(j,\lambda)}\; d^\dagger_{i,\nu}\; p_{j,\lambda} + H.c.\Big) + \sum_{i}\sum_{\nu_1,\nu_2} C_{\nu_1,\nu_2}\; d^\dagger_{i,\nu_1}\; d_{i,\nu_2} + \sum_{i,j}\sum_{\lambda_1,\lambda_2}\; t_{(i,\lambda_1),(j,\lambda_2)}\; p^\dagger_{i,\lambda_1}\; p_{j,\lambda_2}\,. \tag{36}$$

The first term describes hybridization between the $3d$-orbitals and orbitals on other atoms, which are created by $p^\dagger_{j,\lambda}$ where λ is shorthand for some set of quantum numbers which specify these orbitals. The second term contains the orbital energies of the d-electrons and the effects of the crystalline electric field. The third term describes hybridization between orbitals other than the $3d$ orbitals. The matrix elements $t_{(i\nu),(j\lambda)}$ and $t_{(i,\lambda_1),(j,\lambda_2)}$ can be expressed in terms of relatively few parameters such as $(pd\sigma)$, $(pd\pi)\ldots$ by use of the Slater-Coster tables [26]. For a given compound the parameters in (36) can be obtained, e.g., by a fit to an LDA band structure. It was shown in the pioneering work by Fujimori and Minami [27] that the *momentum integrated* photoemission spectra of transition metal oxides can be reproduced very well by considering an octahedron-shaped cluster comprising only a single transition metal ion and its six nearest neighbor oxygen ions. If only the transition metal $3d$ and the oxygen $2p$ shells are taken into account such a cluster has $5 + 6 \cdot 3 = 23$ orbitals per spin direction. This number can be reduced considerably by noting that in octahedral symmetry for each of the five $3d$ orbitals there is precisely one linear combination of O-$2p$ orbitals on the neighboring oxygen atoms that hybridizes with it, so that the number of relevant orbitals is only 10 per spin direction, which is well manageable by the Lanczos algorithm. The Hamiltonian for the cluster reads

$$H = \sum_{\alpha,\sigma}\left(t_\alpha\, d^\dagger_{\alpha,\sigma} p_{\alpha,\sigma} + H.c.\right) + \sum_{\alpha,\beta,\sigma} c_{\alpha,\beta}\, d^\dagger_{\alpha,\sigma} d_{\beta,\sigma} + \sum_{\alpha,\beta\sigma} \tilde{c}_{\alpha,\beta}\, p^\dagger_{\alpha,\sigma} p_{\beta,\sigma} + H_1\,. \tag{37}$$

Here $p^\dagger_{\alpha,\sigma}$ create electrons in the bonding combinations of O-$2p$ orbitals and H_1 is given in (33). The finding of Fujimori and Minami immediately suggests an obvious generalization of Potthoff's treatment of the single-band Hubbard model: instead of a dimer consisting of a single correlated site and a single bath site. see Figure 9 and the Hamiltonian (29), we use an octahedron-shaped cluster comprising the 5 correlated $3d$ orbitals and 5 bath sites corresponding to the the bonding combinations of oxygen $2p$ orbitals, i.e., precisely the Hamiltonian (37) as reference system. The larger size of the clusters makes the calculation more demanding in that the eigenstates of the reference system and the Green's functions now have to be obtained by the Lanczos algorithm. Moreover, the reference system contains more than just one parameter so that (28) actually represents a system of coupled nonlinear equations. The problem still is manageable, however, for the necessary numerical procedures and possible algorithms for the solution of (28) see Refs. [28] and [29].

Here we proceed to some of the results. Figure 13 shows angle-integrated valence band photoemission spectra for the three transition metal oxides NiO, CoO, and MnO. For each compound the figure compares the computed spectral density with transition metal $3d$ character and for oxygen $2p$ character to experimental valence band photoemission spectra obtained with high

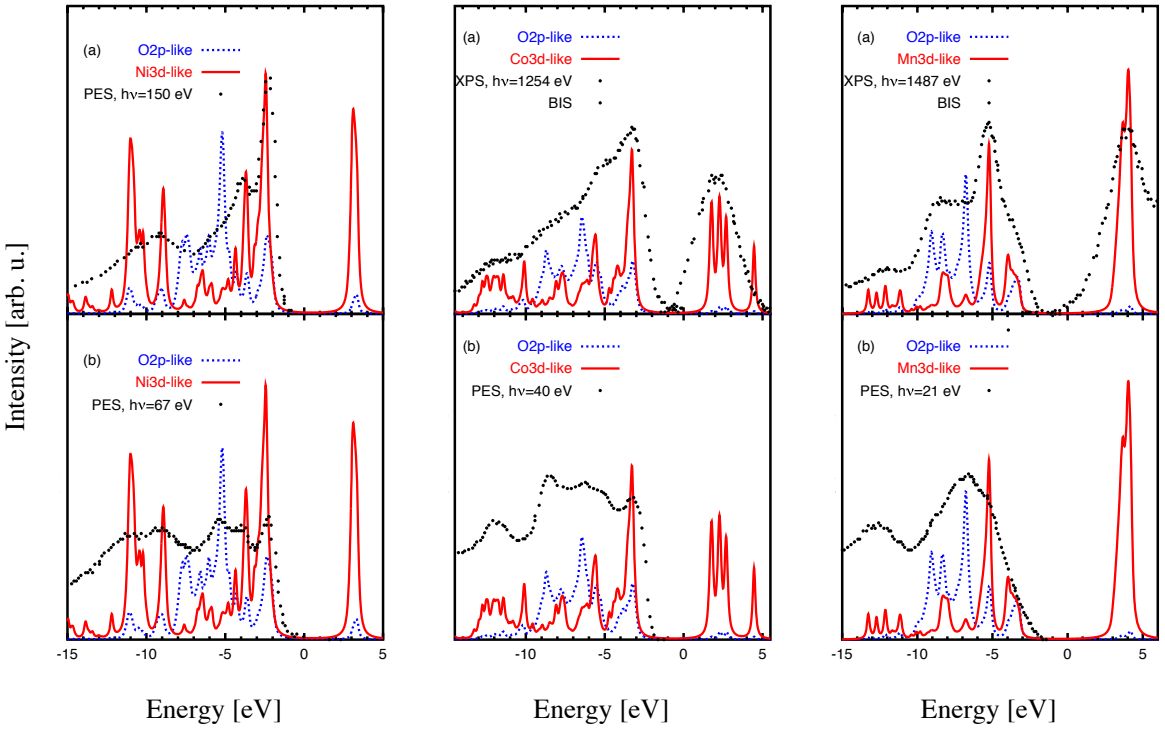

Fig. 13: *Angle integrated valence-band photoemission spectra obtained by the VCA for transition metal oxides NiO (left), CoO (center), and MnO (right), compared to experimental spectra taken with high and low photon energy. Reprinted with permission from [28], Copyright 2008 by the American Physical Society.*

(top) and relatively low photon energy (bottom). It can be seen that the experimental spectra change substantially with photon energy, and the main reason is the dependence of the photoionization cross section on photon energy [30]. As a rule of thumb, one may say that at X-ray energies the spectra show predominantly the transition-metal $3d$-like spectral density, whereas it is the oxygen $2p$-like spectral density at low photon energy. Taking this into account, there is good overall agreement between the theoretical and experimental spectra. One may also compare $\mathbf{k}$-resolved spectra and also find good agreement [28].

6 Summary

In summary, Potthoff's new idea of introducing a reference system to generate trial self-energies [4] allows one to combine the classic field theoretical work of Luttinger and Ward [1] with the numerical technique of exact diagonalization of finite systems, resulting in a method for treating strongly correlated lattice systems by exact diagonalization: the variational cluster approximation. Its variational nature makes the VCA particularly useful as exemplified by the 'dimer-DMFT' description of the metal-insulator transition. Since the VCA always gives an estimate for the grand potential, it is particularly useful for treating ordering transitions. By combining this with the very successful cluster method for transition metal oxides [27], it allows one to perform electronic structure calculations using realistic models of transition metal oxides.

7 Appendix: A theorem on determinants

Here we prove the identity

$$\frac{\partial \ln(\det A)}{\partial A_{ij}} = A^{-1}_{ji} \; .$$

We use Lapalace's formula and expand $\det(A)$ in terms of minors

$$\det(A) \;\; = \;\; \sum_{l=1,n} (-1)^{i+l} \; A_{il} M_{il} \; .$$

Since none of the minors M_{il} contains the element A_{ij} we find

$$\frac{\partial \ln(\det A)}{\partial A_{ij}} = \frac{(-1)^{i+j} M_{ij}}{\det(A)} \; .$$

Next, the i^{th} column of A^{-1} is the solution of the system of equations

$$Ac = e_i \, ,$$

where e_i is the i^{th} column of the unit matrix, which has all elements equal to zero, except for the i^{th}, which is one. We use Kramer's rule and find for the j^{th} element of the i^{th} column

$$A^{-1}_{ji} = \frac{\det(\bar{A}_j)}{\det(A)} \, ,$$

where $\bar{A}_j$ is the matrix where the j^{th} column has been replaced by e_i. Now we use again Laplace's formula for $\det(\bar{A}_j)$ and obtain

$$A^{-1}_{ji} = \frac{(-1)^{i+j} M_{ij}}{\det(A)} \, ,$$

which proves the theorem.
As an application we assume that the matrix elements of A are functions of some parameter α. We then find

$$\frac{\partial \ln(\det A)}{\partial \alpha} \;\; = \;\; \sum_{i,j} \frac{\partial \ln(\det A)}{\partial A_{ij}} \, \frac{\partial A_{ij}}{\partial \alpha} = \sum_{i,j} A^{-1}_{ji} \, \frac{\partial A_{ij}}{\partial \alpha} = \mathrm{Tr} \left(A^{-1} \, \frac{\partial A}{\partial \alpha} \right) \; .$$

References

[1] J.M. Luttinger and J.C. Ward, Phys. Rev. **118**, 1417 (1960)

[2] G. Baym and L.P. Kadanoff, Phys. Rev. **124**, 287 (1961)

[3] L. Hedin, Phys. Rev. **139**, A796 (1965)

[4] M. Potthoff, Eur. Phys. J. B **32**, 429 (2003)

[5] M. Potthoff, M. Aichhorn, and C. Dahnken, Phys. Rev. Lett. **91**, 206402 (2003)

[6] M. Potthoff: *Self-Energy-Functional Theory*
in A. Avella and F. Mancini (eds.): *Strongly Correlated Systems* (Springer (2012); see also preprint arXiv:11082183

[7] A.A. Abrikosov, L.P. Gorkov and I.E. Dzyaloshinksi:
Methods of Quantum Field Theory in Statistical Physics
(Prentice-Hall, New Jersey 1964)

[8] A.L. Fetter and J.D. Walecka: *Quantum Theory of Many-Particle Systems*
(McGraw-Hill, San Francisco, 1971)

[9] J.W. Negele and H. Orland: *Quantum Many-Particle Systems*
(Addison-Wesley, Redwood, 1988)

[10] The main difference as compared to FW is that due to the LCAO-like formulation we are using discrete momentum sums rather than integrals i.e. as compared to FW one has to replace $\frac{1}{(2\pi)^3}\int d\mathbf{k} \to \frac{\mathbf{1}}{\mathbf{N}}\sum_{\mathbf{k}}$. Similarly, the volume of the crystal is simply $V = N$.

[11] G. Baym and N. D. Mermin, J. Math. Phys. **2**,232 (1961)

[12] J.M. Luttinger, Phys. Rev. **121**, 942 (1961)

[13] M. Potthoff, Condens. Mat. Phys. **9**, 557 (2006)

[14] E. Koch: *The Lanczos Method*
in E. Pavarini, E. Koch, A. Lichtenstein, and D. Vollhardt (eds.):
The LDA+DMFT approach to strongly correlated materials
Modeling and Simulation Vol. 1 (Forschungszentrum Jülich, 2011)
`http://www.cond-mat.de/events/correl11`

[15] M. Aichhorn, E. Arrigoni, M. Potthoff, and W. Hanke Phys. Rev. B **74**, 024508 (2006)

[16] M. Potthoff, Eur. Phys. J. B **36**, 335 (2003)

[17] E. Lange, Mod. Phys. Lett. B **12**, 915 (1998)

[18] A. Georges, G. Kotliar, W. Krauth and M.J. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996)

[19] G. Moeller, Q. Si, G. Kotliar, M. Rozenberg, and D.S. Fisher,
Phys. Rev. Lett. **74**, 2082 (1995)

[20] R. Bulla, T.A. Costi, and D. Vollhardt, Phys. Rev. B **64** 045103 (2001)

[21] J. Joo and V. Oudovenko, Phys. Rev. B **64** 103102 (2001)

[22] C. Dahnken, M. Aichhorn, W. Hanke, E. Arrigoni, and M. Potthoff,
Phys. Rev. B **70**, 245110 (2004)

[23] J.C. Slater: *Quantum Theory of Atomic Structure* (McGraw-Hill, New York, 1960)

[24] J.S. Griffith: *The Theory of Transition Metal Ions* (Cambridge University Press, 1961)

[25] R. Eder: *Multiplets in Transition Metal Ions*
in E. Pavarini, E. Koch, F. Anders, and M. Jarrell:
Correlated Electrons: From Models to Materials
Modeling and Simulation Vol. 2 (Forschungszentrum Jülich, 2012)
`http://www.cond-mat.de/events/correl12`

[26] W.A. Harrison: *Electronic Structure and the Properties of Solids* (Dover, 1989)

[27] A. Fujimori and F. Minami, Phys. Rev. B **30**, 957 (1984)

[28] R. Eder, Phys. Rev. B **78**, 115111 (2008)

[29] R. Eder, Phys. Rev. B **81**, 035101 (2010)

[30] D.E. Eastman and J.L. Freeouf, Phys. Rev. Lett. **34**, 395 (1975)