Notes on LDA+U

Tony Paxton

These notes are intended for students and other researchers familiar with the local density approximation to density functional theory and its applications, but who are unfamiliar with standard many-body theory and want to understand LDA+U and "tight-binding+U." In fact the notes are a record of my own attempt to understand LDA+U; since the first papers appeared it seemed to me that these were written in terms that must have been obvious to a many-body theorist, but which contained ideas and logical arguments that were alien to me. Consider these notes to be like a swimming pool. You can choose where to dive in, anywhere between the shallow and the deep end. I shall only assume familiarity with DFT and elementary quantum mechanics.

1. The Simple Harmonic Oscillator

All the textbooks agree that the place to start learning many-body physics, or quantum field theory as it's sometimes called, is with the simple harmonic oscillator. We all know that this classical, one-body system may be *quantised* so that it applies to oscillators that have a very small mass M and so come under the sway of the uncertainty principle, possibly the most fundamental postulate of quantum mechanics, and also possibly the easiest to comprehend. With frequency ω, and p and q being momentum and position, the classical Hamiltonian is

$$H = \frac{1}{2M}p^2 + \tfrac{1}{2}M\omega^2 q^2.$$

You probably know that to quantise, we replace p with $-\mathrm{i}\hbar\,\mathrm{d}/\mathrm{d}q$, and as a consequence there arises the *commutation relation* $[p,q] = pq - qp = -\mathrm{i}\hbar$, which you can easily demonstrate by having pq and qp operate on some arbitrary function and subtracting the results. Later I will reproduce a standard textbook argument for how this can be thought of as ultimately arising as a consequence of the uncertainty principle.

Now, the time independent Schrödinger equation is

$$H\psi_n(q) = E_n\psi_n(q)$$

with solutions for $n = 0, 1, 2\ldots$ and $n = 0$ being the ground state. We have to solve

$$\left(-\frac{\hbar^2}{2M}\frac{\mathrm{d}^2}{\mathrm{d}q^2} + \tfrac{1}{2}M\omega^2 q^2\right)\psi_n(q) = E_n\psi_n(q).$$

Of course this can be done using standard methods for solving differential equations, and you probably know that the solutions are the Hermite polynomials. However there is a neat way that most textbooks demonstrate using simple operator algebra. The big result of this is that it introduces creation and annihilation operators in a clear way; and these then pervade the rest of many-body physics. So it's very useful to get the following development understood. It's not too difficult. Let us define a dimensionless coordinate to simplify the formulas:

$$x = \sqrt{\frac{M\omega}{\hbar}}\,q.$$

Then the Schrödinger equation is

$$\tfrac{1}{2}\hbar\omega\left(-\frac{\mathrm{d}^2}{\mathrm{d}x^2} + x^2\right)\psi_n(x) = E_n\psi_n(x).$$

How do we solve this? Well, note that

$$\left(-\frac{\mathrm{d}^2}{\mathrm{d}x^2} + x^2\right)$$

is a difference of two squares like

$$(-A^2 + B^2) = (-A + B)(A + B)$$

so let's try writing

$$\begin{aligned}\left(-\frac{\mathrm{d}}{\mathrm{d}x}+x\right)\left(\frac{\mathrm{d}}{\mathrm{d}x}+x\right) &= \left(-\frac{\mathrm{d}^2}{\mathrm{d}x^2}+x^2\right)-\left(\frac{\mathrm{d}}{\mathrm{d}x}x-x\frac{\mathrm{d}}{\mathrm{d}x}\right)\\ &= \left(-\frac{\mathrm{d}^2}{\mathrm{d}x^2}+x^2\right)-1;\end{aligned}$$

but this doesn't work because $\mathrm{d}/\mathrm{d}x$ is an operator, not a number. *An extra "minus one" appears* as a consequence of the commutation relation, a little number that would be missing in the classical picture and which ultimately is a manifestation of the uncertainty principle. Now we write

$$\begin{aligned}\frac{1}{2}\hbar\omega\left(-\frac{\mathrm{d}^2}{\mathrm{d}x^2}+x^2\right)\psi_n(x) &= \frac{1}{2}\hbar\omega\left(-\frac{\mathrm{d}}{\mathrm{d}x}+x\right)\left(\frac{\mathrm{d}}{\mathrm{d}x}+x\right)\psi_n(x)+\frac{1}{2}\hbar\omega\psi_n(x)\\ &= \hbar\omega\left(a^\dagger a+\frac{1}{2}\right)\psi_n(x)\end{aligned}$$

that additional $\frac{1}{2}$ coming from our "minus one," and having defined really important operators

$$\begin{aligned}a^\dagger &= \frac{1}{\sqrt{2}}\left(-\frac{\mathrm{d}}{\mathrm{d}x}+x\right)\\ a &= \frac{1}{\sqrt{2}}\left(\frac{\mathrm{d}}{\mathrm{d}x}+x\right).\end{aligned}$$

We easily see that these obey commutation relations

$$aa^\dagger - a^\dagger a = 1$$

or

$$\left[a, a^\dagger\right] = 1.$$

Let us examine the properties of the operators a and $a^\dagger$. We have three ways to write the Hamiltonian now:

$$\begin{aligned}H &= -\frac{\hbar^2}{2M}\frac{\mathrm{d}^2}{\mathrm{d}q^2}+\frac{1}{2}M\omega^2q^2\\ &= \frac{1}{2}\hbar\omega\left(-\frac{\mathrm{d}^2}{\mathrm{d}x^2}+x^2\right)\\ &= \left(a^\dagger a+\frac{1}{2}\right)\hbar\omega.\end{aligned}$$

We consider the operator product

$$\begin{aligned}Ha &= \left(a^\dagger aa+\frac{1}{2}a\right)\hbar\omega\\ &= \left(\left(aa^\dagger-1\right)a+\frac{1}{2}a\right)\hbar\omega\\ &= \left(aa^\dagger a-\frac{1}{2}a\right)\hbar\omega\\ &= a\left(a^\dagger a-\frac{1}{2}\right)\hbar\omega\\ &= a\left(H-\hbar\omega\right).\end{aligned}$$

In the second line, we have used the commutation relation. Operating on an eigenfunction ψ_n we get

$$\begin{aligned} H\left(a\psi_n\right) &= a\left(H - \hbar\omega\right)\psi_n \\ &= \left(E_n - \hbar\omega\right)\left(a\psi_n\right) \end{aligned}$$

since $H\psi_n = E_n\psi_n$. Hence $a\psi_n$ is an eigenfunction with the energy lowered by $\hbar\omega$. This amount of energy is called a *quantum*, and a is called the *quantum annihilation operator*. Similarly you can see that

$$H\left(a^\dagger\psi_n\right) = \left(E_n + \hbar\omega\right)\psi_n$$

and $a^\dagger$ is an operator that *creates* a quantum of energy.

There cannot be negative eigenvalues (since the Hamiltonian is a sum of squares of Hermitian operators). There must exist a ground state ψ_0 with energy $E_0 > 0$. Applying the annihilation operator to this state we get

$$H\left(a\psi_0\right) = \left(E_0 - \hbar\omega\right)\left(a\psi_0\right)$$

which imples that $a\psi_0$ is an eigenfunction with energy *lower* than E_0. Since we have defined E_0 as the lowest energy, to avoid this paradox it must be true that

$$a\psi_0 = 0. \tag{1.1}$$

Therefore

$$\begin{aligned} H\psi_0 &= \left(a^\dagger a + \frac{1}{2}\right)\hbar\omega\psi_0 \\ &= \frac{1}{2}\hbar\omega\psi_0 \end{aligned}$$

So the ground state has energy $\frac{1}{2}\hbar\omega$. This is called the *zero point energy.*

Note, in passing, this neat demonstration that the little “minus one” that arose from the commutation relations is essential if the uncertainty realation is to be upheld. Since

$$H = \frac{1}{2M}p^2 + \frac{1}{2}M\omega^2 q^2$$

then the expectation value of the energy in the ground state is

$$\frac{1}{2}\hbar\omega = \frac{1}{2M}\left\langle p^2\right\rangle + \frac{1}{2}M\omega^2\left\langle q^2\right\rangle.$$

Now if three real numbers are $a + b = c$ then $ab \le \frac{1}{4}c^2$. Cancelling the $\frac{1}{2}$’s we see that

$$M\omega^2\left\langle q^2\right\rangle\frac{1}{M}\left\langle p^2\right\rangle \le \frac{1}{4}\left(\hbar\omega\right)^2$$

that is,

$$\sqrt{\left\langle q^2\right\rangle}\sqrt{\left\langle p^2\right\rangle} \le \frac{1}{2}\hbar.$$

The left hand side is the root mean square position times the r.m.s. momentum and is smaller than or equal to $\frac{1}{2}\hbar$. But by the uncertainty principle

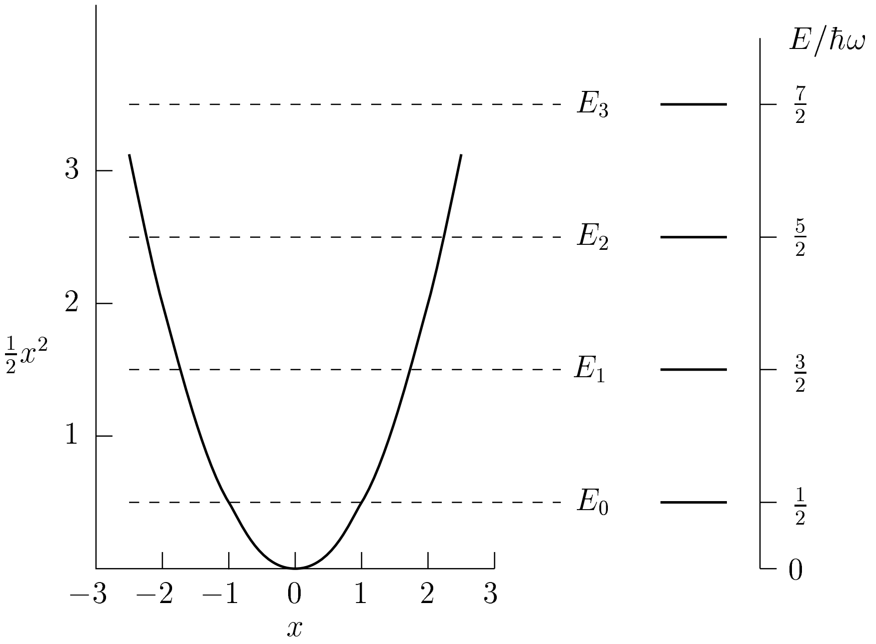
$$\Delta q\Delta p \ge \frac{1}{2}\hbar$$

so $\frac{1}{2}\hbar\omega$ is the *lowest energy the ground state can have without violating the uncertainty principle.*

What are the energy levels of the harmonic oscillator? We know two things:

1. The lowest energy level has energy $E_0 = \frac{1}{2}\hbar\omega$.
2. The operator $a^\dagger$ operating on any eigenfunction ψ_n with energy E_n results in the eigenfunction ψ_{n+1} having the next highest energy E_{n+1}. Successive energy levels are separated by the quantum of energy $\hbar\omega$.

These levels are illustrated below. The diagram shows plotted the classical potential energy $\frac{1}{2}\hbar\omega x^2$ in which the oscillator moves (in units of $\hbar\omega$). On the right are the energy levels, also in units of $\hbar\omega$.



Clearly

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega$$

and the solution to the Schrödinger equation is

$$\begin{aligned} H\psi_n &= \left(a^\dagger a + \frac{1}{2}\right)\hbar\omega\psi_n \\ &= \left(n + \frac{1}{2}\right)\hbar\omega\psi_n. \end{aligned}$$

You can see that ψ_n is an eigenfunction of the operator $a^\dagger a$ having eigenvalue n:

$$a^\dagger a\psi_n = n\psi_n. \tag{1.2}$$

$a^\dagger a$ is called the *number operator* because it counts the number of quanta of energy possessed by the oscillator when it's in the n^{th} excited state.

For completeness, we find the eigenfunctions. For the ground state eigenfunction, we use

$$a\psi_0 = 0 = \left(\frac{\mathrm{d}}{\mathrm{d}x} + x\right)\psi_0$$

which means

$$\frac{\mathrm{d}\psi_0}{\mathrm{d}x} = -x\psi_0$$

and therefore

$$\psi_0 = \mathrm{const.}\, \mathrm{e}^{-\frac{1}{2}x^2}.$$

The successive eigenfunctions are found by applying

$$\psi_n = \left(a^\dagger\right)^n \psi_0$$

and normalising. You use the fact that $(-\mathrm{d}/\mathrm{d}x + x)^n$ operating on $\mathrm{e}^{-x^2/2}$ has the effect of multiplying the exponential by a polynomial and get the well known result

$$\psi_n = (2^n n!)^{-\frac{1}{2}} \left(\frac{M\omega}{\pi\hbar}\right)^{\frac{1}{4}} \mathrm{e}^{-\frac{1}{2}x^2} H_n(x)$$

in terms of the Hermite polynomials H_n.

2. The Schrödinger Equation

It is generally said that by quantising the classical harmonic oscillator, its wave-like solutions are made into particle-like objects which are created and destroyed by the operators $a^\dagger$ and a. In the case of particles, like electrons, *first quantisation* turns the classical equations of motion into a wave equation—the Schrödinger equation. The particle-like nature is recovered when we introduce *second quantisation*; and this is needed to study particle–particle interactions, such as the electron–phonon interaction. Second quantisation is done in the following way.

Let us think of the Schrödinger equation and its Hermitian conjugate as classical wave equations:**1**

$$\begin{aligned} -\frac{\hbar^2}{2m}\nabla^2\psi + V(\mathbf{r})\psi &= \mathrm{i}\hbar\dot{\psi} \qquad (2.1) \\ -\frac{\hbar^2}{2m}\nabla^2\psi^* + V(\mathbf{r})\psi^* &= -\mathrm{i}\hbar\dot{\psi}^* \end{aligned}$$

in terms of independent variables ψ and ψ^* (or equivalently the real and imaginary parts of ψ). If we suppose a Lagrangian

$$L = \int \psi^* \left(\mathrm{i}\hbar\dot{\psi} - V(\mathbf{r})\psi + \frac{\hbar^2}{2m}\nabla^2\psi\right) \mathrm{d}\mathbf{r}$$

then the Lagrangian equation of motion leads directly to the Schrödinger equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\delta L}{\delta\dot{\psi}^*} - \frac{\delta L}{\delta\psi^*} = -\left(\mathrm{i}\hbar\dot{\psi} - V(\mathbf{r})\psi + \frac{\hbar^2}{2m}\nabla^2\psi\right) = 0.$$

We therefore form the canonically conjugate momentum

$$\pi = \frac{\delta L}{\delta \dot{\psi}} = \mathrm{i}\hbar\psi^*$$

in order to form the Hamiltonian

$$H = \int \pi\dot{\psi}\,\mathrm{d}\mathbf{r} - L = \int \left(\mathrm{i}\hbar\dot{\psi}\psi^* - \mathrm{i}\hbar\psi^*\dot{\psi} - \psi^*\frac{\hbar^2}{2m}\nabla^2\psi + \psi^* V(\mathbf{r})\psi\right)\mathrm{d}\mathbf{r}.$$

Since the first two terms in parentheses cancel we find the Hamiltonian is

$$H = \int \psi^*(\mathbf{r})\left(-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r})\right)\psi(\mathbf{r})\,\mathrm{d}\mathbf{r}. \tag{2.2}$$

If you don't like this argument, then you may regard (2.2) as a *fundamental postulate* of quantum mechanics accompanied by the commutation relations to follow. *Now*, by analogy with the harmonic oscillator, we quantise this Hamiltonian by transforming the "classical" quantities $\psi(\mathbf{r})$ and $\psi^*(\mathbf{r})$ into *field operators* $\psi(\mathbf{r})$ and $\psi^\dagger(\mathbf{r})$ which obey commutation relations

$$\begin{aligned}
\left[\psi(\mathbf{r}'), \psi^\dagger(\mathbf{r})\right] &= \psi(\mathbf{r}')\psi^\dagger(\mathbf{r}) - \psi^\dagger(\mathbf{r})\psi(\mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}') \\
\left[\psi(\mathbf{r}), \psi(\mathbf{r}')\right] &= \psi(\mathbf{r})\psi(\mathbf{r}') - \psi(\mathbf{r}')\psi(\mathbf{r}) = 0 \\
\left[\psi^\dagger(\mathbf{r}), \psi^\dagger(\mathbf{r}')\right] &= \psi^\dagger(\mathbf{r})\psi^\dagger(\mathbf{r}') - \psi^\dagger(\mathbf{r}')\psi^\dagger(\mathbf{r}) = 0
\end{aligned}$$

for bosons, and

$$\begin{aligned}
\left[\psi(\mathbf{r}'), \psi^\dagger(\mathbf{r})\right]_+ &= \psi(\mathbf{r}')\psi^\dagger(\mathbf{r}) + \psi^\dagger(\mathbf{r})\psi(\mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}') \\
\left[\psi(\mathbf{r}), \psi(\mathbf{r}')\right]_+ &= \psi(\mathbf{r})\psi(\mathbf{r}') + \psi(\mathbf{r}')\psi(\mathbf{r}) = 0 \\
\left[\psi^\dagger(\mathbf{r}), \psi^\dagger(\mathbf{r}')\right]_+ &= \psi^\dagger(\mathbf{r})\psi^\dagger(\mathbf{r}') + \psi^\dagger(\mathbf{r}')\psi^\dagger(\mathbf{r}) = 0
\end{aligned}$$

for fermions. This fermion commutator was first shown by Jordan and Wigner to be equivalent to the Pauli principle, basically because the number operators always turn out to be zero or one.

It is useful to remember that the "classical" Schrödinger equation (2.1) contains a Hamiltonian operator $-\nabla^2 + V$ and the Hamiltonian in second quantisation

$$H = \int \psi^\dagger(\mathbf{r})\left(-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r})\right)\psi(\mathbf{r})\,\mathrm{d}\mathbf{r} \tag{2.3}$$

is also an operator *but these are not the same thing.* In the first case, the Hamiltonian acts upon an eigenfunction, say $\phi_i(\mathbf{r})$, which is a function of position, yielding its eigenvalue E_i; the second quantised Hamiltonian (2.3) acts upon a (possibly many-body) eigenfunction Φ in an abstract way through the influence of position-dependent field operators. So the second quantised Schrödinger equation is $H\Phi = E\Phi$ and E is the expectation value of the total energy.

Fundamental to many-body physics is the expansion of the field operators in terms of single-particle solutions of the time-independent Schrödinger equation

$$-\frac{\hbar^2}{2m}\nabla^2\phi_i(\mathbf{r}) + V(\mathbf{r})\phi_i(\mathbf{r}) = E_i\phi_i(\mathbf{r}).$$

This is because the one-particle problem is usually thought of as soluble so the $\phi_i(\mathbf{r})$ are known and

$$\begin{aligned}\psi(\mathbf{r}) &= \sum_i a_i\phi_i(\mathbf{r})\\ \psi^\dagger(\mathbf{r}) &= \sum_i a_i^\dagger\phi_i^*(\mathbf{r})\end{aligned}$$

defines creation and annihilation operators for single-particle states which obey the commutation relations

$$\begin{aligned}\left[a_i, a_j^\dagger\right] &= \delta_{ij}\\ \left[a_i, a_j\right] &= 0\\ \left[a_i^\dagger, a_j^\dagger\right] &= 0, \end{aligned}\tag{2.4}$$

with equivalent anticommutation relations for fermions:

$$\begin{aligned}\left[a_i, a_j^\dagger\right]_+ &= \delta_{ij}\\ \left[a_i, a_j\right]_+ &= 0\\ \left[a_i^\dagger, a_j^\dagger\right]_+ &= 0. \end{aligned}\tag{2.5}$$

As I just said, the Schrödinger equation in *second quantisation* is

$$H\Phi = E\Phi,$$

with H given by (2.3); but also by expanding the field operators in the $\phi_i(\mathbf{r})$,

$$\begin{aligned}H &= \sum_{ij} H_{ij}a_i^\dagger a_j \qquad (2.6)\\ &= \sum_i E_i a_i^\dagger a_i.\end{aligned}$$

Here, $H_{ij} = \int\phi_i^*(-\nabla^2 + V)\phi_j\mathrm{d}\mathbf{r}$. The second line follows since the ϕ_i diagonalise the single particle Hamiltonian, but we will use (2.6) in the case that the field operators are expanded in some other orthogonal single particle basis set.

The ground state Φ_0 is called the *vacuum state* in many-body physics and is defined similarly to (1.1):

$$a_i\Phi_0 = 0, \qquad \text{for all } i,$$

and the solutions of the Schrödinger equation, if H is given by (2.6), are constructed from the vacuum state as

$$\Phi = \prod_i \left(a_i^\dagger\right)^{n_i}\Phi_0,$$

where n_i are the occupation numbers of state i, which for fermions are either zero or one. They are expectation values of the number operator, which will be discussed in section 4, below.

At this point it will be very useful to work through a few examples. The *particle density operator* in second quantisation is

$$\rho(\mathbf{r}) = \psi^\dagger(\mathbf{r})\psi(\mathbf{r}),$$

and we can construct its expectation value as the bra-ket

$$\left\langle \Phi \left| \psi^\dagger(\mathbf{r})\psi(\mathbf{r}) \right| \Phi \right\rangle .$$

If Φ is the single particle eigenfunction

$$\Phi = a_k^\dagger \Phi_0, \tag{2.7}$$

then the expectation value of the particle density operator is

$$\left\langle a_k^\dagger \Phi_0 \left| \sum_i a_i^\dagger \phi_i^* \sum_j a_j \phi_j \right| a_k^\dagger \Phi_0 \right\rangle .$$

The trick is to get all the annihilation operators over to the right where they produce zero when they operate on the vacuum state. Freely using the commutation relations (2.4) you will find that the expectation value for this eigenfunction is simply $\phi_k^*(\mathbf{r})\phi_k(\mathbf{r})$ which is what you expect since this is just the single particle density matrix element.

Another example is the position operator in second quantisation,

$$\int \psi^\dagger(\mathbf{r})\, x\, \psi(\mathbf{r})\, \mathrm{d}\mathbf{r}.$$

Using again the eigenfunction (2.7) we soon get for the expectation value of the position operator

$$\int \phi_k^*(\mathbf{r})\, x\, \phi_k(\mathbf{r})\, \mathrm{d}\mathbf{r}$$

as you'd expect. It's more interesting to work out the expectation value of the position operator in the two particle eigenfunction

$$\Phi = a_k^\dagger a_l^\dagger \Phi_0. \tag{2.8}$$

You'll get

$$\int \phi_k^*(\mathbf{r})\, x\, \phi_k(\mathbf{r})\, \mathrm{d}\mathbf{r} + \int \phi_l^*(\mathbf{r})\, x\, \phi_l(\mathbf{r})\, \mathrm{d}\mathbf{r}.$$

Our final example is the most instructive and non trivial. As Mahan remarks in his book, if the Hamiltonian were always bilinear as (2.6), or equivalently contained only single particle operators as (2.3) then the problem would always be solvable by diagonalisation; indeed there'd be no need for second quantisation. *Now* let us introduce particle–particle interactions, a good example being the operator of the Coulomb interaction energy,

$$\frac{1}{2}\int \psi^\dagger(\mathbf{r})\psi^\dagger(\mathbf{r}')\, W(\mathbf{r}\mathbf{r}')\, \psi(\mathbf{r}')\psi(\mathbf{r})\, \mathrm{d}\mathbf{r}\, \mathrm{d}\mathbf{r}'. \tag{2.9}$$

Here, in Gaussian units,

$$W(\mathbf{r}\mathbf{r}') \equiv \frac{e^2}{|\mathbf{r}-\mathbf{r}'|},$$

and the order of the field operators has been chosen carefully so that a particle does not interact with itself, and the operator produces zero if there is only one particle. You can demonstrate this last property by showing that the expectation value of (2.9) in the single particle eigenfunction (2.7) is indeed zero. A quite lengthy exercise is to calculate the expectation value of (2.9) in the two particle eigenfunction (2.8). Eventually you will find

$$\begin{aligned}\frac{1}{2}\int\Big(&\phi_k^*(\mathbf{r})\phi_k(\mathbf{r})W(\mathbf{r}\mathbf{r}')\phi_l^*(\mathbf{r}')\phi_l(\mathbf{r}') + \phi_l^*(\mathbf{r})\phi_l(\mathbf{r})W(\mathbf{r}\mathbf{r}')\phi_k^*(\mathbf{r}')\phi_k(\mathbf{r}')\\ &\pm\phi_l^*(\mathbf{r})\phi_k(\mathbf{r})W(\mathbf{r}\mathbf{r}')\phi_k^*(\mathbf{r}')\phi_l(\mathbf{r}') \pm \phi_k^*(\mathbf{r})\phi_l(\mathbf{r})W(\mathbf{r}\mathbf{r}')\phi_l^*(\mathbf{r}')\phi_k(\mathbf{r}')\Big)\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}', \qquad (2.10)\end{aligned}$$

with the plus sign for bosons and the minus sign for fermions; and you may well believe that for the general N-particle eigenfunction

$$\Phi = a^\dagger_{k_1} a^\dagger_{k_2} \cdots a^\dagger_{k_N}\Phi_0 \qquad (2.11)$$

the expectation value of the Coulomb interaction energy is

$$\begin{aligned}\frac{1}{2}\sum_{i\neq j}\int\Big(&\phi_{k_i}^*(\mathbf{r})\phi_{k_i}(\mathbf{r})W(\mathbf{r}\mathbf{r}')\phi_{k_j}^*(\mathbf{r}')\phi_{k_j}(\mathbf{r}') + \phi_{k_j}^*(\mathbf{r})\phi_{k_j}(\mathbf{r})W(\mathbf{r}\mathbf{r}')\phi_{k_i}^*(\mathbf{r}')\phi_{k_i}(\mathbf{r}')\\ &\pm\phi_{k_i}^*(\mathbf{r})\phi_{k_j}(\mathbf{r})W(\mathbf{r}\mathbf{r}')\phi_{k_j}^*(\mathbf{r}')\phi_{k_i}(\mathbf{r}') \pm \phi_{k_j}^*(\mathbf{r})\phi_{k_i}(\mathbf{r})W(\mathbf{r}\mathbf{r}')\phi_{k_i}^*(\mathbf{r}')\phi_{k_j}(\mathbf{r}')\Big)\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}'.\end{aligned}$$

The first line of (2.10) contains the *direct* Coulomb interaction energy between particles, the second line is called the *indirect* interaction. If the particles are electrons (fermions) the indirect interaction is called *exchange energy.* You will note that the direct interaction contains terms like $|\phi_k(\mathbf{r})|^2\,W(\mathbf{r}\mathbf{r}')\,|\phi_l(\mathbf{r}')|^2$ which is the Coulomb interaction energy between classical charge densities associated with orbitals ϕ_l and ϕ_k. The second line contains interactions between what Slater called *exchange charge densities* such as $\phi_l^*(\mathbf{r})\phi_k(\mathbf{r})$ which have no analogue in classical physics. These terms are usually said to arise from the requirement that the wavefunctions be antisymmetric under exchange of coordinates (hence the word "exchange") as can be seen from the Heitler–London valence bond picture of the hydrogen molecule; but in second quantisation they arise as a consequence of the commutation relations and affect both fermions and bosons.

The *ansatz* for the wavefunction (2.11) in electron theory amounts to the Hartree–Fock approximation; you may well recognise the Hartree and exchange energies in the last expression from your study of Hartree–Fock theory. We shall look at it again in the next section.

3. Electron–electron interaction

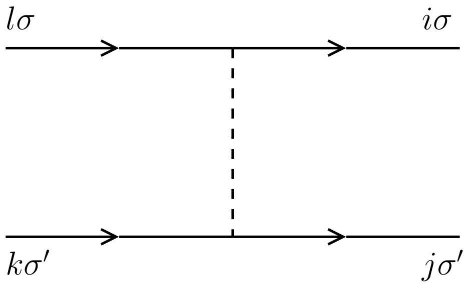
If we include electron–electron interaction then, as we have seen in (2.6) and (2.9), the Hamiltonian in second quantisation is

$$H = \sum_{i\sigma} \varepsilon_i a^\dagger_{i\sigma} a_{i\sigma} + \sum_{\substack{i\neq j\\ \sigma}} t_{ij} a^\dagger_{i\sigma} a_{j\sigma} + \frac{1}{2} \sum_{\substack{ijkl\\ \sigma\sigma'}} V_{ijkl} a^\dagger_{i\sigma} a^\dagger_{j\sigma'} a_{k\sigma'} a_{l\sigma}. \tag{3.1}$$

A spin index σ is included in labelling the operators; written in this way, the first two terms have the look of a tight-binding Hamiltonian, with on-site and hopping terms if the t_{ij} are restricted to nearest neighbour sites. The final term is the only "non-trivial" operator from the many-body point of view. By comparison with (2.9) after expanding the field operators in single particle eigenfunctions we see that

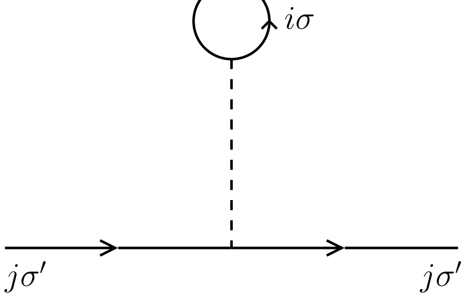
$$\begin{aligned} V_{ijkl} &= \int \phi^*_i(\mathbf{r})\phi_l(\mathbf{r}) W(\mathbf{r}\mathbf{r}')\phi^*_j(\mathbf{r}')\phi_k(\mathbf{r}')\,\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \\ &\equiv \langle ij\,|W|\,lk\rangle\,. \end{aligned} \tag{3.2}$$

Look carefully at the ordering of the indices in (3.1) and (3.2). The electron–electron interaction has a simple interpretation in field theory. A particle in state $l\sigma$ and a particle in state $k\sigma'$ interact and as a result of this interaction, the particles emerge in states $i\sigma$ and $j\sigma'$. This is illustrated in the sketch on the right. The broken line represents the interaction. Note that spin is conserved because the Hamiltonian is nonrelativistic and so does not describe spin flip. The job of the many-body theorist is to construct a perturbation theory which includes terms to first and higher orders in V, to sum up all the diagrams and to find the *self energy* of the electrons. Almost always you have to decide which interactions are important and which you can neglect. For example, in tight-binding we usually ignore all of them, or include just some inter-site electrostatic terms and some on-site Coulomb repulsions.**2** And of course this is only including terms to first order in V.



Certain interaction terms are of special importance in electron theory, namely the direct and exchange Coulomb integrals. If we set $l=i$ and $k=j$ in (3.2) we get

$$\begin{aligned} V_{ijji} &= \int \phi^*_i(\mathbf{r})\phi_i(\mathbf{r}) W(\mathbf{r}\mathbf{r}')\phi^*_j(\mathbf{r}')\phi_j(\mathbf{r}')\,\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \\ &\equiv U_{ij} \end{aligned} \tag{3.3}$$

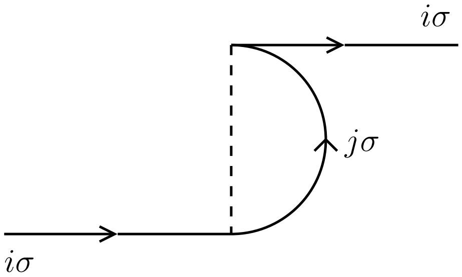


whose diagram is shown on the left. For the lower electron line, the state is the same before and after the interaction, the upper line curls back on itself into what's called a *fermion loop*. U_{ii} is the interaction energy associated with putting an electron of spin up, say, into an orbital where there is already an electron with spin down.

If conversely we set $k = i$ and $l = j$ in (3.2) you immediately see that we obtain the exchange integral

$$V_{ijij} = \int \phi_i^*(\mathbf{r})\phi_j(\mathbf{r})W(\mathbf{rr}')\phi_j^*(\mathbf{r}')\phi_i(\mathbf{r}')\,\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \\ \equiv J_{ij} \tag{3.4}$$

which is depicted on the right. This is sometimes called an "open oyster" diagram. You can see that this can only happen if $\sigma = \sigma'$; in fact it is well known that exchange energy only affects like spin electrons—in Hartree–Fock unlike spin electrons are said to be "uncorrelated," their interaction appears only in the Hartree energy. In density functional theory we talk of "exchange and correlation." In many-body theory, *correlation* is the contribution to the self energy obtained by summing all diagrams to all orders, except the first order direct interaction and the first order open oyster, exchange diagram. In the free electron gas, all higher order open oysters sum to zero.



Some people don't like density functional theory because you can't express the exchange and correlation energy as a sum over diagrams. It is said to have "the structure of a *mean field* theory," but then so does Hartree–Fock. Even the exchange-only energy in DFT is not the "exact exchange" described by the open oyster diagrams. This situation carries over in the local density approximation to DFT. What is missing in Hartree–Fock (*ie*, using exchange only) is that the electron–electron interaction is not screened. The bare Coulomb potential $W(\mathbf{rr}')$ is much too strong; at long distance it is very effectively screened by the electron gas; in fact the electron quasiparticles behave as neutral particles at long range as can be demonstrated using Gunnarsson's sum rule (except in special cases such as at surfaces). One could replace $W(\mathbf{rr}')$ by a screened potential using a fixed, or energy and frequency independent screening length, and this so called *static screened exchange* has been used as a simple extension to LDA to reproduce band gaps in semiconductors. If this screening is made energy and frequency *dependent* we get close to the GW approximation of Hedin, Gunnarsson, Ariasetyawan, Godby and others. The LDA on its own contributes some of these features to the exchange and correlation energy, since the LDA uses a form that is practically exact for the free electron gas. LDA is generally "better" than Hartree–Fock, if less rigorous, because using exact exchange but no correlation can make quite pathological errors, for example very large band gaps in insulators and vanishing density of states at the Fermi level in metals. By treating the direct and exchange integrals U and J as free parameters, one can hope to obtain a simple model, such as the Hamiltonian (3.1) which goes beyond Hartree–Fock by reducing the strength of the bare Coulomb potential.

LDA+U is a first principles attempt to do this. The prescription in outline, is to identify those orbitals, or those regions of space where the electrons are more highly correlated and for which the LDA is leading to large errors (such as making NiO a metal, or $\mathrm{LaCuO_3}$ a metal rather than an antiferromagnetic insulator). For these electrons the theory is to apply the electron–electron interaction from (3.1). First the energy that the LDA is expected to predict for this term is subtracted off, and it's replaced by direct and exchange terms only in (3.1) in a mean-field approximation. The parameters U and J are determined in a local spin density approximation (LSDA) calculation using a supercell to allow these integrals to be screened by the electron gas. In this way the theory is *ab initio* but it goes beyond Hartree–Fock by including screened Coulomb integrals in place of the bare ones. Section 6 will describe this in detail before going on to attempt a tight-binding version of the theory.

4. Number operators and Slater determinants

Often you encounter an ordered product of operators such as in (3.1) and you want its expectation value in some state Φ which for now we will take to be a "Fock" type eigenfunction (2.11)

$$\Phi = a^\dagger_{k_1} a^\dagger_{k_2} \cdots a^\dagger_{k_N} \Phi_0 \tag{2.11}$$

which is a Slater determinant if the particles are Fermions. Often the expectation value is abbreviated as

$$\left\langle a^\dagger_i a^\dagger_j a_k a_l \right\rangle = \left\langle \Phi \left| a^\dagger_i a^\dagger_j a_k a_l \right| \Phi \right\rangle . \tag{4.1}$$

A good way to look at these was shown me by Gleb Gribakin. Think of it as the inner product between the ket

$$a_k a_l \left| \Phi \right\rangle$$

and the bra

$$\left\langle \Phi \left| a^\dagger_i a^\dagger_j \right. .$$

In the ket, a_l destroys an electron in state l then a_k destroys an electron in state k. This always gives zero if $k = l$ because there can't be two fermions in the same state. The ket will also yield zero unless Φ represents a Fock wavefunction in which the states l and k are occupied in which case it yields a new wavefunction in which electrons in these states have been destroyed—let's call it Φ'. The bra behaves in the same way because the the creation operators act to the left and hence they become annihilation operators (by taking their Hermitian conjugates). By the orthogonality of the eigenfunctions, (4.1) is zero unless the bra also results in the same wavefunction Φ' as the ket, that is, the electrons in the same two states must have been destroyed. Hence (4.1) is zero unless either $l = i$ and $k = j$, or $l = j$ and $k = i$, in which case the result is one in the former case since the Slater determinant is normalised, but *minus one* in the latter because the commutation relation is $a_l a_k = -a_k a_l$. Hence we have, in this case,

$$\left\langle a^\dagger_i a^\dagger_j a_k a_l \right\rangle = \left(\delta_{li}\delta_{kj} - \delta_{lj}\delta_{ki} \right), \quad k \neq l, i \neq j. \tag{4.2a}$$

By a similar, and simpler, argument it is clear that

$$\langle a^\dagger_i a_l \rangle \equiv n_{il} = n_i \delta_{il} \tag{4.2b}$$

and n_i are the occupation numbers introduced earlier, which for Fermions are either zero or one. In the case that we are discussing in this section up to now, namely that Φ diagonalises the Hamiltonian, n_{il} appears as a diagonal matrix whose diagonal elements comprise N 'ones' in the first N rows, corresponding to the first N occupied states, all other elements being zero.

Let us now work out the expectation value of the last term in (3.1) in the case where the Coulomb integrals in (3.3) and (3.4) are constants,

$$\begin{aligned} V_{ijji} &= U_{ij} = U \\ V_{ijij} &= J_{ij} = J \end{aligned}$$

and the eigenfunction in (4.1) is a Slater determinant (2.11). These results will be needed to understand LDA+U.

Using (4.2a) we have

$$\begin{aligned} &\frac{1}{2}\sum_{\substack{ijkl\\ \sigma\sigma'}} V_{ijkl}\left\langle a^{\dagger}_{i\sigma}a^{\dagger}_{j\sigma'}a_{k\sigma'}a_{l\sigma}\right\rangle \\ &\qquad = \frac{1}{2}\sum_{\substack{ijkl\\ \sigma\sigma'\\ i\neq j\\ k\neq l}} V_{ijkl}\left(\delta_{i\sigma\, l\sigma}\delta_{j\sigma'\, k\sigma'} - \delta_{i\sigma\, k\sigma'}\delta_{j\sigma'\, l\sigma}\right) \\ &\qquad = \frac{1}{2}\sum_{\substack{ij\\ \sigma\sigma'\\ i\neq j}} V_{ijji} - \frac{1}{2}\sum_{\substack{ij\\ \sigma=+1\\ i\neq j}} V_{ijij} - \frac{1}{2}\sum_{\substack{ij\\ \sigma=-1\\ i\neq j}} V_{ijij} \\ &\qquad = \frac{1}{2}UN(N-1) - \frac{1}{2}JN^{\uparrow}(N^{\uparrow}-1) - \frac{1}{2}JN^{\downarrow}(N^{\downarrow}-1) \end{aligned} \tag{4.3}$$

Put another way, if the electron–electron interaction is the same for all electrons, independent of their quantum numbers, then the electron–electron interaction energy is $\frac{1}{2}UN(N-1)$ since there are N electrons and each one interacts with the remaining $N-1$ electrons. The half is to avoid counting each interaction twice. The same argument applies to the spin up and down electrons in the last two terms. Clearly, while U describes the Coulomb interaction between electrons, J is an interaction between like-spin electrons—it is the "exchange integral." If the total magnetic moment is zero, then $N^{\uparrow} = N^{\downarrow} = \frac{1}{2}N$ and the final two terms become $-\frac{1}{4}JN(N-2)$.

In the general case that the wavefunction is not a single Slater determinant (2.11) but is in fact a linear combination of these (see below at the end of this section), we can make what is called a generalised *mean field* approximation to arrive at a similar result to (4.2). The object is to reduce a two electron operator to a sum of one electron operators before taking the expectation value. Including spin indices we want an approximation to

$$a^{\dagger}_{i\sigma}a^{\dagger}_{j\sigma'}a_{k\sigma'}a_{l\sigma}$$

which is obtained as a sum of two terms: the first involves moving the fourth operator back two places to put it to the right of the first (the commutation relations require two changes of sign to do this) and the second involves moving the third operator back once place to put it next to the first (this changes the sign of the second term). In each term we then take the expectation value of the first two operators and leave the other alone to obtain

$$\begin{aligned} a^{\dagger}_{i\sigma}a^{\dagger}_{j\sigma'}a_{k\sigma'}a_{l\sigma} &\approx \left\langle a^{\dagger}_{i\sigma}a_{l\sigma}\right\rangle a^{\dagger}_{j\sigma'}a_{k\sigma'} - \left\langle a^{\dagger}_{i\sigma}a_{k\sigma'}\right\rangle a^{\dagger}_{j\sigma'}a_{l\sigma} \\ &= n^{\sigma\sigma}_{il}\, a^{\dagger}_{j\sigma'}a_{k\sigma'} - n^{\sigma\sigma'}_{ik}\, a^{\dagger}_{j\sigma'}a_{l\sigma}. \end{aligned}$$

The approximation amounts to expressing the two electron operator on the left in terms of a one electron operator acting alone but in the averaged field of the other; hence *mean field approximation.* The result is that the expectation value of the two electron operator becomes

$$\begin{aligned}\left\langle a^\dagger_{i\sigma}a^\dagger_{j\sigma'}a_{k\sigma'}a_{l\sigma}\right\rangle &\approx \left\langle a^\dagger_{i\sigma}a_{l\sigma}\right\rangle\left\langle a^\dagger_{j\sigma'}a_{k\sigma'}\right\rangle - \left\langle a^\dagger_{i\sigma}a_{k\sigma'}\right\rangle\left\langle a^\dagger_{j\sigma'}a_{l\sigma}\right\rangle \\ &= n^{\sigma\sigma}_{il}n^{\sigma'\sigma'}_{jk} - n^{\sigma\sigma'}_{ik}n^{\sigma'\sigma}_{jl}.\end{aligned} \tag{4.4}$$

If we apply (4.2b) and only permit the occupation numbers to be either one or zero then (4.4) reduces to (4.2a). However in the general case that the wavefunction is not a single Slater determinant, and the single particle states, $\phi_i(\mathbf{r})$ do not diagonalise H_0 as in (2.6), then the $n^{\sigma\sigma'}_{ij}$ are *generalised* occupation numbers, in fact density matrix elements with non zero off-diagonal elements. We come back to these in section 6.

Let us continue the theme that the many-body wavefunction of our system is not of the form (2.11). It is very useful to make a digression here, to examine a simple problem in which the solution is quite evidently not a single Slater determinant; that is, in which the wavefunction is correlated. You have doubtless studied the molecular orbital and valence bond pictures of the hydrogen molecule. To study H_2, we begin with atomic orbitals which are the *s*-wavefunctions of the hydrogen atoms, labelled a and b which form the molecule:

$$\begin{aligned} a(1) \equiv \phi_a(\mathbf{r}_1) &= \frac{1}{\sqrt{\pi}}\mathrm{e}^{-\alpha r_a} \\ b(2) \equiv \phi_b(\mathbf{r}_2) &= \frac{1}{\sqrt{\pi}}\mathrm{e}^{-\alpha r_b}\end{aligned} \tag{4.5}$$

in atomic units. $\mathbf{r}_1$ and $\mathbf{r}_2$ are the coordinates of electrons 1 and 2; r_a is the distance of the electron from the nucleus labelled a, and similarly r_b; α is a variational parameter included for completeness, take it to be one. Here $a(1)$ is the wavefunction of electron 1 localised at hydrogen nucleus a and $b(2)$ is the wavefunction of electron 2 localised at hydrogen nucleus b. Of course, when the molecule is formed from its atoms, electron 1 may become associated with hydrogen nucleus b and *vice versa* in which case we will need $a(2)$ and $b(1)$ defined in the same way. In the *molecular orbital* picture of Hund and Mulliken, we begin by constructing the bonding molecular orbital of the H_2^+ molecular ion,

$$u_g(1) = \left[\frac{2}{\alpha^3}(1+S)\right]^{-\frac{1}{2}}[a(1)+b(1)] \tag{4.6}$$

with the overlap integral

$$\begin{aligned} S &= \frac{\alpha^3}{\pi}\int a(1)\,b(1)\,\mathrm{d}\mathbf{r} \\ &= \mathrm{e}^{-\alpha R}\left(1+\alpha R+\frac{1}{3}\alpha^2R^2\right).\end{aligned}$$

R is the distance between the hydrogen nuclei. Note that u_g is a *bonding* MO; the *antibonding* MO, u_u is made by replacing the two ‘plus’ signs by ‘minus’ signs in (4.6). As you know, to make the molecular orbital picture for H_2 we define spin functions, or *spinors*, as follows. The electron spin s is either $+1$ or -1 (in units of $\frac{1}{2}\hbar$) and the function $\alpha(i) \equiv \alpha(s_i)$ is defined to be one if electron i has spin $+\frac{1}{2}\hbar$ and zero if it has spin $-\frac{1}{2}\hbar$; $\beta(i)$ is zero if electron i has spin $-\frac{1}{2}\hbar$ and zero otherwise:

$$\alpha(1) = 1 \qquad \alpha(-1) = 0 \qquad \beta(1) = 0 \qquad \beta(-1) = 1.$$

We could construct the *product* or Hartree wavefunction

$$u_g(1)\,\alpha(1)\,u_g(2)\,\beta(2) \tag{4.7}$$

by putting electrons 1 and 2 into a bonding orbital, having opposite spins, but this wavefunction does not have the fermion antisymmetry property. A better wavefunction is the Slater determinant,

$$\frac{1}{\sqrt{2}}\begin{vmatrix} u_g(1)\,\alpha(1) & u_g(2)\,\alpha(2) \\ u_g(1)\,\beta(1) & u_g(2)\,\beta(2) \end{vmatrix} = u_g(1)\,u_g(2)\,\Xi_1, \tag{4.8}$$

where

$$\Xi_1 = \frac{1}{\sqrt{2}}\left[\alpha(1)\,\beta(2) - \beta(1)\,\alpha(2)\right]$$

is the *singlet* spin function; that is, the total spin is zero. The point is that the MO picture amounts to the Hartree–Fock approximation since the wavefunction is a single Slater determinant. Instead of using bare atomic orbitals you could vary the parameter α in (4.5) until you get the lowest energy: this is what Slater called a *self consistent field* (SCF). Quantum chemists make the distinction between an SCF and the *Hartree-Fock limit* which is the lowest energy that can be obtained using a single Slater determinant. At all events this wavefunction is *uncorrelated*, that's what Hartree–Fock means.

In contrast, the well known Heitler–London or *valence bond* wavefunction for the H_2 singlet ground state is

$$[\,a(1)\,b(2) + a(2)\,b(1)\,]\,\Xi_1, \tag{4.9}$$

and the first excited, triplet, states are

$$[\,(a(1)\,b(2) + a(2)\,b(1)\,] \times \begin{cases} \alpha(1)\,\alpha(2), \\ \alpha(1)\,\beta(2) + \beta(1)\,\alpha(2), \\ \beta(1)\,\beta(2). \end{cases}$$

The physical picture is that if electron 1 is associated largely with nucleus a, then electron 2 will be found near nucleus b, and *vice versa.* So instead of the electrons sharing a bonding MO, they are keeping well apart *in a correlated state.* Mathematically, this is reflected in the fact that (4.9) cannot be written as a single determinant. Indeed, consider the four determinants

$$\frac{1}{\sqrt{2}}\begin{vmatrix} a(1)\,\chi_a(1) & b(1)\,\chi_b(1) \\ a(2)\,\chi_a(2) & b(2)\,\chi_b(2) \end{vmatrix}$$

arising when (*i*) χ_a and χ_b are both α; (*ii*) both β; (*iii*) α, β and (*iv*) β, α respectively. You will find that (4.9) is the sum of determinants *iii* and *iv*. The triplet states which are degenerate in the absence of a magnetic field are, respectively: determinant *i*; the difference of *iii* and *iv*; and *ii*.

So, the valence bond is an "ionic," correlated wavefunction having the right properties upon dissociation that each nucleus ends up with one electron of a different spin. This is always the best starting point for studying a correlated system, and the wavefunction cannot be written as a single Slater determinant and equation (4.2) does not apply.

The molecular orbital $u_g(1)\,u_g(2)$ can be expanded using (4.6) to give

$$\frac{1}{2(1+S)}\left[a(1)\,b(2) + b(1)\,a(2) + a(1)\,a(2) + b(1)\,b(2)\right]$$

so the electron spends equal amounts of time in the "covalent" state as in the "ionic" state and there's no guarantee that on dissociation each nucleus will get one electron: there's an equal probability that both electrons will end up on the same nucleus. One way to "fix up" the MO picture is to include some amount of the antibonding MO determinant $u_u(1)\,u_u(2)$ in a variational calculation. This is called *configuration interaction.*

It is interesting that if the Hartree approximation is used in the SCF, in other words the wavefunction is (4.7) and the energy is compared to that obtained with the Hartree–Fock SCF using (4.8), the energy in the Hartree–Fock approximation is *lower*, as you might expect at first sight. On the other hand, since the two electrons are of opposite spin, there is no exchange energy, so why isn't the energy unchanged? The answer is that there is an energy lowering due to electrostatic interaction of the exchange density with the nuclear charges; Slater called these *exchange forces.*

After this long digression, let's go back to ask what if the the wavefunction is not a single Slater determinant. We hope never to have to write down what it is, but have in our minds that it may be the sum or difference of two or more Slater determinants as is the valence bond wavefunction; or some general linear combination. In configuration interaction, to the SCF function $\Phi_{\rm SCF}$ is added excited state wavefunctions constructed by destroying one or more electrons in occupied levels and creating an equal number in unoccupied levels, for example if i, j refer to levels unoccupied in the ground state and μ, ν refer to occupied levels then

$$a_i^\dagger\,a_\mu\,\Phi_{\rm SCF}$$

is called a single substitution, and

$$a_i^\dagger\,a_j^\dagger\,a_\mu\,a_\nu\,\Phi_{\rm SCF}$$

is called a double substitution.

5. Janak's theorem and the discontinuity in the potential

You may not know about Janak's theorem.[3] He considered the question of a *non integral* number of electrons occupying a particular state. This can be done if you imagine an open system (*ie*, one that does not posess a fixed number of electrons); and then consider this system as belonging to a grand canonical ensemble.

Let us recall that in density functional theory the charge density is obtained from the Kohn–Sham orbitals from the summation

$$\rho(\mathbf{r}) = \sum_i^N |\phi_i(\mathbf{r})|^2 .$$

Then according to Hohenberg–Kohn and Kohn–Sham, the total ground state energy is

$$\begin{aligned} E[\rho] &= T_s[\rho] + U[\rho] + E_{\rm xc}[\rho] \\ &= \sum_i^N \int \phi_i^* \left(-\nabla^2 + V_{\rm es} + \varepsilon_{\rm xc}\right) \phi_i \end{aligned}$$

in Rydberg atomic units, with the usual symbols and omitting d$\mathbf{r}$ under the integral sign, as usual. You recall, then, that minimising the energy functional with respect to each of the ϕ_i^* subject to their normalisation recovers a Kohn–Sham Schrödinger equation for each orbital, and exposes the one-electron eigenvalues as Lagrange multipliers:

$$\begin{aligned} &\frac{\delta}{\delta\phi_i^*}\left\{E - \varepsilon_i\left(\int \phi_i^*\phi_i - 1\right)\right\} = 0 \\ &\quad\Rightarrow \left(-\nabla^2 + V_{\rm es} + \mu_{\rm xc}\right)\phi_i = \varepsilon_i\phi_i . \end{aligned}$$

On the other hand minimising with respect to ρ, subject to the number of electrons being N, results in the Euler–Lagrange equation,

$$\begin{aligned} &\frac{\delta}{\delta\rho}\left\{E - \mu\left(\int\rho - N\right)\right\} = 0 \\ &\Rightarrow \frac{\delta E}{\delta\rho} = \mu = V_{\rm es} + \mu_{\rm xc} + \frac{\delta T_s}{\delta\rho}. \end{aligned} \tag{5.1}$$

Under a perturbation δV in the external potential, the first order change in the total energy is

$$\delta E = \int \rho(\mathbf{r})\delta V. \tag{5.2}$$

Now, $E = E[\rho, V]$ so

$$\begin{aligned} {\rm d}E &= \int \left(\frac{\delta E}{\delta\rho}\right)_V {\rm d}\rho + \int \left(\frac{\delta E}{\delta V}\right)_\rho {\rm d}V \\ &= \mu\,{\rm d}N + \int \rho(\mathbf{r}){\rm d}V \end{aligned}$$

using (5.1) for the first, and (5.2) for the second term in the second line. This identifies the Lagrange multiplier in (5.1) as the chemical potential of an open system of electrons in an external potential (usually, of course the potential due to the nuclei),

$$\mu = \left(\frac{\partial E}{\partial N}\right)_V = -\chi, \tag{5.3}$$

χ is called the *Mulliken electronegativity.* If our system is an atom *in vacuo* then the total energy as a function of the number of electrons is shown as the filled dots in fig. 5.1 for the neutral and positively and negatively charged ions. By definition, $E(N-1) - E(N)$ is the ionisation energy I, and $E(N) - E(N+1)$ is the electron affinity A. These are also minus the values of the slopes of the two straight lines drawn between the filled dots. These slopes, by (5.3) are also the chemical potential μ which clearly has a *discontinuity* at the atomic number Z:

$$\mu = \begin{cases} -I, & Z-1 < N < Z; \\ -A, & Z < N < Z+1. \end{cases}$$

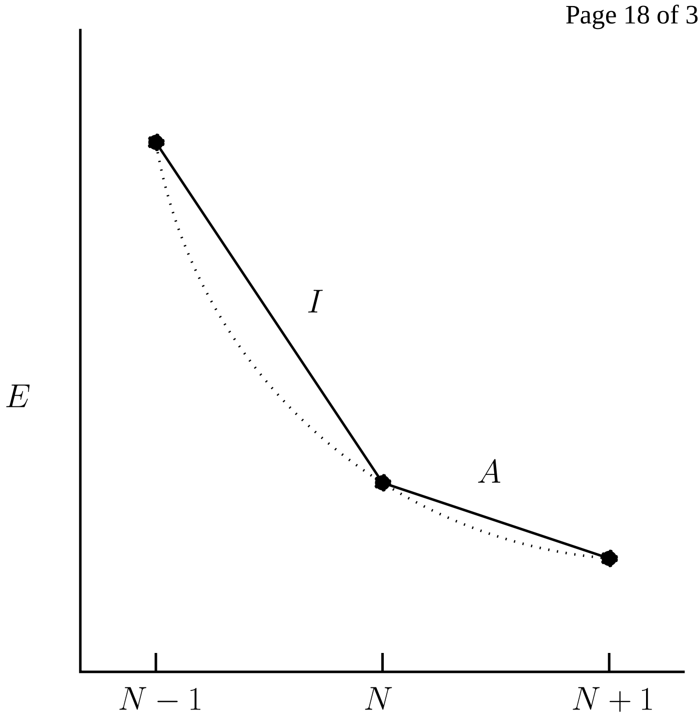


Fig. 5.1

You may think of a quantum system with fractional occupation $N \pm x$ $(x < 1)$ as in a time average of the pure states with N and $N \pm 1$ electrons having a total energy

$$E = (1-x)E(N) + xE(N \pm 1)$$

and this is the situation indicated by the solid lines joining the filled dots if we now regard fig. 5.1 as representing any open quantum system, not particularly an atom. The dotted line is the energy as a function of N that would be predicted by local density functional theory: there is no discontinuity in the chemical potential. In a semiconductor, the band gap should be the energy change in adding an electron take away the energy change in removing an electron (analogous to $-A+I$) hence the gap is

$$E_g = E(N+1) + E(N-1) - 2E(N).$$

Often the failure to predict the correct band gap is attributed to a failure of density functional theory itself (not the local density approximation). It is suspected that the discontinuity in (5.1) must be in the exchange–correlation potential, since it is certainly not in the electrostatic potential which must be smooth as a function of N. One of the aims behind LDA+U was to reintroduce a discontinuity into the potential due to the electron–electron interaction.

Janak's theorem is obtained by rederiving (5.1) in an open quantum system with a variable occupancy of electronic states. We then write

$$\rho(\mathbf{r}) = \sum_i n_i \left|\phi_i(\mathbf{r})\right|^2 \tag{5.4}$$

where n_i are occupation numbers. Then there is an energy functional that depends on ρ and the $\{n_i\}$:

$$\begin{aligned} \tilde{E}[\rho, \{n_i\}] &= \tilde{T} + U[\rho] + E_{\text{xc}}[\rho] \\ &= \sum_i n_i \int \phi_i^* \left(-\nabla^2 + V_{\text{es}} + \mu_{\text{xc}}\right) \phi_i \end{aligned}$$

which upon minimisation with respect to ϕ_i^* yields the Schrödinger equation

$$\begin{aligned}&\frac{\delta}{\delta\phi_i^*}\left\{\tilde{E}-\varepsilon_i'\left(\int\phi_i^*\phi_i-1\right)\right\}=0\\&\Rightarrow n_i\left(-\nabla^2+V_{\rm es}+\mu_{\rm xc}\right)\phi_i=\varepsilon_i'\phi_i.\end{aligned}$$

As long as $n_i \neq 0$ then this is the Kohn–Sham equation with $\varepsilon_i = \varepsilon_i'/n_i$. Janak points out that $\tilde{E} = E$ when the $\{n_i\}$ are those of the Fermi–Dirac distribution. He also shows that

$$\begin{aligned}\frac{\partial\tilde{E}}{\partial n_i}&=\left(\frac{\partial\tilde{E}}{\partial n_i}\right)_\rho+\int\frac{\delta\tilde{E}}{\delta\rho}\frac{\partial\rho}{\partial n_i}\\&=\varepsilon_i.\end{aligned}$$

The proof is left to the reader, and can also be found in Parr and Yang, page 165.[4] The point is that the $\{n_i\}$ may simply be regarded as just some variables, but still $\tilde{E}$ is a functional that is stationary with respect to variations of the density (5.4). Furthermore if the $\{n_i\}$ are occupation numbers of Kohn–Sham orbitals in an open quantum system that are taken from the Fermi–Dirac distribution, then $\tilde{E} = E$ and

$$\frac{\partial E}{\partial n_i}=\varepsilon_i \tag{5.5}$$

which is Janak's theorem and gives a clear physical meaning to the eigenvalues of the Kohn–Sham orbitals, namely they are the change in total energy upon adding an infinitessimal amount of charge into that orbital.

Janak's theorem is the basis of Slater's *transition state formula.* For example, the ionisation energy is

$$\begin{aligned}E(N-1)-E(N)&=\int_1^0\frac{\partial E}{\partial n_i}\,{\rm d}n_i\\&=\int_1^0\varepsilon_i(n_i)\,{\rm d}n_i\\&\approx-\varepsilon_i\left(\frac{1}{2}\right),\end{aligned}\tag{5.6}$$

where, here, ε_i is the eigenvalue of the highest occupied state. The first line is just a formal identity; the second line is Janak's theorem, it says get the ionisation energy in a thought experiment by slowly removing an electron from the highest occupied state measuring the change in eigenvalue as the charge is removed. The third line is Slater's approximation obtained by supposing the eigenvalue ε_i varies linearly with occupation number in the Taylor series

$$\begin{aligned}\varepsilon(n)&=\varepsilon(0)+nU\\&=\varepsilon(0)+n\frac{\partial^2E}{\partial n^2}.\end{aligned}$$

Slater's approximation says you don't have to do that thought experiment: just calculate the eigenvalue of the system when that state is *half occupied.* Strange, but true. Later we will see the connection between U, the so called chemical hardness of the system (its willingness in chemical reactions to give up an electron) and the direct Coulomb interaction in the Hubbard model (3.1). We will also see how U can provide the discontinuity in the potential as the occupation is changed.

6. LDA+U

After all this preamble we are ready to understand LDA+U. The easiest paper to read about this is Anisimov *et al.*[7] which I will follow here. A comprehensive review is Anisimov *et al.*[8]. Following equations (2.3) and (2.9) and including a specific spin index, $\sigma = \pm 1$ the Hamiltonian is

$$
\begin{aligned}
H &= \sum_{\sigma} \int \psi_\sigma^\dagger(\mathbf{r}) \left(-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r}) \right) \psi_\sigma(\mathbf{r})\, \mathrm{d}\mathbf{r} && (6.1a)\\
&+ \frac{1}{2}\sum_{\sigma\sigma'} \int \psi_\sigma^\dagger(\mathbf{r})\psi_{\sigma'}^\dagger(\mathbf{r}')\, W(\mathbf{r}\mathbf{r}')\, \psi_{\sigma'}(\mathbf{r}')\psi_\sigma(\mathbf{r})\, \mathrm{d}\mathbf{r}\, \mathrm{d}\mathbf{r}' && (6.1b)\\
&= H_0 + H' && (6.1c)
\end{aligned}
$$

in which the field operators may be expanded in single particle orbitals,

$$
\psi_\sigma^\dagger(\mathbf{r}) = \sum_i a_{i\sigma}^\dagger \phi_i^*(\mathbf{r}), \tag{6.2}
$$

in terms of which the electron–electron interaction is

$$
H' = \frac{1}{2} \sum_{\substack{ijkl\\ \sigma\sigma'}} V_{ijkl} a_{i\sigma}^\dagger a_{j\sigma'}^\dagger a_{k\sigma'} a_{l\sigma}
$$

and V_{ijkl} is given by equation (3.2):

$$
\begin{aligned}
V_{ijkl} &= \int \phi_i^*(\mathbf{r})\phi_l(\mathbf{r}) W(\mathbf{r}\mathbf{r}')\phi_j^*(\mathbf{r}')\phi_k(\mathbf{r}')\, \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}'\\
&\equiv \langle ij\, |W|\, lk\rangle\,. && (3.2)
\end{aligned}
$$

Now DFT reduces the ground state of the many-body problem to an *independent electron* scheme in which non interacting "electrons" occupy Kohn–Sham orbitals which are eigenfunctions of H_0 in which $V(\mathbf{r})$ in (6.1a) is the Kohn–Sham *effective potential*,

$$
V_{\text{eff}}(\mathbf{r}) = V_{\text{es}}(\mathbf{r}) + \mu_{\text{xc}}(\mathbf{r}).
$$

All the electron–electron interactions are included in the exchange and correlation potential

$$
\mu_{\text{xc}}(\mathbf{r}) = \frac{\delta E_{\text{xc}}}{\delta \rho(\mathbf{r})}
$$

in a *mean field* way. The philosophy behind LDA+U is to describe the total energy as the Hohenberg–Kohn energy associated with the Hamiltonian (6.1a) *taking away* the electron–electron interaction energy as the LDA has described it and then *replacing it* with the expectation value of (6.1b) obtained in an approximate way. The assumptions are these,

1. Not all the electrons are correlated. These are separated out from the others, identified either by the orbitals they occupy (*e.g.*, the d or f electrons) or by the region of space they occupy (or both).

2. The LDA has described their electron–electron interaction as a mean field, independent of the quantum numbers of the electrons and hence as in equation (4.3) which must be subtracted from the LDA total energy.
3. When replacing this with an expectation value of (6.1b) it is assumed that strong correlations exist *only between electrons whose orbitals are on the same atomic site.* Furthermore it is assumed that these electrons all belong to one angular momentum manifold, *ie* they are all either d or f electrons.

Now LDA+U is itself a generalised mean field theory, hence at all events we will have for the expectation value of H', as in (4.4)

$$\frac{1}{2}\sum_{\substack{ijkl\\ \sigma\sigma'}} V_{ijkl}\left\langle a^\dagger_{i\sigma}a^\dagger_{j\sigma'}a_{k\sigma'}a_{l\sigma}\right\rangle \approx \frac{1}{2}\sum_{\substack{ijkl\\ \sigma\sigma'}} V_{ijkl}\left(\left\langle a^\dagger_{i\sigma}a_{l\sigma}\right\rangle\left\langle a^\dagger_{j\sigma'}a_{k\sigma'}\right\rangle - \left\langle a^\dagger_{i\sigma}a_{k\sigma'}\right\rangle\left\langle a^\dagger_{j\sigma'}a_{l\sigma}\right\rangle\right). \tag{6.3}$$

The very simplest LDA+U scheme is very instructive. Let us suppose $\langle a^\dagger_i a_l\rangle = n_{il} = n_i\delta_{il}$ and let us furthermore neglect exchange and spin-dependence. Then (6.3) becomes simply

$$\frac{1}{2}\sum_{m\neq m'} U\, n_m\, n_{m'}.$$

If this is, say, a d electron metal then we are treating the d orbitals as degenerate having magnetic quantum numbers $m = 1 - 5$ and spin degeneracy is assumed. Furthermore we regard the d band of the atom as an open system interacting with the sp bands, so that the number of d electrons is a continuous variable n_m $(i = 0 - 1)$—the same for all m in this simple degenerate example, up to a total of 10 d electrons.

Now we write down the total energy as

$$\begin{aligned} E &= E_{\text{LDA}} - \frac{1}{2}UN(N-1) + \frac{1}{2}\sum_{i\neq j} U\, n_i\, n_j \\ &= E_{\text{LDA}} + \frac{1}{2}U\sum_i n_i(1-n_i) \end{aligned} \tag{6.4}$$

employing assumptions 2 and 3 above. The second line follows rather easily after using

$$N = \sum_i n_i$$

and completing the sum in the third term. It serves to expose the point that the n_i are not Fermion occupation numbers equal only to zero or one: in that case this second term in the second line would vanish taking us back to the LDA!

According to Janak's theorem (5.5), the orbital energies are

$$\varepsilon_i = \frac{\partial E}{\partial n_i} = \varepsilon_{\text{LDA}} + U\left(\frac{1}{2} - n_i\right). \tag{6.5}$$

If we take the functional derivative of the total energy with respect to orbital occupancy we get the potential seen by that electron,

$$V_i(\mathbf{r}) = \frac{\delta E}{\delta n_i} = V_{\text{LDA}}(\mathbf{r}) + U\left(\frac{1}{2} - n_i\right). \tag{6.6}$$

This is an *orbital dependent* potential since it depends through the second term on the occupancy. Recalling the discussion of open quantum systems in section 5, we can see that equations (6.5) and (6.6) provide the discontinuity in the eigenvalue and the potential that is missing in the LDA. The simplest way to see this is by looking at eigenvalues of the hydrogen atom *in vacuo*.

To begin with, table 1 shows the total energy and eigenvalues of H and H$^-$ in the LDA (and in the generalised gradient approximation, GGA, of Perdew, Burke and Ernzerhof).

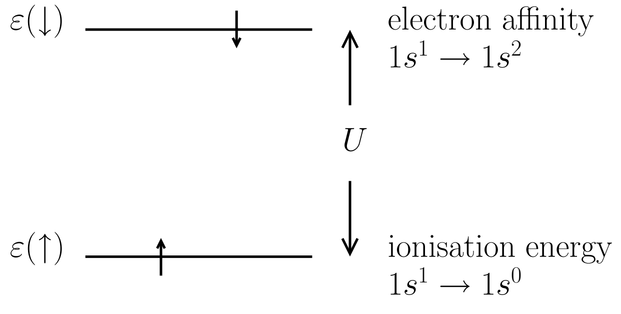
TABLE 1: Total energy and eigenvalues of H and H$^-$ in the local density approximation (see figure 5.1). All energies are in Rydberg. C–A and vB–H refer to the LDA recipes of Ceperley and Alder, and von-Barth and Hedin. PBE refers to the GGA of Perdew et al.

	H			H$^-$		
	C–A	vB–H	PBE	C–A	vB–H	PBE
$\varepsilon(\uparrow)$	–0.5379	–0.5569	–0.5586	–0.0780	–0.1939	–0.3391
$\varepsilon(\downarrow)$	–0.2562	–0.1886	–0.1351	–0.0780	–0.1939	–0.3391
E_{tot}	–0.9544	–0.9723	–0.9952	–1.0810	–1.1484	–1.2683
I	0.9544	0.9723	0.9952			
A				0.1266	0.1760	0.2731

The first thing to note is that the total energy of the neutral H–atom, which is by definition the ionisation potential I, is rendered fairly accurately in the LDA; in reality it is exactly 1 Ry. However the total energy of the H$^-$–atom is very much too low; that is, the electron affinity A which is the difference in the two total energies should be just 0.055 Ry from observations. Another big error is in the eigenvalue of the spin up electron (which is the occupied state in H here) which in the LDA is about –0.54 Ry, but should actually be exactly –1 Ry. We see here a typical feature of the LDA: total energies are rather accurately rendered, whereas orbital energies are not. Let us now see what happens if we apply the formula (6.5) to the H–atom. We need a value for U, which by definition is $I - A$. We could use the LDA value, $U_{\text{LDA}} \approx 0.8$ *Ry* or the observed value, $U = 0.945$. We'll use the latter. Then in the H–atom, we have $n_\uparrow = 1$, $n_\downarrow = 0$ and (6.5) gives,

$$\varepsilon(\uparrow) = -0.54 - \frac{1}{2}U = -1.0125$$

$$\varepsilon(\downarrow) = -0.54 + \frac{1}{2}U = -0.0675$$



Fig, 6.1

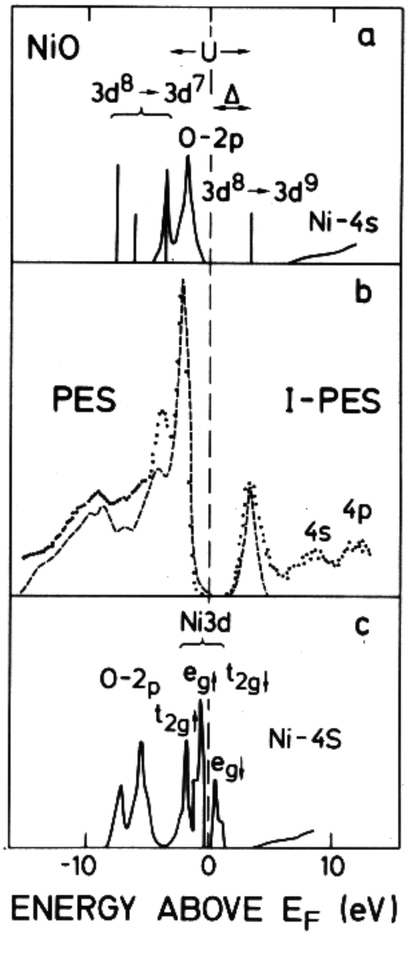
which is illustrated at the right in figure 6.1. The first point is that the LDA+U eigenvalues are behaving properly as removal and addition energies and so can be identified directly with I and A; this cannot be said of the LDA eigenvalues. The second point is that the *potential* in equation (6.6) behaves similarly. That means there is a *discontinuity* of an amount U in the potential as an electron is added to the H–atom. A third, rather subtle, point is that the eigenvalues in fig. 6.1

are split by the exchange interaction because the electrons have opposite spins (this is ultimately the origin of magnetism); however the exchange splitting is given here by the Coulomb integral U, not the exchange integral J. This seems paradoxical but in fact it is a consequence of there being just one orbital. In the same way we will see below that in a *single band* the spins are split by U and not J. A final interesting point is this. According to Slater's transition state formula the ionisation energy is predicted by equation (5.6), $I = -\varepsilon_i(\frac{1}{2})$, but we can write using Taylor's theorem

$$\begin{aligned} I = -\varepsilon_i(\tfrac{1}{2}) &= \varepsilon_i(1) - \frac{1}{2}\frac{\partial \varepsilon}{\partial N} \\ &= \varepsilon_i(1) - \frac{1}{2}U \end{aligned}$$

using Janak's theorem again, which is consistent with (6.5) as we have just seen.

Let me just emphasise again that although the LDA total energy of the H–atom is close to the true ionisation energy (especially when using the GGA) the LDA eigenvalues in table 1 cannot be interpreted as experimental observables. In spectroscopy one measures the energy to add or subtract an electron and hence, among other things, the band gap. Table 1 shows that the "band gap" in hydrogen is about 0.3 Ry in the LDA (0.4 Ry in the GGA) which is as usual much less than the 1 Ry observed. This is corrected by LDA+U which properly includes the discontinuity in the exchange and correlation potential. It has been a matter of debate whether the neglect of this discontinuity is a failure of density functional theory itself or only of the local density approximation. This is because it is extremely difficult to do an "exact" density functional calculation. However Gunnarsson and Schönhammer**[5]** constructed a linear chain model for which an "exact" DFT solution could be obtained and this *did not* produce the expected energy gap.



A much more complicated example than the H–atom is the highly correlated oxide NiO reproduced in figure 6.2 (right) from Zaanen *et al.***[6]** This shows in panel **c** the lamentable LSDA density of states with the almost vanishing band gap of 0.3 eV separating different manifolds of d-states. The middle panel shows experimental spectra revealing a 4 eV band gap which in fact separates largely oxygen p-states from unoccupied d-states. The upper panel compares with our figure 6.1 showing addition and removal energies calculated using a model of localised d-electrons hybridising with a ligand band of oxygen p-orbitals. An extension of this using a many-body Hamiltonian (the Anderson impurity model) results in a theoretical prediction of the spectra very close to the observed as seen in the middle panel. Recent LDA+U calculations also agree with the experimental picture.

Let us now go back to equation (6.3). We have seen that the simplest LDA+U neglecting exchange, spin and orbital non sphericity is very instructive; now we have to construct a complete LDA+U energy functional. We will maintain, however, assumption 3, that the N strongly interacting electrons are all on one site and have one angular momentum. Hence we replace the indices $i - l$ with $m - m'''$. Using the generalised mean field approximation (6.3) the interaction energy or expectation value of H' is

$$\begin{aligned}
E^U &= \frac{1}{2}\sum_{\substack{\sigma\sigma'\\ \{m\}}} V_{mm'm''m'''}\left(n^{\sigma\sigma}_{mm'''}n^{\sigma'\sigma'}_{m'm''} - n^{\sigma\sigma'}_{mm''}n^{\sigma'\sigma}_{m'm'''}\right)\\
&= \frac{1}{2}\sum_{\substack{\sigma\sigma'\\ \{m\}}} V_{mm'm''m'''}n^{\sigma}_{mm'''}n^{\sigma'}_{m'm''} - \frac{1}{2}\sum_{\substack{\sigma\\ \{m\}}} V_{mm'm''m'''}n^{\sigma}_{mm''}n^{\sigma}_{m'm'''}\\
&= \frac{1}{2}\sum_{\substack{\sigma\\ \{m\}}} V_{mm'm''m'''}n^{\sigma}_{mm'''}n^{-\sigma}_{m'm''} - \frac{1}{2}\sum_{\substack{\sigma\\ \{m\}}} V_{mm'm''m'''}\left(n^{\sigma}_{mm'''}n^{\sigma}_{m'm''} - n^{\sigma}_{mm''}n^{\sigma}_{m'm'''}\right)\\
&= \frac{1}{2}\sum_{\substack{\sigma\\ \{m\}}} \left(V_{mm'm''m'''}n^{\sigma}_{mm'''}n^{-\sigma}_{m'm''} + \left(V_{mm'm''m'''} - V_{mm'm'''m''}\right)n^{\sigma}_{mm'''}n^{\sigma}_{m'm''}\right). \qquad (6.7)
\end{aligned}$$

This is the most general expression for the energy of interacting electrons in LDA+U and it is intended to replace the estimated LDA expression (4.3) in the LDA+U functional as was done above for the simplest LDA+U functional (6.4). We do not assume that the number operators are diagonal because the single particle orbitals in (6.2) do not necessarily diagonalise H_0 in (6.1c). However we do require them to be diagonal in the spin since the Hamiltonian does not contain any terms that can flip the spin. The expectation values of number operators can be obtained in a straighforward way in a bandstructure context such as tight-binding or LMTO from the eigenvectors which may be used to form matrix elements of the Green function, †

$$G^{\sigma}_{\mathbf{R}\ell m,\mathbf{R}'\ell' m'}(z) = \left\langle \mathbf{R}\ell m\sigma \left| (z - H)^{-1} \right| \mathbf{R}'\ell' m'\sigma \right\rangle, \qquad (6.8)$$

where the matrix elements are labelled by site $\mathbf{R}$, and quantum numbers ℓ, m and σ. We then can construct matrix elements of the number operator,

$$n^{\sigma}_{mm'} = -\pi^{-1}\Im\int^{E_{\text{Fermi}}} G^{\sigma}_{\mathbf{R}\ell m,\mathbf{R}\ell m'}(E)\,\mathrm{d}E. \qquad (6.9)$$

In the same way as we obtained (6.4) in the case that exchange and spin were neglected we can write the total energy as

$$\begin{aligned}
E &= E_{\text{LDA}} - E_{\text{dc}} + E^U \qquad (6.10)\\
&= E_{\text{LDA}} - \frac{1}{2}UN(N-1) + \frac{1}{4}JN(N-2) + E^U
\end{aligned}$$

where the “double counting” term is found using (4.3). The effective potential of the Kohn–Sham equations will now contain a term in addition to the LDA potential which is the derivative of the second two terms in (6.10) and these provide the discontinuity that we have discussed earlier. The new term in the potential is also *orbital dependent* as we shall see now.

---

† This is the Green function in an orthogonal basis. One can also be defined in terms of an overlap matrix, and the number operators so defined can be used exactly as follows as long as their trace is equal to the number of electrons.

The potential is found by differentiating E with respect to one of the elements of the number operator. Firstly, then, we need the derivative of the terms in (6.7). After a little thought, remembering that a partial derivative regards all other variables as constants, you can convince yourself of the following result. The sum in (6.7) is over all values of the quantum numbers m, m', m'', m''' and we ask for a derivative with respect to some $n^{\sigma'}_{m_1m_2}$ (I use m_1 and m_2 simply because I don't want any more than three primes on a symbol). We get

$$\begin{aligned}
\frac{\partial}{\partial n^{\sigma'}_{m_1m_2}} \frac{1}{2} \sum_{\substack{\sigma \\ mm'm''m'''}} & V_{mm'm''m'''}\, n^{\sigma}_{mm'''}\, n^{-\sigma}_{m'm''} \\
&= \frac{1}{2} \sum_{\substack{\sigma \\ mm'm''m'''}} V_{mm'm''m'''} \left(n^{\sigma}_{mm'''}\, \delta_{m'm_1}\, \delta_{m''m_2}\, \delta_{-\sigma'\sigma} + n^{-\sigma}_{m'm''}\, \delta_{mm_1}\, \delta_{m'''m_2}\, \delta_{\sigma'\sigma}\right) \\
&= \frac{1}{2} \sum_{mm'''} V_{mm_1m_2m'''}\, n^{-\sigma'}_{mm'''} + \frac{1}{2} \sum_{m'm''} V_{m_1m'm''m_2}\, n^{-\sigma'}_{m'm''} \\
&= \sum_{mm'} V_{m_1mm'm_2}\, n^{-\sigma'}_{mm'} .
\end{aligned}$$

At the last line we have relabelled indices and used the fact that

$$V_{mm'm''m'''} = V_{m'mm'''m''}$$

which follows from (3.2) by interchanging the dummy variables of integration, $\mathbf{r}$ and $\mathbf{r}'$. Coming now to the second term in (6.7) the same reasoning leads to the result,

$$\begin{aligned}
\frac{\partial}{\partial n^{\sigma'}_{m_1m_2}} \quad \frac{1}{2} \sum_{\substack{\sigma \\ mm'm''m'''}} & \left(V_{mm'm''m'''} - V_{mm'm'''m''}\right) n^{\sigma}_{mm'''}\, n^{\sigma}_{m'm''} \\
&= \sum_{mm'} \left(V_{m_1mm'm_2} - V_{m_1mm_2m'}\right) n^{\sigma'}_{mm'} .
\end{aligned}$$

In order to differentiate E_{dc} we note that

$$N = \sum_{m\sigma} n^{\sigma}_{m} ,$$

where $n^{\sigma}_{m} = n^{\sigma}_{mm}$, and use the same tricks as before, noting that N^2 expands as a double sum,

$$N^2 = \sum_{mm'\sigma\sigma'} n^{\sigma}_{m}\, n^{\sigma'}_{m'} = \sum_{mm'\sigma} \left(n^{\sigma}_{m} n^{-\sigma}_{m'} + n^{\sigma}_{m} n^{\sigma}_{m'}\right) .$$

The result is

$$\begin{aligned}
E_{\mathrm{dc}} &= \frac{1}{2} U N\,(N-1) - \frac{1}{4} J N\,(N-2) \\
&= \frac{1}{2} U \sum_{mm'\sigma} \left(n^{\sigma}_{m} n^{-\sigma}_{m'} + n^{\sigma}_{m} n^{\sigma}_{m'}\right) - \frac{1}{2} U \sum_{m\sigma} n^{\sigma}_{m} \\
&\quad - \frac{1}{4} J \sum_{mm'\sigma} \left(n^{\sigma}_{m} n^{-\sigma}_{m'} + n^{\sigma}_{m} n^{\sigma}_{m'}\right) + \frac{1}{2} J \sum_{m\sigma} n^{\sigma}_{m} \\
\frac{\partial E_{\mathrm{dc}}}{\partial n^{\sigma}_{m}} &= U \sum_{m'} \left(n^{-\sigma}_{m'} + n^{\sigma}_{m'}\right) - \frac{1}{2} U - \frac{1}{2} J \sum_{m'} \left(n^{-\sigma}_{m'} + n^{\sigma}_{m'}\right) + \frac{1}{2} J \\
&= U \left(N - \frac{1}{2}\right) - \frac{1}{2} J\,(N-1) .
\end{aligned}$$

We put all this together and obtain the effective Kohn–Sham potential as

$$V = V_{\text{LDA}} + V^{\sigma}_{mm'}$$

where

$$\begin{aligned} V^{\sigma}_{mm'} &= \frac{\partial E^U}{\partial n^{\sigma}_{mm'}} - \frac{\partial E_{\text{dc}}}{\partial n^{\sigma}_{mm'}}\,\delta_{mm'} \\ &= \sum_{m''m'''} \left(V_{mm''m'''m'}\, n^{-\sigma}_{m''m'''} + \left(V_{mm''m'''m'} - V_{mm''m'm'''}\right) n^{\sigma}_{m''m'''} \right) \\ &\qquad\qquad - \left[U\left(N - \frac{1}{2}\right) - \frac{1}{2} J\,(N-1) \right] \delta_{mm'}. \end{aligned} \tag{6.11}$$

Clearly this is a spin and orbital dependent potential, and Anisimov *et al.*[7] point out that (6.11) constitutes a natural extension of the LDA formalism towards the exact density functional. The progression LDA→LSDA→LDA+U introduces *spin* polarisation and then *orbital* polarisation as successive refinements of the theory. The discontinuity of the potential is less obviously clear in (6.11) as it is in (6.6) and the subsequent discussion. We can make this explicit and at the same time simplify things by taking the special case that the number operator is diagonal. This provides the clearest picture of LDA+U; in any calculation, however, one needs to retain the full indexing of (6.7) and (6.11) or the formulation may not be rotationally invariant. We will see below how the $V_{mm'm''m'''}$ can be defined in terms of Gaunt coefficients to guarantee rotational invariance. We should point out here that as long as the number operator is non diagonal there are terms in the hamiltonian that couple orbitals of different magnetic quantum number on the same site, reminiscent of the crystal field terms in self consistent tight binding. But it would be illusory to suppose that $V^{\sigma}_{mm'}$ can properly lift the degeneracy between atomic multiplets. Such splitting can only be described correctly in a theory in which the wavefunction is a combination of several Slater determinants (*ie*, in configuration interaction) while LDA+U is still a Hartree–Fock-like theory. No, the off diagonal hamiltonian matrix elements are here merely a reflection that the number operator is not diagonal in a general representation; they need to be retained simply to ensure rotational invariance. We don't throw out any physics by constructing the following model.

So, in the case that $n^{\sigma}_{mm'} = n^{\sigma}_{m}\,\delta_{mm'}$, (6.7) becomes, rather more simply,

$$E^U = \frac{1}{2} \sum_{mm'\sigma} U_{mm'}\, n^{\sigma}_{m}\, n^{-\sigma}_{m'} + \frac{1}{2} \sum_{mm'\sigma} \left(U_{mm'} - J_{mm'}\right) n^{\sigma}_{m}\, n^{\sigma}_{m'}. \tag{6.12a}$$

Here, $U_{mm'} = V_{mm'm'm}$ is a direct Coulomb integral (3.3) and $J_{mm'} = V_{mm'mm'}$ is an indirect or exchange integral (3.4). If we preferred to assume average, orbital independent U and J, we could then write

$$E^U = \frac{1}{2} U \sum_{mm'\sigma} n^{\sigma}_{m}\, n^{-\sigma}_{m'} + \frac{1}{2}\left(U - J\right) \sum_{\substack{mm'\sigma \\ m\neq m'}} n^{\sigma}_{m}\, n^{\sigma}_{m'}. \tag{6.12b}$$

As in (6.4) we have to explicitly exclude the case $m = m'$ on physical grounds as this corresponds to an energy associated with having two electrons with the same spin and quantum numbers on the same atomic site. This is not necessary in (6.12a) since $(U_{mm} - J_{mm}) = 0$ as is clear from (3.2). In a single band model, or an s-band, the second term in (6.12) would, of course, vanish and we'd recover equation (6.4). Either by direct calculation or from (6.11) using $n^{\sigma}_{mm'} = n^{\sigma}_m\,\delta_{mm'}$ you find that the interaction potential associated with (6.12a) is

$$\frac{\partial E^U}{\partial n^{\sigma}_m} = \sum_{m'} U_{mm'} n^{-\sigma}_{m'} + \sum_{m'} \left(U_{mm'} - J_{mm'}\right) n^{\sigma}_{m'}, \tag{6.13}$$

This is the interaction potential that an electron of spin σ in orbital m sees due to the other electrons of spins σ and $-\sigma$ in orbitals m' at the same atomic site; the occupations of these are given by the occupation numbers $n^{\sigma}_{m'}$ and $n^{-\sigma}_{m'}$. The interaction potential energy of our "test" electron with electrons in orbital m' of opposite spin is $U_{mm'}$. This energy is reduced to $(U_{mm'} - J_{mm'})$ if the interaction is with a like spin electron. Physically this is because the Pauli principle keeps them, on average, further apart, so their electrostatic interaction energy is smaller.

Historically LDA+U was presented in the reverse sequence to these notes. Anisimov and co-workers**7** (as a revision of the original LDA+U with Jan Zaanen and Ole Andersen) first wrote down equation (6.4) which they immediately generalised to (6.12b) and then (6.12a). In a later paper, the rotationally invariant form of (6.7) was conceived. In fact in that paper the formulation is rather LSDA+U since they use the spin polarised expression of (4.3) for the double counting of the interactions in LSDA, after which the final term in (6.11) is $J(N^{\sigma} - \frac{1}{2})$. However, for the purposes of proposing a "tight-binding+U" formalism it is convenient for me to relegate *all* the spin and orbital polarisation to the interaction term H' in the hamiltonian.

In order to explicitly expose the discontinuity in the LDA+U potential, Anisimov *et al.*[7] first added the double counting contribution from the last line of (6.11) to (6.13) and then with some clever manipulation found the following result,

$$\begin{aligned} V^{\sigma}_m &= \sum_{m'} U_{mm'} n^{-\sigma}_{m'} + \sum_{m'} \left(U_{mm'} - J_{mm'}\right) n^{\sigma}_{m'} - U\left(N - \frac{1}{2}\right) + \frac{1}{2} J\,(N-1) \\ &= \sum_{m'} \left(U_{mm'} - U_{\text{eff}}\right) n^{-\sigma}_{m'} + \sum_{\substack{mm' \\ m\neq m'}} \left(U_{mm'} - J_{mm'} - U_{\text{eff}}\right) n^{\sigma}_{m'} + U_{\text{eff}}\left(\frac{1}{2} - n^{\sigma}_m\right) - \frac{1}{4} J \end{aligned}$$

as you can verify by direct substitution using $U_{\text{eff}} = U - \frac{1}{2}J$. (You'll have to complete the sum in the second term on the second line.) Now the last but one term exposes the discontinuity just as in (6.6). In comparison with (6.6) we see that in LDA+U when exchange is included, the Coulomb or "Mott-Hubbard" splitting is reduced from U to U_{eff}.

The simpler model of (6.12b) yields the following expression for the one-electron potential to compare with (6.6)

$$V^{\sigma}_m = (U - J)\left(\frac{1}{2} - n^{\sigma}_m\right) - \frac{1}{2} J\left(N^{\sigma} - N^{-\sigma}\right)$$

where

$$N^{\sigma} - N^{-\sigma} = \sum_m \left(n^{\sigma}_m - n^{-\sigma}_m\right)$$

is the magnetic moment. This shows that even assuming orbital averaged values of U and J does not take us back to LDA. Instead, we have a Coulomb splitting of $(U - J)$ and an additional potential proportional to the magnetic moment. This second term shows a reduced potential for a spin up electron at a site where the spin up population is already greater than the spin down. This is the atomic origin of Hund's rule: the electron sees a lower Coulomb interaction with opposite spin electrons because the Pauli exclusion keeps them, on average, further apart.

Of course both these simple features are retained in the complete LDA+U of (6.7) and (6.11). There remains only to state briefly how the parameters U and J, or more generally, $V_{mm'm''m'''}$ are obtained. In the limit of localised electrons, for which LDA+U is intended, these are essentially atomic Coulomb integrals and fall within the purview of atomic physics. They can be found in textbooks such as Slater or Griffith.**[9]** Looking at equation (3.2), if we suppose that the single particle eigenfunctions are atomic wavefunctions comprising a radial part $u_{n\ell}(r)$ depending on principal and angular momentum quantum numbers $n\ell$ and an angular part depending on the angular momentum and magnetic quantum numbers ℓm, then we can write them as

$$\phi(\mathbf{r}) = u_{n\ell}(r)\, Y_{\ell m}(\hat{\mathbf{r}}).$$

We prefer to work with real orbitals, so we will use real spherical harmonics and as a shorthand will use capital K and L as composite indices, *e.g.* $L = \{\ell m\}$. Note that the real spherical harmonics are related to the complex harmonics Y_ℓ^m of, say, Jackson,**[10]** p. 99 as follows,

$$Y_L = Y_{\ell m} = \begin{cases} (-1)^m \sqrt{2}\, \Re\, Y_\ell^m & m > 0 \\ Y_\ell^m & m = 0 \\ (-1)^m \sqrt{2}\, \Im\, Y_\ell^{-m} & m < 0. \end{cases}$$

We'll first try and work out the most general Coulomb integral as in (3.2)

$$V_{ijkl} = \int \phi_i^*(\mathbf{r})\phi_l(\mathbf{r}) W(\mathbf{r}\mathbf{r}')\phi_j^*(\mathbf{r}')\phi_k(\mathbf{r}')\, \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \tag{3.2}$$

where, since the ϕ are atomic orbitals, the quantum numbers are n, ℓ and m. Next we use an expansion of $W(\mathbf{r}\mathbf{r}') = e^2/\left|\mathbf{r} - \mathbf{r}'\right|$ in terms of spherical harmonics as found in Jackson,**[10]** equation 3.70:

$$\frac{1}{|\mathbf{r} - \mathbf{r}'|} = \sum_K \frac{4\pi}{2k+1} \frac{r_<^k}{r_>^{k+1}}\, Y_K(\hat{\mathbf{r}}) Y_K(\hat{\mathbf{r}}')$$

where $r_>$ is the greatest and $r_<$ the smallest of r and r'. Using this we can write (3.2) as

$$\begin{aligned} V_{nL\,n'L'\,n''L''\,n'''L'''} &= \sum_K e^2 \int r^2 \mathrm{d}r\, u_{n\ell}^*(r)\, u_{n'''\ell'''}(r) \int r'^2 \mathrm{d}r'\, u_{n'\ell'}^*(r')\, u_{n''\ell''}(r') \frac{r_<^k}{r_>^{k+1}} \\ &\times \frac{4\pi}{2k+1} \int \mathrm{d}\Omega\, Y_K(\hat{\mathbf{r}})\, Y_L(\hat{\mathbf{r}})\, Y_{L'''}(\hat{\mathbf{r}}) \int \mathrm{d}\Omega'\, Y_K(\hat{\mathbf{r}}')\, Y_{L'}(\hat{\mathbf{r}}')\, Y_{L''}(\hat{\mathbf{r}}') \\ &= \sum_k R^k \left(n\ell, n'\ell', n''\ell'', n'''\ell'''\right) A_k \left(LL'L''L'''\right) . \end{aligned} \tag{6.14}$$

The last line serves to define the radial integrals R^k and angular terms respectively, on the first and second lines,

$$A^k \left(LL'L''L'''\right) = \frac{4\pi}{2k+1} \sum_{p=-k}^{k} C_{KLL'''}\, C_{KL'L''} ,$$

remembering the composite angular momentum, $K = \{kp\}$ and defining the Gaunt coefficients

$$C_{LL'L''} = \int \mathrm{d}\Omega\, Y_L\, Y_{L'}\, Y_{L''}.$$

These angular terms are just numbers which can be found tabulated, say, in Griffith or Condon and Shortly. They are also routinely computed in LMTO and self consistent tight binding programs. The values of ℓ, ℓ' and ℓ'' are restricted by the triangle inequality: each has to be less than or equal to the sum of the other two, otherwise the integral vanishes. It also vanishes if $\ell + \ell' + \ell''$ is odd. I believe that whereas the radial integrals may be regarded as disposable parameters in a model calculation or in a tight-binding hamiltonian, it is essential to include the angular terms correctly if a calculation is to be rotationally invariant.

The radial integrals then carry all the atomic information. Let us now consider the two cases important to LDA+U. Remembering assumption 3 on page 20, we can take it that we are only interested in those R^k whose arguments take a single value of n and ℓ. But to proceed in stages, let us first assume only that we have just one principal quantum number for any angular momentum in our basis (*e.g.*, we have $3d$ electrons, but no $4d$ electrons; or $4p$ but no $3p$ or $5p$.) Then we can supress the n index in R^k and find that in order to calculate both direct and exchange integrals we are dealing with the *Slater–Condon parameters*,

$$\begin{aligned} F^k\left(\ell,\ell'\right) &= R^k\left(\ell,\ell',\ell,\ell'\right) \\ G^k\left(\ell,\ell'\right) &= R^k\left(\ell,\ell',\ell',\ell\right), \end{aligned}$$

and in the case that we allow just one value of ℓ (a single manifold of correlated states), everything can be expressed in terms of the *Slater integrals*,

$$\begin{aligned} F^k &= F^k(\ell,\ell) = G^k(\ell,\ell) \\ &= e^2 \int_0^\infty r^2 \mathrm{d}r \int_0^\infty r'^2 \mathrm{d}r' \, \frac{r_<^k}{r_>^{k+1}} \, \phi_{n\ell}^2(r)\, \phi_{n\ell}^2(r'), \end{aligned}$$

as before, $r_>$ being the greater and $r_<$ the smaller of r and r'. We then have that in a particular angular momentum manifold ℓ,

$$V_{mm'm''m'''} = \sum_k{}' \frac{4\pi}{2k+1} F^k \sum_{p=-k}^{k} C_{KLL'''}\, C_{KL'L''}.$$

The primes on the L, of course, refer only to the m component since we are dealing with a single value of ℓ. The prime on the summation is to remind us of the restrictions on the Gaunt coefficients, namely $0 \le k \le 2\ell$ and k even (since $k + 2\ell$ must not be odd).

The direct and exchange integrals that enter the model of (6.12a) are therefore

$$V_{mm'm'm} = U_{mm'} = \sum_k{}' \frac{4\pi}{2k+1} F^k \sum_{p=-k}^{k} C_{KLL}\, C_{KL'L'} \tag{6.15a}$$

$$V_{mm'mm'} = J_{mm'} = \sum_k{}' \frac{4\pi}{2k+1} F^k \sum_{p=-k}^{k} \left(C_{KLL'}\right)^2. \tag{6.15b}$$

To make contact between (6.12a) and (6.12b) U and J must be constructed by a proper averaging over the quantum numbers $\{m\}$. For example, in the case of a d-manifold ($\ell = 2$) we could write

$$U = \frac{1}{(2\ell+1)^2} \sum_{mm'} U_{mm'} = F^0, \tag{6.16a}$$

since there are $(2\ell+1)^2$ terms in the double sum over m and m'. It is not at all obvious that the sum is simply equal to F^0, but I was able to obtain this result by direct computation of the Gaunt coefficients using (6.15a). If I were to make an average over $J_{mm'}$ in the same way I would obtain, say,

$$\overline{J} = \frac{1}{(2\ell+1)^2} \sum_{mm'} J_{mm'} = \frac{1}{5}F^0 + \frac{2}{35}\left(F^2 + F^4\right).$$

However, to do the averaging properly we need to look carefully at (6.12a) and (6.12b). Note firstly that for the like spin interaction energy the parameters appear in the combination $U - J$, and secondly that in this combination one should exclude the terms in the sum for which $m = m'$. This is because there cannot be an interaction between two electrons with exactly the same spin and quantum numbers on account of the Pauli principle. (The reason this is not explicitly excluded in (6.12a) is that the corresponding term vanishes anyway because $U_{mm} = J_{mm}$, but this cancellation does not happen in (6.12b). This is a lesson that the correct physics must be taken into account when constructing an approximate model, as was done in (6.4).) Therefore the correct average to take is

$$\begin{aligned} U - J &= \frac{1}{2\ell(2\ell+1)} \sum_{\substack{mm' \\ m \neq m'}} \left(U_{mm'} - J_{mm'}\right) \\ &= F^0 - \frac{1}{14}\left(F^2 + F^4\right), \end{aligned} \tag{6.16b}$$

since there are $2\ell(2\ell+1)$ terms in the restricted sum. Again, I obtained the second line by direct computation and (6.16) agrees with a result by Anisimov *et al.*[7], who may well have obtained it analytically or by inspection. Combining (6.16a) and (6.16b) we then find

$$\begin{aligned} U &= F^0 \\ J &= \frac{1}{14}\left(F^2 + F^4\right). \end{aligned}$$

This completes our description of LDA+U at three levels of sophistication, (6.7), (6.12a) and (6.12b). It seems that the fundamental parameters are the Slater integrals, rather than the U and J. Furthermore one could use the more general (6.3) and (6.14) to construct a multi-band LDA+U or tight-binding+U by treating the R^k as parameters. But note, as mentioned earlier, we mustn't believe that we can treat multiplet splitting in this way. It would not be a good idea to take F^0 from atomic data because the Coulomb interaction between, say, the d-electrons in a metal, are strongly screened by the *sp* electrons. A better approach is to calculate the average U as the second derivative of the total LSDA energy (or the first derivative of the eigenvalue) as a function of electron number on one site in a large supercell and identify this with the Slater integral F^0. The physical meaning of U and how to calculate it is discussed with great clarity by Anisimov and Gunnarsson,[11] and by Cox *et al.*[12] These authors point out that F^2 and F^4 may be taken from atomic data since these are not at all strongly screened in a solid. For d-electrons in transition metals atomic calculations show that $F^2 = 0.625F^4$ to a good approxmation, so for a d-band LDA+U we now just need to find J. This can be found in terms of the Stoner parameter I, which has been calculated for the transition metals using LSDA. The relation between J and I is deduced next. Once the Slater integrals are chosen all the $V_{mm'm''m'''}$ can be found by multiplication through with the Gaunt coefficients.

6. Stoner theory of magnetism and some insights into U, J and I

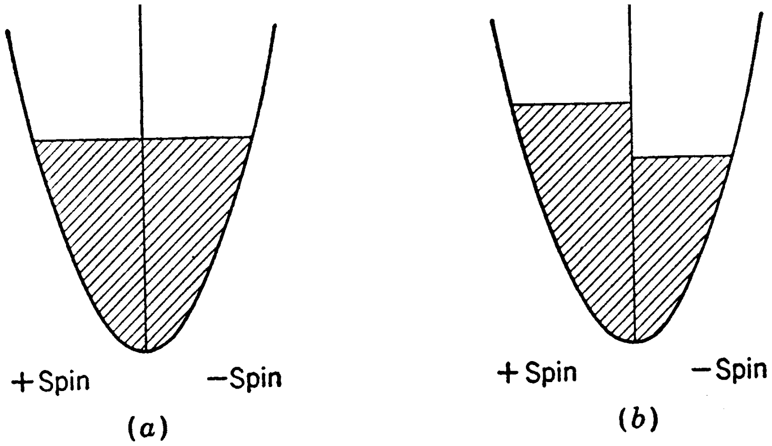
One can make the connection between LDA+U theory and the theory of *itinerant magnetism.* Historically, the first quantum theory of magnetism was based on solutions to the "Heisenberg exchange hamiltonian:"

$$H_{\mathrm{ex}} = \mathcal{E}_0 - 2\sum_{i<j} J_{ij}\,\sigma_i \cdot \sigma_j.$$

Here, the electrons are found on sites i, and J_{ij} are exchange coupling integrals. It can be shown that this is an exact reworking of the hamiltonian of the Heitler–London hydrogen molecule, with $\mathcal{E}_0$ the energy average averaged over the singlet and three triplet states and J the *intersite* exchange integral.[14] This model is the starting point for the treatment of "localised" magnetism, the Ising model, and so on. You can see that this is essentially a picture of spins localised on atomic sites.

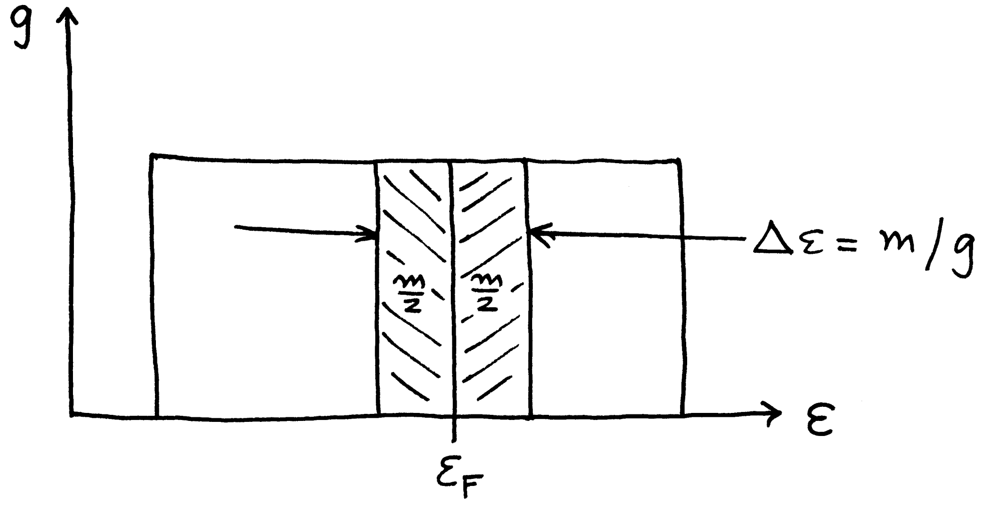
Unfortunately, the Heisenberg hamiltonian is not suitable for describing ferromagnetism in metals. In fact, whereas $J < 0$ in the Heitler–London H_2 molecule—which has the antiferromagnetic ground state—you can see that you will need $J > 0$ to stabilise a ferromagnetic state. It turns out that invariably J is *negative* when calculated between one-electron states on neighbouring atomic sites, and it is difficult to see how this picture can describe ferromagnetism.

The solution is to move away from the atomic picture and think of the electrons in itinerant, band states. The obvious first question to address then is, can the free-electron gas become spontaneously spin polarised at any density? In other words, what is the ground state in the following figure, (a) or (b)?



The figure shows very schematically, back-to-back, the free-electron densities of states for spin-up ($\sigma = +1$) and spin-down ($\sigma = -1$) electrons. If we transfer $\frac{1}{2}m$ spin-down electrons into the spin-up band (*ie*, flip the spins of $\frac{1}{2}m$ electrons), there will be m more spin-up than spin-down electrons—a magnetic moment of $\mu_B m$, where $\mu_B = e\hbar/2mc$ is the Bohr magneton, the magnetic moment of an electron. The system will gain exchange energy as a result of the lowering due to exchange (because it costs less energy to put like spin electrons together as they keep further apart due to the Pauli principle and the electrostatic repulsion energy is reduced) but the process will cost kinetic energy since higher energy band states need to be occupied. For the case of Jellium, both contributions can be worked out in the Hartree–Fock approximation and it turns out that only at very low densities does the exchange term win out, and the Jellium becomes ferromagnetic.**13** The critical density is so low, however, that it is only realised in very few metals (for example Cs) but at these low densities correlation effects that have been ignored become overwhelmingly important.

The d-electrons in transition metals are neither fully localised as are f-electrons nor free-electron like as are the s and p band states. But we need to treat the d-electrons in a band, or itinerant picture. As in the previous paragraph we have to think of the energy bands as two degenerate sets, spin-up and spin-down, which can split under the action of a magnetic or exchange field. The "exchange field," comes from *within* the crystal, due to collective alignment of neighbouring spins. This brings us, then, to the "Stoner model" of itinerant ferromagnetism that draws on the ideas given up to now. As in the previous paragraph, we imagine flipping the spins of $\frac{1}{2}m$ electrons. The magnetic moment (in units of μ_B) is then $m = n^\uparrow - n^\downarrow$ where $n^\uparrow = \frac{1}{2}(n+m)$ is the number of up spins and $n^\downarrow = \frac{1}{2}(n-m)$ if $n = n^\uparrow + n^\downarrow$ is the total number of electrons. The easiest way to calculate the change in the kinetic energy is to consider a rectangular, *ie* constant, density of states $g(\varepsilon) = g$.



We need to calculate

$$\int^{\varepsilon_F-\Delta\varepsilon} \varepsilon g(\varepsilon)\mathrm{d}\varepsilon + \int^{\varepsilon_F+\Delta\varepsilon} \varepsilon g(\varepsilon)\mathrm{d}\varepsilon - 2\int^{\varepsilon_F} \varepsilon g(\varepsilon)\mathrm{d}\varepsilon$$

which is simply

$$g(\Delta\varepsilon)^2 = \frac{1}{4}\frac{m^2}{g(\varepsilon_F)} \tag{7.1}$$

and this is also approximately true if $g(\varepsilon)$ does not vary too much in the range of energy splitting $2\Delta\varepsilon$ around ε_F, the Fermi level.

This is an energy penalty; what about the energy gain from exchange? Here we postulate a term in addition to the one-electron hamiltonian that represents a positive repulsive energy between pairs of *un*like spins,

$$H^I = I\, n^\uparrow\, n^\downarrow$$

where I represents roughly the exchange energy penalty when unlike spin electrons find themselves, say, together in the d-shell of the same atom, in comparison to two like spin electrons whose interaction energy will be smaller because the Pauli exclusion keeps them, on average, further apart. From this term in the hamiltonian it is easy to work out the energy gain from flipping $\frac{1}{2}m$ spins:

$$I\frac{1}{2}(n+m)\,\frac{1}{2}(n-m) - I\frac{1}{2}n\frac{1}{2}n = -\frac{1}{4}Im^2. \tag{7.2}$$

This is an energy gain, so the total (kinetic plus exchange) energy change on flipping $\frac{1}{2}m$ spins is

$$\frac{1}{4}\frac{m^2}{g(\varepsilon_F)} - \frac{1}{4}Im^2 = \frac{1}{4}\frac{m^2}{g(\varepsilon_F)}\left[1 - Ig(\varepsilon_F)\right] \tag{7.3}$$

which leaves a net energy *gain* if

$$Ig(\varepsilon_F) > 1$$

which is called the Stoner Criterion for ferromagnetism. What is new here, compared to the argument over the free electron gas, is that the electron has a smaller potential energy in the presence of like spin electrons. This is similar to the situation described at the top of page 28. Indeed we can take the analogy much further. Consider that the electron sees a potential as in (6.13) and apply the model of constant, average U and J. The electron–electron interaction energy is then given by (6.12b):

$$E^U = \frac{1}{2}U\sum_{mm'\sigma} n_m^\sigma\, n_{m'}^{-\sigma} + \frac{1}{2}(U-J)\sum_{\substack{mm'\sigma\\ m\neq m'}} n_m^\sigma\, n_{m'}^\sigma. \tag{6.12b}$$

Now suppose we consider a degenerate, rigid d-band of states ($\ell = 2$) so the occupation numbers are the same for all m:

$$n_m^\sigma = \frac{1}{2\ell+1}n^\sigma = \frac{1}{5}n^\sigma.$$

We recall that there are $(2\ell+1)^2$ terms in the first sum and $2\ell(2\ell+1)$ terms in the second, so (6.12b) becomes

$$E^U = \frac{1}{2}U\sum_\sigma n^\sigma n^{-\sigma} + \frac{1}{2}(U-J)\,\frac{4}{5}\sum_\sigma n^\sigma n^\sigma \tag{7.4}$$

If we can identify a term in this that looks like $-\frac{1}{4}Im^2$ then we can write down I in terms of U and J. To do this we develop these simple identities:

$$\begin{aligned} n^2 = (n^\uparrow + n^\downarrow)^2 &= n^\uparrow n^\uparrow + n^\downarrow n^\downarrow + n^\uparrow n^\downarrow + n^\downarrow n^\uparrow \\ &= \sum_\sigma n^\sigma n^\sigma \quad + \sum_\sigma n^\sigma n^{-\sigma} \\ m^2 = (n^\uparrow - n^\downarrow)^2 &= n^\uparrow n^\uparrow + n^\downarrow n^\downarrow - n^\uparrow n^\downarrow - n^\downarrow n^\uparrow \\ &= \sum_\sigma n^\sigma n^\sigma \quad - \sum_\sigma n^\sigma n^{-\sigma}, \end{aligned}$$

from which it follows that

$$\begin{aligned}\sum_\sigma n^\sigma n^{-\sigma} &= \frac{1}{2}\left(n^2 - m^2\right)\\ \sum_\sigma n^\sigma n^\sigma &= \frac{1}{2}\left(n^2 + m^2\right).\end{aligned}$$

Putting this into (7.4) we have

$$\begin{aligned}E^U &= \frac{1}{2}U\frac{1}{2}\left(n^2 - m^2\right) + \frac{1}{2}(U - J)\frac{4}{5}\,\frac{1}{2}\left(n^2 + m^2\right)\\ &= \frac{1}{20}(9U - 4J)\,n^2 - \frac{1}{4}\frac{U + 4J}{5}m^2\end{aligned}$$

from which we immediately identify

$$I = \frac{U + 4J}{5} = \frac{U + 2\ell J}{2\ell + 1} \tag{7.5}$$

where the last result is a straightforward extension to any single degenerate band with angular momentum ℓ. Note that for an s-band ($\ell = 0$) and $I = U$.

We may use (7.2) to put some bounds onto U and I, which now emerge as the fundamental quantities to be determined. Fortunately they can both be calculated from first principles using the local spin density approximation.[11,15] We expect $U - J > 0$, and also $J > 0$ follows from (6.15b). Using (7.5) we see that

$$\begin{aligned}J &= \frac{1}{2\ell}\left((2\ell + 1)\,I - U\right) > 0\\ U - J &= \frac{2\ell + 1}{2\ell}(U - I) > 0\end{aligned}$$

so that

$$I < U < (2\ell + 1)\,I.$$

As I have mentioned, U can be estimated from a local density functional calculation in which the eigenvalue is followed as a function of d-electron number under certain conditions that allow the screening by the sp-electrons to be properly accounted for.[11] There are two interesting consequences.

1. Since this is a local density calculation, both exchange *and* correlation are included in the one electron potential, so that whereas LDA+U is in form a Hartree–Fock like theory, correlation can be included by this "renormalisation" of U.
2. Although the local spin density approximation is known to fail in predicting the electronic structure of highly correlated solids because the potential does not contain a term like (6.13), the parameter U *can* be extracted from an LDA calculation. Somehow the correct physics is contained, hidden in the LDA potential.

In a similar way the Stoner I can be extracted from an LSDA calculation.[15] In local spin density, the spin dependent exchange and correlation potential is very reminiscent of the potential described by (6.13) and discussed just below equation (7.3). In the parameterisation of von Barth and Hedin,[15] for a uniform electron gas of density ρ and magnetic moment m, the exchange correlation potential seen by up and down spin electrons is

$$\mu^{\pm}_{\mathrm{XC}}(\rho) = \mu^{P}_{\mathrm{X}}(\rho)\left[B(\rho) \pm \frac{2}{3}\frac{D(\rho)m/n}{1 \pm \gamma m/n}\right], \quad \text{[Ryd.]}$$

where the plus and minus signs refer respectively to the up and down spins. The constants, in terms of the density parameter

$$r_s = \left(\frac{3}{4\pi\rho}\right)^{\frac{2}{3}},$$

are

$$\begin{aligned} B(r_s) &= 1 + 0.0545 r_s \ln(1 + 11.4/r_s) \\ D(r_s) &= 1 - 0.036 r_s + 1.36 r_s/(1 + 10 r_s) \\ \gamma &= 0.297 \end{aligned}$$

Here, $D = B = 1$ are the exchange-only potentials, and μ_{x}^{P} is the paramagnetic exchange potential. You can see that in exchange-only the exchange splitting is exaggerated as found in Hartree–Fock calculations (which, for example, predict a large number of non magnetic transition metals like Pd to be ferromagnetic). So again a calculation of I using the LSDA potential results in a renormalised parameter containing effects of correlation that are absent in Hartree–Fock. The formula for calculating I discovered by Gunnarsson is

$$I(\varepsilon) = -\int_0^{r_{\mathrm{WS}}} r^2 \mathrm{d}r \left(\frac{\mu_{\mathrm{x}}^{P}(r_s)}{6\pi}\right) D(r_s) \frac{\phi_d^2(r,\varepsilon)\,\phi_d^2(r,\varepsilon_{\mathrm{F}})}{\rho(r)}.$$

Here, ϕ_d is the radial part of the self-consistent $\ell = 2$ Kohn–Sham orbital and the integral is carried from the nucleus to the radius of the Wigner–Seitz sphere. The rather weak energy dependence can be removed by averaging over the Fermi surface. It is notable that the calculated I is typically only very weakly dependent on crystal structure or even atomic volume (through r_{WS}). It thereby turns out to be essentially an atomic property.

1. H. Haken, “Quantum field theory of solids,” (North Holland, 1976)
2. Walter Harrison, *Phys. Rev. B*, **31**, p2121 (1985))
3. J. F. Janak, *Phys. Rev. B*, **18**, 7165 (1978)
4. R. G. Parr and W. Yang, “Density functional theory of atoms and molecules,” (OUP, 1989)
5. O. Gunnarsson and K. Schönhammer, *Phys. Rev. Lett.*, **56**, 1969 (1986)
6. J. Zaanen, O. Jepsen, O. Gunnarsson, A. T. Paxton, O. K. Andersen and A. Svane, *Physica C*, **153–155**, 1636 (1988)
7. V. I. Anisimov, I. V. Solovyev, M. A. Korotin, M. T. Czyżyk and G. A. Sawatzky, *Phys. Rev. B*, **48**, 16929 (1993)
8. V. I. Anisimov, F. Aryasetiawan and A. I. Lichtenstein, *J. Phys. : Condens. Matter*, **9**, 767 (1997)
9. J. S. Griffith, “The theory of transition-metal ions,” (CUP, 1961)
10. J. D. Jackson, “Classical Electrodynamics,” 2nd Edition (Wiley, 1975)
11. V. I. Anisimov and O. Gunnarsson, *Phys. Rev. B*, **43**, 7570 (1991)
12. B. N. Cox, M. A. Coulthard and P. Lloyd, *J. Phys. F*, **4**, 807 (1974)
13. J. C. Slater, “Quantum Theory of Matter,” 1st Edition (McGraw-Hill, 1951) pp. 418ff
14. For a demonstration of this, and an amplified version of the description of the Stoner model, see `http://titus.phy.qub.ac.uk/group/Tony/ESC` (section 7)
15. O. Gunnarsson, *J. Phys. F*, **6** , 587 (1976); J. F. Janak, *Phys. Rev. B*, **16**, 255 (1977)