INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

SMR.764 - 5

RESEARCH WORKSHOP ON CONDENSED MATTER PHYSICS
13 JUNE - 19 AUGUST 1994

WORKING GROUP ON "DISORDERED ALLOYS"
8 - 19 AUGUST 1994

"Topological Disorder, Cohesive Energy and the Recursion Method"

Roger HAYDOCK
Department of Physics & Materials Science Institute
University of Oregon
Eugene, Oregon 99352
U.S.A.

These are preliminary lecture notes, intended only for distribution to participants

Lecture 1 Topological Disorder

This is the first of three lectures on topological disorder, the properties of amorphous graphite, and nearly conserved quantum numbers in topologically disordered systems. The main point of the first lecture are that systems such as amorphous carbon show a surprising, near conservation of momentum which is a consequence of chemical order, a property captured in the Thorpe-Weaire model. The second lecture is about how the local order in amorphous carbon can be used to obtain a nearly conserved quantum number for amorphous graphite and the spectroscopic consequences of this near conservation law. The third lecture is about the general relationship between local order and conservation laws, with an application to photoemission from glassy materials

A. Different Kinds of Disorder

The simplest kind of disorder is that of random substitution between different kinds of atoms on a lattice. The most common example of substitutional disorder is a random alloy such as CuAu. In pure form, both Cu and Au are FCC metals with atoms of similar size. As a result it is easy for one kind of atom to substitute for the other without needing to create defects in the lattice.

The simplest model for a random alloy is that of a lattice, FCC in this example, whose sites are occupied by one kind of atom, Cu, with probability p and by the other, Au, with probability $1-p$. In this model, every site is occupied with probability one.

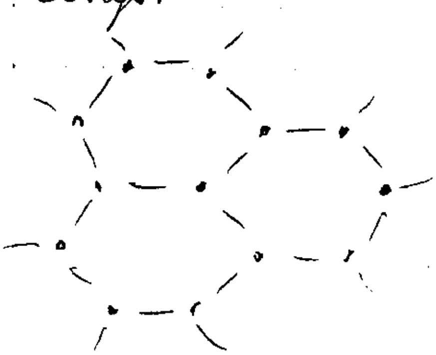
In this kind of disorder, it is the geometric packing of the atoms which is most important. The energy of the system is lowered by the close packing of the atoms irrespective of which atoms end up next to one another. In the case of Cu-Au the chemical similarity of the two atoms, one s electron per atom, keeps them in an FCC structure rather than a random stacking of close packed layers which occurs in some other systems.

In random alloys the bonding is usually metallic in which the dominant forces are due to electron density, and there are only weak covalent forces.

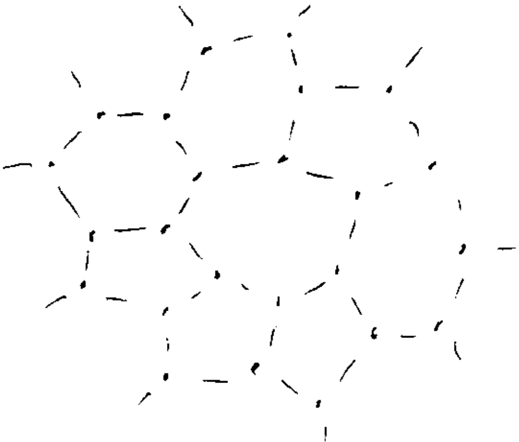
At the other extreme of bonding, to the right-hand side of the periodic table, the laws of valency, or chemical forces dominate the formation of structure. These forces are due to the energetic advantages of exactly filled shells which even out over electron density. The structures tend to be open with one kind of atom strongly preferring particular arrangements of other kinds for neighbours.

The simplest example of such a covalent material is Si which strongly prefers to have each atom bonded to four neighbors in a nearly tetrahedral arrangement. Although the lowest structural energy seems to be achieved by the diamond lattice, a crystalline structure, there are many non-crystalline structures in which the neighbor relationships are nearly ideal, at least over some region of space.

These disordered structures satisfy the chemical requirements by having each atom bonded to four neighbours in a nearly tetrahedral arrangement, but rings occur with 5 and 7 members instead of the six membered rings of the crystal. Graphite is a 2-dimensional analogue of diamond in which each atom is three-coordinated with neighbors at the equilateral angles. Again, the crystal has only 6-membered rings, while ~~the~~ other structures can have 5 and seven-membered rings.



Crystalline Graphite



Amorphous Graphite

This kind of disorder where there is local chemical order, but no long range order is called topological disorder because the topology of the network has been changed from that of the crystal. In contrast to substitutional disorder in which the overall structure or topology survives, in topological disorder the structure is changed and stabilized by the strong chemical forces which prevent the bonds from breaking and reforming the crystal.

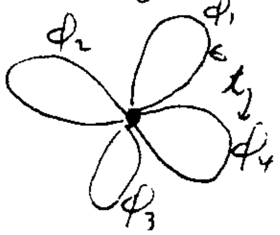
A precise definition of topological disorder is that an (infinite) macroscopic number of bonds must be broken and reformed in order to make the structure crystalline. If we assume that there is a finite energy barrier to ~~[illegible]~~ rebonding each atom, then the total barrier to crystallization is macroscopic and requires the system to be heated to recrystallize

B. Thorpe-Weaire Model

The obvious question to ask about amorphous Si and other topologically disordered materials is how their electronic (and vibrational) structure differs from that of the crystal.

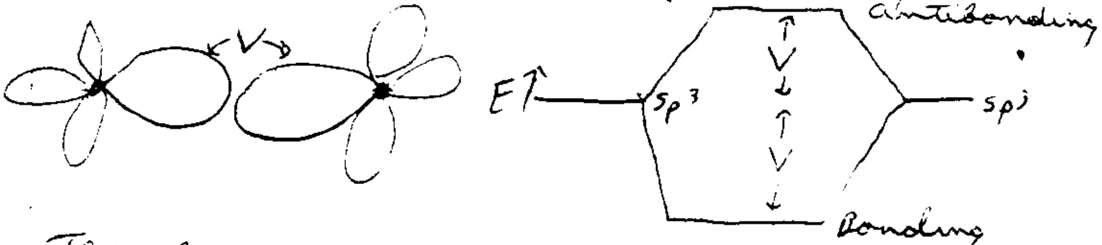
At one time it was thought that long-range (crystalline) order was necessary for the existence of a gap. That this is not so was shown by Thorpe and Weaire using a simple model for the electronic structure of topologically disordered materials. Cry~~stalline~~

Crystalline silicon has a valence band separated from the conduction band by a relatively narrow gap. This electronic structure can be understood in terms of sp^3 hybrid orbitals on each atom - the combinations of the valence s and three valence p orbitals which form lobes along the tetrahedral directions. The sum of the four sp^3 hybrids is the s orbital.



These orbitals are occupied when it is energetically favorable to promote one of the atomic s electrons to the p shell in order to increase the coordination of the atoms.

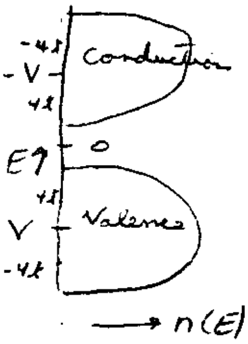
The structure is formed by strong hybridization of sp^3 orbitals directed toward one another from neighbouring atoms.



This bonding lowers the energy of two electrons by $|V|$ in each band to give a cohesive energy of 4V per atom (V is negative due to attractive potentials).

In addition to the interatomic hybridization which separates the bonding and antibonding levels in each bond, the sp^3 hybrids on one atom are coupled by the atomic potential, because the atomic s and p levels are not really degenerate, with a matrix element t. Without t, the electronic spectrum of the crystal or any other 4-coordinated structure, consists of the fully occupied bonding level at V and the unoccupied antibonding level at -V, relative to the atomic level.

The intraatomic hybridization t spreads the bonding and antibonding levels out into band and somewhat increases the energy by the promotion energy. This then gives the usual band structure of a tetrahedral semiconductor as shown to the left.

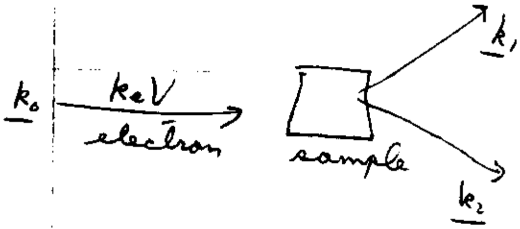


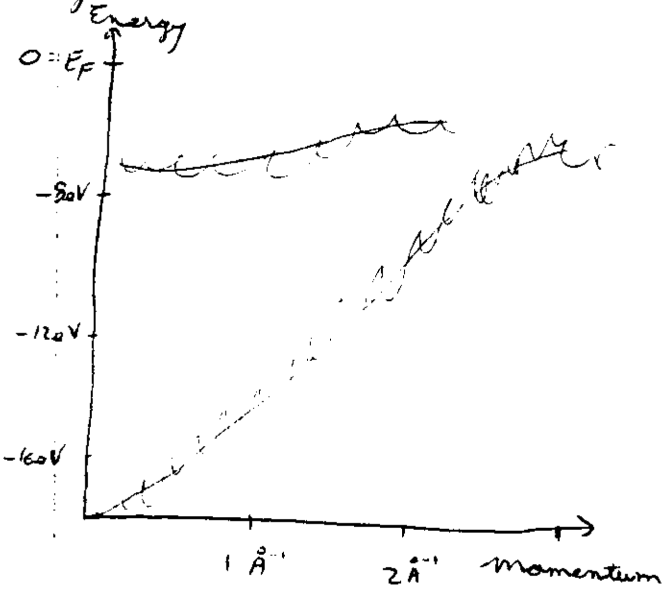
The crux of the Thorpe-Weaire argument is that as long as $4t > V$ (t and V both neg.) there must be a gap between the conduction and valence bands. This is a consequence of 4-fold coordination of atoms that the maxima and minima of the valence bands and conduction bands cannot exceed $4|t|$ from the bonding and antibonding levels.

Provided the mixing between different bonds is weak, the disorder cannot smear the valence and conduction bands into the gap. This is why window glass is transparent despite being topologically disordered Si-O-Si, and when there are gap states in disordered tetrahedral materials, they must be due to dangling bonds - defects in the coordination

C. e-2e Spectroscopy of Amorphous Carbon

The last topic today is the surprising results of e-2e spectroscopy on amorphous carbon. This spectroscopy is carried out by shooting an electron of several kilovolts into the sample, knocking a second electron out and detecting both the scattered electrons measuring their wave vectors. Using the initial wave~~number~~ vector $\underline{k}_0$ and the wavevectors of the scattered electrons $\underline{k}_1$ and $\underline{k}_2$, the energy and wave~~number~~ vector E_v, k_v of the electron originating in the sample can be determined. This experiment has been done by Ritter and Gao PRL ~~PR~~ 62 p945 (1989) and what they got was narrow bands very like those of graphite, out into the 3rd Brillouin zone.





It seems that despite the disorder present in the amorphous carbon, the momenta ~~to~~ of the valence electrons is well defined. The remaining two lectures deal with understanding this surprising result

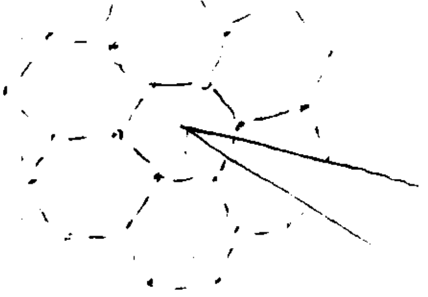
It should be noted that the coincident detection of the two scattered electrons makes this a very time consuming experiment. attempts to carry out similar experiments on amorphous silicon simply melted the sample when the electron flux was raised to obtain a reasonable rate of detection.

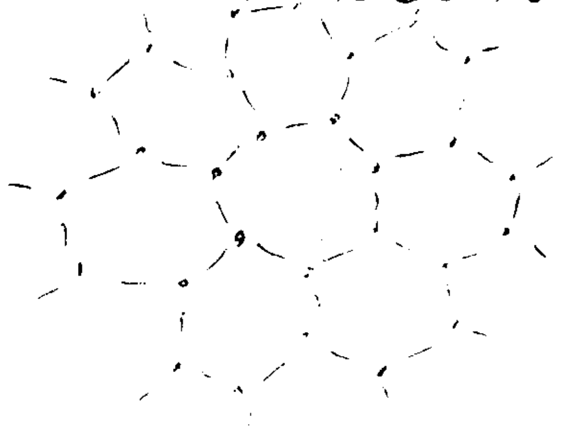
2 Amorphous Graphite

In this lecture I will use the property of local chemical order to show that in amorphous graphite, an idealization of amorphous carbon, there is a momentum-like quantum number which is almost conserved. This is surprising since one expects conservation of momentum to be one of the most obvious casualties of topological disorder, which destroys the long range order needed for Bloch's theorem.

A. A topological Defect in Graphite

The simplest example of a topological defect in graphite is an odd membered ring. In order to create such a ring, take the graphite structure and cut bonds from the center of a six-membered ring straight out to infinity, as shown below. Then add an extra 1/6th sector to make the structure shown on the right which has a single seven-membered ring at the center, corresponding to the origin of the cut in the crystal. All the surrounding rings are six-membered, and the structure can be extended infinitely with six-membered rings.





The first important conclusion from the above construction is that an infinite number of bonds must be broken and 'reconnected' in order to insert a single seven or odd-membered ring into the structure. ~~If~~ In terms of the electronic hamiltonian such as the Weaire Thorpe model, this is an infinite or macroscopic perturbation of the system. Yet, it only changes one ring - not really, it inserts or removes an infinite wedge of material.

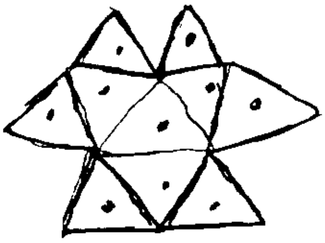
The second conclusion is that although the change in the system is macroscopic, it ends up looking just like the crystal except when one sees the ~~oo~~ whole of the odd-membered ring. If just part of the defect is seen, it still looks like part of the standard six-membered ring. The defect does change the structure even in the regions where all rings are six-membered, as can be seen by measuring the number of hops to complete an orbit or circle around the defect. The relation between the radius and circumference of this circle is different than in the crystal.

Such a structure can actually be realized. The addition of the wedge turns the flat plane of graphite into a saddle shape.

B. Local Order

When we say that the above defective structure looks the same locally as the crystal, we should be precise and define local order more carefully.

In this case local order means that there is a geometric relationship between nearby atoms. It is the same relationship, in this case but not necessarily in others, as that in the crystal, namely that each atomic cell has six, second-neighbour atomic cells to which it is equivalent. The cell and its six equivalent second neighbors ~~to~~ are illustrated below. The edges of the triangular atomic cells equivalent



to the central cell have been made darker.

There are several important points about this relationship. The first is that the arrangement of ten atoms includes no rings. The central atom is where three rings meet, but the relationship between the central atom and its neighbors does not involve the entirity of any ring. As a result, the above figure can be superimposed on <u>any</u> atom in the defective structure and the others will match up with only distortion no breaking of bonds. This same property also holds for the graphite structure.

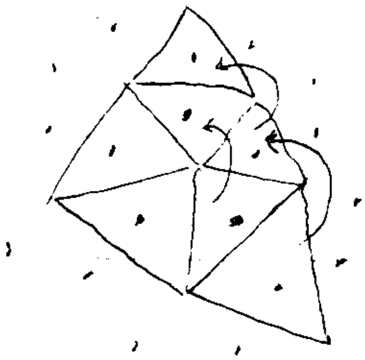
The property that the ten atoms can be superimposed anywhere in the defective structure or any three-coordinated structure with no rings of less than ~~[illegible]~~ five members defines the local order. In this case the local order is not just the 3-fold coordination of atoms, but includes the constraint on minimum ring size.

The local order is a mathematical formulation of the idea of chemical order, namely the chemical tendency to form structures in which every atom has six second neighbors with almost identical orientation. Because we allow small distortions in the structure, including the closing of the rings of first and second neighbors to make five-membered rings, this local order is a topological property of the structure

C The Network Operator

Having identified the local order as the local geometric relationships between cells, the next step is to use this to define a transformation of the system, and then a transformation of wavefunctions.

The transformation suggested by the local order is to map atomic cells into their second neighboring atomic cells, just as the translation of a crystal maps atomic cells to equivalent atomic cells. If we start by mapping one cell to one of its second neighbors, then we can extend the transformation by mapping bonded cells to bonded cells as illustrated below. The mapping can be extended



in a way which preserves the bonding and second neighbor relationships except on the five membered ring.

If the mapping of cells to equivalent second neighbor cells is extended to the two atoms on the left of the 5-ring, then in order to preserve the bonding relations, ~~a~~ one cell must be mapped to its neighbor, or the preserving of bonding must be broken. The latter is preferable, and in this sense, the odd-membered rings are true defects in the extension of the mapping.

A different thing happens if the mapping is extended around an odd-membered ring without being extended onto the ring. If the mapping started ~~with~~ by relating a particular cell to one of its neighbors, then after extending the mapping once around the defect, the original cell is now mapped to a different one of its neighbors. Another circuit of the defect maps it to the third of the six, and continued extension of the mapping does not relate the particular atom to the same second neighbor until the defect has been encircled six times.

What this means is that the mapping has six branches corresponding to each of the six local translations of the graphite crystal. The odd-membered rings couple these branches in the same way that analytic continuation of the complex analytic function $z^{1/6}$ connects its 6 branches into a Rieman surface.

Although in the crystal, these different mappings are decoupled, in the topologically disordered structure they are coupled by defects. The global mapping generated by extending the mapping between two second neighbor cells, couples each cell to all six of its second neighbors. This global transformation, called the Network Transformation is a geometric mapping of ~~the~~ each atomic cell onto each of its second neighbors.

Like a symmetry transformation, the network transformation preserves geometric relationships - the bonding between atoms - except at the defects. The network transformation is simply a generalization of a symmetry transformation to being multivalued and to breaking geometric relationships at isolated defects. This is like the application of the crystal symmetries when some lattice sites are occupied by wrong atoms making defects; as is often done in the theory of alloys

Just as the symmetry transformations of a crystal transform wavefunctions and vibrational modes into one another, the network transformation does the same, generating a network operator which is the representation of the network transformation on the electronic wavefunctions.

In the simplest case we can suppose there is just one s-orbital on each atom. The network operator simply replaces the ~~comp~~ coefficient of each s-orbital by the sum of the coefficients of the six, second neighbor orbitals. Written as a matrix, the network operator N ~~is yes~~ has ~~only~~ exactly six unit elements in each row and column with all the rest zero. These six non-zero elements are simply those corresponding to pairs of second neighbor orbitals in the structure.

The network operator can be extended to any choice of basis for the electrons or any other excitation in each cell. If there are twenty orbitals per cell, then each row and column of the matrix for N has 6 times 20 1's correspon to equivalent pairs of orbitals. The rest of the elements are zero.

D Conservation of Network Momentum

Since the network transformation preserves geometric relationships, except on odd membered rings, and since the hamiltonians for electrons depend only on geometric relationship the network operator N commutes with the hamiltonian H except on the odd-membered rings.

$$[H, N] = \Delta$$

where Δ is only modifies orbitals on odd-membered rings.

This commutation, or almost commutation means that the invariants or representations of N change only slowly with time, proportional to the number of defects. The invariants of N are called states of definite network momentum and their invariant values, scalar in this case, are the values of moment

3 Approximate Conservation Laws

In this third lecture of the series on topological disorder I will discuss the consequences of approximate commutation of the network operator with the Hamiltonian. The first of these is that the states of approximate network momentum can be obtained by diagonalyzin the network operator. These are isotropic because there is usually only one component of network momentum. The second consequence is that an approximate or effective medium hamiltonian can be constructed to commute exactly with the network operator. The true hamiltonian can be viewed as the approximate hamiltonian with substitutional rather than topological disorder so that the introduction of network momentum has greatly reduced the problem. Finally the conservation of network momentum implies a selection rule for transitions such as photoemission.

A. Representations of Network Momentum

Just as symmetry transformations can be represented as operators which act on electronic wavefunction, so the network transformation produces an operator which acts on wavefunction. The example of graphite can be generalized to any structure which has local order, namely a mapping ~~from~~ between atomic cells which are locally equivalent such as second neighbor cells for amorphous silicon and so on. When these local mappings are extended by requiring bonded cells map to bonded cells except at isolated defects, they produce global transformations which act on electronic orbitals to make a representati of the network transformation which is called the network operator N.

N is analogous to the translation or rotation operators except that it does peculiar things to orbitals near the topological defects. ~~Also~~ In the general case N couples all the orbitals of the system together and so is a reducible representation of the network transformation.

Irreducible representations can be constructed by finding linear combinations of orbitals which transform into themselves. A Bloch wave is an irreducible representation of the translation operator because translation of a Bloch wave only changes its phase. The network operator N is particularly simple because it is usually symmetric - if it maps orbitals from one cell into another, then the reverse mapping is also part of N. Because it is Hermitian, N can be diagonalized to produce one-dimensional representations which are therefore irreducible,

$$N \phi_s = s \phi_s$$

where the invariant wave is ϕ_s and s ~~it~~ is its ~~eigenval~~ invariant value. The real number s is not itself a momentum, but is easily converted to a momentum by dividing it by Z, the number of equivalent neighbors (6 for α·c) and then getting a phase angle ~~an angle, dividing~~

$$\theta = \cos^{-1} s/Z,$$

dividing by the distance between equivalent neighbor a, and multiplying by $\hbar$ to get the network momentum

$$n_s = (\hbar/a)\cos^{-1}(s/Z).$$

In the case where the locally ordered material is actually crystalline, the network operator separates into three components whose values are related to crystal momentum by the above formula, and whose invariant states or irreducible representations are Bloch waves. In the case of non-crystalline materials, the irreducible representations ϕ_s look like superpositions of Bloch waves travelling in different directions corresponding to the different directions between equivalent sites. This is the origin of the phase relationship, but unlike Bloch waves, the network waves vary in amplitude from cell to cell because of the disorder. They might also ~~be thought of as~~ be localized in which case the momentum might be thought of as complex.

The irreducible representations of N have similar properties to bands in that the network momentum is periodic in the different kinds of orbitals in each atomic cell. In general the ~~the~~ analogue of the Brillouin zone, the network zone is one dimensional because N has only one component, put this one dimensional reduced zone has an infinite number of bands corresponding to each of the inequivalent degrees of freedom in the structural cells.

B. Effective Hamiltonian

In the last lecture I showed that the network operator N does not commute exactly with the Hamiltonian because the bonding relationships are not preserved at the topological defects;

$$[H, N] = \Delta,$$

where Δ is localized on the defects.

The Hamiltonian can be split into two parts,

$$H = H_0 + V$$

where H_0 commutes with N and V does not. The effective medium Hamiltonian H_0 can be constructed from H and N using a variation of the recursion method.

What we wish to do is find V, the network momentum changing potential whose commutator with N is Δ,

$$[N, V] = \Delta.$$

In order to do this, expand V in the commutators of N with Δ,

$$V = c_0 \Delta + c_1 [N, \Delta] + c_2 [N, [N, \Delta]] + \dots$$

Now the requirement for V is that

$$\Delta = c_0 [N, \Delta] + c_1 [N, [N, \Delta]] + c_2 [N, [N, [N, \Delta]]] + \dots$$

It is more convenient to generate an operator basis, $U_{-1} = 0$, $U_0 = \Delta$,

$$U_{n+1} = [N, U_n] - \beta_n U_{n-1},$$

and the diagonal elements of the recurrence are zero because commutation of an operator with N changes it from Hermitian to anti hermitian or the opposite. For the same reason, the expansion of V can contain only Hermitian terms and

$$V = \rho_1 U_1 + \rho_3 U_3 + \dots + \rho_{2n+1} U_{2n+1} + \dots$$

The equation is now solved by requiring

$$[N, V] = \Delta = \beta_1 \rho_1 \Delta + (\rho_1 + \beta_3 \rho_3) U_2 + (\rho_3 + \beta_5 \rho_5) U_4 + \dots$$

or $\rho_1 = 1/\beta_1$, $\rho_3 = -\frac{1}{\beta_1 \beta_3}$, $\rho_5 = \frac{1}{\beta_1 \beta_3 \beta_5}$, ...

This series converges very quickly to a momentum changing potential V which is exponentially localized about the defects.

The effective Hamiltonian H_0 is

$$H_0 = H - V$$

now commutes exactly with N and its invariant functions are approximately stationary states of the amorphous material.

The first test of these states of the effective Hamiltonian is to see whether they give the same momentum energy distribution as the experiment for amorphous carbon. The results of calculations using atomic orbitals and Robertson's parameters for graphite gave results very similar to experiment with the most noticeable difference being that the calculated π-band is more dispersive than the experimental π-band. This may be due to the calculations neglecting interactions between graphitic sheets which could tend to flatten the π-band.

In other calculations the local density of states was compared for the effective and full hamiltonians at a site on an five-membered ring. The results were very similar despite the potential V being greatest on these sites. The main difference was oscillations in the effective hamiltonian density of states near the top of the valence band. This can be understood in terms of the unstable exclusion of these states from five, or odd-membered rings because the phase of the state must alternate + and - around the ring which it cannot do on an odd ring

C Selection Rule for Photoemission

The most stringent test for conservation of network momentum is whether it leads to an observable selection rule for photoemission. If ~~it is an~~ this conservation law is an important effect then it should prevent network momentum non-conserv transitions from contributing to a photocurrent.

Ultraviolet photoemission in the 10 - 100 eV range uses photons whose momentum is negligible compared to that of the electrons because the electromagnetic wavelengths are

are 1000Å to 100Å compared to electron wavelengt of the order of 1 Å.

As a result, the initial and final wavefunct of the electron, before and after excitation, must have the same reduced momentum. In the case of photoemission from a crystal, the ~~state~~ crystal momentum of the two states must be the same, and in the case of an amorphous material, the network momentum must be the same, if network momentum is conserved.

Because the photon also has a definite energy, the emission process only occurs if there are initial an final states having ~~the~~ ~~correct~~ an energy difference equal to the energy of the photon, and both having the same network momentum. This selection rule should dramatically restrict the states into which electrons are photoexcited as well as allowing the initial states to be mapped out much as is done in photoemission from crystals and in e-2e spectroscopy of amorphous materials. Note that unlike the selection rule for crystal momentum, the conservation of network momentum still leads to isotropic emission, but with very big changes in the emission as the photon energy is changed, like those from crystals

__Acknowledgement__ This work was done in collaboration with Andrew Gibson at the University of Oregon and has been published in Phys. Rev. Lett 69 3793 (1992) and J. Non. Cryst. Sol 164-6 1135 (1993). ~~These lectures were written while enjoying the hospitality of the Cavendish Lab. Cambridge, Pembroke College Camb~~ and the assistance of NSF.

4. Cohesive Energy Functionals

Density functional theory and the local density approximation have been an important part of recent work on the electronic structure of solids. In this lecture I will review density functional theory explaining in simple terms why it works to the extent it does and then describe how it can be formulated in terms of variation of the electronic wavefunctions, the electronic potential, and the electronic charge density.

A Density Functional Theory

Density functional theory is related to the Thomas-Fermi approach in that the total energy of an electronic system is separated into various terms which are expressed as functionals ~~in terms~~ of the electronic density $\rho(r)$. The ~~total~~ ~~as function~~ total energy

$$U = \underbrace{\frac{1}{2}\sum_{\alpha\beta}\frac{Z_\alpha Z_\beta}{|R_\alpha - R_\beta|}}_{\text{ionic repulsion}} - \underbrace{\sum_\alpha \int \rho(r)\frac{Z_\alpha}{|r-R_\alpha|}\,dr^3}_{\text{electron-ion attraction}}$$

$$+ \underbrace{\frac{1}{2}\iint \frac{\rho(r)\rho(r')}{|r-r'|}\,dr^3 dr'^3}_{\text{electron-electron repulsion}} + \underbrace{T[\rho]}_{\text{electronic kinetic energy}} + \underbrace{U_{xc}[\rho]}_{\text{exchange and correlation attraction}}$$

One of the main disadvantages of this approach is clear from the above formula in that there are enormous cancellations between the various terms. Some of these are between the ionic repulsion and electron-ion attraction, and between the electron-electron repulsion and the exchange correlation attraction. It would be far better to make the cancellations implicit as is done later.

The approach is classical in that the classical electron-electron repulsion is calculated from the total charge density and then explicitly corrected for the quantum effects of exchange and correlation. The kinetic energy is viewed classically as the lowest kinetic energy for a system of the appropriate number of independent electrons constrained to have charge density ρ. The effects of interactions ~~and~~ on the kinetic energy are all put into U_{xc}.

All the difficult parts of the problem of electronic energies are relegated to U_{xc} which may be thought of as ~~the~~ mainly the electrostatic binding energy of each electron to its exchange and correlation holes - the absence of other electrons of the same and different spin from the vicinity of one electron. The kinetic corrections to U_{xc} are the energies required for the formation of the exchange and correlation holes.

B. How the Local Density Approximation Works.

The local density approximation (LDA) is the approximation of U_{xc} by the exchange and correlation energy density of the homogeneous electron gas locally of the same density. This is the only approximation possible because the homogeneous electron gas is the only electron system for which these energies can be calculated (by monte carlo simulation or earlier by perturbation theory). The criterion for the applicability of LDA is that the change in electron density over one fermi wavelength be small, and this is rarely statisfied, consistent with ~~the~~ binding energies from LDA which are about 10% to large because the exchange and correlation is local encouraging electrons to pile up on ~~p~~ bonds and in atoms.

Obsession with interaction effects obscures the reason the density functional theory is useful, namely because it does a good job on kinetic energy. The electronic kinetic energy is very sensitive to structure as can be seen in bandstructure calculations. Small structural changes have big effects on band gaps and hence on the kinetic energy contributions. Electrostatic energies which dominate U_{xc} are long-ranged, slowly varying, and hence insensitive to structure implying that it doesn't much matter what goes into U_{xc} provided that it gets the average interactions right - basically the atomic volume.

C Selfconsistent Solution

The usual method for solution of the density functional equations is to guess an electronic potential $V(r)$, then solve the electronic Schroedinger equation in order to find independent electron states $\{\Psi_\alpha(r)\}$ which minimize the hamiltonian

$$H = -\hbar^2/2m\, \nabla^2 + V(r)$$

The associated charge density $\rho(r)$ is obtained as

$$\rho(r) = \sum_{\alpha=1}^{N} |\Psi_\alpha(r)|^2$$

where the wavefunctions summed ~~are the number~~ accomodate all ~~equal to the number of~~ the electrons, with lowest energy. The kinetic energy,

$$T[\rho] = \sum_{\alpha=1}^{N} \epsilon_\alpha - \int V(r)\rho(r)\,dr^3$$

where ϵ_α is the energy of $\Psi_\alpha(r)$. The rest of the functionals in the total energy can be evaluated directly from $\rho(r)$ to give the energy of the system for which the electronic density is $\rho(r)$.

This energy $U[\rho]$ is not the ground state energy, ~~which~~ because that can only be found by varying $\rho(r)$, or in this approach, by varying $V(r)$. The correctio to $V(r)$ can be estimated by evaluating the functiona derivative $\frac{\delta U_{ee}[\rho]}{\delta\rho}$ where

$$U_{ee}[\rho] = -\sum_\alpha \int \rho(r)\frac{Z_\alpha}{|r-R_\alpha|}\,dr^3 + \frac{1}{2}\iint \frac{\rho(r)\rho(r')}{|r-r'|}\,dr^3 dr'^3 + U_{xc}[\rho].$$

The potential $V(r)$ is varied until $U[\rho]$ is minimized, wi is when $V(r) = \frac{\delta U_{xc}[\rho]}{\delta \rho}$. This iteration to self-consistently, is notoriously unstable because it is a loop in which some of the quantities are much more sensitive to errors than others. Small errors in a sensitive quantity produce big errors further around the loop.

$$\begin{tikzcd}
 & \{\Psi(r)\} \arrow[dl] & \\
\rho(r) \arrow[rr] & & V(r) \arrow[ul]
\end{tikzcd}$$

D. A General Functional

There are three quantities which are obviously important in the calculation of electronic energies. The first are the electronic wavefunctions $\{\Psi_\alpha(r)\}$, the second is the electronic potential $V(r)$, and the third is the electronic charge density $\rho(r)$. The relationships between these quantities are complicated and prone to error, so an ideal approach would be to treat them as independent variables on which the total energy depends, rather than as dependent on one another as in the self-consistent method.

Such a functional is the following

$$U_T[\Psi, V, \rho] = \sum_\alpha \int \Psi_\alpha(r)^* H \Psi_\alpha(r) dr^3 - \int \rho(r) V(r) dr^3 + U_{xc}[\rho] + \frac{1}{2} \sum_{\alpha\beta} \frac{Z_\alpha Z_\beta}{|R_\alpha - R_\beta|}.$$

We can see that this functional is stationary at the values of Ψ, V, and ρ which define the groundstate with respect to what ever exchange and correlation functional is used in $U_{xc}[\rho]$. Suppose Ψ differs from the ground state wavefunction by $\delta\Psi$. By the Hellman-Feynman Theorem, $\sum_\alpha \int (\Psi_\alpha(r) + \delta\Psi_\alpha)^* H (\Psi_\alpha + \delta\Psi_\alpha) dr^3$ has no linear variation with $\delta\Psi$ when ~~to~~ Ψ ~~are~~ is the ground state of the non interacting H. No other terms involve Ψ so there is no linear variation of $\delta\Psi$. Now vary $V(r)$ by $\delta V(r)$ from its ground state value.

$$\frac{\delta U_T}{\delta V} = \sum_\alpha \int \Psi_\alpha \delta V \Psi_\alpha dr^3 - \int \rho \delta V dr^3 = 0$$

at the value of $\rho = \sum_\alpha |\Psi_\alpha|^2$. Finally, vary ρ by $\delta\rho$, and again the $-\int \delta\rho V dr^3$ cancels the $\delta U_{el}[\rho]$ because $\delta U_{el}[\rho]/\delta\rho = \int V(r)\rho(r) dr^3$

The physical interpretation of the functional $U_T[\Psi, V, \rho]$ is that the first term or band structure term $\int \Psi_\alpha^* H \Psi_\alpha dr^3$ is the sum of the independent ~~pa~~ electron energies. This energy overestimates the interaction because it counts all the interactions of each particle independently. For pair forces this double counts the interactions, once for each ~~p~~ electron of every pair and so the interaction energies must be subtracted. $-\int \rho V dr^3$ subtracts double the interaction energies, but $U_{el}[\rho]$ adds the interaction energies back to give a net of subtracting them once. The ionic repulsion term should be added into the bandstructure term to cancel some of the electron-ion attraction. The stationarity of the functional results from the electrons finding a state which minimizes their total energy given the position of the ions. Because this is a minimum, a small ~~change~~ errors in the electronic state, $\delta\Psi$, δV, or $\delta\rho$ produces no first order change in the total energy.

There are two quantities whose variation does produce first order changes in the total energy. The first of these is the ionic positions, which is easily seen, produces a change in the total energy of $\int \delta V_{ion} \rho dr^3$. The second is $U_{xc}[\rho]$, the exchange and correlation functional. A change in this is like a change in the interaction potential between electrons and so is like a change in the ionic potential which contributes to the total energy in first order. The only reason that structures derived from density function theory are insensitive to the choice of U_{xc} is that being long range and ~~cost~~ slowly varying, so that the

~~structural energy~~ is insensitive to U_{xc} structure. The total energy does manifest first order variation with U_{xc} in that LDA values overbind by about 10%, ~~as~~ characteristic of a 1st order dependence

E Calculating Total Energies with U_T

Since the total energy is stationary with respect to the independent variation of Ψ, V, and ρ, the ~~s~~ calculation of the total energy is strictly a variational problem. One of the best ways of finding stationary points of functions of many variables is the Conjugate Gradient method in which particular degrees of freedom are varied for which others remain critical.

In approaching a variational solution of the total energy it is useful to know which variable have the biggest effect on the total energy. Intuition and experience both show that structural energies are most sensitive to the electronic kinetic energy which is determined by Ψ. Next to this, the band structure energy is important and depends on both Ψ and V making V the second most important variable. Finally the charge density ρ is the least important despite beliefs that charge transfer and polarizability have big effects on cohesive energies.

Recently M. Methfessel has ~~s~~ shown for ~~was~~ which variables U_T is maximal and for which minimal. For fixed V and ρ, U_T is clearly minimal at the ground state value of Ψ. For fixed ~~Ψ and~~ ρ Methfessel shows that U_T is maximal with respect to variation of V. ~~For~~ fixed Ψ and ~~ρ~~ V, U_T is then minimal with respect to ρ.

Acknowledgement Much of this work was done in collaboration with Matthew W. C. Foulkes and was published in Phys Rev. B 39 12520 (1989)

5. Dynamic Recursion

This fifth lecture is about how the recursion method can be used to calculate the projected density of states for systems with many more degrees of freedom than there are memory locations in the largest computer. The idea is that small components of states are neglected in favor of the largest components which can be stored in the computer. The surprising property of recursion is that the neglect of small components of states does not produce a cumulative error, but rather the error is proportional to the single largest neglected component.

A. Recursion Method

The recursion method is a way of tridiagonalizing large Hermitian matrices which represent the time reversal invariant equations of motion for some system. A basis of states for the system $\{\phi_0, \phi_1, \ldots, \phi_n, \ldots\}$ is chosen so that in terms of this basis, H is as sparse a matrix as possible. The basis is usually localized is space ~~sac like~~ as are atomic orbitals which can be used for solutions of the electronic equations of motion.

Recursion begins by tridiagonalizing H with the construction of a new basis starting with u_0 which I take equal to ϕ_0, $u_{-1} = 0$, and the subsequent $u_1, u_2, \ldots, u_n$, constructed to satisfy,

$$H u_n = a_n u_n + b_{n+1} u_{n+1} + b_n u_{n-1}$$

where $a_n = (u_n, H u_n)$, the inner product in which H is Hermitian, indicated by $(\,,\,)$. The normalizations $b_{n+1}^2 = ((H-a_n)u_n - b_n u_{n-1}, (H-a_n)u_n - b_n u_{n-1})$ together with the choice of a_n make the $\{u_n\}$ orthonormal to one another.

The projected density of states is the imaginary or singular part of the continued fraction,

$$R_{00}(E) = 1 \rfloor E - a_0 - b_1^2 \rfloor E - a_1 - b_2^2 \rfloor E - a_2 - \dots - b_n^2 \rfloor E - a_n - \dots$$

Physically this ~~system~~ continued fraction describes the different frequency components of the correlation within of the system after some time c ~~[illegible]~~

B Numerical Recursion

In addition to analytic calculation of $\{a_n\}$ and $\{b_n\}$, they can be determined numerically using floating point arithmetic on a computer. The $\{u_n\}$ are represented in the computer by their components ~~of~~ on the $\{\phi_n\}$. If the computer ~~has~~ can store N floating point numbers, then the recursion can be carried out within a basis of $N/2$ elements $\{\phi_0, \phi_1, \dots, \phi_{N/2}\}$. ~~Instead~~

In such calculations there are several sources of error. The most prominant is the truncation of the full equations of motion to an $N/2$ by $N/2$ matrix. The second source of error is is the finite precision of the floating point arithmetic. As the new u_n's are constructed, their components contain errors which are initially due to rounding in their ~~[illegible]~~ calculation. Subsequently, these errors are multiplied by H and so grow exponentially as the recursion continues.

Both the errors due to truncation of the basis and those due to rounding grow exponentially in the tridiagonal basis $u_0, u_1, \dots, u_n, \dots$

C. Dynamic Recursion

In the error analysis to follow, it will be shown that it is in fact advantageous to reduce the truncation errors at the expense of greater errors in the components of the $\{u_n\}$. The truncation of the basis ~~let~~ limits the degrees of freedom which can be explored by the calculation, whereas the errors in components seem much less dramatic.

The idea of dynamic recursion is to use a large or even infinite basis set $\{\phi_0, \phi_1, \ldots, \phi_n, \ldots\}$ but to approximate each u_n only by its $N/2$ largest components assuming that the computer can still store N floating point numbers. The above equation for u_{n+1} in terms of u_n and u_{n-1} is no longer satisfied, even to the precision of the arithmetic, but only to the accuracy of the $N/2$th largest component of u_{n+1}.

Implementation of this simple idea requires more sophisticated programming of the recursion. First, the basis $\{\phi_n\}$ must be indirectly addressed in the sense that both the component of ϕ_n and the index n of this basis element must be stored because, as the calculation progresses, ϕ_n may no longer have a significant component and the storage location may be needed for the component of a different basis element.

In dynamic recursion, the error due to basis truncation is converted into errors in the components of the u_n. The effects of this error is analyzed below.

D. Error Theory

Assuming that errors are made due to neglect of components as well as rounding, the recurrence for the tridiagonalization may be written as

$$Hu_n = a_n u_n + b_{n+1} u_{n+1} + b_n u_{n-1} + \delta_n$$

where $\{\delta_n\}$ are now the vectors of errors in each recurrence. It is clearer to write these recurrences in matrix notation as,

$$HU = UJ + \Delta$$

where H is an $M \times M$ matrix (M can be infinite) describing the equations of motion, U is an M by L matrix whose columns are the M - components of each u_n from u_0 to u_{L-1}, J is an L by L tridiagonal matrix whose diagonal elements are $a_0, a_1, \ldots, a_{L-1}$, and whose symmetric subdiagonal elements are $b_1, b_2, \ldots, b_{L-1}$, and Δ is the $M \times L$ matrix of errors whose columns are M components of δ_0 to δ_{L-1}.

Now define two resolvent matrices,

$$R(E) = (EI-H)^{-1}$$

an M by M matrix which is exact in the sense that it inverts the original equations of motion in the $\{\phi_n\}$ basis, and

$$G(E) = (EI-J)^{-1}$$

an L by L matrix which inverts the tridiagonal matrix. In terms of these two resolvents, the relation between the other matrices can be written as

$$R(E)\,U = U\,G(E) + R(E)\,\Delta\,G(E)$$

which is an analogue of Dyson's Equation.

The projected resolvent $R_0(E)$ whose imaginary part is the projected density of states, is the ϕ_0, u_0 matrix element of this equation, where we now assume that ~~[illegible]~~ $u_0 = \phi_0$. This may be written as

$$R_{00}(E) = \sum_{l=0}^{L-1} u_{l,0}\, G_{l,0}(E) + \sum_{l=0}^{L-1} \sum_{m=0}^{M-1} R_{0m}(E)\, \delta_{l,m}\, G_{l,0}(E)$$

where $u_{l,0}$ is the ϕ_0 component of u_l, ~~and~~ $\delta_{l,m}$ is the ϕ_m component of δ_l, $R_{0,m}(E)$ and $G_{l,0}(E)$ are the $\phi_0 \cdot \phi_m$ and $u_l - u_0$ elements of the two resolvents

This can be thought of as saying that the diffence between the exact $R_{00}(E)$ and the approximate $G_{00}(E)$ is a sum of other components $G_{l,0}(E)$, which can be made zero by making $u_{l,0} = 0$ for $l > 0$, and perturbations of the form $R_{0m}(E)\, \delta_{l,m}\, G_{l0}(E)$ which describe the propagation from ϕ_0 to ϕ_m, the error $\delta_{l,m}$ and the propagation from u_l back to u_0. This error dies away exponentially with the number of matrix elements of H needed to couple ϕ_0 to ϕ_m plus the number of matrix elements of J needed to couple u_l to u_0, which is l.

In order to see that these contributions decay exponentially with m and l, we consider two extreme cases. Suppose that there is a δ-distribution in the density of states of H at E and also of ~~&~~ J at E. A δ-distribution ~~describes~~ represents a localized state whose amplitude decreases exponentially with distance from its center. As a result, $G_{l,0}(E)$ and $R_{0m}(E)$ also decrease exponentially with when both l and m are beyond the centers of each δ-distribution in H and J.

If the density of states of H and J are both smooth at E, then the argument is more subtle ~~co~~ and depends on the averaging of sin kx and cos kx over a small range of x for large k. Smooth density of states at E means that the states of H and J near E are plane-wave like. Because of this, they vary like $\cos(m\sqrt{E}+\theta)$ and $\cos(l\sqrt{E}+\theta')$ which average to zero exponentially when m and l are large, and E is averaged over even a small range.

As a result of the exponentially small contribution to the total error from $\delta_{l,m}$ when l and m are large, the $\delta_{l,m}$ for small l and m dominate the sum and those for large l and m make no difference. Hence it is much more important to carry out the first few levels of iteration to as great an accuracy as possible, than to do the later stages accurately.

E Moment Theorems

The imaginary part of $R_{00}(E)$ for E on the real axis is equal to the projected density of states, and it is the quantity most closely associated with the physics of the recursion. A simple way to incorporate [illegible] of [illegible] which then introduce dynamic recursion in terms of the projected density of states by means of the moment theorems. The moments of the projected density of states are

$$\mu_n = \frac{1}{2\pi i}\oint_C R_{00}(E)\, E^n\, dE$$

where the contour C encloses the whole real E-axis. If an error is made in the calculation of T so that $G_{00}(E)$ differs from $R_{00}(E)$ in that its n^{th} moment is wrong, then the difference between the two projected densities of states is an oscillatory function of E with n+2 zeros counting the two zeros at $E = \pm\infty$ where the two ~~[illegible]~~ projected densities of states must both go to zero.

Returning back to the error theory, $\delta_{\ell m}$ produces an error in the moment whose index is ℓ plus the number of powers of the H needed to connect 0 to ℓm. Each $\delta_{\ell m}$ contributes an oscillatory function to the projected density of states. The exponential decay with increasing ℓ and m of the effect of $\delta_{\ell m}$ is due to either an exponential decay in the amplitude of these oscillatory corrections or the exponential decrease in the average of these oscillations over a small interval of E.

In physics [illegible] the decay of these oscillatory corrections is called the [illegible] theorem – the reason that [illegible] as a [illegible]. [illegible] characteristic of the [illegible] of the [illegible] [illegible] to [illegible] [illegible] that the projected density of ~~[illegible]~~ states [illegible] to distant parts of the system.

Acknowledgment This work was done in collaboration with [illegible] [illegible] 3 [illegible], [illegible]-[illegible].
All these [illegible] were written while enjoying the hospitality of the Cavendish Laboratory and Pembroke College in Cambridge. This work was supported by the [illegible] DMR.