TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Excitation energies in semiconductors

Citation for published version (APA):
Farid, B., Lenstra, D., & van Haeringen, W. (1986). *Excitation energies in semiconductors*. Technische Hogeschool Eindhoven.

Document status and date:
Published: 01/01/1986

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Download date: 04. Oct. 2023

EXCITATION ENERGIES IN SEMICONDUCTORS

by

Behnam Farid, Daan Lenstra, Willem van Haeringen

Department of Physics, Eindhoven University of Technology,

P.O.Box 513, 5600 MB Eindhoven, The Netherlands.

Abstract

The quasi-particle excitation structure in a semiconductor is strongly connected to and determined by the -dynamically screened- Coulomb interaction. Part of this interaction may be included in a one- electron effective potential while the remaining effects can best be described with the aid of the so-called mass operator M(1,2). Starting from a diagrammatic expression for M(1,2) we give a rederivation of Hedin's equations, relating M to a polarization function P and a vertex function Γ. Various approximation schemes with particular choices of one- electron effective potential and mass operator are discussed. The ultimate goal of the paper is to show how the excitation structure can in principle be obtained within the so-called GW (bubble) approximation, which is generally advocated to be a promising scheme.

PA 71.10

71.25R

71.25T

71.45

Internal report nr. 1986-14

Contents

Excitation energies in semiconductors

1. *The energy gap in a density functional approach*

In an earlier paper[1] an extensive discussion has been devoted to a (re)derivation of the expression

$$\Delta_g = \hbar \int d^3r d^3r' \; \psi^*_{N+1}(r,N)\; \tilde{M}(r,r';\epsilon_{N+1}(N))\; \psi_{N+1}(r';N), \qquad (1.1)$$

being the energy that has to be added to the Kohn-Sham (KS) energy gap ϵ_g of a semiconductor in order to obtain the "true" energy gap $E_g = \epsilon_g + \Delta_g$. In this expression, obtained earlier by Perdew and Levy[2] and Sham and Schlüter[3], the function $\psi_{N+1}(r;N)$ is the KS eigenfunction belonging to the *lowest* unoccupied KS level, the energy eigenvalue of which is equal to $\epsilon_{N+1}(N)$; N is the number of electrons in the charge-neutral semiconducting crystal; $\tilde{M}(r,r';\epsilon)$ is the Fourier transform (with respect to time) taken at the frequency $\epsilon/\hbar$ of the *improper* mass operator $\tilde{M}(1,2)$ defined by means of the relation (compare with eq. (6.21) of ref. 1)

$$G(1,2) = G_o(1,2) + \int d(3)d(4)\; G_o(1,3)\; \tilde{M}(3,4)\; G_o(4,2). \qquad (1.2)$$

Here $G_o(1,2)$ and $G(1,2)$ are one-particle Green functions belonging to the unperturbed KS system and to the fully interacting system, respectively. The arguments $j = 1,2,3,4 \ldots$ in the functions G_o, G and $\tilde{M}$ stand for the space-time points r_j, t_j. In ref. 1 the relation (1.2) was given in short-hand notation

$$G = G_o + G_o \tilde{M} G_o. \tag{1.2a}$$

In order to determine $\tilde{M}(1,2)$ it is useful to consider the *proper* mass operator $M(1,2)$ first. It is defined by Dyson's equation

$$G = G_o + G_o M G. \tag{1.3}$$

A relation between $\tilde{M}$ and M, not explicitly involving G was given in eq. (6.24) of ref. 1 and reads

$$\tilde{M} = M(1 - G_o M)^{-1}. \tag{1.4}$$

The Green functions G and G_o fulfil the equations

$$\left[i\hbar \frac{\partial}{\partial t_1} + \frac{\hbar^2}{2m} \nabla_1^2 - v_{eff}(r_1;N)\right] G_o(1,2) = \hbar\, \delta(1,2), \tag{1.5}$$

and

$$\left[i\hbar \frac{\partial}{\partial t_1} + \frac{\hbar^2}{2m} \nabla_1^2 - v_{eff}(r_1;N)\right] G(1,2)$$

$$- \hbar \int d(3)\, M(1,3)\, G(3,2) = \hbar\, \delta(1,2), \tag{1.6}$$

where the effective potential $v_{eff}(r_1;N)$ in the above-considered local density functional (LDF), (or KS) scheme[4], reads (compare eq. (3.6) of ref.1)

$$v_{eff}(r_1;N) = u(r_1) + v_H(r_1;N) + v_{xc}(r_1;N). \qquad (1.7)$$

In (1.7) $u(r_1)$ is the *external* potential felt by an electron, which, e.g., in a pseudopotential approach is to be identified with a sum over all ionic pseudopotentials; $v_H(r_1;N)$ is the Hartree potential ; $v_{xc}(r_1;N)$ is the exchange-correlation potential.

The above mass operator M(1,2) can formally be expressed in terms of diagrams (see for instance appendix B of ref. 1), giving M in terms of Green functions G and interaction functions $v(i,j) = v(r_i,r_j)\delta(t_i-t_j)$ where $v(r_i,r_j) = e^2/(4\pi\epsilon_o|r_i-r_j|)$. Together with the solution $\psi_{N+1}(r;N)$ of the KS equation

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + v_{eff}(r;N)\right]\psi_j(r;N) = \epsilon_j(N)\,\psi_j(r;N), \qquad (1.8)$$

and the above equations (1.2) to (1.7) this contains the complete information necessary to obtain the quantity Δ_g of eq. (1.1). In the above expressions both $v_H(r;N)$ and $v_{xc}(r;N)$ are functionals of the electron density

$$\rho(r;N) = -\,i\,G(rt,rt^+), \qquad (1.9)$$

where the notation t^+ stands for $t + \eta$, with $\eta > 0$ but infinitesimally small. In the above-assumed LDF framework the right-hand side of (1.9) can in a good approximation be written $-i\,G_o(rt,rt^+)$, expressing the fact that the Hohenberg-Kohn (HK) theory[5] leads to the exact ground-state density.

2. *The mass operator and quasi-particles*

Determining (calculating) the energy gap E_g of a semiconductor constitutes only part of determining the more general excitation structure of the semiconductor. The question therefore arises whether this excitation structure may also be determined. In this section we will indicate in how far the mass-operator concept may be of help in evaluating the excitation structure.

Before doing so we emphasize that the mass operator M(1,2) as defined by Dyson's relation $G = G_o + G_o M G$ is dependent on the *choice* of the unperturbed Green function G_o (as a matter of course the exact Green function G is independent of this choice). In the previous section the function G_o was chosen to fulfil (1.5), which is special in that it involves the (local) one-electron effective potential $v_{eff}(r;N)$ *as chosen in the LDF context*. One may, however, choose other one-electron effective potentials instead. Well-known in this respect is for instance the (local) one-electron effective potential

$$v^H_{eff}(r) = u(r) - i\int d^3r'\ v(r,r')G(r't,r't^+), \qquad (2.1)$$

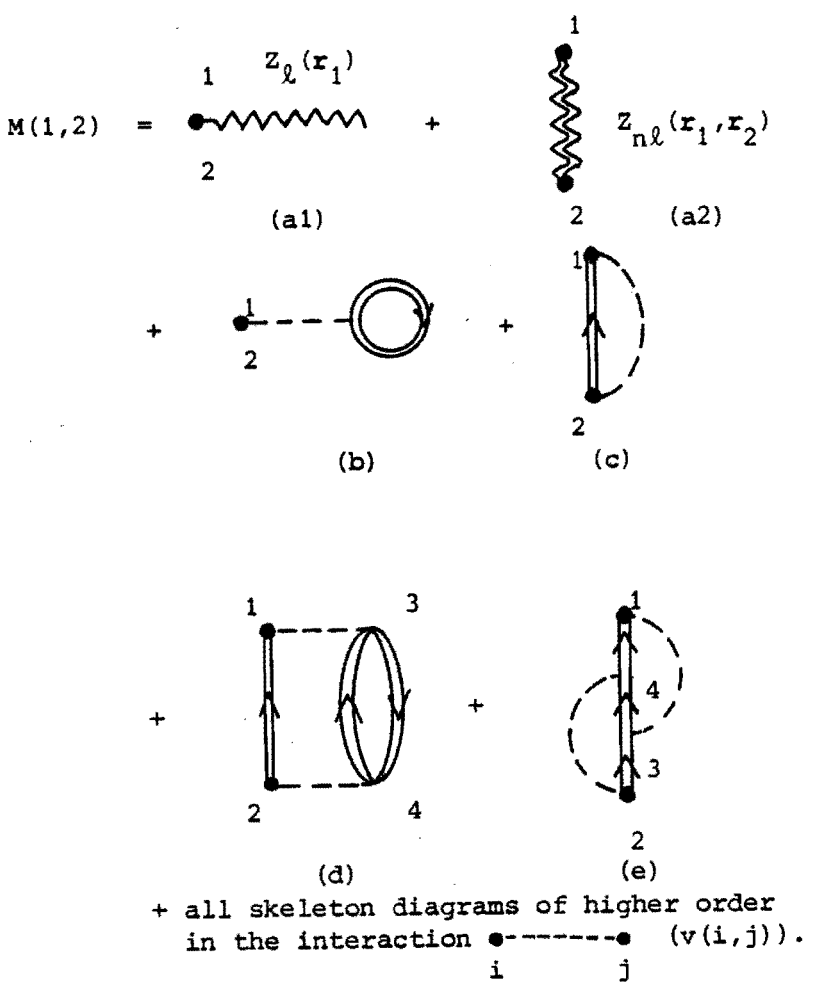
where the second term in the right-hand side of (2.1) is the Hartree potential. Note that we define this potential in terms of the exact Green function G. In actual practice $v^H_{eff}(r)$ is always used in an approximated way, by applying approximate versions of G.

As there is no necessity to restrict to local effective potentials one may introduce (non-local) one-electron effective potentials as well. Well-known in this respect is the addition to (2.1) of the non-local Hartree-Fock potential

$$v_{eff}^{HF}(r,r') = i\, v(r,r')\, G(rt,r't^+). \qquad (2.2)$$

Also other local or non-local effective potentials may be chosen (see for instance section 4). Each particular choice is representative for some subdivision of the total Hamiltonian $\hat{H}$ of the interacting system of many particles in an unperturbed term $\hat{H}_o$ describing independent particles and a perturbation term $\hat{H}_1$.

Starting from any such independent-particle picture, implying a choice of G_o, the related mass operator M(1,2) is formally given by the *complete* diagrammatic expansion (compare Fig. B5 of ref. 1)



(2.3)

In (2.3) the contribution to M(1,2) of the first diagram (a_1) is defined by

$$M^{(a_1)}(1,2) = -\frac{1}{\hbar}\,\delta(1,2)\; z_\ell(r_1), \tag{2.4}$$

where the function $z_\ell(r_1)$ is equal to $v_{eff}(r) - u(r)$, in which $v_{eff}(r)$ represents a *particular choice* of local one-electron effective potential. If, for instance, $v_{eff}(r)$ is chosen equal to (2.1), the diagram (a_1) compensates the Hartree diagram (b) completely. The contribution of diagram (a_2) is defined by

$$M^{(a_2)}(1,2) = -\frac{1}{\hbar}\,\delta(t_1-t_2)\; z_{n\ell}(r_1,r_2), \tag{2.5}$$

where $z_{n\ell}(r_1,r_2)$ stands for a particular choice of non-local effective potential $v_{eff}(r_1,r_2)$. The function $z_{n\ell}(r_1,r_2)$ might for instance be chosen equal to (2.2), just compensating the Hartree-Fock diagram (c). Double full lines in the diagrams stand for Green functions G(i,j); dotted lines indicate interaction functions v(i,j). The only extension with respect to the diagrammatic expansion in Fig. B5 of ref. 1 lies in the inclusion of the diagram (a_2) involving the contribution to M(1,2) due to a non-local one-electron effective potential. Note that, in spite of the various choices to be made for $z_\ell(r_1)$ and $z_{n\ell}(r_1,r_2)$, the contribution of each individual diagram (b), (c), (d), etc., is independent of these choices, as these diagrams involve G and v functions only. It is important to note, however, that this statement is correct as long as the *exact* G is meant, as obtained with the help of the *complete* expansion (2.3). In an approximation scheme such as the GW scheme, to be discussed in section 4, in which M(1,2) is

approximated by restricting to a *fixed* subset of diagrams, while choosing $z_\ell(r_1) = -i \int d^3r' \cdot v(r,r')G(r't,r't^+)$ and $z_{n\ell}(r_1,r_2) = 0$, the resulting approximate G (and thus the resulting quasi-particle excitation structure) will generally be different if a different choice for z_ℓ and $z_{n\ell}$ is made. It might be argued, however, in this *particular* GW approximation scheme, that the subset of diagrams is large enough to ensure that such differences are of minor importance.

Just as in (1.4) the *improper* mass operator $\tilde{M}(1,2)$ can be obtained from M and G_o. Unless this function is obtained in the LDF context, it cannot be used in the expression (1.1) for Δ_g, as will be obvious from the earlier discussion in this section.

Let us now outline the connection between the mass operator (starting from any one-electron effective potential) and the excitation spectrum. This connection is based on eq. (1.6):

Let us first introduce the Fourier transform of $G(1,2) = G(r_1,r_2;t_1-t_2)$ with respect to t_1-t_2 at frequency $\epsilon/\hbar$, which is given by

$$G(r_1,r_2;\epsilon) = \int d(t_1-t_2)\, G(1,2)\, e^{i\epsilon(t_1-t_2)/\hbar}. \qquad (2.6)$$

If we then Fourier transform eq. (1.6) with respect to t_1-t_2 (note that we replace $v_{eff}(r_1;N)$ in (1.6) by a more general local or non-local one-electron effective potential v_{eff}) one directly obtains

$$(\epsilon + \frac{\hbar^2}{2m}\nabla_1^2 - v_{eff})\, G(r_1,r_2;\epsilon)$$

$$- \hbar \int d^3r_3\, M(r_1,r_3;\epsilon)\, G(r_3,r_2;\epsilon) = \hbar\, \delta(r_1-r_2). \qquad (2.7)$$

We will prove[6,7] that the solution of this equation may be written as

$$G(r_1,r_2;\epsilon) = \hbar \sum_n \frac{\varphi_n(r_1;\epsilon)\ \psi_n^*(r_2;\epsilon)}{\epsilon - E_n(\epsilon)}, \tag{2.8}$$

where the functions $\varphi_n(r;\epsilon)$ and $\psi_n(r;\epsilon)$ are solutions of

$$(E_n(\epsilon) + \frac{\hbar^2}{2m}\nabla^2 - v_{eff})\ \varphi_n(r;\epsilon) - \hbar \int d^3r' M(r,r';\epsilon)\ \varphi_n(r';\epsilon) = 0 \tag{2.9}$$

and

$$(E_n^*(\epsilon) + \frac{\hbar^2}{2m}\nabla^2 - v_{eff})\ \psi_n(r;\epsilon) - \hbar \int d^3r' M^\dagger(r,r';\epsilon)\ \psi_n(r';\epsilon) = 0, \tag{2.10}$$

respectively. In (2.9) and (2.10) the energy $E_n(\epsilon)$ is generally complex-valued, while $M^\dagger$ is the Hermitian adjoint of M. Writing (2.9) and (2.10) in short-hand notation as $(E_n(\epsilon) - \mathcal{L}(\epsilon))\ \varphi_n = 0$ and $(E_n^*(\epsilon) - \mathcal{L}^\dagger(\epsilon))\psi_n = 0$, respectively, we easily obtain the following identities for scalar products involving the functions ψ_m and φ_n:

$$\begin{aligned}(\psi_m,\ \mathcal{L}(\epsilon)\ \varphi_n) &= E_n(\epsilon)\ (\psi_m,\varphi_n) \\ &= (\mathcal{L}^\dagger(\epsilon)\ \psi_m,\varphi_n) = E_m(\epsilon)\ (\psi_m,\varphi_n),\end{aligned} \tag{2.11}$$

where the scalar product is defined by

$$(\psi,\varphi) = \int d^3r\ \psi^*(r;\epsilon)\ \varphi(r;\epsilon). \tag{2.12}$$

Eq. (2.11) gives

$$(E_n(\epsilon) - E_m(\epsilon))\ (\psi_m,\varphi_n) = 0, \tag{2.13}$$

implying

$$(\psi_m,\varphi_n) = 0, \qquad \text{if } E_n(\epsilon) \neq E_m(\epsilon). \tag{2.14}$$

The case $E_n(\epsilon) = E_m(\epsilon)$ deserves additional attention in case of degeneracy. It can be proven[8] that the freedom of choice in functions φ_n and ψ_m in case of degeneracy makes it possible to choose them such that *all* functions φ_n and ψ_m are *bi-orthonormal* in the sense that

$$(\psi_m,\varphi_n) = \delta_{m,n}. \tag{2.15}$$

Assuming completeness of the functions φ_n, we may now write any function $F(r)$ as

$$F(r) = \sum_n \left[\int d^3r'\ \psi_n^*(r';\epsilon)F(r') \right] \varphi_n(r;\epsilon)$$

$$= \int d^3r'\ F(r') \left[\sum_n \varphi_n(r;\epsilon)\ \psi_n^*(r';\epsilon) \right], \tag{2.16}$$

which implies the closure relation

$$\sum_n \varphi_n(r;\epsilon)\ \psi_n^*(r';\epsilon) = \delta(r-r'). \tag{2.17}$$

In order to prove (2.8) to be the solution of (2.7) we now proceed by substituting (2.8) in (2.7). Application of (2.9) is then easily shown to lead to the identity (2.17), which proves (2.8) to be correct.

We note that the functions $\psi_n^*(r;\epsilon)$ occuring in (2.8) are the solutions of

$$(E_n(\epsilon) + \frac{\hbar^2}{2m}\nabla^2 - v_{eff})\ \psi_n^*(r;\epsilon) - \hbar\int d^3r' M^{\dagger *}(r,r';\epsilon)\ \psi_n^*(r';\epsilon) = 0. \tag{2.18}$$

This equation is obtained by taking the complex conjugate of eq. (2.10). As

$$M^{\dagger}(r_1,r_2;\epsilon) = M^*(r_2,r_1;\epsilon), \tag{2.19}$$

we may replace $M^{\dagger *}(r,r';\epsilon)$ in (2.18) by $M(r',r;\epsilon)$ which makes equation (2.18) for the functions $\psi_n^*(r;\epsilon)$ very similar to eq. (2.9) for the functions $\varphi_n(r;\epsilon)$, but the equations are not equal a priori, unless the property $M(r_1,r_2;\epsilon) = M(r_2,r_1;\epsilon)$ holds. This leads us to the conclusion that, quite generally, in the above-considered case of degeneracy, one has to find (i.e. to construct), according to (2.8), at each couple of energy levels $E_n(\epsilon)$ and $E_n^*(\epsilon)$ in eqs. (2.9) and (2.10), two sets of functions $\varphi_n^{(i)}$ and $\psi_n^{(i)}$ with $i = 1,2,\ldots,N$ such that $(\psi_n^{(i)}, \varphi_n^{(j)}) = \delta_{ij}$. Here N is the number of linearly

independent eigenfunctions $\varphi_n^{(i)}$ of eq. (2.9) at energy $E_n(\epsilon)$, or, equivalently, the number of linearly independent eigenfunctions $\psi_n^{(i)}$ of eq. (2.10) at energy $E_n^*(\epsilon)$.

We will now discuss the quasi-particle interpretation of $G(r_r, r_2; \epsilon)$. First, we emphasize that (2.8), although it is an *exact* representation of the Fourier-transformed Green function, is not very useful in an actual numerical scheme. In the quasi-particle approximation[6] we *assume* $G(r_1, r_2; \epsilon)$ in (2.8) to have simple poles ϵ_n for which holds

$$\epsilon_n = E_n(\epsilon_n), \tag{2.20}$$

while the (possible) singularities due to the non-analyticity of $\varphi_n(r_1; \epsilon)$, $\psi_n(r_2; \epsilon)$ or $E_n(\epsilon)$ will be neglected. The corresponding approximation to G is then obtained by putting

$$G(r_1, r_2; \epsilon) \simeq \hbar \sum_n g_n \frac{\varphi_n(r_1; \epsilon_n)\, \psi_n^*(r_2; \epsilon_n)}{\epsilon - E_n(\epsilon_n)}, \tag{2.21}$$

where

$$g_n^{-1} = 1 - \frac{dE_n(\epsilon)}{d\epsilon}\Big|_{\epsilon=\epsilon_n}. \tag{2.22}$$

The addition of g_n in (2.21) assures, in an approximate way, equal pole contributions in (2.8) and (2.21).

The set of complex energy values ϵ_n defined through (2.20) is to be interpreted as the quasi-particle spectrum. The imaginary parts of ϵ_n define

the quasi-particle lifetimes $\hbar|\mathrm{Im}(\epsilon_n)|^{-1}$. The general believe is that the values of $|\mathrm{Im}(\epsilon_n)|$ are small, such that the real parts of the poles ϵ_n can be interpreted in the usual way as the single quasi-particle energies, provided that a chemical potential μ exists such that $\mathrm{Im}(\epsilon_n) > 0$ when $\mathrm{Re}(\epsilon_n) < \mu$ and $\mathrm{Im}(\epsilon_n) < 0$ when $\mathrm{Re}(\epsilon_n) > \mu$. This chemical potential μ separates the occupied quasi-particle states from the empty ones.

In the special case that the mass operator M is Hermitian, i.e. $M = M^\dagger$, the procedure of finding the functions $\varphi_n^{(i)}$ and $\psi_n^{(i)}$ is easy. As the equations (2.9) and (2.10) are identical in this case, we may start by orthonormalizing the functions $\varphi_n^{(i)}$ according to $(\varphi_n^{(i)}, \varphi_n^{(j)}) = \delta_{ij}$. This can always be achieved if $M = M^\dagger$. Realizing that the functions $\psi_n^{(i)}$ are also solutions of eq. (2.9) we may choose them equal to $\varphi_n^{(i)}$ in this case, such that eq. (2.15) is satisfied. Eq. (2.8) then reduces to

$$G(r_1,r_2;\epsilon) = \hbar \sum_n \frac{\varphi_n(r_1;\epsilon)\, \varphi_n^*(r_2;\epsilon)}{\epsilon - E_n(\epsilon)} \,. \tag{2.23}$$

The more general case, in which $M \neq M^\dagger$ turns out to be particularly simple in the cases to be considered in this paper, in which the mass operator has the two important properties:

$$M(r_1,r_2;\epsilon) = M(r_2,r_1;\epsilon), \tag{2.24}$$

and

$$M(r_1,r_2;\epsilon) = M(r_1+R, r_2 + R;\epsilon) \tag{2.25}$$

where R is any lattice vector belonging to a given Bravais lattice. These properties apply when dealing with crystal periodicity and can generally be proven in such cases. It can easily be shown that condition (2.25) allows that the solutions of (2.9) may be chosen to be of the Bloch type, to be denoted by $\varphi_{\ell,\mathbf{k}}(\mathbf{r};\epsilon)$; here $\mathbf{k}$ is a wavevector in the first Brillouin zone (1BZ) and ℓ is a bandindex. It can also be shown, if $\varphi_{\ell,\mathbf{k}}(\mathbf{r};\epsilon)$ is an eigenfunction of (2.9) at eigenvalue E, that there exists a function $\varphi_{\ell,-\mathbf{k}}(\mathbf{r};\epsilon)$, with opposite k-vector, which is an eigenfunction of (2.9) at the same energy. Starting now from the complete set of Bloch functions $\varphi_{\ell,\mathbf{k}}(\mathbf{r};\epsilon)$ belonging to the level with energy E, the dual set of ψ^* functions, which, due to the property (2.24) must be linear combinations of the $\varphi_{\ell,\mathbf{k}}$ functions (observe that eqs. (2.18) and (2.9) are identical equations), can be simply obtained by choosing the ψ^* functions equal to the functions $\varphi_{\ell,-\mathbf{k}}(\mathbf{r};\epsilon)$. Namely $(\varphi^*_{\ell,\mathbf{k}'}, \varphi_{\ell,\mathbf{k}}) = 0$ for any $\mathbf{k}' \neq -\mathbf{k}$ as a direct consequence of the Bloch property, while $(\varphi^*_{\ell,-\mathbf{k}}, \varphi_{\ell,\mathbf{k}}) \neq 0$ (this scalar product is necessarily different from zero as otherwise we would be in conflict with (2.15)). The functions $\varphi_{\ell,\mathbf{k}}$ and $\varphi^*_{\ell,-\mathbf{k}}$ can be normalized such that their scalar product yields 1. Eq. (2.8) then assumes the form

$$G(\mathbf{r}_1,\mathbf{r}_2;\epsilon) = \hbar \sum_{\ell,\mathbf{k}} \frac{\varphi_{\ell,\mathbf{k}}(\mathbf{r}_1;\epsilon)\, \varphi_{\ell,-\mathbf{k}}(\mathbf{r}_2;\epsilon)}{\epsilon - E_\ell(\mathbf{k};\epsilon)} . \qquad (2.26)$$

We emphasize that $\varphi_{\ell,-\mathbf{k}}$ equals $\varphi^*_{\ell,\mathbf{k}}$ only in the case of a Hermitian mass operator M; if $\varphi_{\ell,\mathbf{k}}(\mathbf{r};\epsilon)$ is an eigenfunction of (2.9) at energy $E_\ell(\mathbf{k};\epsilon)$, the function $\varphi^*_{\ell,\mathbf{k}}(\mathbf{r};\epsilon)$ is generally no eigenfunction of eq. (2.9) but belongs to eq. (2.10) with energy $E^*_\ell(\mathbf{k};\epsilon)$.

The quasi-particle approximation of G in (2.26) can be written as

$$G(r_1,r_2;\epsilon) = \hbar \sum_{\ell,k} g_{\ell,k} \frac{\varphi_{\ell,k}(r_1)\varphi_{\ell,-k}(r_2)}{\epsilon - \epsilon_\ell(k)} \qquad (2.27)$$

where $\epsilon_\ell(k)$ satisfies $\epsilon_\ell(k) = E_\ell(k;\epsilon_\ell(k))$. The φ-functions in (2.27) are to be taken at $\epsilon = \epsilon_\ell(k)$ but in the corresponding notation this has been suppressed. Again the residue factor $g_{\ell,k}$ is similarly defined as in (2.22).

In principle the equations (2.8) for $G(r_1,r_2;\epsilon)$, and (2.9) for $\varphi_n(r;\epsilon)$ have to be solved selfconsistently in order to reveal all characteristics of the quasi-particle spectrum. In this procedure, it is well understood that the mass operator M, occurring in (2.7), is related to G and G_o according to Dyson's equation, while $G_o(r_1,r_2;\epsilon)$ is the solution of

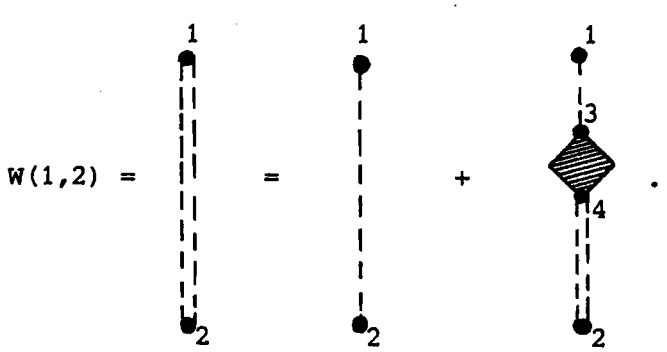
$$(\epsilon + \frac{\hbar^2}{2m}\nabla_1^2 - v_{eff})\, G_o(r_1,r_2;\epsilon) = \hbar\delta(r_1-r_2). \qquad (2.28)$$

As this task is too formidable, one is obliged to adopt some drastically simplifying approximation scheme. Several schemes have been proposed until now and some of them will be discussed in section 4. Before describing these schemes we first discuss a reordering of the diagrams in the expression (2.3) for M(1,2) in terms of a *dynamically screened interaction function* W(1,2). In this connection an analytical expression for M(1,2) is obtained, expressing M in terms of Green functions G(i,j) and interaction functions W(i,j). This expression has first been given by Hedin[9] and is rederived in section 3 by starting from the diagrammatic expansion (2.3).

3. *Diagrammatic derivation of Hedin's equations for M(1,2)*

According to (2.3) the mass operator M(1,2) can always be represented by an infinite series of skeleton Feynman diagrams in which internal (double) full lines represent Green function G(i,j) , dotted lines represent interaction functions v(i,j), while the first two diagrams (the diagrams (a_1) and (a_2) in (2.3)) contain wiggled and (double) wiggled lines representing local and non-local functions $z_{\ell}(r_1)$ and $z_{n\ell}(r_1,r_2)$, respectively. It is our aim to show that the contribution to M(1,2) due to all diagrams (but the first two) can also be expressed in a *closed* analytical form, in terms of functions G and v. The resulting expression, however, has to be supplemented with three additional equations, as it turns out to be helpful to introduce three additional functions, i.e. a screened interaction function W, a polarization function P and a vertex function Γ. The resulting equations are known as Hedin's equations[9].

We first introduce the dynamically screened interaction function W(1,2) by means of the symbolic equation

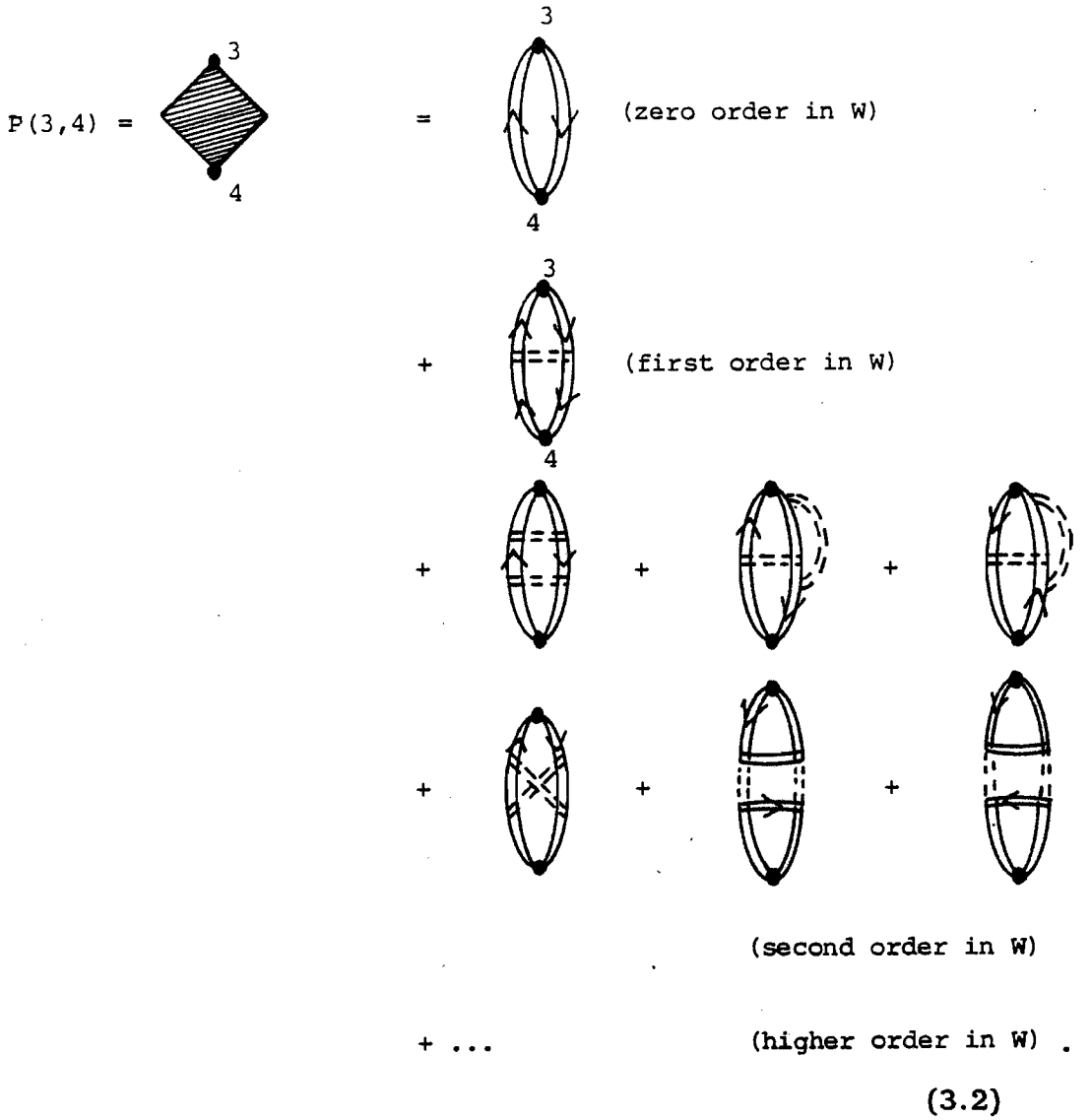


(3.1a)

In analytical form this implicit relation, defining W(1,2) reads

$$W(1,2) = v(1,2) + \int d(3)d(4)\; v(1,3)\; P(3,4)\; W(4,2). \qquad (3.1b)$$

Here the so-called polarization function P(1,2) is symbolically represented by



(3.2)

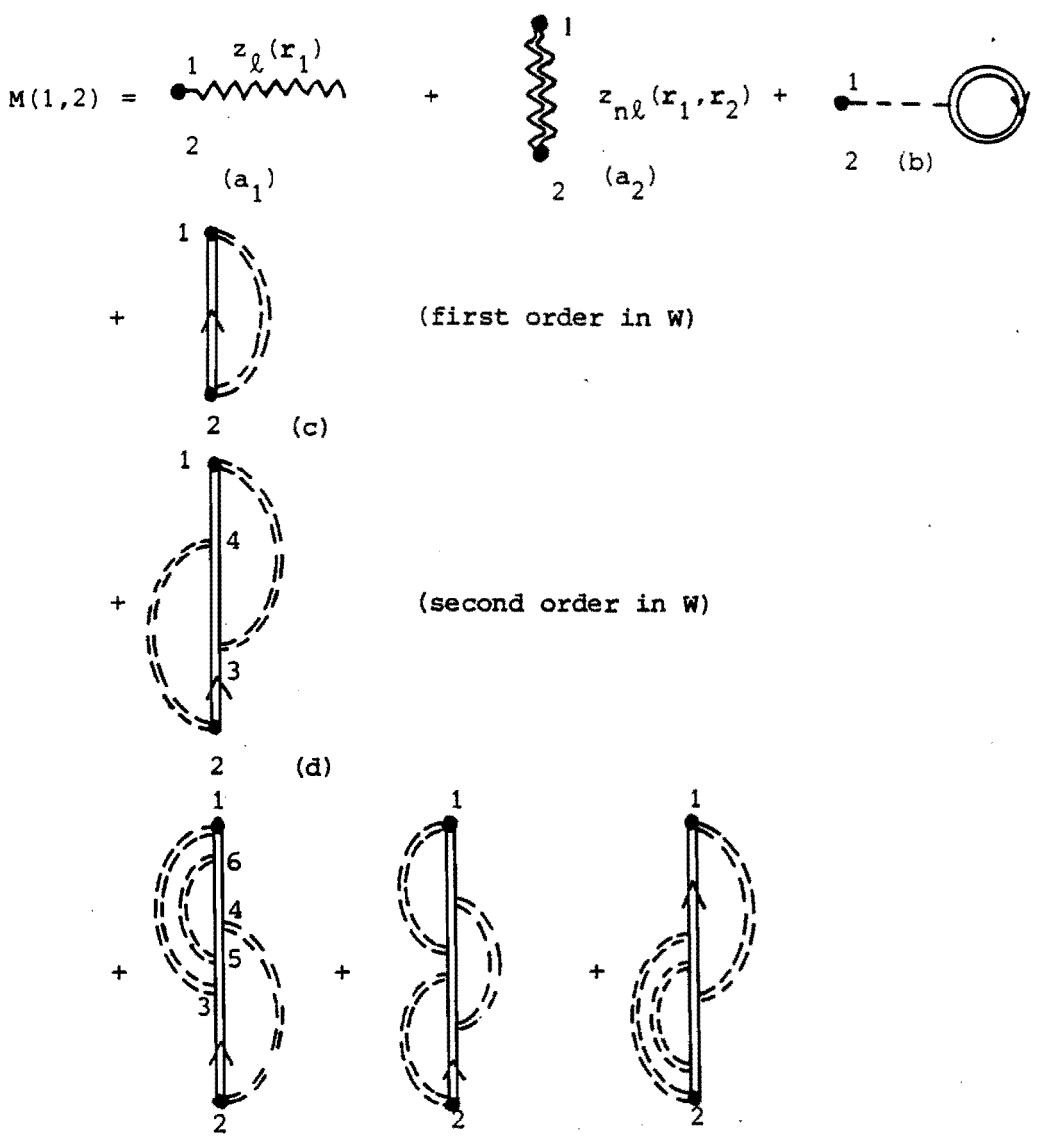
Note that due to the definition of the double dotted W-interaction lines, no P diagram in the expansion (3.2) may contain a P diagram in one of its W interaction lines, nor is it possible that a W interaction line separates a P diagram into parts which are P diagrams themselves.

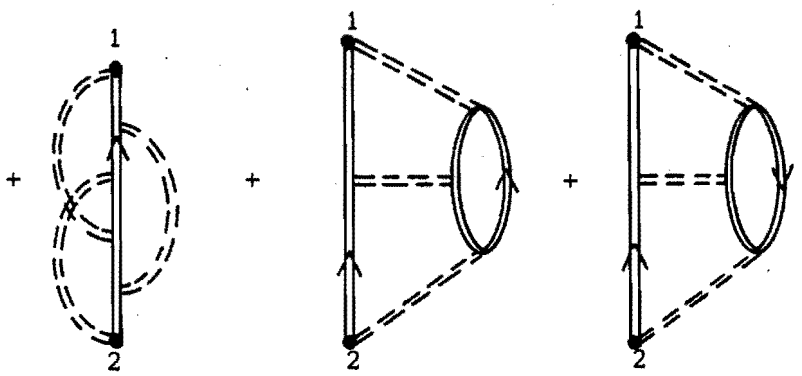
The interaction function W(1,2) is in fact a *dynamically screened* interaction, as it takes the polarization effects inherent in the function P into account in such a way that it leads to non-vanishing contributions also

if the time arguments in 1 and 2 are different, unlike in the unscreened interaction function v(1,2). The function W(1,2) can alternatively be written

$$W(1,2) = \int d(3)\, \epsilon^{-1}(1,3)\, v(3,2), \tag{3.3}$$

thus introducing the inverse dielectric screening function ϵ^{-1}. By the introduction of the double dotted lines representing the screened interaction W the number of diagrams contributing to M(1,2) is reduced: Up to the third order in W we now have





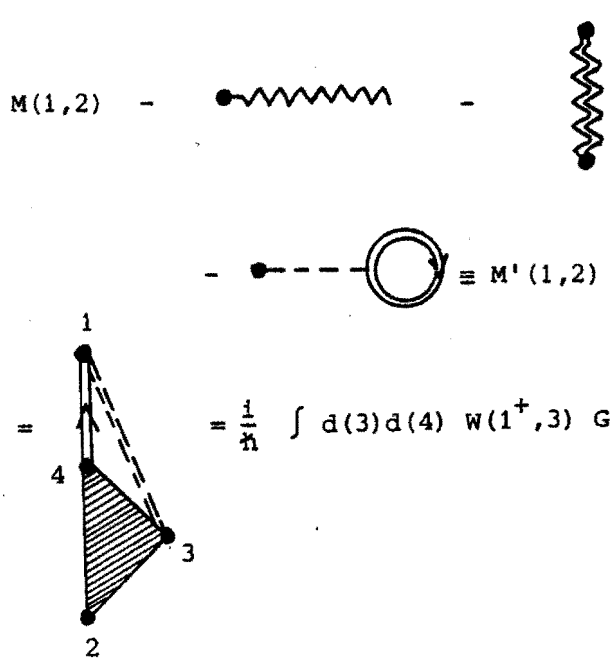
(third order in W)

+ ... (higher order in W).

(3.4)

It should be noted that the third diagram in (3.4) (the so-called Hartree diagram) still contains an unscreened interaction line. The Hartree diagram can easily be suppressed by choosing (part of) the local function $z_\ell(r_1)$ in the first diagram of (3.4) equal to $v^H_{eff}(r_1) - u(r_1)$ (see (2.1)) such that the diagram (b) is compensated by (part of) the first diagram in (3.4).

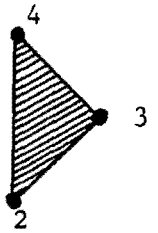
It is observed from (3.4) that all diagrams except the first three ones can formally be represented by a diagram with a *vertex part*:



$$= \frac{i}{\hbar} \int d(3)d(4)\ W(1^+,3)\ G(1,4)\ \Gamma(4,2;3), \qquad (3.5)$$

where the vertex part, according to (3.4) is formally represented by

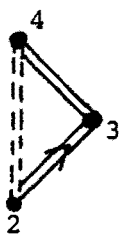
$\Gamma(4,2;3)$ =



=

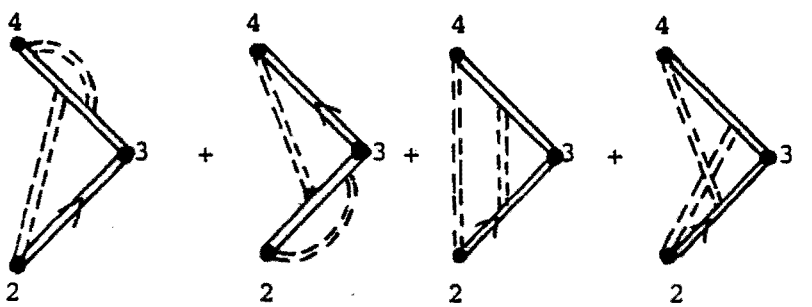
(zero order in W)

+

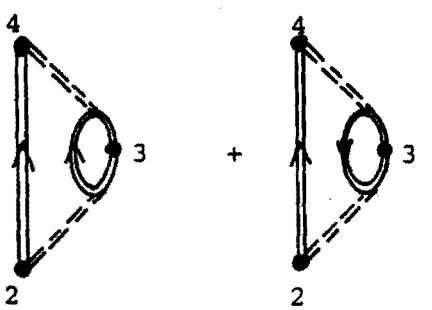


(first order in W)

+



+



(second order in W)

+ ... (higher order in W).

(3.6)

The argument 1^+ in the function $W(1^+,3)$ in the right-hand side of eq. (3.5) does not simply follow from the diagrammatic equation (3.4). Its origin lies in eq. (5.26) of ref. 1 being an equation relating the Green function G to an expectation value of a product of Heisenberg field operators, among which some have equal time arguments. In the procedure of expressing such expectation values quite generally in terms of one-particle Green functions it is inevitable to introduce infinitesimal time differences, such that products of operators with equal time arguments keep the order in which they were put before the time ordering operation is introduced. In a number of cases, e.g., cases in which an integration over some space-time variable j has to be performed, the additional + in j^+ can safely be omitted. Note that the prescription to have the argument 1^+ in eq. (3.5) is consistent with, e.g., the prescription given in eq. (B.27) of ref. 1 for calculating the contribution of the diagram (c) of Fig. B.5 (see also prescription (i) at page 49 of ref. 1).

From the diagrammatic structure of the vertex part Γ in (3.6) it is straightforward to obtain the corresponding analytical expression for the vertex function: The zero order contribution to $\Gamma(4,2;3)$ is clearly equal to $\delta(4,2)\ \delta(4,3)$ as can be deduced from the $M(1,2)$ diagram in (3.4) contributing to first order in W. As far as the higher order contributions to $\Gamma(4,2;3)$ are concerned it should be realized that the structure of the vertex diagrams in (3.6) (with the exception of the zero order diagrams in W) is such that each Γ-diagram can be obtained by starting from any skeleton diagram contributing to $M'(4,2)$, in which, however, one internal particle line representing $G(k,\ell)$ is replaced by two particle lines representing $G(k,i)$ and $G(j,\ell)$, and a diagram contributing $\Gamma(i,j;3)$. Note that this replacement has to take place in the polarization parts contributing to

double dotted interaction lines as well (see for instance two of the diagrams in (3.6) which are of second order in W).The elimination from a M'(4,2) diagram of subsequently *all* internal particle lines representing the function G(k,ℓ) is in fact equivalent with taking the functional derivative δM'(4,2)/δG(k,ℓ). Therefore, taking the above structural property of the vertex part Γ into account, we conclude to the analytical expression

$$\Gamma(4,2;3) = \delta(4,2)\ \delta(4,3)$$

$$+ \int d(i)d(j)d(k)d(\ell)\ \frac{\delta M'(4,2)}{\delta G(k,\ell)}\ G(k,i)G(j,\ell)\Gamma(i,j;3). \qquad (3.7)$$

Until now, the polarization function P(3,4) is the only function which has not yet been given in analytical form. In view of the above definition (3.6) of Γ it is straightforward, however, to conclude to*) (see also (3.2))

$$P(3,4) = -\frac{i}{\hbar} \int d(i)d(j)\ G(4,i)G(j,4^+)\Gamma(i,j;3). \qquad (3.8)$$

This completes our attempt to write M'(1,2) in closed analytical form. We succeeded in doing so by means of eqs. (3.5), (3.1b), (3.7) and (3.8).

It is tempting to assume that due to the introduction of W(1,2), or ϵ^{-1}(1,2), the series of diagrams as given in (3.4) leads to contributions which will converge much more rapidly than the series of diagrams as given in (2.3). However, it has not yet rigorously been proven whether the subsequent contributions of the diagrams in (3.4) indeed show this convergence property.

---

*) As far as the argument 4^+ in the right-hand side of (3.8) is concerned, the reader is referred to the discussion just below eq. (3.6).

We conclude this section by emphasizing that the above equations (3.5), (3.1b), (3.7) and (3.8) express M'(1,2) in terms of the Green function G and the interaction function v. These equations have to be supplemented with Dyson's equation expressing G in terms of G_o and M, where G_o is completely determined by eq. (1.5) in which $v_{eff}(r_1;N)$ has to be replaced by the effective potential one wants to start with. As a matter of course it is a considerable task to accurately determine the function M(1,2) starting from a known G_o function. In subsequent sections we will elaborate on this subject.

4. *Approximation schemes*

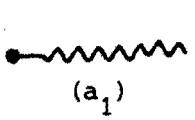
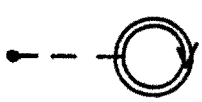
In this section we give a brief description of some approximation schemes that have been, or may be, considered in order to obtain a quasi-particle excitation structure. According to eq. (2.9) it will always be necessary to specify both v_{eff} and $M(r_1,r_2;\epsilon)$. Specifying v_{eff} is equivalent to giving the functions $z_\ell(r_1)$ and $z_{n\ell}(r_1,r_2)$, introduced in eq. (2.3) and below. Specification of $M(r_1,r_2;\epsilon)$ may, for instance, be achieved by specifying which (parts of) diagrams are to be retained in the expansion (2.3) for M(1,2).

(i) *The Hartree approximation*

In this scheme we may choose

$$z_\ell(r_1) = -i \int d^3r' v(r_1,r') \, G(r't,r't^+), \tag{4.1a}$$

$$z_{n\ell}(r_1,r_2) = 0, \tag{4.1b}$$

$$M(1,2) = \quad \underset{(a_1)}{\text{  }} \quad + \quad \text{  } \quad = 0. \tag{4.1c}$$

This choice is such that the Hartree diagram (b) in (2.3) is exactly compensated by diagram (a_1), while diagram (a_2) does not contribute. Eq. (4.1c) implies that the contribution of diagrams (c), (d), etc., is taken equal to zero. Eq. (2.9) reduces to

$$\left[-\frac{\hbar^2}{2m}\nabla_1^2 + u(r_1) - i\int d^3r' v(r_1,r')\, G(r't,r't^+)\right] \varphi_s^{\pm}(r_1;\mu\pm\epsilon_s^{\pm})$$

$$= (\mu\pm\epsilon_s^{\pm})\ \varphi_s^{\pm}(r_1;\mu\pm\epsilon_s^{\pm}). \qquad (4.2)$$

Here we distinguish between solutions $\varphi_s^{+}(r_1;\mu+\epsilon_s^{+})$ with energy eigenvalues $\mu + \epsilon_s^{+}$ and $\epsilon_s^{+} > 0$, and solutions $\varphi_s^{-}(r_1;\mu-\epsilon_s^{-})$ with energy eigenvalues $\mu - \epsilon_s^{-}$ and $\epsilon_s^{-} > 0$. It is well known that the Green function $G(r_1,r_2;\epsilon)$ connected with the system of one-particle equations (4.2) can generally be written

$$G(r_1,r_2;\epsilon) = \hbar \sum_{s\pm} \frac{\varphi_s^{\pm}(r_1;\mu\pm\epsilon_s^{\pm})\varphi_s^{\pm *}(r_2;\mu\pm\epsilon_s^{\pm})}{\epsilon - (\mu \pm \epsilon_s^{\pm}) \pm i\eta}\ , \qquad (4.3)$$

where η is a infinitesimally small positive quantity, "formally" giving the poles for which $\mathrm{Re}(\epsilon) < \mu$ a positive imaginary part, and those for which $\mathrm{Re}(\epsilon) > \mu$ a negative imaginary part (see the discussion around (2.21)). Eq. (4.3) is nothing but a special case of (2.21).

From (4.3) and (2.6) one easily derives $-iG(r_1t,r_1t^+)$ $= \sum_{s^-} |\varphi_s^{-}(r_1;\mu-\epsilon_s^{-})\ |^2 = \rho(r_1)$, representing the density of electrons at r_1. The functions $\varphi_s^{\pm}(r;\mu\pm\epsilon_s^{\pm})$ form a complete orthonormal set due to the Hermiticity of the potential occurring in (4.2). The quasi-particle excitation structure (i.e. the energies $\epsilon_s^{\pm}$ and the functions $\varphi_s^{\pm}(r;\mu\pm\epsilon_s^{\pm})$) follows in principle by solving (4.2) self-consistently, using (4.3). As no exchange and correlation effects are taken into account the obtained excitation structure is

generally considered to be inaccurate, at least for semiconductors. As $M(1,2) = 0$ according to (4.1c), the implication of Dyson's equation is that the above G coincides with the unperturbed G_o.

It is instructive to realize that the Hartree scheme may alternatively be obtained by choosing

$$z_\ell(r_1) = 0, \tag{4.4.a}$$

$$z_{n\ell}(r_1, r_2) = 0, \tag{4.4.b}$$

$M^H(1,2) =$ 

$$= -\frac{i}{\hbar}\,\delta(t_1-t_2)\;\delta(r_1-r_2)\int d^3r'\,v(r_1,r')G(r't,r't^+), \tag{4.4.c}$$

where (4.4c) precisely equals the contribution of diagram (b) in (2.3) : one easily checks that the mass-operator term in (2.9) indeed reduces to

$$\hbar\int d^3r_3\; M^H(r_1,r_3;\mu\pm\epsilon_s^\pm)\;\varphi_s^\pm(r_3;\mu\pm\epsilon_s^\pm)$$

$$= -\,i\int d^3r'\;v(r_1,r')\;G(r't,r't^+)\;\varphi_s^\pm(r_1;\mu\pm\epsilon_s^\pm). \tag{4.5}$$

just as in (4.2), while the effective potential v_{eff} in (2.9) is reduced to $u(r_1)$. Dyson's equation now reads $G = G_o' + G_o' M^H G$, where G_o' belongs to the unperturbed system with $v_{eff} = u(r)$. The function $M^H(r_1, r_3; \epsilon)$ is clearly local in space and independent of ϵ.

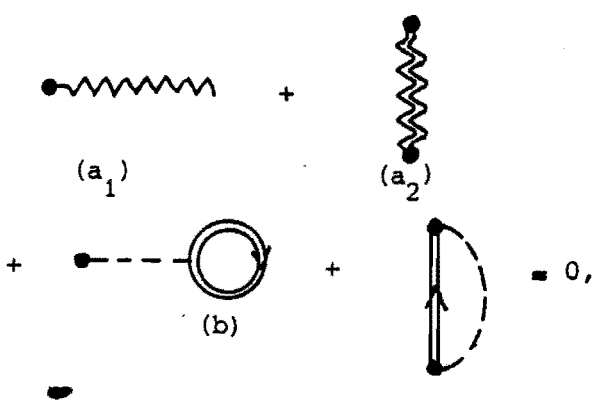
(ii) *The Hartree-Fock approximation*

In this scheme we may choose

$$z_\ell(r_1) = - i \int d^3r' \; v(r_1, r') \; G(r't, r't^+), \tag{4.6a}$$

$$z_{n\ell}(r_1, r_2) = i \; v(r_1, r_2) \; G(r_1 t, r_2 t^+), \tag{4.6b}$$

$M(1,2) =$



(4.6c)

leading in (2.3) to the exact compensation of diagrams (a_1) and (b), as well as (a_2) and (c), respectively. (Note that in diagram (c) of (2.3) the interaction v is unscreened, unlike in diagram (c) of (3.4).) The further prescription, in accordance with (4.6c) is to disregard the contributions of all other diagrams in (2.3). Eq. (2.9) reduces to

$$\left[-\frac{\hbar^2}{2m}\nabla_1^2 + u(r_1) - i\int d^3r'\; v(r_1,r')\; G(r't,r't^+)\right] \varphi_s^{\pm}(r_1;\mu\pm\epsilon_s^{\pm})$$

$$+ i\int d^3r_3\; v(r_1,r_3)\; G(r_1t,r_3t^+)\; \varphi_s^{\pm}(r_3;\mu\pm\epsilon_s^{\pm})$$

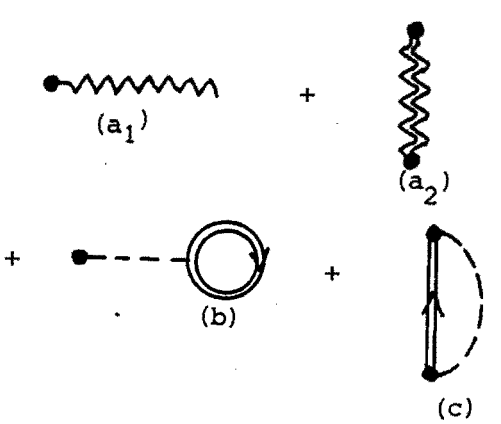
$$= (\mu \pm \epsilon_s^{\pm})\; \varphi_s^{\pm}(r_1;\mu\pm\epsilon_s^{\pm}), \qquad (4.7)$$

while G is given in terms of $\varphi_s^{\pm}(r;\mu\pm\epsilon_s^{\pm})$ and $\epsilon_s^{\pm}$, as in (4.3). The functions $\varphi_s^{\pm}$ form again a complete set of orthonormal functions due to the Hermiticity of the non-local potential in (4.7). Solving (4.7) self-consistently, using the expression (4.3) for $G(r_1,r_2;\epsilon)$ in terms of $\varphi_s^{\pm}$ and $\epsilon_s^{\pm}$, yields in principle the quasi-particle excitation structure in this scheme. As a general result[10] the energy gap obtained in the Hartree-Fock (HF) scheme overestimates the true energy gap. As $M(1,2) = 0$, according to (4.6c), the implication of Dyson's equation is that the above G coincides with the unperturbed G_o. The HF scheme can alternatively be arrived at by choosing

$$z_\ell(r_1) = 0, \qquad (4.8a)$$

$$z_{n\ell}(r_1,r_2) = 0, \qquad (4.8b)$$

$M^{HF}(1,2) =$



$$= M^{H}(1,2) + \frac{i}{\hbar}\,\delta(t_1-t_2)\; v(r_1,r_2)\; G(r_1t,r_2t^{+}), \quad (4.8c)$$

where (4.8c) is precisely equal to the contribution of diagrams (b) and (c) in (2.3). One easily checks that the mass-operator term in (2.9) indeed reduces to

$$\hbar \int d^3r_3\; M^{HF}(r_1,r_3,\mu\pm\epsilon_s^{\pm})\; \varphi_s^{\pm}(r_3\mu\pm\epsilon_s^{\pm})$$

$$= -\,i \int d^3r'\; v(r_1,r')\; G(r't,r't^{+})\; \varphi_s^{\pm}(r_1;\mu\pm\epsilon_s^{\pm})$$

$$+\,i \int d^3r_3\; v(r_1,r_3)\; G(r_1t,r_3t^{+})\; \varphi_s^{\pm}(r_3;\mu\pm\epsilon_s^{\pm}), \quad (4.9)$$

just as in (4.7), while the effective potential v_{eff} in (2.9) is reduced to $u(r_1)$. Dyson's equation now reads $G = G_o' + G_o'M^{HF}G$ where G_o' belongs to the unperturbed system with $v_{eff} = u(r)$. The function $M^{HF}(r_1,r_3;\epsilon)$ has local as well as non-local parts in space and is independent of ϵ.

(iii) *The Pratt scheme*

In this scheme, originally due to G.W. Pratt[11] we choose

$$z_{\ell}(r_1) = -\,i \int d^3r'\; v(r_1,r')\; G(r't,r't^{+}), \quad (4.10a)$$

$$z_{n\ell}(r_1,r_2) = i\; v(r_1,r_2)\; G(r_1t,r_2t^{+}) + v'_{eff}(r_1,r_2), \quad (4.10b)$$

$$M(1,2) = 0. \quad (4.10c)$$

The choice is such, as in the HF scheme, that diagrams (a_1) and (b), as well as (a_2) and (c) in the expansion (2.3) compensate exactly. The function $v'_{eff}(r_1,r_2)$ is chosen such as to compensate for *all those parts* of the contributions of *all* remaining diagrams (d), (e), (f), ..., which can be written in factorized form $(1/\hbar)\delta(t_1-t_2)f(r_1,r_2)$. Clearly the function $v'_{eff}(r_1,r_2)$ should be chosen equal to the sum of all those functions $f(r_1,r_2)$. Disregarding the remaining *parts* of the contributions of the diagrams, the mass operator M(1,2) equals zero, in accordance with (4.10c). Eq. (2.9) reduces to

$$\left[-\frac{\hbar^2}{2m}\nabla_1^2 + u(r_1) - i\int d^3r' v(r_1,r')\, G(r't,r't^+)\right]\varphi_s^{\pm}(r_1;\mu\pm\epsilon_s^{\pm})$$

$$+\int d^3r_3\left\{i\, v(r_1,r_3)\, G(r_1t,r_3t^+) + v'_{eff}(r_1,r_3)\right\}\varphi_s^{\pm}(r_3;\mu\pm\epsilon_s^{\pm})$$

$$= (\mu \pm \epsilon_s^{\pm})\varphi_s^{\pm}(r_1;\mu\pm\epsilon_s^{\pm}), \qquad (4.11)$$

where G is given in terms of $\varphi_s^{\pm}(r;\mu\pm\epsilon_s^{\pm})$ and $\epsilon_s^{\pm}$, as in (4.3). The functions $\varphi_s^{\pm}$ form again a complete set of orthonormal functions due to the Hermiticity of the non-local potential in (4.11) (the Hermiticity of the non-local operator $f(r_1,r_2)$ can easily be proven). Solving (4.11) self-consistently, using the expression for G in terms of $\varphi_s^{\pm}$ and $\epsilon_s^{\pm}$, as in (4.3), yields in principle the quasi-particle excitation structure. The Pratt scheme can be considered as a logical extension of the above HF scheme, but has to our knowledgde not been

investigated in any detail. The reason for this lies undoubtedly in the complicated character of the non-local potential $v'_{eff}(r_1,r_2)$. As $M(1,2) = 0$ according to (4.10c), the implication of Dyson's equation is that the above G coincides with the unperturbed G_o. The Pratt scheme is alternatively arrived at by choosing

$$z_\ell(r_1) = 0, \tag{4.12a}$$

$$z_{n\ell}(r_1,r_2) = 0, \tag{4.12b}$$

$$M^P(1,2) = M^{HF}(1,2) + \frac{1}{\hbar}\,\delta(t_1-t_2)v'_{eff}(r_1,r_2), \tag{4.12c}$$

where (4.12c) is precisely equal to the sum of contributions of all parts of M diagrams (d), (e), (f),... in (2.3) which are proportional to $\delta(t_1-t_2)$. One can again easily check that $M^P(1,2)$ leads to the mass-operator term in (2.12), while the effective potential v_{eff} in (2.12) is reduced to $u(r_1)$. Dyson's equation now reads $G = G'_o + G'_oM^PG$ where G'_o belongs to the unperturbed system with $v_{eff} = u(r)$. The function $M^P(r_1,r_3;\epsilon)$ has local as well as non-local parts in space and is independent of ϵ. The above Pratt scheme is as far as one can get within a scheme of quasi-particles with infinite lifetime (due to the Hermiticity of the operator in (4.11) all energy eigenvalues are real). Inclusion of *additional* parts of M diagrams will introduce the ϵ dependence of $M(r_1,r_2;\epsilon)$; it introduces complex energy values in (2.9), the *real* parts of which may essentially be different from the quasi-particle energies to be obtained in the above Pratt scheme. For that reason there is no guarantee at all that the obtained energy

band structure in the Pratt scheme will indeed be close to the experimentally known structure in semiconductors, though this might be so under circumstances. This implies in fact the incompleteness of any scheme in which M(1,2) is approximated by taking it proportional to $\delta(t_1-t_2)$; such schemes lack the incorporation of dynamical screening effects. Such effects are included for instance in the GW scheme, to be discussed at the end of this chapter. It is instructive, however, to describe the so-called Slater Xα[13] and LDF[4] schemes first.

(iv) *The Slater Xα scheme*

In this scheme we choose

$$z_\ell(r_1) = -\int d^3r'\, v(r_1,r')\, G(r't,r't^+) - \frac{3\alpha e^2}{8\pi\epsilon_o}\left[\frac{3\rho(r_1)}{\pi}\right]^{1/3}. \qquad (4.13a)$$

$$z_{n\ell}(r_1,r_2) = 0, \qquad (4.13b)$$

$$M(1,2) = 0. \qquad (4.13c)$$

Here the second term in the right-hand side of (4.13a) which we call $\mu_{xc}(r_1)$ is introduced in order to simulate the correction to quasi-particle energies due to exchange and correlation. The constant α is usually chosen in between the values 1 and 2/3. The choice $\alpha = 1$

is due to Slater[13] and is based on a calculation of the average exchange energy per particle for a uniform electron gas in which the wave functions of the electrons are represented by plane waves. The choice $\alpha = 2/3$ is due to Kohn and Sham[4] and is based on a local effective exchange one-electron potential obtained within the LDF scheme (see next section)

$$\mu_x(r_1) = d(\rho(r_1)\ \epsilon_x(\rho(r_1)))/d\rho, \tag{4.14}$$

where $\rho(r_1) = -\ i\ G(r_1 t, r_1 t^+)$ and where $\epsilon_x(\rho)$ is approximated by the expression obtained for a uniform electron gas[4]

$$\epsilon_x(\rho) = \frac{-e^2}{4\pi\epsilon_o} \cdot \frac{3}{4} \left[\frac{3\rho}{\pi}\right]^{1/3}. \tag{4.15}$$

A choice for α in between 1 and 2/3 is often made, as it is believed to account both for exchange and correlation effects in an acceptable way. The choice (4.13) is such that the contribution of diagrams (a_1) and (b) in (2.3) cancel, while diagram (a_2) contributes zero. The remaining diagrams in (2.3) are disregarded, such that $M(1,2) = 0$. Eq. (2.9) reduces to

$$\left[-\frac{\hbar^2}{2m}\nabla_1^2 + u(r_1) - i\int d^3r' v(r_1,r') G(r't,r't^+)\right.$$

$$\left. - \frac{3\alpha e^2}{8\pi\epsilon_o}\left[\frac{3\rho(r_1)}{\pi}\right]^{1/3}\right] \quad \varphi_s^{\pm}(r_1;\mu\pm\epsilon_s^{\pm})$$

$$= (\mu \pm \epsilon_s^{\pm})\varphi_s^{\pm}(r_1;\mu\pm\epsilon_s^{\pm}), \tag{4.16}$$

while G is given in terms of $\varphi_s^{\pm}(r;\mu\pm\epsilon_s^{\pm})$ and $\epsilon_s^{\pm}$, as in (4.3). The functions $\varphi_s^{\pm}$ form a complete orthonormal set due to the Hermiticity of the potential in (4.16). Solving (4.16) self-consistently, using the expression for G in terms of $\varphi_s^{\pm}$ and $\epsilon_s^{\pm}$ yields, in principle, the quasi-particle excitation structure. As M(1,2) = 0, according to (4.13c) , the implication of Dyson's equation is that the above G coincides with the unperturbed G_o.The obtained quasi-particle excitation structures within this scheme, generally deviate more or less significantly from the experimentally known structures. Due to the particular way in which the extra potential term in (4.13a) is arrived at, it is not possible to indicate precisely which (parts of) M diagrams in the expansion (2.3) may be held responsible for the related quasi-particle excitation structure. It is true that the quasi-particle excitation structure may alternatively be obtained by means of the *formal* choice

$$z_{\ell}(r_1) = 0, \tag{4.17a}$$

$$z_{n\ell}(r_1,r_2) = 0, \tag{4.17b}$$

$$M^{X\alpha}(1,2) = \frac{1}{i\hbar}\,\delta(t_1-t_2)\;\delta(r_1-r_2)$$

$$\cdot\left\{-\,i\int d^3r'v(r_1,r')G(r't,r't^+) - \frac{3\alpha e^2}{8\pi\epsilon_o}\left[\frac{3\rho(r_1)}{\pi}\right]^{1/3}\right\}, \tag{4.17c}$$

but it is not easy if not impossible to decide which (parts of) M diagrams in either (2.3) or (3.4) are accounted for by means of the expression (4.17c) . Dyson's equation now reads $G = G_o' + G_o' M^{X\alpha} G$ where G_o' belongs to the unperturbed system with $v_{eff} = u(r)$.

(v) *The density functional (DF) scheme*

In this scheme, due to Hohenberg, Kohn and Sham[4,5], we choose

$$z_\ell(r_1) = - i \int d^3r' v(r_1,r') \, G(r't,r't^+) + \frac{\delta E_{xc}[\rho]}{\delta \rho(r)}\Big|_{\rho=\rho(r_1)}, \tag{4.18a}$$

$$z_{n\ell}(r_1,r_2) = 0, \tag{4.18b}$$

$$M(1,2) = 0. \tag{4.18c}$$

The exchange-correlation-energy functional $E_{xc}[\rho]$ is often approximated by

$$E_{xc}[\rho] = \int d^3r \; \rho(r) \; \epsilon_{xc}(\rho(r)), \tag{4.19}$$

in which case $\delta E_{xc}[\rho]/\delta\rho$ equals $d(\rho \; \epsilon_{xc}(\rho))/d\rho$. The exchange-correlation energy per particle is then approximated by

$$\epsilon_{xc}(\rho) = \epsilon_x(\rho) - \frac{e^2}{4\pi\epsilon_o} \frac{0.44}{(7.8 + r_s(\rho)/a_o)a_o}, \tag{4.20}$$

where $4\pi\ (r_s(\rho))^3/3 = \rho^{-1}$ and a_o is the Bohr radius, while $\epsilon_x(\rho)$ is given by (4.15). The second term in the right-hand side of (4.20) is the well-known Wigner interpolation expression[14] for the correlation energy at intermediate particle densities. This approximative scheme is called the local density functional (LDF) scheme.

The choice in (4.18) is such that the contribution of diagrams (a_1) and (b) in (2.3) or (3.4) cancel, while diagram (a_2) contributes zero; the remaining diagrams in (2.3) or (3.4) are disregarded such that M(1,2) = 0. Eq. (2.9) reduces to

$$\left[-\frac{\hbar^2}{2m}\nabla_1^2 + u(r_1) - i\int d^3r' v(r_1,r')\, G(r't,r't^+) + \frac{\delta E_{xc}[\rho]}{\delta\rho}\Big|_{\rho=\rho(r_1)}\right] \varphi_s^{\pm}(r_1;\mu\pm\epsilon_s^{\pm}) = (\mu\pm\epsilon_s^{\pm})\varphi_s^{\pm}(r_1;\mu\pm\epsilon_s^{\pm}), \tag{4.21}$$

where ρ is given in terms of G by (1.9) , while G is given in terms of $\varphi_s^{\pm}(r;\mu\pm\epsilon_s^{\pm})$ and $\epsilon_s^{\pm}$, as in (4.3). The functions $\varphi_s^{\pm}(r;\mu\pm\epsilon_s^{\pm})$ form a complete set of orthonormal functions due to the Hermiticity of the potential in (4.21). Solving (4.21) self-consistently, using the expression for G in terms of $\varphi_s^{\pm}$ and $\epsilon_s^{\pm}$, yields in principle the DF quasi-particle excitation structure. The special character of the DF scheme shows up in the property[4] that the obtained ground-state electron density $\rho(r) = -i\ G(rt,rt^+)$ is equal to the *exact* ground-state electron density. In spite of this property, there is no reason at all to expect a correct reproduction of the excitation structure. Results obtained thusfar within the LDF scheme substantially underestimate, e.g., the experimentally known value of the energy gap[15].

As $M(1,2) = 0$, according to (4.18c) , the implication of Dyson's equation is that the above G coincides with the unperturbed G_o. Again the quasi-particle excitation structure may alternatively be obtained by means of the *formal* choice

$$z_\ell(r_1) = 0, \tag{4.22a}$$

$$z_{n\ell}(r_1,r_2) = 0, \tag{4.22b}$$

$$M^{DF}(1,2) = \frac{1}{\hbar}\,\delta(t_1-t_2)\,\delta(r_1-r_2)$$

$$\cdot\left\{-i\int d^3r'v(r_1,r')\,G(r't,r't^+)\right.$$

$$\left.+\frac{\delta\,E_{xc}[\rho]}{\delta\rho}\,\Big|_{\rho=\rho(r_1)}\right\}, \tag{4.22c}$$

but it is again not easy to decide which (parts of) M diagrams are accounted for by means of the expression (4.22c). Dyson's equation now reads $G = G_o'+G_o'M^{DF}G$ where G_o' belongs to the unperturbed system with $v_{eff} = u(r)$. If we denote the Green function G occurring in (4.22c) by G^{DF} and the exact Green function by G^{ex}, we have

$$G^{DF}(rt,rt^+) = G^{ex}(rt,rt^+), \tag{4.23}$$

while, according to (2.7)

$$\left[\epsilon+\frac{\hbar^2}{2m}\nabla_1^2-u(r_1)+i\int d^3r'v(r_1,r')\,G^{DF}(r't,r't^+)\right.$$

$$\left.-\frac{\delta\,E_{xc}[\rho]}{\delta\rho}\,\Big|_{\rho=\rho(r_1)}\right]G^{DF}(r_1,r_2;\epsilon) = \hbar\,\delta(r_1-r_2), \tag{4.24}$$

and

$$\left[\epsilon + \frac{\hbar^2}{2m}\nabla_1^2 - u(\mathbf{r}_1) + i\int d^3\mathbf{r}'v(\mathbf{r}_1,\mathbf{r}')\right.$$

$$\left.\cdot\, G^{ex}(\mathbf{r}'t,\mathbf{r}'t^+)\right]G^{ex}(\mathbf{r}_1,\mathbf{r}_2;\epsilon)$$

$$-\hbar\int d^3\mathbf{r}_3\, M_{xc}(\mathbf{r}_1,\mathbf{r}_3;\epsilon)\, G^{ex}(\mathbf{r}_3,\mathbf{r}_2;\epsilon) = \hbar\,\delta(\mathbf{r}_1-\mathbf{r}_2). \tag{4.25}$$

Here M_{xc} is the exchange-correlation mass operator, which is obtained if the local one-electron effective potential is chosen as in the Hartree scheme, (i), while no approximations are applied (consistent with G^{ex} being the exact Green function). Using (4.23) and Dyson's equation

$$G^{ex} = G^{DF} + G^{DF}(M_{xc} - V_{xc})G^{ex}, \tag{4.26}$$

where $V_{xc}(\mathbf{r}_1,\mathbf{r}_2)$ stands for $\hbar^{-1}\delta(\mathbf{r}_1-\mathbf{r}_2)\ \delta E_{xc}[\rho]/\delta\rho(\mathbf{r}_1)$, one easily obtains by transforming (4.26) back to the t_1-t_2 domain and by taking the limit $t_2\downarrow t_1$:

$$\int d^3\mathbf{r}'\ \delta E_{xc}[\rho]/\delta\rho(\mathbf{r}')\int d\epsilon\ G^{DF}(\mathbf{r},\mathbf{r}';\epsilon)\ G^{ex}(\mathbf{r}',\mathbf{r};\epsilon)$$

$$= \int d^3\mathbf{r}_1 d^3\mathbf{r}_2 \int d\epsilon\ G^{DF}(\mathbf{r},\mathbf{r}_1;\epsilon)\ M_{xc}(\mathbf{r}_1,\mathbf{r}_2;\epsilon)\ G^{ex}(\mathbf{r}_2,\mathbf{r};\epsilon). \tag{4.27}$$

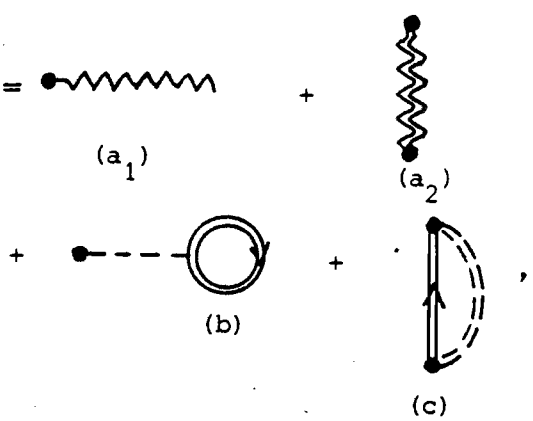
Eq. (4.27) gives the connection between $\delta E_{xc}/\delta\rho(r')$ and the exact functions M_{xc} and G^{ex}, and can for instance be used to construct it[3,16].

(vi) *The GW approximation*

In this scheme[9] we choose

$$z_\ell(r_1) = -\,i \int d^3r'\; v(r,r')\; G(r't,r't^+), \tag{4.28a}$$

$$z_{n\ell}(r_1,r_2) = 0, \tag{4.28b}$$

$M^{GW}(1,2) =$ 

, (4.28c)

leading in (2.3) and (3.4) to the exact compensation of diagrams (a_1) and (b); the diagram (a_2) contributes zero, while the further prescription is to take diagram (c) in the expansion (3.4) fully into account (We emphasize that it is the screened interaction W that operates in this diagram, unlike in the expansion (2.3) where the bare interaction v operates in diagram (c)). Eq. (4.28c) therefore reduces to

$M^{GW}(1,2) = M'^{GW}(1,2) =$  $= \frac{i}{\hbar}\, G(1,2)W(1^+,2),$ (4.29)

in accordance with (3.5), where the function $\Gamma(4,2;3)$ has been put equal to $\delta(4,2)\ \delta(4,3)$. As is clear from (4.29) the function $M^{GW}(1,2)$ factorizes into a product of two functions G and W (which explains the name GW approximation). The scheme heavily *anticipates* on the "weakness" of W, assuming therefore all higher order diagrams (in W) in (3.4) to be negligible. There is, however, no reason *a priori* why the GW scheme would be a valuable scheme. Note, that if in (4.29) the function $W(1^+,2)$ is replaced by $v(1^+,2)$, expression (4.28) reduces to $M^{HF}(1,2)$ of eq. (4.8c), such that the HF scheme is recovered. Note furthermore, that the expression (4.29) does not factorize with $\delta(t_1-t_2)$ due to the more general time dependence in $W(1^+,2)$ (see (3.1a), (3.1b) and (3.3)) which introduces the effects of dynamical screening.

At this point we might ask whether the above choice of diagrams in (4.28c) would lead to the same quasi-particle excitation structure if the functions $z_\ell(r_1)$ and $z_\ell(r_1,r_2)$ in (4.28a) and (4.28b) had been chosen differently. Suppose for instance that the functions z_ℓ and $z_{n\ell}$ had been chosen as in the HF scheme (i.e. equal to (4.6a) and (4.6b), respectively). In order to answer this question one has to realize that the particular choice of z_ℓ and $z_{n\ell}$ *in the HF case* was made because it led to contributions of diagrams (a_1) and (a_2) in (2.3) which had their *exact* counterparts in diagrams (b) and (c) of the expansion (2.3) of M(1,2). The same HF scheme will result (i.e. the quasi-particle equations (4.7) are equal) if z_ℓ and $z_{n\ell}$ are chosen equal to (4.8a) and (4.8b), respectively. Any choice of z_ℓ and $z_{n\ell}$ "in between" those two extremes (i.e. representing only *fractions of the* z_ℓ *and* $z_{n\ell}$ *of* (4.6a) and (4.6b)) is such that the exact

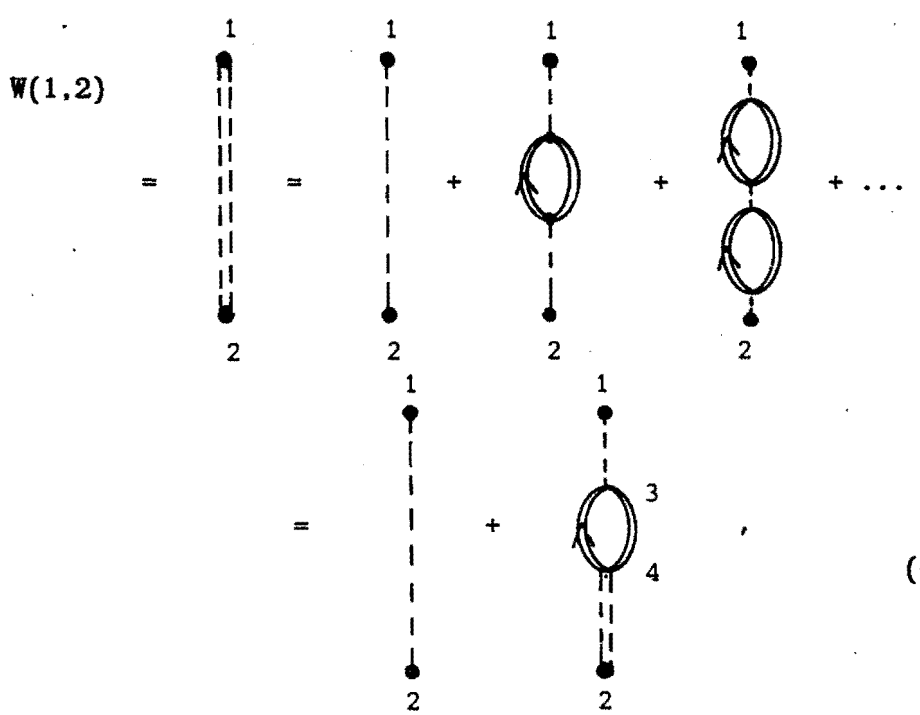
counterparts are fully *contained* in diagrams (b) and (c) of (2.3) and will therefore lead to precisely the same approximate quasi-particle excitation structure. Put otherwise: the quasi-particle equations (4.7) are completely equal in all the above cases, as an arbitrary subdivision of the operator occuring in (4.7) in an "effective-potential" part and a "mass-operator" part is of no influence to the solutions $\varphi_s^{\pm}(r;\mu\pm\epsilon_s^{\pm})$ and eigenvalues $\epsilon_s^{\pm}$. Therefore, in answering the above question concerning the choice of z_ℓ and $z_{n\ell}$ in the GW scheme one has to investigate whether the exact counterparts of the contributions of diagrams (a_1) and (a_2) in (4.28c) are indeed *contained* in diagrams (b) and (c). It will be clear from the above that, e.g., a choice for z_ℓ and $z_{n\ell}$ as in the HF case leads to the same quasi-particle equations as the exact counterparts of (a_1) and (a_2) are indeed contained in (b) and (c). But suppose that instead we choose z_ℓ and $z_{n\ell}$ as in the LDF scheme, i.e. z_ℓ and $z_{n\ell}$ equal to (4.18a) and (4.18b) , respectively. The question has then to be put whether the exact counterparts of (a_1) and (a_2) are indeed contained in (b) and (c). The answer is obviously negative: the first (Hartree) term in the right hand side of (4.18a) is indeed such that diagrams (a_1) and (b) compensate exactly; the exchange-correlation term in (4.18a) , however, has its counterpart undoubtedly in a class of M diagrams larger than the single diagram (c) in (4.28c); this was already made plausible in the earlier discussion of the LDF scheme. The conclusion is therefore that GW schemes with different choices of one-electron effective potentials may lead to different excitation structures. From a practical point of view this conclusion is very likely to be of minor importance.

Namely, the idea behind the GW scheme is in fact to neglect in expression (3.4) all diagrams with more than one W-interaction line. If W indeed *proves* to be a "weak" interaction function, this truncation in the expansion (3.4) leads to "small" deviations from the true M(1,2) function, implying that the main correction terms are included. Consequently, the above-mentioned counterpart of the exchange-correlation term in (4.18a) will be *mainly* contained in diagram (c) of (4.28c). All this points to the necessity of investigating whether or not GW schemes with different effective potentials lead to different results.

In actual practice the GW scheme is further restricted by using approximate diagrammatic relations for W. We discuss first the so-called bubble approximation scheme.

(vi a) *The bubble approximation*

In this scheme the screened interaction function W(1,2) is approximated by



(4.30a)

or

$$W(1,2) = v(1,2) - \frac{i}{\hbar}\int d(3)d(4)v(1,3)G(3,4)G(4,3^+)W(4,2)$$

$$\equiv v(1,2) + W_p(1,2), \qquad (4.30b)$$

approximating the polarization function P(3,4) of (3.2) by the bubble diagram of zero order in W. In view of the anticipated "weakness" of W, this approximation is certainly not unreasonable. It leads, according to (4.29) to (b stands for bubble)

$$M_b'^{GW}(1,2) = \frac{i}{\hbar}\left[v(1^+,2) + W_p(1^+,2)\right]G(1,2). \qquad (4.31)$$

In order to relate this scheme to the earlier HF and Pratt schemes it is instructive to consider the *approximation* to $W_p(1,2)$ in which only those contributions to W_p are retained that can be written as $\bar{W}(r_1,r_2)\,\delta(t_1-t_2)$. The related $M'(1,2)$ then reduces to

$$M_{b,static}'^{GW}(1,2) = \frac{i}{\hbar}\,\delta(t_1-t_2)\left[v(r_1,r_2) + \bar{W}(r_1,r_2)\right]G(r_1t_1,r_2t_1^+), \qquad (4.32)$$

where the argument $t_1^+-t_2$ in the δ function is replaced by t_1-t_2. When comparing (4.32) to the second term in the expression (4.8c) for $M^{HF}(1,2)$ it is observed that the interaction function $v(r_1,r_2)$ in (4.8c) is replaced by a *statically* screened interaction function

$v(r_1,r_2) + \bar{W}(r_1,r_2)$. Note furthermore that (4.32) will be close to the second term in the right hand side of expression (4.12c) for $M^P(1,2)$ in the Pratt scheme, due to the "weakness" of the screened interaction function $W(1,2)$.

In subsequent sections we will concentrate mainly on the bubble approximation scheme with *dynamically* screened interaction as it apparently shows the ability to reproduce the main characteristics of the quasi-particle structure of a semiconductor[15]. Before doing so we shortly discuss the so-called ladder-bubble approximation scheme.

(vi b) *The ladder-bubble approximation*

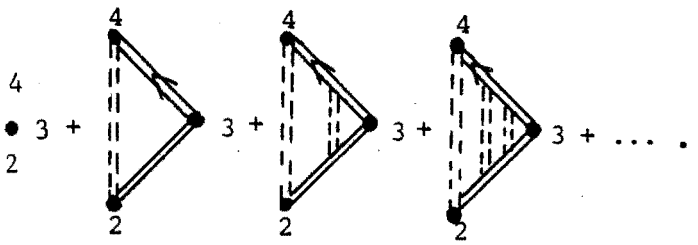
In this scheme the starting point is again the GW approximation (4.29) for $M(1,2)$. Unlike in the previous bubble approximation scheme, the expression for $\Gamma(4,2;3)$ in (3.7) is not approximated by $\delta(4,2)\ \delta(4,3)$, but is written

$$\Gamma(4,2;3) = \delta(4,2)\ \delta(4,3) + \frac{i}{\hbar}\int d(i)d(j)W(4^+,2)G(4,i)G(j,2)\Gamma(i,j;3). \qquad (4.32)$$

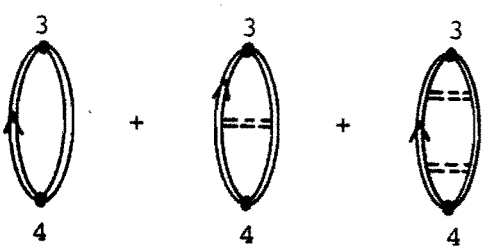
This approximate expression is arrived at by putting

$$\frac{\delta\, M^{GW}(4,2)}{\delta\, G(k,\ell)} \simeq \frac{i}{\hbar}\, W(4^+,2)\ \delta(4,k)\ \delta(2,\ell), \qquad (4.33)$$

which is obtained by neglecting the implicit dependence of W on G. Diagrammatically, eq. (4.32) can be written

$\Gamma(4,2;3) =$  (4.34)

Substitution of this expansion in the expression (3.8) for the polarization P yields (cf. (3.2)) the so-called ladder-bubble expansion P_{lb} (lb stands for ladder bubble)

$P_{lb}(3,4) =$  , (4.35)

such that

$$M_{lb}^{GW}(1,2) = \frac{i}{\hbar} W_{lb}(1^+,2)\, G(1,2), \tag{4.36}$$

with

$$W_{lb}(1,2) = v(1,2) + \int d(3)d(4)\, v(1,3)\, P_{lb}(3,4)\, W_{lb}(4,2). \tag{4.37}$$

The scheme has been applied by Strinati et al.[17] In view of the assumed "weakness" of W one may expect, however, that improvements with respect to the earlier-discussed bubble approximation scheme will be of limited significance.

5. *Relating the mass operator and the quasi-excitation structure in the bubble approximation*

The mass operator M(1,2) in the GW approximation given by eq. (4.29) can easily be Fourier transformed with respect to $t_1 - t_2$, leading to

$$M^{GW}(r_1,r_2;\epsilon) = \frac{i}{\hbar}\int_{-\infty}^{+\infty}\frac{d\epsilon'}{2\pi\hbar}\,G(r_1,r_2;\epsilon-\epsilon')\;W(r_1,r_2;\epsilon')e^{-i\epsilon'\eta/\hbar}, \tag{5.1}$$

where η is a positive infinitesimally small quantity. The functions G, W, M as well as the related functions P, ϵ and $\epsilon^{-1}(r_1,r_2;\epsilon)$ all have the translational property

$$F(r_1,r_2;\epsilon) = F(r_1+R,r_2+R;\epsilon), \tag{5.2}$$

where **R** is a lattice vector. For that reason, when Fourier transforming with respect to r_1 and r_2 we may write such functions as

$$F(r_1,r_2;\epsilon) = \frac{1}{\Omega}\sum_{q}\sum_{K,K'} e^{i(q+K)\cdot r_1}\,F_{K,K'}(q;\epsilon)e^{-i(q+K')\cdot r_2}, \tag{5.3}$$

with

$$F_{K,K'}(q;\epsilon) = \frac{1}{\Omega}\int_{\Omega} d^3r_1 d^3r_2\; e^{-i(q+K)\cdot r_1}F(r_1,r_2;\epsilon)\; e^{i(q+K')\cdot r_2}, \tag{5.4}$$

where Ω is the volume of the crystal; **q** is a wave vector in 1BZ; K and K' are reciprocal lattice vectors. Fourier transforming (5.1) in this way leads to

$$M_{K,K'}(q;\epsilon) = \frac{i}{\hbar\Omega} \sum_{q'} \sum_{G,G'} \int_{-\infty}^{+\infty} \frac{d\epsilon'}{2\pi\hbar}$$

$$\cdot\, G_{G,G'}(q',\epsilon-\epsilon')\, W_{K-G-K_q,K'-G'-K_q}(q-q'+K_q,\epsilon')\, e^{-i\epsilon'\eta/\hbar}, \tag{5.5}$$

where G, G' and K_q are reciprocal lattice vectors; K_q is introduced to ensure that $q - q' + K_q$ lies in 1BZ. It might be very difficult if not impossible to actually compute all the matrix elements of G,W, and M that are involved. Let us, however, in order to show how the matrix elements $M_{K,K'}(q;\epsilon)$ are related to the excitation spectrum, assume that the computation of these quantities can indeed be achieved. By considering the eigenvalue equations (2.9) we might then proceed as follows:

Take v_{eff} to be equal to $v^H_{eff}(r)$ of (2.1) and try to solve the equations for the *Bloch functions* $\varphi_n(r;\epsilon)$ by expanding them in plane waves $\exp(i(k+G)\cdot r)/\Omega^{1/2}$ for given k points in 1BZ. (see the discussion below (2.25), where it was argued that (2.9) indeed admits Bloch type solutions). Due to the non-Hermiticity of M, the obtained set of functions will not be orthonormal, while the eigenvalues will generally be complex valued. Eqs. (2.9) can then be written

$$\left[-\frac{\hbar^2}{2m}\nabla_1^2 + v^H_{eff}(r_1) - E_n(k;\epsilon)\right]\varphi_{n,k}(r_1;\epsilon)$$

$$+ \frac{\hbar}{\Omega}\int d^3r_3 \sum_{q} \sum_{K,K'} e^{i(q+K)\cdot r_1} M_{K,K'}(q;\epsilon) e^{-i(q+K')\cdot r_3} \varphi_{n,k}(r_3;\epsilon) = 0, \tag{5.6}$$

where n is the band index. In the quasi-particle approximation we have to find ϵ values ϵ_n such that $\epsilon_n = E_n(k;\epsilon_n)$ (see eq. (2.20)). We may now try o determine the eigenfunctions $\varphi_{n,k}(r;E_n(k))$ and the eigenvalues $E_n(k)$. (Note that we dropped the argument ϵ_n in $E_n(k)$). If we expand $\varphi_{n,k}(r;E_n(k))$ in plane waves, i.e.

$$\varphi_{n,k}(r;E_n(k)) = \sum_G d_{n,k}(G;E_n(k))e^{i(k+G)\cdot r}, \tag{5.7}$$

it is easily shown that the coefficients $d_{n,k}(G;E_n(k))$ fulfil the set of equations

$$\left[\frac{\hbar^2}{2m}(k+G)^2 - E_n(k)\right] d_{n,k}(G;E_n(k))$$

$$+ \sum_{G'} \left[v^H_{eff,G,G'} + \hbar M_{G,G'}(k;E_n(k))\right] d_{n,k}(G';E_n(k)) = 0. \tag{5.8}$$

Diagonalization of this system of equations may be achieved by standard means, leading to the quasi-particle spectrum $E_n(k)$; the real part of $E_n(k)$ yields the so-called band-structure within this scheme.

Let us now outline the possibilities (and difficulties) in obtaining the matrix elements $M_{K,K'}(q;\epsilon)$ of (5.5). Clearly, there is a need to determine the matrix elements $G_{G,G'}(q;\epsilon)$ and $W_{G,G'}(q;\epsilon)$ first. Let us (in an iterative scheme) start by calculating the matrix elements $G_{o,G,G'}(q;\epsilon)$ of *some* unperturbed Green function G_o. As the LDF scheme, (v) in section 4, may be considered as an approximation scheme in which important exchange-correlation effects are already contained in its unperturbed Green function, we will take the LDF Green function G_o as a first approximation to G. This function G_o, according to (4.3), can be written

$$G_o(r_1,r_2,;\epsilon) = \hbar\Big\{ \sum_{s=N+1}^{\infty} \frac{\varphi_s(r_1)\varphi_s^*(r_2)}{\epsilon-\epsilon_s^+-\mu+i\eta} + \sum_{s=1}^{N} \frac{\varphi_s(r_1)\varphi_s^*(r_2)}{\epsilon+\epsilon_s^--\mu-i\eta} \Big\}$$

$$= \hbar \sum_{k,\ell} \frac{\varphi_{\ell,k}(r_1)\ \varphi^*_{\ell,k}(r_2)}{[\epsilon-\epsilon_\ell(k) - i\ \eta\ \mathrm{sgn}(\mu-\epsilon_\ell(k))]} \quad . \tag{5.9}$$

Here the functions $\varphi_s(r)$ ($s = N+1,\ldots$) are the conduction-band KS eigenfunctions, while the functions $\varphi_s(r)$ ($s=1,2,\ldots N$) are the valence-band KS eigenfunctions (note that we dropped the energy argument in $\varphi_n(r)$). From now on we will choose for them the Bloch functions $\varphi_{\ell,k}(r)$, where k is in 1BZ and ℓ is the band index. The quasi-particle energies $\epsilon_s^+ + \mu$ and $-\epsilon_s^- + \mu$ are equal to the KS eigenvalues $\epsilon_j > \mu$ and $\epsilon_j < \mu$, respectively, which, henceforth we indicate by $\epsilon_\ell(k)$. The function $\mathrm{sgn}(\mu - \epsilon_\ell(k))$ enables us to write $G_o(r_1,r_2;\epsilon)$ in compact form. As $\varphi_{\ell,k}(r) = \varphi^*_{\ell,-k}(r)$ (due to time-reversal symmetry) and $\epsilon_\ell(k) = \epsilon_\ell(-k)$ we easily deduce from (5.9) the property

$$G_o(r_1,r_2;\epsilon) = G_o(r_2,r_1;\epsilon). \tag{5.10}$$

From this property it can easily be shown that *any* diagram contributing to $M(r_1,r_2;t_1 - t_2)$ has the same symmetry property. This leads to the observation that all functions G_o, G, M, $W(r_1,r_2;\epsilon)$ have the symmetry property (5.10).

The Fourier transform $G_{o,K,K'}(q;\epsilon)$, according to (5.4) and (5.9), can be written

$$G_{o,K,K'}(q;\epsilon)=\frac{\hbar}{\Omega}\sum_{k,\ell}\frac{\left\{\int_{\Omega}d^3r_1 e^{-i(q+K)\cdot r_1}\varphi_{\ell,k}(r_1)\right\}\left\{\int_{\Omega}d^3r_2 e^{i(q+K')\cdot r_2}\varphi^{*}_{\ell,k}(r_2)\right\}}{[\epsilon-\epsilon_\ell(k)-i\,\eta\ \ \mathrm{sgn}(\mu-\epsilon_\ell(k))]}, \tag{5.11}$$

and is accessible for numerical calculations[18]. The accuracy of the result is of course related to the mesh of k-points in 1BZ and the number of energy bands taken into account.

The second step is to determine the matrix elements $W_{G,G'}(q;\epsilon)$ in the GW approximation (which in the present treatment is reduced to the bubble approximation). In view of our first step in which G in (5.1) is replaced by G_o of (5.9), it is not unlogical to consider in the calculation of W in the expansion (4.30a), the bubble diagram involving G_o functions instead of G functions. In order to find the matrix elements, we first recall expression (3.3), which in Fourier transformed form reads

$$W(r_1,r_2;\epsilon) = \int d^3r_3\ \epsilon^{-1}(r_1,r_3;\epsilon)\ v(r_3,r_2). \tag{5.12}$$

where $\epsilon^{-1}(r_1,r_2;\epsilon)$ is the Fourier transform with respect to time of the inverse dynamical dielectric screening function $\epsilon^{-1}(1,2)$. This function as well as $W(r_1,r_2;\epsilon)$ itself cannot easily be expressed in terms of the (bubble) polarization function P. However, using (3.1) and (5.12), formally written as $W = v + vPW$ and $W = \epsilon^{-1}v$, we may write $(1-vP)\epsilon^{-1}v = v$ leading to $\epsilon = 1 - vP$, or

$$\epsilon(1,2) = \delta(1,2) - \int d(3)\ v(1,3)\ P(3,2), \tag{5.13}$$

which in Fourier transformed form reads

$$\epsilon(r_1,r_2;\epsilon) = \delta(r_1-r_2) - \int d^3r_3\, v(r_1,r_3)\, P(r_3,r_2;\epsilon). \qquad (5.14)$$

Eqs. (5.13) or (5.14) relate the dielectric function ϵ in a relatively simple way to the polarization function P. The strategy must therefore be to determine $\epsilon(r_1,r_2;\epsilon)$ first, in terms of $P(r_1,r_2;\epsilon)$, after which the obtained $\epsilon(r_1,r_2;\epsilon)$ is inverted in order to obtain $W(r_1,r_2;\epsilon)$. We will take P(1,2) in this first iteration cycle equal to (compare (3.8))

$$P^{(o)}(1,2) = -\frac{i}{\hbar} G_o(1,2^+)\, G_o(2,1), \qquad (5.15)$$

or, in Fourier transformed form with respect to t_1-t_2

$$P^{(o)}(r_1,r_2;\epsilon) = -\frac{i}{\hbar}\int \frac{d\epsilon'}{2\pi\hbar} G_o(r_1,r_2;\epsilon')\, G_o(r_2,r_1;\epsilon'-\epsilon)e^{i\eta\epsilon'/\hbar}. \qquad (5.16)$$

Similar to the procedure leading from $M(r_1,r_2;\epsilon)$ in (5.1) to $M_{K,K'}(q;\epsilon)$ in (5.5), we may express the Fourier transform of $P^{(o)}$ in (5.16) in terms of a convolution integral containing Fourier transforms of G_o. We will outline another way in which the ϵ'-integration in (5.16) is done first (this possibility exists in this case, as the function $G_o(r_1,r_2;\epsilon)$ is explicitly given by (5.9)). Using (5.9), eq. (5.16) can be written

$$P^{(o)}(r_1,r_2;\epsilon) = \sum_{m,n}^{\infty} \varphi_m(r_1)\,\varphi_m^*(r_2)\,\varphi_n(r_2)\,\varphi_n^*(r_1)$$

$$\cdot\,\frac{1}{2\pi i}\int_{-\infty}^{+\infty} d\epsilon'\,\frac{e^{i\eta\epsilon'/\hbar}}{[\epsilon'-\epsilon_m-i\eta\ \mathrm{sgn}(\mu-\epsilon_m)][\epsilon'-\epsilon-\epsilon_n-i\eta\ \mathrm{sgn}(\mu-\epsilon_n)]}. \tag{5.17}$$

Applying the residue theorem, use of $1-\mathrm{sgn}(\mu-\alpha) = 2(1-\theta(\mu-\alpha))$, and $1/(x \pm i\eta) = P(1/x) \mp i\pi\delta(x)$ lead to the expression

$$P^{(o)}(r_1,r_2;\epsilon) = 2\sum_{k,k',\ell,\ell'} \frac{\{\theta(\mu-\epsilon_\ell(k))-\theta(\mu-\epsilon_{\ell'}(k'))\}}{[\epsilon-\epsilon_{\ell'}(k')+\epsilon_\ell(k)+i\eta\{\theta(\mu-\epsilon_\ell(k))-\theta(\mu-\epsilon_{\ell'}(k'))\}]}$$

$$\cdot\,\varphi_{\ell',k'}(r_1)\,\varphi_{\ell',k'}^*(r_2)\,\varphi_{\ell,k}(r_2)\,\varphi_{\ell,k}^*(r_1). \tag{5.18}$$

The factor 2 in front of (5.18) originates from the fact that each ℓ,k state is degenerate with respect to spin. In fact the bubble diagram from which (5.18) originates, consists of two particle lines, both representing Green functions with *equal* spin (this is a result of the interaction $v(1,2)$ being spin independent). Taking both spin possibilities into account gives the factor 2. We note that the reasoning that led to property (5.10) also leads to the property

$$P^{(o)}(r_1,r_2;-\epsilon) = P^{(o)}(r_1,r_2;\epsilon), \tag{5.19}$$

which therefore holds for $\epsilon^{-1(o)}(r_1,r_2;\epsilon)$ and $\epsilon^{(o)}(r_1,r_2;\epsilon)$ as well. More generally, this property will hold for P, ϵ^{-1} and ϵ.

The Fourier transform $P^{(o)}_{K,K'}(q;\epsilon)$ follows from (5.4) and (5.18) and reads

$$P^{(o)}_{K,K'}(q;\epsilon)$$
$$= \frac{2}{\Omega} \sum_{k,\ell,\ell'} \frac{\{\theta(\mu-\epsilon_\ell(k))-\theta(\mu-\epsilon_{\ell'}(k+q+K_q)\}}{[\epsilon-\epsilon_{\ell'}(k+q+K_q)+\epsilon_\ell(k)+i\eta\{\theta(\mu-\epsilon_\ell(k))-\theta(\mu-\epsilon_{\ell'}(k+q+K_q))\}]}$$

$$\cdot \int_\Omega d^3r_1 \, \varphi^*_{\ell,k}(r_1) \, e^{-i(q+K)\cdot r_1} \varphi_{\ell',k+q+K_q}(r_1)$$

$$\cdot \int_\Omega d^3r_2 \, \varphi^*_{\ell',k+q+K_q}(r_2) \, e^{i(q+K')\cdot r_2} \, \varphi_{\ell,k}(r_2). \qquad (5.20)$$

Note that the summation in (5.18) over k' reduces to one term for which $k'=k+q+K_q$, where K_q is such that k' lies in 1BZ. Note furthermore that (5.19) implies

$$P^{(o)}_{K,K'}(q;-\epsilon) = P^{(o)}_{K,K'}(q;\epsilon), \qquad (5.21)$$

and similarly for the matrix elements of ϵ and ϵ^{-1}.

Eq. (5.20) is *in principle* accessible for calculations: If the KS-eigenfunctions and eigenvalues for a sufficient number of k points and band indices ℓ are known, the computation involves, apart from summations over ℓ,ℓ', a summation over 1BZ. This has to be done, however, for a variety of combinations of K,K',q and ϵ.

By Fourier transforming the expression (5.14) for $\epsilon(r_1,r_2;\epsilon)$, according to (5.4), it is observed that it is an easy task to obtain the related Fourier coefficients of the dielectric function. We have

$$\epsilon^{(o)}_{K,K'}(q;\epsilon) = \delta_{K,K'} - \tilde{v}(q+K)\, P^{(o)}_{K,K'}(q;\epsilon), \qquad (5.22)$$

in which

$$\tilde{v}(q+K) = \frac{e^2}{\epsilon_o} \frac{1}{|q+K|^2} . \qquad (5.23)$$

The desired matrix elements $W^{(o)}_{K,K'}(q;\epsilon)$ in (5.5), however, according to (5.12), read

$$W^{(o)}_{K,K'}(q;\epsilon) = \epsilon^{-1(o)}_{K,K'}(q;\epsilon)\, \tilde{v}(q+K'). \qquad (5.24)$$

This implies that, in order to obtain the matrix elements of $\epsilon^{-1(o)}$, there is a need to invert the matrix consisting of the elements $\epsilon^{(o)}_{K,K'}(q;\epsilon)$ given in (5.22). Eq. (5.24) then yields the matrix elements $W^{(o)}_{K,K'}(q;\epsilon)$ in an easy way. If a sufficient number of matrix elements of G_o and $W^{(o)}$ may indeed be obtained in this way, the matrix elements $M^{(o)}_{K,K'}(q;\epsilon)$ can be obtained by (5.5). From the behaviour of the functions $G^{(o)}(r_1,r_2;\epsilon)$ and $P^{(o)}(r_1,r_2;\epsilon)$ it can be expected that the integrand in (5.5) will tend to zero, if $|\epsilon| \to \infty$, in a sufficiently fast way, such that the actual integration range can be limited. By introducing the function [17]

$$G_{o,G,G'}(q;\epsilon,\epsilon') \equiv G_{o,G,G'}(q;\epsilon-\epsilon') + G_{o,G,G'}(q;\epsilon+\epsilon') \qquad (5.25)$$

and using the property (5.19) for $W^{(o)}_{G,G'}(q;\epsilon)$ the ϵ'-integration in (5.5) can furthermore be restricted to positive ϵ' values only. Previous treatments[15] suggest that there is no need for an ϵ'-mesh finer than 0.25 eV in such an integral. The obtained $M^{(o)}_{K,K'}(q;\epsilon)$ may now be used in the calculation of an energy structure $E^{(1)}_n(k)$ with the aid of eq. (5.8); here the superscript (1)

indicates that only a first step has been achieved in the iteration scheme to calculate the excitation structure in the bubble approximation scheme. Strictly spoken, one has to repeat the outlined calculational scheme until convergence is obtained. There are indications, however, that the first step is already sufficiently accurate[15]. Note that, if more iteration steps are needed, the expression (2.27) for the Green function applies instead of a less general expression like (2.23). In that case one is confronted with the principal difference between functions $\varphi^{*}_{\ell,k}$ and $\varphi_{\ell,-k}$ on the one hand, and the additional factor $g_{\ell,k}$ on the other. However, there are strong indications[15], that these additional features are of minor importance. The physical implications of such differences still remain to be investigated.

We end up this section by emphasizing that the above-outlined procedure has not yet proven to be a really *practical* scheme. It involves an enormous amount of data handling, which may make the scheme impractical.

The main problem is very probably the choice of (plane wave) representation. A representation in terms of local functions[15] seems to be a more plausible choice, as $M(r_1,r_2;\epsilon)$ presumably falls off strongly with increasing distance $|r_1-r_2|$. The above description may serve, however, as a guide for the underlying theory, showing how the mass operator on the one hand is related to the quasi-particle excitation structure on the other. Most schemes until now make use of model expressions for matrix elements of the ϵ^{-1} operator[13] (i.e. the so-called plasmon pole model), the parameters of which are determined by demanding that certain sum rules are fulfilled. Though such an approach indeed does not contain adjustable parameters, it does force the dielectric function to be of a special form. The ultimate theory has, in our opinion, to be free of such models. This requires the calculation of the mass operator in a fully *ab initio* (i.e. model-free) way.

This enables one to judge rigorously whether the bubble approximation scheme is indeed sufficiently general in order to yield the quasi-particle excitation structure in a semiconductor. We are currently investigating whether $E_n(k)$ can indeed be obtained by following such a procedure.

REFERENCES

1 B.Farid, D.Lenstra and W. van Haeringen, Internal report nr. 1986-4, Department of Physics, Eindhoven University of Technology (1986).

2 J.P.Perdew and M.Levy, Phys. Rev. Lett. 51, 1884 (1983).

3 L.J.Sham and M.Schlüter, Phys. Rev.Lett. 51, 1888 (1983).

4 W.Kohn and L.J.Sham, Phys. Rev. 140, A 1133 (1965).

5 P.Hohenberg and W.Kohn, Phys. Rev. 136, B 864 (1964).

6 A.J.Layzer, Phys. Rev. 129, 897 (1963).

7 P.M. Morse and H. Feshbach, "*Methods of Theoretical Physics*", McGraw Hill, New York 1953, p. 884-886; see also G. Goertzel and N. Tralli, "*Some mathematical methods of Physics*", McGraw Hill, New York 1960, chapters 4 and 8.

8. E. Kreyszig, "*Introductory Functional Analysis with Applications*", John Wiley and Sons, New York 1978, p. 444.

9. L.Hedin, Phys. Rev. 139, A 796 (1965).

10. R.Dovesi, M. Causà and G. Angonoa, Phys. Rev. B24, 4177 (1981).

11 G.W.Pratt, Jr. Phys.Rev. 118, 462 (1960).

12 J.C.Slater,"*The Self-Consistent Field for Molecules and Solids*", McGraw Hill, New York, (1974).

13 J.C.Slater, Phys. Rev. 81, 385 (1951).

14 E.Wigner, Phys. Rev. 46, 1002 (1934).

15 M.S. Hybertsen and S.G.Louie, Phys. Rev. Lett. 55, 1418 (1985).

16 M.Lannoo and M.Schlüter, Phys. Rev. B32, 3890 (1985).

17 G. Strinati, H.J.Mattausch and W.Hanke, Phys. Rev. B25, 2867 (1982).

18. G.A.Baraff and M. Schlüter, Phys. Rev. B19, 4965 (1979).