The following note is based on the 2nd edition of *Computational Materials Science: An Introduction* by **June Gunn Lee**

▾ Chapter 6: Treating solids

In solids, the number atoms becomes infinite which is not possible to tackle. That's why we use *supercell* and *Periodic Boundary Condition (PBC)* to reduce the size of the system. Then we pseudize the wave function and the corresponding potential to conveniently express them which also eliminates a large number of electrons. We further reduce this by introducing k-point sampling. So that at the end of this chapter, we will be able to treat a solid by only handling a few k-points.

▾ *Pseudopotentials (PP):*

Concepts: We remove the core (nucleus + core electrons) from our calculation and only deal with valence electrons. This is possible because the core electrons are highly localized and only valence electrons are active in forming bonds, being ionized, conducting electricity, etc. This is called the *frozen-core approximation.*

Pseudizing procedure: This process refers to the generation of PP (via pseudo wave function). We generate the pseudo wave fucntion by removing the inner nodes of all electron (AE) wave functions. These nodes occurs when valence electron wave function passes by the core region because of orthogonality condition. However, because of *frozen-core approximation*, we can neglect any electron in the core. And that's why we can remove such nodes (no electron in the core means no wave function for which the valence wave function needs to be orthogonal, i.e., no nodes/wiggles/oscillation). The process is:

1. Calculate the exact AE potential, wave function, and energy by DFT using spherical symmetry. Separately perform a calculation for the core electrons.
2. Choose a core-valence boundary (r_c) and obtain $\psi_{PP}(r)$ by making $\psi_{AE}(r)$ smooth and nodeless whenever $r < r_c$
3. Ensure the following conditions:
$$\psi_{AE}(r)'|_{r=r_c} = \psi_{PP}(r)'|_{r=r_c}$$
$$\psi_{AE}(r)''|_{r=r_c} = \psi_{PP}(r)''|_{r=r_c}$$
4. $\psi_{AE}(r) = \psi_{PP}(r)'$ whenever $r > r_c$
5. Make the eigenvalues of the smooth ψ_{PP} and original ψ_{AE} the same
6. Generate a PP from ψ_{PP} and valence electron $\rho(r)$.

PP helps by reducing the number of planewaves(PW) needed for exapanding a wave function. It also eliminates the relativistic effect since core electrons are more prone to such effect. And also the errors involved are less than a couple of percentages. At this stage, we can rewrite the KS equation as:

$$\left[-\frac{1}{2}\nabla^2 + U_{PP}[\rho(r)] \right] \psi_i^{PP}(r) = \varepsilon_i \psi_i^{PP}(r)$$

$$\rho(r) = \sum_i |\psi_i^{PP}(r)|^2$$

Note, that this charge density and the wave functions are the pesudized values.

Various pseudopotential:

- *Norm-conserving PP:* Here, the pseudo and AE charge densities within the core are set to be equal. These PPs give us only the valence $\rho(r)$, not the total one.
  $\int_0^{r_c} |\psi_{PP}(r)|^2 \, dr = \int_0^{r_c} |\psi_{AE}(r)|^2 \, dr$

- *UltraSoft PP (USPP):* The ψ_{PP} is made very flat. These PPs need even less number of PW expansion.

- *Projector-Augmented Wave PP (PAW):* PAW approaches give us the AE charge density, unlike norm-conserving or ultrasoft PPs. To obtain the PAW wave function, we add the valence wavefunction with core wavefunction and then subtract the overlapping wave function:
  $\psi_{PAW} = \psi_{inter} + \psi_{core} - \psi_{net}$

▼ *Reducing calculation size:*

Supercell: By PP approach, we conveniently express the potential. Now, we will reduce the number of atoms to consider by using the idea of supercell under PBC consisting of several unit cells. The supercell is duplicated periodically throghout the space.

If there are non-periodic entities such as vacancies, we have to repeat the vacancies as well in all the supercells. We have to be careful that there is no interaction between these non-periodic entities by making the supercell sufficiently large. Also, the supercell should be charge neutral and of zero dipole moment. Otherwise a correction is necessary.

Reciprocal lattice: Mathematically, reciprocal lattice is the Fourier transform of real/direct/physical Bravais lattice. In direct lattice, all the points are physical location, x, y, z - of the atom/molecule/particle, but in reciprocal lattice, it is the spatial frequency, k_x, k_y, k_z. Reciprocal space is introduced as an abstract device since it is convenient to work with wave vectors in there.

For a direct lattice vector (in 3D), $\mathbf{R} = n_1 \mathbf{a_1} + n_2 \mathbf{a_2} + n_3 \mathbf{a_3}$, the corresponding reciprocal lattice vector is given as: $\mathbf{G} = m_1 \mathbf{b_1} + m_2 \mathbf{b_2} + m_3 \mathbf{b_3}$, where all the n_i and m_j are integers. Some of the key mathematical formulas:

$$f(\mathbf{r}) = f(\mathbf{r} + \mathbf{R}) \;\; ; \;\; f(\mathbf{G}) = \sum_{\mathbf{G}} c(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{r}}$$

$$\mathbf{b_1} = 2\pi \frac{a_2 \times a_3}{V} \;\; ; \;\; \mathbf{b_2} = 2\pi \frac{a_3 \times a_1}{V} \;\; ; \;\; \mathbf{b_3} = 2\pi \frac{a_1 \times a_2}{V}$$

where $V = |a_1 \cdot a_2 \times a_3|$, the volume of the direct lattice. Also, the unit vectors satisfy the *bi-orthogonality* property: $a_i \cdot b_j = 2\pi \delta_{ij}$

Note that the reciprocal unit vectors have the unit of $1/length$, so by increasing the volume of the real suppercell, we can decrease the volume of the reciprocal supercell. The relationship between real and reciprocal lattices are not straightforward. A BCC (FCC) direct lattice becomes FCC (BCC) in reciprocal space, while SC remains SC. The volume of the reciprocal lattice, $\Omega = \frac{(2\pi)^3}{V}$.

First Brillouin zone (BZ) and IBZ: The first Brillouin zone is defined as the volume inside reciprocal space containing k vectors whose distance from $\mathbf{G} = 0$ is smaller than the distance from any other G-vector. It is a primitive cell analogue to real space. In 2D:

1. Draw a reciprocal lattice first.
2. Draw lines from any reference point to its all nearest neighbors and bisect the lines perpendicularly.
3. The formed square (polygon in 3D) with the bisecting lines is the first BZ.

There can be other BZ at further distance but since they are periodic in nature, we only need to be concerned with the first BZ. Solution of KS equations can be written inside the first BZ. Also, note that there are some high symmetry points in the first BZ that we use most often such as Γ (center point of the BZ), H (corner point), etc.

The first BZ can be further reduced to irreducible BZ (IBZ) by exploiting their full symmetry (rotation and inversion). Any point in IBZ can represent a k-point.

k-**points:** The wave function and other properties vary very smoothly inside the IBZ. So, we can sample a few k- points and evaluate our expected properties only in these points. The wave vector contains rich information such as wave length ($\lambda = 2\pi/k$), kinetic energy ($E = k^2/2$), etc. If we plot all the incoming energies on each k-point, it will give us the band structure.

Different k-points in IBZ have different weight factor which comes from the fact of how many times that particular k-point has been folded during the first BZ to IBZ transformation. The best single k-point is the Γ-point ($k = 0$) which is sufficient to describe any property in IBZ for a big supercell.

Lastly, the Monkhorst-Pack (MP) method generates evenly spaced k-points in IBZ. In QE, the (automatic) options set k-points sampling method as MP method.

▾ *Bloch theorem:*

In repeatedly patterned solids, any quantity that depends on r is also periodic such as potential or electron density: $U(r) = U(r + R)$, $\rho(r) = \rho(r + R)$, where R is the real lattice vector. The wave function is also periodic but with a phase factor. Now, the valence electrons wave functions have the plane wave form of:

$$\psi(r) = C \, \exp(i\mathbf{k} \cdot \mathbf{r})$$

However, under a periodic potentials, Bloch says that:

$$\psi_k(r) = u_k(r) \exp(i\mathbf{k} \cdot \mathbf{r})$$

where $u_k(r)$ is a periodic function with the same periodicity as the potential. Hence:

$$\psi_k(r + \mathbf{R}) = u_k(r + \mathbf{R}) \exp[(i\mathbf{k} \cdot (\mathbf{r} + \mathbf{R})] = \psi_k(r) \exp(i\mathbf{k} \cdot \mathbf{R})$$

Therefore, the solution of the KS equation can be represented as a plane wave differs from the free electron PW only by a periodic modulation. This theorem enables us wokring only with a small piece of solid instead of the whole solid.

▾ *Plane wave expansion:*

Concept: The KS orbitals could take any shape. But we can approximate them as a linear combination of known simple functions such as PW. This is called the PW basis set. Three types of basis set are used commonly: local, nonlocal, and augmented. Gaussian basis set is local, while PW basis set is nonlocal. Gaussian basis set is used for atoms and molecules with highly localized electron.

PW basis: PW basis set is the most common basis set in DFT codes. They have several advantages. A few of them are:

- PWs are solutions for a free electron satisfying Bloch condition, and the basis set can represent any smoothly function very well.
- It has the form: $\phi_{pw}(\mathbf{r}) = C \exp(i\mathbf{K} \cdot \mathbf{r})$. So, taking derivative reduces to taking simple product. In DFT, taking gradient is very common.
- A systematic improvement of *convergence* and *accuracy* is possible by increasing the size of the basis set.
- The number of PWs needed for expansion is reduced by PP approach and a much smaller E_{cut} can be used (discussed later).

However, PWs cover all space equally. So, the same number of basis set is needed even for vacuum. Also, more than a hundred PWs are needed per atom.

PW expansion for charge density: Any observable can be calculated for KS quantities by plugging in the expanded PW expression for the KS orbitals and then summing over the sampled k-points within the IBZ. The charge density in terms of $\mathbf{G}$ vectors is:

$$\rho(\mathbf{r}) = \sum_{\mathbf{G}} \rho(\mathbf{G}) \exp(i\mathbf{G} \cdot \mathbf{r})$$

Here, $\rho(\mathbf{G})$ is the Fourier coefficient for charg density.

Band structure: In solid, electrons form bands. A band structure diagram means the energy vs wave vector diagram. For free electrons, the energy curves are continuous and any state with k beyond the first BZ is the same to a state inside the first BZ with a different band index n. The presentation that shows only the first BZ, i.e., $k = -\pi/a$ to π/a, is called the reduced presentation.

In solid, the case is very similar to the free electron case except when electrons with $k = n\pi/a$ meets at the BZ boundaries. Then these electrons get diffracted. They get perturbed by lattice potential, reflected back and become standing waves. These will contribute differently to charge density.

Density of States: The density of states (DOS) is the number of electronic states per unit energy range. For example, for a free-electron in one-dimension, the energy, $\varepsilon_k = \frac{\hbar^2}{2m}k^2$ and the total number of states, $n(\varepsilon) = \frac{V}{3\pi^2}k_F^3 = \frac{V}{3\pi^2}\frac{2m\varepsilon}{\hbar^2}^{3/2}$. So,

$$D(\varepsilon) = \frac{dn}{d\varepsilon} = \frac{V}{2\pi^2}\frac{2m}{\hbar^2}^{3/2}\varepsilon^{1/2} \propto \sqrt{\varepsilon}$$

Also,

$$\int_0^{E_F} D(\varepsilon)d\varepsilon = n$$

where E_F is the Fermi energy. At finite temperature, some electrons will be excited and will cross the Fermi line in DOS. In actual solid, both the DOS and band structure will be much complicated but have similar features

▼ *Some practical topics:*

Cutoff energy: In solid, the unbound free electrons occupy the higher energy levels. To us, the lower energy electrons are more important because we are aiming for ground-state energy. We thus cut the energy at some upper limit, i.e., $E_{cut} \geq \frac{1}{2}(k + G)^2$.

We only include PWs with kinetic energy smaller than this cutoff energy for calculation. But also we have to be careful that E_{cut} is high enough to give accurate result. Therefore, it changes the wave function expression by modifying the summation range:

$$\psi_k(r) = \sum_{|k+G|_{cut}} c_{k+G} \exp[i(\mathbf{k} + \mathbf{G}) \cdot \mathbf{r}]$$

k-vectors become finite by k-point sampling in IBZ and \mathbf{G} vectors become finite by energy cutoff.

smearing: In metals, charge density occupation is not as smooth as insulators or semiconductors. The sharp drop of occupation makes the resolution with PW expansions very difficult. That's why we make this drop smooth and this process of replacing the delta function with a smooth function is called smearing.

Smearing can be thought of as creating a fictitious electronic temperature in the system. After the calculation is done, the artificially introduces temperature is brought back to 0K by extrapolation. A proper balance is required in smearing. Too much smearing will cause a wrong total energy, whereas a too small smearing will cause computational inefficiency.

There are different types of smearing such as Gaussian smearing (Gaussian type delta function), Fermi smearing (Fermi–Dirac distribution function), Methfessel-Paxton smearing (a single parameter σ, a step function expanded into polynomials) etc.

Ionic minimizations: In DFT, we seek not only an electronically relaxed system, but also ionically. Thus in each self-consistent loop, there is a outer loop that precedes electronic minimization, where we seek relaxed ionic position and/or unit cell shape/size. There are three common methods: the quasi-Newton method, the conjugate gradient (CG) method, and the damped MD method.

Hellmann-Feynman theorem states that, if an exact $\hat{H}$ and the corresponding ϕ_i are calculated, the force on an atom is the expectation value of the partial derivative of $\hat{H}$ with respect to atomic position r_I. So,

$$F_I = -\frac{dE}{dr_I} = -\left\langle \phi_i \left| \frac{\partial \hat{H}}{\partial r_I} \right| \phi_i \right\rangle = -\frac{\partial U_{IJ}}{\partial r_I} - \int \frac{\partial U_{ext}}{\partial r_I} \rho(r) \, dr$$

So, after an electronic SC loop, the force can be calculated by above given formula. Based on this calculation, we will direct which way the atom should move (downward direction of force). If the basis set depends on atomic position, the Hellmann-Feynman equation involves another term called Pulay forces. But PW approach has no such term since it's completely nonlocal.