

Lecture Notes (**draft**):
Electronic Structure of Solids
(SISSA-Trieste, 2002)

Paolo Giannozzi
Scuola Normale Superiore, Piazza dei Cavalieri 7
I-56126 Pisa, Italy

Hellmann-Feynman forces

Hellmann-Feynman forces are the derivative of the total energy with respect to atomic positions \mathbf{R}_i . For many-body Hamiltonians and wavefunctions, only terms containing *explicit* derivatives in the Hamiltonian contribute (Hellmann-Feynman theorem). The terms containing *implicit* derivatives through the wavefunctions, that we indicate with $\tilde{\mathbf{F}}_i$, vanish:

$$\mathbf{F}_i = -\frac{d}{d\mathbf{R}_i}\langle\Psi|H|\Psi\rangle = -\langle\Psi|\frac{\partial H}{\partial\mathbf{R}_i}|\Psi\rangle - \tilde{\mathbf{F}}_i \quad (1)$$

with

$$\tilde{\mathbf{F}}_i = \langle\frac{d\Psi}{d\mathbf{R}_i}|H|\Psi\rangle + \langle\Psi|H|\frac{d\Psi}{d\mathbf{R}_i}\rangle = E\langle\frac{d\Psi}{d\mathbf{R}_i}|\Psi\rangle + E\langle\Psi|\frac{d\Psi}{d\mathbf{R}_i}\rangle = E\frac{d}{d\mathbf{R}_i}\langle\Psi|\Psi\rangle. \quad (2)$$

The last term vanish because it is the derivative of a constant quantity. Note that partial derivative are used to indicate explicit derivation, otherwise the total derivative is used.

In DFT the same applies, thanks to the variational character of the energy. Let us write the force as

$$\mathbf{F}_i = -\frac{dE}{d\mathbf{R}_i} = -\int n(\mathbf{r})\frac{\partial V(\mathbf{r})}{\partial\mathbf{R}_i}d\mathbf{r} - \frac{\partial E_{II}}{\partial\mathbf{R}_i} - \tilde{\mathbf{F}}_i \quad (3)$$

where the first term comes from explicit derivation of the energy functional, E_{II} is the ion-ion (classical) interaction energy, and the $\tilde{\mathbf{F}}_i$ contains the implicit derivation through Kohn-Sham (KS) orbitals:

$$\tilde{\mathbf{F}}_i = \sum_i \int \left(\frac{d\psi_i^*(\mathbf{r})}{d\mathbf{R}_i} \frac{\delta E}{\delta\psi_i^*(\mathbf{r})} + \frac{d\psi_i(\mathbf{r})}{d\mathbf{R}_i} \frac{\delta E}{\delta\psi_i(\mathbf{r})} \right) d\mathbf{r}. \quad (4)$$

Using the expression for the functional derivative of the energy functional and the identity

$$0 = \frac{d}{d\mathbf{R}_i} \int n(\mathbf{r})d\mathbf{r} = \sum_i \left(\int \frac{d\psi_i^*(\mathbf{r})}{d\mathbf{R}_i} \psi_i(\mathbf{r})d\mathbf{r} + \int \psi_i^*(\mathbf{r}) \frac{d\psi_i(\mathbf{r})}{d\mathbf{R}_i} d\mathbf{r} \right), \quad (5)$$

the term $\tilde{\mathbf{F}}_i$ can be recast as

$$\tilde{\mathbf{F}}_i = \sum_i \int \left(\frac{d\psi_i^*(\mathbf{r})}{d\mathbf{R}_i} (H - \epsilon_i) \psi_i(\mathbf{r}) + \frac{d\psi_i(\mathbf{r})}{d\mathbf{R}_i} (H - \epsilon_i) \psi_i^*(\mathbf{r}) \right) d\mathbf{r}. \quad (6)$$

This term vanishes on the ground state. Finally, one finds that, in perfect analogy to the many-body case, the forces acting on atoms are the matrix element on the ground state of the gradient of the external potential plus an ion-ion term:

$$\mathbf{F}_i = -\int n(\mathbf{r})\frac{\partial V(\mathbf{r})}{\partial\mathbf{R}_i}d\mathbf{r} - \frac{\partial E_{II}}{\partial\mathbf{R}_i}. \quad (7)$$

Pulay forces

Unfortunately Eq.6 vanishes only if we have ground state charge density and wavefunctions at perfect convergence. In the real world, this is never the case. In particular, the wavefunctions are expanded on a finite basis set that is never complete. This may produce a nonzero value of Eq.6, called *Pulay force*.

Let us write the expansion of wavefunctions into a basis set, taken to be orthonormal for simplicity:

$$\psi_i(\mathbf{r}) = \sum_n c_n^{(i)} \phi_n(\mathbf{r}). \quad (8)$$

This will yield a secular equation

$$\sum_m (H_{nm} - \epsilon_i) c_m^{(i)} = 0, \quad H_{nm} = \int \phi_n^*(\mathbf{r}) H \phi_m(\mathbf{r}) d\mathbf{r}. \quad (9)$$

By inserting the expansion of the KS orbitals into Eq.6 one finds

$$\tilde{\mathbf{F}}_i = \sum_i \sum_{mn} \frac{\partial c_n^{(i)}}{\partial \mathbf{R}_i} (H_{nm} - \epsilon_i) c_m^{(i)} + \sum_{mn} c_n^{(i)} c_m^{(i)} \int \frac{\partial \phi_n^*(\mathbf{r})}{\partial \mathbf{R}_i} (H - \epsilon_i) \phi_m(\mathbf{r}) d\mathbf{r} + c.c. \quad (10)$$

At perfect self-consistency (which is never achieved in practical calculations, though) the first term vanishes exactly even if the basis set is not complete (see Eq.9). The second term instead vanishes only i) if the basis set is complete, or ii) if $\partial \phi_n^*(\mathbf{r})/\partial \mathbf{R}_i$ has no component outside the subspace defined by the $\phi_n(\mathbf{r})$, or iii) if the basis set does not depend explicitly on the atomic positions. The latter is the case of Plane Waves (PW). Pulay forces do not arise just because the basis set is incomplete, but because it is “incomplete in a different way” when atoms are moved. Using PW one has also an incomplete basis set, but it is “equally incomplete” for all atomic positions in the unit cell.

In practical calculations with localized basis sets, Pulay forces must be taken into account, otherwise the error on the forces is quite large. If one wants to minimize the energy, or to do molecular dynamics simulations, it is crucial that the forces are the derivative of the energy with high accuracy. Although much progress has been done in the last years towards reliable calculation of forces with localized basis sets, PW are still much more used than localized basis sets for all applications in which forces are important.

Structural optimization

In the following we assume that ions are classical objects and that the Born-Oppenheimer approximation holds.

At zero temperature the equilibrium atomic positions \mathbf{R}_i , $i = 1, \dots, N$ are determined by the minimum of the *total energy* E_{tot} of the system, that is, the sum of the electronic (DFT) energy E and of the ion-ion interaction (electrostatic) energy E_{II} . If we consider the electrons in their ground state for any given configuration of \mathbf{R}_i (collectively indicated by $\{\mathbf{R}\}$), the total energy will be a function of the atomic positions:

$$E_{tot}(\{\mathbf{R}\}) = E(\{\mathbf{R}\}) + E_{II}(\{\mathbf{R}\}). \quad (11)$$

The procedure to find the atomic configuration yielding the minimum energy is usually called *structural optimization* or *relaxation*.

For an infinite system we must distinguish between atomic displacements that change the form and volume of the unit cell (related to *elastic* modes) and atomic displacements internal to the unit cell (related to *phonon* modes). Such distinction does not exist for a finite system. The optimization of the lattice and that of atomic positions have to be done separately, or in any case, using different procedures.

In the following we assume that the unit cell is fixed. Our problem reduces to finding the minimum of a function of $3N$ variables \mathbf{R}_i , where N is the number of atoms in the unit cell. The gradients of this function with respect to the variables are easily calculated (at least for a PW basis set): these are the Hellmann-Feynman forces.

Several well-known and well-studied algorithms exist for finding the minimum of a function, knowing its gradients: conjugate gradient, quasi-Newton methods, DIIS (Direct Inversion in Iterative Space). All these algorithms are *iterative*: they are repeated until the precision of the solution is judged satisfactory, but there is no way in general to know in advance how many iterations will be needed.

Note that the potential energy surface (PES) defined by Eq.(11) may have many *local* minima separated by barriers. Minimization algorithms will only find the closest local minimum, with no warranty that it is a global minimum. One will need to try at least a few starting points, corresponding to different candidate structures, before starting to trust the results. Another feature (sometimes a limitation, sometimes an advantage) of such algorithms is that the starting symmetry of the system (if any) will not be broken.

It is instructive to examine in some detail one of these algorithms: the conjugate gradient method.

Conjugate Gradient

In the following let us consider a function $f(\mathbf{x})$ of the variables $\mathbf{x} \equiv (x_1, \dots, x_{3N})$ and its gradients $\mathbf{g}(\mathbf{x}) = -\nabla_{\mathbf{x}} f(\mathbf{x})$.

The first obvious minimization algorithm that comes to mind is *steepest descent* (SD). This consists in minimizing $f(\mathbf{x})$ along the direction $\mathbf{g}(\mathbf{x})$. Once the minimum along such direction is reached, the gradient is recalculated, a new minimum is sought along the new direction of the gradient, and so on.

SD is a prototypical *direction set* method: the gradient is eliminated one component at the time along a set of directions. In SD every direction is orthogonal to the previous one (by construction). SD is not bad far from convergence, but it becomes very bad very quickly. A reason for bad convergence is that the set of directions in SD is not optimal. Let us consider such aspect in more detail in the following.

The function in the region not far from the minimum is approximately quadratic:

$$f(\mathbf{x}) \simeq \frac{1}{2} \mathbf{x} \cdot \mathbf{A} \cdot \mathbf{x} - \mathbf{b} \cdot \mathbf{x} + f_0, \quad \mathbf{g}(\mathbf{x}) = -\mathbf{A} \cdot \mathbf{x} + \mathbf{b} \quad (12)$$

where \mathbf{A} is a matrix, \mathbf{b} is a vector (not necessarily known).

An optimal set of directions should ensure that when we search for a minimum along the new direction, we do not lose what we have gained in the preceding step. Let us assume that at step n we reached the minimum along line \mathbf{h}_n . This implies: $\mathbf{g}(\mathbf{x}_n) \cdot \mathbf{h}_n = 0$. We move from \mathbf{x}_n along direction \mathbf{h}_{n+1} . The gradient change $\delta \mathbf{g}$ is proportional to $\mathbf{A} \cdot \mathbf{h}_{n+1}$. If we impose that this change has no component along all previous minimization directions \mathbf{h}_n , we get the condition

$$\mathbf{h}_n \cdot \mathbf{A} \cdot \mathbf{h}_m = 0 \quad (13)$$

that defines *conjugate* directions. The simpler *conjugate gradient* (CG) algorithm is as follows:

1. start by minimizing along $\mathbf{h}_0 = \mathbf{g}_0 = -\nabla f(\mathbf{x}_0)$. If the function is quadratic, the minimum can be found analytically: $\mathbf{x}_1 = \mathbf{x}_0 + \lambda_0 \mathbf{h}_0$, where $\lambda_0 = -\mathbf{h}_0 \cdot \mathbf{g}_0 / \mathbf{h}_0 \cdot \mathbf{A} \cdot \mathbf{h}_0$.
2. find the next direction $\mathbf{h}_1 = \mathbf{g}_1 + \gamma_1 \mathbf{h}_0$ and impose that it is conjugate to the preceding one, Eq.13. One finds $\gamma_1 = \mathbf{g}_1 \cdot \mathbf{g}_0 / \mathbf{g}_0 \cdot \mathbf{g}_0$.
3. iterate the procedure until the desired convergence. The sequence of gradients \mathbf{g}_n and of conjugate gradients \mathbf{h}_n is found to obey $\mathbf{g}_n \cdot \mathbf{g}_m = 0$, $\mathbf{g}_n \cdot \mathbf{h}_m = 0$, and Eq.13, for all n, m .

If the problem is not quadratic, so that \mathbf{A} is not a priori known, the algorithm remain the same, but the analytical determination of the line minimum in step 1) is not performed. A numerical minimization along the \mathbf{h} directions is performed and the gradient \mathbf{g} is calculated at the line minimum.

CG converges much better than SD, with negligible additional effort. If the problem is purely quadratic, *exact* convergence is guaranteed to be reached in $3N$ steps. This would take $\mathcal{O}[(3N)^3]$ operations, not better than the inversion of \mathbf{A} . *Approximate* convergence, however, can be reached in a much smaller number of steps. Moreover CG can be applied in presence of large matrices \mathbf{A} for which inversion is impractical or impossible. Only the results of operator \mathbf{A} on trial vectors $\mathbf{A} \cdot \mathbf{x}$ are required, not the entire operator.

In general, the rate of convergence is determined by the ratio between the largest and smallest eigenvalues of the matrix \mathbf{A} : the closer to 1, the better. Since in real-life example such ratio may considerably differ from 1, a technique known as *preconditioning* is often used to produce an equivalent problem for which such ratio is closer to 1, thus yielding better convergence properties.

The CG method, in many variants, is much used not only for structural optimization but also as an alternative method to self-consistency for finding the minimum of the energy functional at fixed ions (“electronic” minimization). In this case the variables \mathbf{x} are the expansion coefficients of KS orbitals into the PW or any other basis set. The algorithm becomes slightly more complicated because orthonormality constraints between orbitals must be taken into account.

Other minimization methods

The CG method does not use explicitly the second derivative matrix \mathbf{A} or its inverse \mathbf{A}^{-1} . This is an advantage if the number of variables in the problem is large (as i.e. in the electronic minimization problem mentioned above): the storage of an approximate \mathbf{A} or \mathbf{A}^{-1} would be prohibitely large. For structural optimization however the number of variables never exceeds a few hundreds. Moreover it is conceivable to find with little effort reasonable approximations to \mathbf{A} , which is related to the force constant matrix. *Quasi-Newton* methods make use of some guess for \mathbf{A} and produce an iterative refinement of \mathbf{A} (or of \mathbf{A}^{-1}) using the available information.

DFT and Molecular Dynamics

Let us assume also in this section that ions behave as classical particles and that the BO approximation holds. Under these assumptions the dynamical behavior of ions can be described by a classical Lagrangian

$$L = \frac{1}{2} \sum_i M_i \dot{\mathbf{R}}_i^2 - E_{tot}(\{\mathbf{R}\}) \quad (14)$$

where M_i are the mass of ions. The electrons are assumed to stay on the *BO surface*, that is, in the ground state corresponding to their instantaneous positions. The corresponding equations of motion:

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{\mathbf{R}}_i} - \frac{\partial L}{\partial \mathbf{R}_i} = 0, \quad \mathbf{P}_i = \frac{\partial L}{\partial \dot{\mathbf{R}}_i} \quad (15)$$

are nothing but Newton's equations.

It is tempting to use Eq.14 as the basis for a *molecular dynamics* (MD) study. In classical MD, the forces are generated by an interatomic potential (often a sum of two-body terms like Lennard-Jones potentials) and the Newton equations are *discretized* and numerically solved. The discrete interval of time is called *time step*. A sequence of atomic coordinates and velocities is generated starting from a suitable initial set of coordinates and velocities. The sequence can be used to calculate thermodynamical averages. Straightforward MD will sample the microcanonical ensemble: constant energy at fixed volume, but it is possible to build a dynamics at constant temperature (canonical ensemble) using a *Nosé thermostat* that simulates a thermal bath, or at constant pressure, by adding a fictitious dynamics on the volume, and even more complex cases.

MD can also be used to find the global minima using the *simulated annealing* technique. The configuration space is sampled at equilibrium, then the kinetic energy is gradually removed from the systems that has the possibility (but is not guaranteed to do so) to reach the global minimum. Such procedure is sometimes the only practical way to find the global minimum for especially hard problems. In mathematical terms, "easy" problems are exactly solvable by computer algorithms in polynomial time, that is, in a number of steps that is a polynomial function of the dimension N of the problem; "hard" problems are solved in exponential time. A problem is NP (nondeterministic polynomial) if its solution (if one exists) can be guessed and verified in polynomial time. This is the kind of problems for which the simulated annealing has been devised. The determination of the structure in clusters is believed to be a NP-hard problem.

Classical Molecular Dynamics

Let us consider the most basic MD : a purely mechanical system of N atoms, enclosed in a volume V (usually with periodical boundary conditions, PBC, for a condensed-matter system), having mechanical energy $E = T + E_p$, where $T = \frac{1}{2} \sum_i M_i \dot{\mathbf{R}}_i^2$ is the kinetic energy of ions, $E_p = E_p(\{\mathbf{R}\})$ is the interatomic potential energy. This is known as the *NVE*, or microcanonical, ensemble.

Discretization of the equation of motion

The numerical solution (integration) of the equations of motions is generally performed using the *Verlet algorithm*. This is obtained from the following basic and very simple equations :

$$\mathbf{R}_i(t + \delta t) = \mathbf{R}_i(t) + \delta t \mathbf{V}_i(t) + \frac{\delta t^2}{2M_i} \mathbf{f}_i(t) + \frac{\delta t^3}{6} \mathbf{b}_i(t) + \mathcal{O}(\delta t^4) \quad (16)$$

$$\mathbf{R}_i(t - \delta t) = \mathbf{R}_i(t) - \delta t \mathbf{V}_i(t) + \frac{\delta t^2}{2M_i} \mathbf{f}_i(t) - \frac{\delta t^3}{6} \mathbf{b}_i(t) + \mathcal{O}(\delta t^4) \quad (17)$$

where $\mathbf{V}_i = \dot{\mathbf{R}}_i$ are velocities, \mathbf{f}_i forces acting on ion i . By summing and subtracting Eqs. (16) and (17) we get the Verlet algorithm:

$$\mathbf{R}_i(t + \delta t) = 2\mathbf{R}_i(t) - \mathbf{R}_i(t - \delta t) + \frac{\delta t^2}{M_i} \mathbf{f}_i(t) + \mathcal{O}(\delta t^4) \quad (18)$$

$$\mathbf{V}_i(t) = \frac{1}{2\delta t} [\mathbf{R}_i(t + \delta t) - \mathbf{R}_i(t - \delta t)] + \mathcal{O}(\delta t^3). \quad (19)$$

The velocities are one step behind the positions, but they are not used to update the positions. It is possible to recast the Verlet algorithm into an equivalent form (one giving exactly the same trajectories) in which both velocities and positions are updated in the same step. By combining Eq.(16) with Eq.(17) displaced in time by $+\delta t$, one finds

$$\mathbf{V}_i(t + \delta t) = \mathbf{V}_i(t) + \frac{\delta t}{2M_i} [\mathbf{f}_i(t) + \mathbf{f}_i(t + \delta t)] \quad (20)$$

$$\mathbf{R}_i(t + \delta t) = \mathbf{R}_i(t) + \delta t \mathbf{V}_i(t) + \frac{\delta t^2}{2M_i} \mathbf{f}_i(t). \quad (21)$$

Note that the update of velocities requires the forces for the new positions. This algorithm is known as *Velocity Verlet*. Its equivalence to the Verlet algorithm may not seem evident, but it can be proved quite simply.

In spite of his simplicity, the Verlet algorithm, in any incarnation, is efficient and numerically stable. In particular, it yields trajectories that conserve to a very good degree of accuracy the energy E . A small loss of energy conservation, due both to numerical errors and to the discretization, is unavoidable, but a systematic drift of the energy is not acceptable. In this respect Verlet is superior to apparently better (i.e. higher-order) schemes. In one of the following sections we will see one reason why this happen.

Thermodynamical averages

In the following we will use the *phase space* canonical variables, collectively indicated as \mathbf{R}, \mathbf{P} , instead of coordinates and velocities. From a practical point of view, the calculation of thermodynamical averages in classical MD is an average over many time steps:

$$A_T = \frac{1}{T} \int_0^T A(\mathbf{R}(t), \mathbf{P}(t)) dt \simeq \frac{1}{M} \sum_{n=1}^M A(t_n), \quad t_n = n\delta t, \quad t_M = M\delta t = T. \quad (22)$$

For an ergodic system (that is, one whose trajectories in a sufficiently long time pass arbitrary close to any point in the phase space), it is believed that:

$$\lim_{T \rightarrow \infty} A_T \rightarrow \langle A \rangle \quad (23)$$

where $\langle \rangle$ is the average over the corresponding ensemble:

$$\langle A \rangle = \int \rho(\mathbf{R}, \mathbf{P}) A(\mathbf{R}, \mathbf{P}) d\mathbf{R} d\mathbf{P} \quad (24)$$

where ρ is the probability of a microscopic state. In NVE MD the microcanonical ensemble is sampled:

$$\rho_{NVE}(\mathbf{R}, \mathbf{P}) = \frac{f(N)}{\Omega} \delta(H - E) \quad (25)$$

where H is the Hamiltonian corresponding to the Lagrangian of Eq.(14), E is the mechanical energy (including kinetic energy of ions) $f(N) = (h^{3N} N!)^{-1}$ for N indistinguishable atoms, and Ω , related to the entropy S by the Boltzmann relation $S = k_B \log \Omega$, is the total number of microscopic states:

$$\Omega = f(N) \int d\mathbf{R} d\mathbf{P} \delta(H - E). \quad (26)$$

The time step must be as big as possible in order to sample as much phase space as possible, but at the same time it must be small enough to allow to follow the motion the ions with little loss of accuracy (which usually appears as a drift in the energy). Typically $\delta t \sim 0.01 - 0.1 \delta t_{max}$, where δt_{max} is the period of the fastest phonon mode: $\delta t_{max} = 1/\omega_{max}$.

Verlet algorithm as unitary discretization of the Liouvillian

In the following we will use the *phase space* canonical variables, collectively indicated as \mathbf{R}, \mathbf{P} , instead of coordinates and velocities. Let us consider an observable $A = A(\mathbf{R}, \mathbf{P}, t)$. Its time evolution can be written as

$$\frac{dA}{dt} = \sum_i \left(\dot{\mathbf{R}}_i \frac{\partial A}{\partial \mathbf{R}_i} + \dot{\mathbf{P}}_i \frac{\partial A}{\partial \mathbf{P}_i} + \frac{\partial A}{\partial t} \right) = i\mathcal{L}A + \frac{\partial A}{\partial t} \quad (27)$$

where the operator \mathcal{L} is called the *Liouvillian*. Assuming that $A = A(\mathbf{R}, \mathbf{P})$ does not depend explicitly on the time, the Liouvillian determines entirely the time evolution of A : formally,

$$A(t) = e^{i\mathcal{L}t} A(t=0) = U(t)A(t=0). \quad (28)$$

It can be shown that \mathcal{L} is an Hermitian operator and thus U is a unitary operator (as it should be: time-reversal symmetry must hold). We can write \mathcal{L} as

$$i\mathcal{L} = \sum_i \left(\dot{\mathbf{R}}_i \frac{\partial}{\partial \mathbf{R}_i} + \dot{\mathbf{P}}_i \frac{\partial}{\partial \mathbf{P}_i} \right) = \sum_i \left(\dot{\mathbf{R}}_i \frac{\partial}{\partial \mathbf{R}_i} + \mathbf{f}_i \frac{\partial}{\partial \mathbf{P}_i} \right) \quad (29)$$

and finally as a sum of two terms, one acting on coordinates and one on momenta: $i\mathcal{L} = i\mathcal{L}_p + i\mathcal{L}_r$, where

$$i\mathcal{L}_p = \sum_i \mathbf{f}_i \frac{\partial}{\partial \mathbf{P}_i}, \quad i\mathcal{L}_r = \sum_i \dot{\mathbf{R}}_i \frac{\partial}{\partial \mathbf{R}_i}. \quad (30)$$

Until now, we have just recast the classical equation of motion into an elegant but not especially useful formalism. Let us discretize now the time evolution operator, by dividing t into N small intervals $\delta t = t/N$, and apply the *Trotter* approximation:

$$e^{i(\mathcal{L}_p + \mathcal{L}_r)t} = \left[e^{i(\mathcal{L}_p + \mathcal{L}_r)\delta t} \right]^N = \left[e^{i\mathcal{L}_p\delta t/2} e^{i\mathcal{L}_r\delta t} e^{i\mathcal{L}_p\delta t/2} + \mathcal{O}(\delta t^3) \right]^N. \quad (31)$$

Remember that \mathcal{L}_p and \mathcal{L}_r are operators: the Trotter approximation is not trivial. Let us apply the operator between square brackets to a point $(\mathbf{R}_i(t), \mathbf{P}_i(t))$ in phase space at time t . We will use the known result

$$e^{a\partial/\partial x} f(x) = f(x+a) \quad (32)$$

if a does not depend on x . Since \mathcal{L}_p and \mathcal{L}_r are sums of terms acting on each particle separately, we can consider their action on each particle independently.

$$e^{i\mathcal{L}_p\delta t/2} (\mathbf{R}_i, \mathbf{P}_i) = \left(\mathbf{R}_i, \mathbf{P}_i + \frac{\delta t}{2} \mathbf{f}_i(\mathbf{R}) \right) \equiv (\mathbf{R}'_i, \mathbf{P}'_i) \quad (33)$$

$$e^{i\mathcal{L}_r\delta t} (\mathbf{R}'_i, \mathbf{P}'_i) = \left(\mathbf{R}'_i + \frac{\delta t}{m} \mathbf{P}'_i, \mathbf{P}'_i \right) \quad (34)$$

$$= \left(\mathbf{R}_i + \frac{\delta t}{m} \mathbf{P}_i + \frac{\delta t^2}{2m} \mathbf{f}_i(\mathbf{R}), \mathbf{P}_i + \frac{\delta t}{2} \mathbf{f}_i(\mathbf{R}) \right) \equiv (\mathbf{R}''_i, \mathbf{P}''_i) \quad (35)$$

$$e^{i\mathcal{L}_p\delta t/2} (\mathbf{R}''_i, \mathbf{P}''_i) = \left(\mathbf{R}''_i, \mathbf{P}''_i + \frac{\delta t}{2} \mathbf{f}_i(\mathbf{R}'') \right) \quad (36)$$

$$= \left(\mathbf{R}_i + \frac{\delta t}{m} \mathbf{P}_i + \frac{\delta t^2}{2m} \mathbf{f}_i(\mathbf{R}), \mathbf{P}_i + \frac{\delta t}{2} [\mathbf{f}_i(\mathbf{R}) + \mathbf{f}_i(\mathbf{R}'')] \right) \quad (37)$$

Noting that $\mathbf{f}_i(\mathbf{R}) = \mathbf{f}_i(t)$, $\mathbf{f}_i(\mathbf{R}'') = \mathbf{f}_i(t + \delta t)$, the last expression is nothing but the velocity Verlet algorithm for $(\mathbf{R}_i(t + \delta t), \mathbf{P}_i(t + \delta t))$.

In conclusion: the Verlet algorithm may be derived by a discretization of the time evolution operator that conserves unitarity. Such property is crucial for any well-behaved algorithm one can think of.

Canonical ensemble in MD

We are often interested in systems in thermal equilibrium with a thermal bath at temperature T : the NVT or *canonical* ensemble, for which

$$\rho_{NVE}(\mathbf{R}, \mathbf{P}) = \frac{f(N)}{Z} e^{-H(\mathbf{R}, \mathbf{P})/k_B T} \quad (38)$$

where Z is the partition function:

$$Z = f(N) \int d\mathbf{R} d\mathbf{P} e^{-H(\mathbf{R}, \mathbf{P})/k_B T}. \quad (39)$$

Integration over \mathbf{P} gives for the partition function of N identical atoms:

$$Z = Z_r / (N! \lambda^{3N}) \quad (40)$$

where λ is the thermal wavelength:

$$\lambda = \frac{h}{\sqrt{2\pi m k_B T}} \quad (41)$$

and Z_r is the configurational partition function:

$$Z_r = \int d\mathbf{R}_1 \dots \int d\mathbf{R}_n e^{-E_p(\mathbf{R})/k_B T}. \quad (42)$$

In the canonical ensemble, the temperature is related to the expectation value of the kinetic energy:

$$\left\langle \sum_{i=1}^N \frac{\mathbf{P}_i^2}{2M_i} \right\rangle_{NVT} = \frac{3}{2} N k_B T. \quad (43)$$

The canonical ensemble can be simulated using what is called *Nosé-Hoover thermostat*: an additional fictitious degree of freedom produces a dynamical friction force having the effect of heating ions when the kinetic energy is lower than the desired value, cooling them in the opposite case. Specifically, the equations of motion become

$$\ddot{\mathbf{R}}_i = \frac{\mathbf{f}_i}{M_i} - \zeta \dot{\mathbf{R}}_i \quad (44)$$

$$\ddot{\zeta} = \frac{1}{Q} \left[\sum_{i=1}^N M_i \dot{\mathbf{R}}_i^2 - 3N k_B T \right] \quad (45)$$

where Q plays the role of “thermal mass”. The constant of motion for this system is

$$\tilde{H} = H + \frac{Q}{2} \dot{\zeta}^2 + 3N k_B T \zeta \quad (46)$$

but \tilde{H} does not generate the dynamics (the dynamics is *non-canonical*). It can be shown that such dynamics samples the canonical ensemble.

Although all thermodynamical properties could in principle be determined from the free energy F , it is not possible to calculate directly F from a MD simulation. The free energy (like the partition function and the entropy) cannot be simply expressed as a thermodynamical average (like the energy). Specialized algorithms are needed for free energy calculation.

Constant-pressure MD

Very often we are interested in simulating systems kept at a given pressure P rather than occupying a fixed volume V . *Constant-pressure* MD can be obtained by adding the volume V or, in a more general case, the cell parameters, to the dynamical variables. In the simple case of a liquid, one defines a Lagrangian:

$$\tilde{L} = \frac{1}{2} \sum_{i=1}^N M_i \left(V^{1/3} \dot{\sigma}_i \right)^2 - E_p(\{V^{1/3} \sigma\}) + \frac{1}{2} W \dot{V}^2 - PV \quad (47)$$

where $\sigma_i = \mathbf{R}_i/V^{1/3}$ are scaled variables, P is the desired external pressure, and W is a (fictitious) mass for V .

For a solid, we may be interested in knowing the equilibrium unit cell volume and form under a given stress state (typically a constant external hydrostatic pressure) rather than working at fixed cell and calculating the corresponding stress. In this case one introduces a matrix \mathbf{h} , formed by the unit cell vectors \mathbf{a}_i : $\mathbf{h} = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$, and defines scaled variables \mathbf{S}_i as $\mathbf{S}_i = \mathbf{h}^{-1}\mathbf{R}_i$. The extended lagrangian becomes

$$\tilde{L} = \frac{1}{2} \sum_{i=1}^N M_i \dot{\mathbf{S}}_i \mathcal{G} \dot{\mathbf{S}}_i - E_p(\{\mathbf{h}\mathbf{S}\}) + \frac{1}{2} W \text{Tr} \mathbf{h}^t \mathbf{h} - PV \quad (48)$$

where $\mathcal{G} = \mathbf{h}^t \mathbf{h}$ is the metric tensor. The interest of variable-cell dynamics for solid-state systems reside in the possibility to simulate structurale phase transitions (under applied pressure but also as a function of temperature).

Car-Parrinello Molecular Dynamics

Implementations of MD using first-principle interatomic potential calculated from DFT, as in Eq.(14), are widely used. All the MD machinery developed for classical interatomic potentials can be used. However these implementation suffer from a serious drawback. MD is quite sensitive to the quality of forces. If the forces are not the derivatives of the energy with high accuracy, the MD simulation will have problems, appearing as a drift of quantities that should be conserved (like e.g. the energy) from their values. The error on DFT forces is *linear* in the selfconsistency error of the charge density (while for the DFT energy it is *quadratic*). As a consequence, a very good and expensive convergence to self-consistency is required at every time step.

In 1985 Car and Parrinello (CP) proposed a different approach. They introduced a Lagrangian for both electronic and ionic degrees of freedom:

$$L = \frac{\mu}{2} \sum_k \int d\mathbf{r} |\dot{\psi}_k(\mathbf{r})|^2 + \frac{1}{2} \sum_i M_i \dot{\mathbf{R}}_i^2 - E_{tot}(\{\mathbf{R}\}, \{\psi\}) + \sum_{k,l} \Lambda_{kl} \left(\int \psi_k^*(\mathbf{r}) \psi_l(\mathbf{r}) d\mathbf{r} - \delta_{kl} \right) \quad (49)$$

which generates the following set of equations of motion:

$$\mu \ddot{\psi}_k = H \psi_k - \sum_l \Lambda_{kl} \psi_l, \quad M_i \ddot{\mathbf{R}}_i = - \frac{\partial E_{tot}}{\partial \mathbf{R}_i} \quad (50)$$

where μ is a fictitious electronic mass, and the Lagrange multipliers Λ_{kl} enforce orthonormality constraints.

The electronic degrees of freedom are, in the typical implementation, expansion coefficients of KS orbitals into PW. The forces acting on them at each time step are determined by the KS Hamiltonian calculated from the current values of ψ_k and of \mathbf{R}_i . The sum over orbitals for an insulating system of n electrons includes $n/2$ states, assuming that spin polarization is neglected (every orbital is occupied by two electrons). Most CP calculations are done for aperiodic systems or for systems having a large unit cell (or supercell), so that typically only the Γ point ($\mathbf{q} = 0$) is used to sample the Brillouin Zone. Note that the entire Hamiltonian operator is not required: only products $H\psi_i$ are.

The forces acting on ions have the Hellmann-Feynman form:

$$\frac{\partial E_{tot}}{\partial \mathbf{R}_i} = \sum_k \langle \psi_k | \frac{\partial V}{\partial \mathbf{R}_i} | \psi_k \rangle \quad (51)$$

where V is the electron-ion interaction (pseudo-)potential. Note however that Hellmann-Feynman theorem holds only on the exact ground state. The relation of Car-Parrinello forces to Hellmann-Feynman forces is explained in the next section.

Orthonormality constraints are exactly imposed to the ψ at each time step, using an iterative procedure that exploits the fact that the loss of orthonormality at each time step is small.

The simulation starts by bringing the electrons to the BO surface (that is, to the ground state) at fixed ions and proceeds, using classical MD technology, on both electronic and ionic degrees of freedom. With appropriate values of μ and δt , the electrons always remain close to the BO surface, while the ions follow a trajectory that is close to the trajectory they would follow in the BO approximation.

The Car-Parrinello dynamics has turned out to be very successful especially in the study of low-symmetry situations: surfaces, clusters, liquids, disordered materials, and for the study of chemical reactions.

Why Car-Parrinello works

The reasons why the Car-Parrinello dynamics works so effectively are quite subtle. The dynamics for the electrons is purely classical (and fictitious: it has nothing to do with real electron dynamics). As a consequence the energy would tend to equipartition between electronic and ionic degrees of freedom, causing an energy transfer from ionic to electronic degrees of freedom. This does not happen (and must not happen, otherwise the electrons will leave the BO surface) even on long simulation times. If we analyze the dynamics in terms of oscillators, we find that the typical

frequencies associated to the fictitious electron dynamics are given by $\omega^{el} \sim \sqrt{(\epsilon_i - \epsilon_j)/\mu}$, if there is a gap in the electronic spectrum. For ions, the oscillator frequencies are the typical phonon frequencies. It turns out that, for reasonable values of the gap and of the fictitious electron mass μ , the maximum phonon frequency is much smaller than the minimum electron frequency: $\omega_{max}^{ph} \ll \omega_{min}^{el}$. The energy transfer from ionic to electronic degrees of freedom is as a consequence very small even on long times.

This situation generates a fast electron dynamics that keeps the electrons close to the BO surface and averages out the error on the forces, so that the much slower ionic dynamics turns out to be correct (that is, very close to the BO dynamics one would obtain from highly converged self-consistency). A detailed explanation is contained in a 1991 paper by Pastore, Smargiassi, and Buda.

If there is no gap in the electronic spectrum, or if the gap is too small, the above picture breaks down. It may be needed to add separate thermostats to ionic and electronic degrees of freedom in order to prevent the flow of energy from the former to the latter.

Choice of the parameters

The choice of the electronic mass μ must strike a compromise between conservation of adiabaticity (favoured by small values of μ , see above) and maximum admissible time step (that is limited by the maximum electronic frequency, so that the heaviest μ , the smaller ω_{max}^{el} , the larger δt_{max}). Typically $\mu \sim 200$ amu (1 amu=1 electron mass). For large gap systems, such as SiO₂ or H₂O, in which adiabaticity problems are minor, μ may be increased up to ~ 500 -700 amu and even more. Such values of μ correspond to a typical timestep of $\sim 0.1 - 0.2$ fs.

In order to increase the time step, it is customary to introduce the so-called *mass preconditioning*. In a PW basis set, the time step is limited by high-frequency components with the largest \mathbf{G} vector. These components are dominated by the kinetic energy $\hbar^2 \mathbf{G}^2 / 2\mu$. Since electronic masses are fictitious, it is advantageous to introduce a mass that for high-frequency components goes like $\mu(\mathbf{G}) \simeq \mu(1 + \mathbf{G}^2)$. The corresponding equations of motions are only slightly more complex.

It should be noticed, however, that too heavy electron masses adversely affect the quality of simulation via an “electron drag” effect. The electron motion follow the ionic motion with some delay, thus introducing a drag force that appears as if the ions were heavier than their masses. This “mass renormalization” must be taken into account when extracting vibrational frequencies from MD runs. In some cases, this effect can introduce a nonnegligible deviation from the true ionic dynamics.

Structural phase transitions in solids

Introduction

The field of phase transitions (PT) is a huge one. We will deal only with structural (not magnetic or electronic) PT between two solid structures (not liquid-solid PT). Among the different types of solid to solid PT we will distinguish between:

- PT between two structures having no symmetry relationship (for instance: cubic Si to β -tin Si under pressure, see fig.1). There is no way to go from one structure to the other with continuity: such transitions are definitely discontinuous (*first order*).
- PT in which one can go with continuity from one structure to the other. Such PT may happen both in a continuous way (*second order* PT) or in a discontinuous way. The continuity or discontinuity refers to some variable, the *order parameter*, describing the transition.

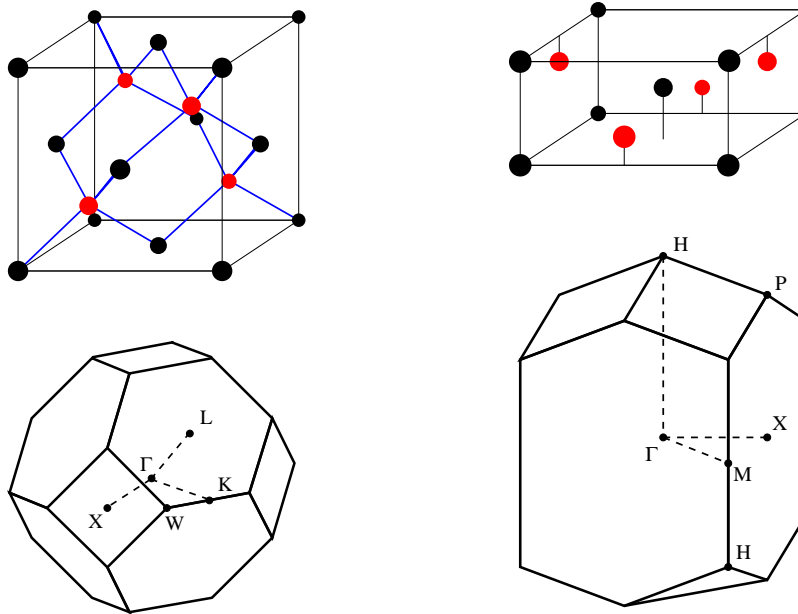


Fig.1. A phase transition between the diamond and the β -tin structure occurs in Si under applied pressure.

Moreover one can distinguish between

- order to order PT (from an ordered crystal phase to another). Examples: the many phases of ice and of SiO_2 .
- order/disorder PT :
 - in metallic alloys (for instance: CuZn) or in semiconducting alloys (GaAlAs, SiGe,...)
 - amorphization PT (ice, SiO_2 ,...) and PT to incommensurate phases (that is: to a structure that is not compatible with translation symmetry) (it may be difficult to distinguish between the two)
 - orientational PT (typically in molecular solids: C_{60} , H_2)

PT can be induced by temperature, pressure, or by other factors like irradiation. We are going to deal mainly with continuous PT between crystal phases (or to amorphous phases) induced by pressure. We will see in more detail the case of Cs halides with a combination of first-principle techniques and of Landau phenomenological theory of PT.

Pressure-induced transition have a considerable fundamental interest, and a specific interest in geophysics. Experiments are usually performed using diamond anvil cells (a high-tech nutcracker, basically), more seldom shock waves. The behavior of samples is monitored using X-Ray Diffraction (XRD), Raman and IR spectroscopy.

Some thermodynamics and some examples

A PT between two phases may take place when the chemical potential μ of the two phases is equal, assuming that the pressure P and the temperature T are the relevant thermodynamical quantities. μ is the Gibbs free energy G per unit mass (or per particle): $\mu = dG/dN$, and $G = E - TS + PV = H - TS = F + PV$ (F is the Helmholtz free energy, S is the entropy, H is the enthalpy, E is the energy, V is the volume).

For a solid the main contribution to the entropy comes from the vibrational entropy. If we assume $T = 0$ (or equivalently if we neglect the vibrational entropy, or neglect its variation between the two phases) the Gibbs construction between two $E(V)$ curves for two different phases yields the transition pressure and the transition volumes. The Gibbs construction consists in finding the common tangent to two $E(V)$ curves. As $P = -dE/dV$ at $T = 0$, this is equivalent to compare the enthalpies $E + PV$ (the Gibbs free energy at $T = 0$) of the two phases. In this case P is obtained from the equation of state $E(V)$ but it can also be calculated directly using the *stress theorem* of Nielsen and Martin: $P = -\text{Tr}\sigma/3$, where the stress σ is the derivative of the energy with respect to the strain ϵ . An early example of this approach is the study of the stability of different phases of Si under applied pressure: M.T.Yin and M.Cohen, Phys.Rev. **B26**, 568 (1982)

In temperature-induced PT the vibrational entropy may play an important role. The free energy (F : at ambient pressure, $G \simeq F$ for a solid) is the sum of an electronic and a vibrational term, $F = F_{el} + F_{ph}$, with $F_{el} \simeq E_{el}$. The vibrational term F_{ph} can be in many cases calculated within the *quasi-harmonic approximation*: one calculates F_{ph} in the harmonic approximation, retaining the dependency on the volume V (such dependency is clearly an anharmonic effect). F_{ph} depends on T and V through the (harmonic) phonon spectra:

$$F_{ph}(T, V) = -k_B T \log \left(\text{Tr} e^{-\mathcal{H}_{ph}/k_B T} \right) \quad (52)$$

where \mathcal{H}_{ph} is the phonon Hamiltonian at a given volume V . One calculates the complete phonon spectra (the force constants in real space) for different phases and integrates on the Brillouin Zone (BZ) the vibrational free energy:

$$F_{ph}(T, V) = -k_B T \sum_{i\mathbf{k}} \log \left(\sum_n e^{-(n+1/2)\hbar\omega_{i\mathbf{k}}/k_B T} \right) = -k_B T \sum_{i\mathbf{k}} \log \left(e^{\hbar\omega_{i\mathbf{k}}/2k_B T} (1 - e^{-\hbar\omega_{i\mathbf{k}}/k_B T}) \right) \quad (53)$$

which finally yields

$$F_{ph}(T, V) = k_B T \sum_{i\mathbf{k}} \log [2\sinh(\hbar\omega_{i\mathbf{k}}/2k_B T)]. \quad (54)$$

Then one compares the calculated $F(T, V)$ for different phases and finds where the free energies cross. This yields the transition temperature. With the current phonon technology, calculations of this type are straightforward. A good example is the PT from grey tin (α -tin) at low temperature to white tin (β -tin) at $T = 13\text{C}$ (the same two structures of fig.1): see Phys.Rev.B **57**, 10421 (1998). Note that in this paper the dependency of $F_{ph}(T, V)$ on the volume is neglected.

The two examples mentioned above refer to first order transitions: at a given P or T , the (Gibbs) free energy curves of two different phases cross, and this is the end of the story. There are also other types of PT, in which a lattice instability develops (so that the free energy curve of one phase has no longer a minimum after the transition). In crystal there are mainly two different types of lattice instabilities:

- phonon instabilities (*soft phonons*). Usually these are phonon modes having a \mathbf{k} vector at the border of the BZ, whose energy goes to zero (they become “soft”). As a consequence a distortion sets up that causes an increase of the dimension of the unit cell. In temperature-induced PT (for instance: ferroelectric transitions, BaTiO₃) anharmonicity may play a fundamental role. In pressure-induced PT, instead, anharmonicity is often irrelevant: it is the harmonic part that becomes soft as a consequence of the changes in volume and atomic positions caused by the pressure.
- elastic instabilities, leading to a change of shape of the unit cell. In a finite system there would be no need to distinguish between the two cases, but in a system with periodic boundary

conditions one has to distinguish between normal and elastic modes (the latter are the limit of the former for $\mathbf{k} \rightarrow 0$). This kind of PT is sometimes called *martensitic*.

The two types of instability are often interdependent. In particular, a phonon instability may couple with strain, as we are going to see.

Second-order Phase Transitions

Let us try to better understand what are the factors that drive a PT and how they can be described. In particular, we want to understand under which conditions a PT can or cannot be continuous.

A continuous PT may happen between two phases having different symmetry, provided that one phase has a higher symmetry than the other. This is a necessary condition: at the transition point the two phases coexists, so that the symmetry of both phases must be satisfied. In a first-order PT the symmetries of the two phases may have no relationship between them. In temperature-induced PT the high-temperature phase is usually more symmetric than the low-temperature phase. In pressure-induced PT there is no similar general rule.

The PT between a high-symmetry and a low-symmetry phase may be described by an *order parameter* whose value is zero in the high-symmetry phases and different from zero in the low-symmetry phase. A simple example is given by the magnetization in a magnetic transition. For a simple order-disorder PT (AB alloy in a BCC lattice like ZnCu) one may define an order parameter η :

$$\eta = \frac{w(A) - w(B)}{w(A) + w(B)} \quad (55)$$

where $w(A)$ is the probability to find A on a given sublattice. In this case the “high symmetry” phase is the disordered phase: the statistical-mechanics meaning of the term is exactly the opposite of the solid-state meaning! It is likely that this derives from magnetic transitions, in which the high-temperature, paramagnetic phase, has higher symmetry than the ordered phase.

Landau Theory of Phase Transitions

The Landau theory of PT is a phenomenological theory. It is basically a *mean-field* theory and as such neglects fluctuations, so that its predictions for the critical properties interesting to statistical mechanics (critical exponents etc) are not always accurate. It offers however a useful framework for the classification of PT and its parameters can sometimes be derived from microscopic (also first-principles) calculations.

Let us consider the Gibbs free energy $G(P, T)$ around the transition point. We can expand G around $\eta = 0$ and write:

$$G(P, T, \eta) = G_0 + \alpha\eta + A\eta^2 + C\eta^3 + B\eta^4 + \dots \quad (56)$$

(A, B, C are also functions of P, T). For given values of P and T the stable phase is determined by the minimization of G wrt η . It is easily demonstrated that a second-order PT between two structures of different symmetry, one with $\eta = 0$ and the other with $\eta \neq 0$, can happen only if

1. $\alpha = 0$ (so that the $\eta = 0$ phase is stable)
2. $A > 0$ for the $\eta = 0$ phase, A changes sign at the transition (so that something actually happens),
3. $C = 0$ (so that the order parameter changes continuously), and
4. $B > 0$ (so that the minimum point with $\eta \neq 0$ is stable).

C can be zero by symmetry (for instance: in the magnetic case, the free energy cannot depend on the sign of the magnetization) or it can be only for some special values of P e T , or can be different from zero, in which case a transition can still take place but it is first-order.

Let us assume that the high-symmetry phase ($\eta = 0$) has a symmetry group \mathcal{G}_0 . A necessary condition for a second-order PT is that the symmetry group \mathcal{G} of the low-symmetry phase ($\eta \neq 0$) is a subgroup of \mathcal{G}_0 . We can always write the charge density ρ (or any other relevant symmetric

function involved in the transition) as a sum of basis functions of irreducible representations (irreps) of \mathcal{G}_0 :

$$\rho = \sum_n \sum_i \eta_i^{(n)} \phi_i^{(n)} \quad (57)$$

where n is the irrep index, $\phi_i^{(n)}$ is the i -th basis functions of the n -th irrep. Let us extract the totally symmetric irrep:

$$\rho = \rho_0 + \delta\rho, \quad \delta\rho = \sum_n' \sum_i \eta_i^{(n)} \phi_i^{(n)} \quad (58)$$

where \sum_n' excludes the totally symmetric irrep. The symmetry \mathcal{G} of $\delta\rho$ is necessarily inferior to the symmetry \mathcal{G}_0 of ρ_0 . Let us expand the Gibbs free energy as a function of $\eta_i^{(n)}$. The expansion must contain only invariant quantities wrt the symmetry \mathcal{G}_0 . From group theory we know that there are no invariants formed by linear terms in $\eta_i^{(n)}$ (otherwise the irrep n would be reducible), and that there is only one second-order invariant that is always reducible to the sum of squares: $\sum_i (\eta_i^{(n)})^2$. So, up to second order:

$$G = G_0 + \sum_n' A_n \sum_i (\eta_i^{(n)})^2 \quad (59)$$

where $A_n = A(P, T)$ and all A_n are positive (if the phase is stable). At the transition one of the A_n vanishes and changes sign, the corresponding $\eta_i^{(n)}$ become nonzero, the symmetry of the system is lowered. Let us omit the index n and introduce the notations

$$\eta_i = \eta \gamma_i, \quad \eta^2 = \sum_i \eta_i^2 \quad \left(\sum_i \gamma_i^2 = 1 \right) \quad (60)$$

so that G may be written as

$$G = G_0 + \eta^2 A(P, T) + \eta^3 \sum_\alpha C_\alpha(P, T) f_\alpha^{(3)}(\gamma_i) + \eta^4 \sum_\alpha B_\alpha(P, T) f_\alpha^{(4)}(\gamma_i) + \dots \quad (61)$$

where $f^{(3)}, f^{(4)}$ etc are third-, fourth-, etc, order invariants built with γ_i . There are as many m -th order invariants as many times the totally symmetric irrep appears in $[\Gamma^m]$, where $[\Gamma^m]$ is the *symmetric product* (made with symmetric combinations of the basis functions) $\Gamma \times \Gamma \times \Gamma \times \dots$ m times of the irrep Γ that is *active* at the transition. If $[\Gamma^3]$ does not contain the totally symmetric irrep, there are no third-order invariants and the transition can be continuous (second order). Otherwise the transition is discontinuous (first order).

Landau theory in crystals: soft phonons

In many PT the frequency of a phonon mode goes to zero: there is a *soft phonon*. Physically, this means that a distortion along the soft mode will cost zero energy, so that the crystal will tend to distort along that mode. It is natural to take the amplitude of the soft phonon as the order parameter for the PT.

With respect to the simple case of the preceding section, PT induced by soft phonons have an additional complication: the existence of a quantum number \mathbf{k} related to translational symmetry. The order parameter is in most cases multidimensional: if a phonon with wavevector \mathbf{k}_0 goes soft, all the phonons whose wavevectors is in the *star* of \mathbf{k}_0 (the set of vectors that transform into each other under the symmetry operations of the point group of the crystal) will form the order parameter. Such multidimensional order parameters can give raise to many different structures, having different symmetry and different unit cell, according to which combination minimizes the free energy. Very often just one \mathbf{k}_0 is present, but not always. In the case of magnetic PT, the magnetic structures in which more than one \mathbf{k}_0 component in the star are present are called *multiple- \mathbf{k} structures*.

The conditions for the existence and stability of the new phases are however very stringent. Let we assume that the A coefficients depend on \mathbf{k} : $A = A(\mathbf{k})$. The new structure can be a minimum of the free energy only if $A(\mathbf{k}_0)$ has a minimum in \mathbf{k}_0 . That is, the linear terms in $\mathbf{k} - \mathbf{k}_0$ of the

expansion of $A(\mathbf{k})$ around \mathbf{k}_0 must be zero. One can demonstrate that a second-order PT is possible only if the antisymmetric square (the representation formed with antisymmetric combinations of the basis functions: requires two different sets of basis functions) $\{\Gamma^2\}$ of the active irrep Γ does not contain irreps according to which the components of a vector transform.

The formal demonstration of such conditions is quite boring. A more physical demonstration can be found in the book of Landau-Lifschitz, Statistical mechanics, chapter 14. In summary: expanding in powers of $\mathbf{k} - \mathbf{k}_0$ is equivalent to consider the η_i as slowly varying functions (slowly wrt a unit cell) of coordinates. Let us expand the free energy around the transition point into a series of powers of η_i and of their gradients $\nabla\eta_i$. The stability with respect to macroscopic distortion, once all irrelevant terms are ruled out, will translate into the non existence of terms $\eta_i\nabla\eta_j - \eta_j\nabla\eta_i$ in the expansion. This is exactly what the above condition guarantees.

The above condition cannot be verified for a generic \mathbf{k}_0 in the BZ. By appropriately combining the basis functions for the s vectors in the star of a generic \mathbf{k}_0 (s is the number of symmetry rotations in the point group of the crystal) one can generate a reducible representation (the so-called *regular* representation) that contains all the irreps of the group.

It can be demonstrated that only \mathbf{k}_0 points whose symmetry group contains intersecting axis, or axis intersected by symmetry planes, or an inversion, can verify the above condition on $\{\Gamma^2\}$. For practical purposes, this means the only the so-called “high-symmetry points” of the BZ (zone-boundary points with special symmetries) can give raise to second-order transitions. These are in general given as $\mathbf{k}_0 = \mathbf{G}/m$, where \mathbf{G} is a reciprocal lattice vector and m is a small number: $m = 2$, in some cases (for instance hexagonal and fcc lattices) $m = 3$ and $m = 4$, are allowed. This means that second-order PT can only double (triple or quadruple in some cases) the unit cell in one or more directions, leading to a increase of a factor 2,4, or 8 times of the unit cell volume (special cases excepted).

Note that the above conditions are sufficient conditions for a second-order PT to take place. A transition can be continuous for other reasons that the symmetry: it may be such at some values of P and T , or the discontinuity can be very small (so that it is impossible to observe the discontinuity) because the terms that hinder the continuity of the PT are very small.

Landau theory in crystals: elastic instabilities

The soft phonon can also be an acoustic phonon at the Γ point of the BZ ($\mathbf{k} = 0$), or equivalently, an elastic mode. In the latter case the order parameter is a macroscopic strain $\epsilon_{\alpha\beta}$. Even in this case the rules for second-order transitions derived for soft phonons apply. The crystal is deformed and the shape (and symmetry) of the unit cell changes accordingly.

The condition for mechanical stability us that the energy E is a positive definite quadratic form of the strains (that is, all the eigenvalues of the matrix of elastic constants are positive). This yields the *Born conditions* for the stability of a crystal with respect to elastic deformations. For a cubic crystal: $C_{44} > 0$, $C_{11} - C_{12} > 0$, $C_{11} + 2C_{12} > 0$. For an hexagonal crystal the first two conditions hold, the third becomes $(C_{11} + C_{12})C_{33} > 2C_{13}^2$. If such conditions are violated there can be an elastic soft mode for which the strain is the order parameter.

The strain tensor components generate a 6-dim. reducible representation that can be decomposed into irreps of the cubic symmetry group as follows : $\Gamma_1^+ + \Gamma_3^+ + \Gamma_5^+$ (a symmetric tensor has an s and a d part: the s part is Γ_1^+ , the d part is split into $\Gamma_3^+ + \Gamma_5^+$). The functions that realize the irreps are easily determined:

$$\epsilon(\Gamma_1^+) = \frac{\epsilon_{xx} + \epsilon_{yy} + \epsilon_{zz}}{\sqrt{3}} \quad (62)$$

$$\epsilon_1(\Gamma_3^+) = \frac{\epsilon_{xx} - \epsilon_{yy}}{\sqrt{2}}, \quad \epsilon_2(\Gamma_3^+) = \frac{2\epsilon_{zz} - \epsilon_{xx} - \epsilon_{yy}}{\sqrt{6}} \quad (63)$$

$$\epsilon_1(\Gamma_5^+) = \epsilon_{xy}, \quad \epsilon_2(\Gamma_5^+) = \epsilon_{xz}, \quad \epsilon_3(\Gamma_5^+) = \epsilon_{yz} \quad (64)$$

In terms of these functions, one can rewrite the elastic energy as

$$\frac{E}{V} = \frac{1}{2}(C_{11} + 2C_{12})\epsilon^2(\Gamma_1^+) + \frac{C_{11} - C_{12}}{2} (\epsilon_1^2(\Gamma_3^+) + \epsilon_2^2(\Gamma_3^+)) + C_{44} (\epsilon_1^2(\Gamma_5^+) + \epsilon_2^2(\Gamma_5^+) + \epsilon_3^2(\Gamma_5^+)). \quad (65)$$

Let us consider first Γ_1^+ strain. The quantity $B = (C_{11} + 2C_{12})/3$ is known as *bulk modulus*. The Born condition $C_{11} + 2C_{12} > 0$ requires the crystal to be stable against an isotropic strain (dilatation) and does not correspond to any phase transition.

Let us consider now Γ_3^+ strain. The quantity $C_s = (C_{11} - C_{12})/2$ is known as *shear elastic constant*. The strain $\epsilon_1(\Gamma_3^+)$ induces a cubic to orthorhombic transition; the strain $\epsilon_2(\Gamma_3^+)$ induces a cubic to tetragonal transition. It is easily demonstrated that these transitions cannot be second order. In fact an invariant of $[\Gamma_3^+]^3$ exists and is given by

$$f_3 = \frac{1}{4}\epsilon_1(\Gamma_3^+) (\epsilon_1^2(\Gamma_3^+) - 3\epsilon_2^2(\Gamma_3^+)) = (\epsilon_{zz} - \epsilon_{yy} - \epsilon_{xx})(\epsilon_{yy} - \epsilon_{xx} - \epsilon_{zz})(\epsilon_{xx} - \epsilon_{zz} - \epsilon_{yy}). \quad (66)$$

The invariant character of f_3 is evident by direct inspection of the right-hand side.

Coupling of soft phonons with the strain

The coupling of soft phonons with the strain can play an important role in phase transitions. In fact the strain state of the crystal in general is not fixed and the strain will adjust itself so as to minimize the free energy. Assuming a simplified model in which one soft phonon η is coupled with a given strain component ϵ , one can write down the free energy expansion as

$$G = G_0 + A\eta^2 + C\eta^3 + B\eta^4 + D\epsilon\eta^2 + \frac{1}{2}c\epsilon^2 + .. \quad (67)$$

provided that an invariant coupling term $\epsilon\eta^2$ can be constructed (terms of the form $\epsilon\eta$ are not allowed by symmetry if η is a soft phonon of finite wavevector). The last term is the elastic energy and c is the appropriate elastic constant. Minimization of G with respect to ϵ for a given η yields

$$\sigma = \frac{\partial G}{\partial \epsilon} = D\eta^2 + c\epsilon = 0 \quad (68)$$

where σ is the corresponding stress. This means $\epsilon = -D\eta^2/c$ and finally

$$G = G_0 + A\eta^2 + C\eta^3 + \left(B - \frac{D^2}{2c}\right)\eta^4 + ... \quad (69)$$

The coefficient of the 4th-order term B' is “renormalized”: $B' = B - D^2/2c$, always smaller than the “bare” value B . If $B > 0$ but $B' < 0$ the interaction with the strain may drive a 2nd-order transition to become 1st-order.

Applications: Cesium Halides under pressure

The complete story of Cesium Halides under pressure can be found in Phys.Rev.B51, 8060 (1995), and in the Sissa PhD thesis of Marco Buongiorno Nardelli. The following is a summary of some relevant steps.

At normal pressure Cesium Halides crystallize in the B2 cubic structure (a simple cubic lattice with two atoms in the unit cell, one at the origin and the other in the center of the cube). CsI is the softest among alkali halides and has the smallest gap. As a consequence it is a good candidate for metallization under pressure, achieved at ~ 110 GPa according to some experimental data. In 1984 a PT at ~ 39 GPa for CsI and at ~ 54 GPa was reported. This was attributed to a continuous cubic to tetragonal transition. A further transition to a simple orthorhombic lattice was found in one study but not confirmed by others. Different theoretical analysis found an instability of the shear constant $c_s = (C_{11} - C_{12})/2$, leading to a first-order cubic to tetragonal transition.

Further experimental studies in CsI showed that the onset of the transition was at lower pressures than previously assumed ($P \sim 15$ GPa) and that the new structure was orthorhombic with doubled unit cell ($\sqrt{2} \times \sqrt{2}$ unit cell) and a gliding of the cubic $\langle 110 \rangle$ planes.

The atomic displacement pattern derived from XRD is reminiscent of an acoustic phonon mode at $\mathbf{k}_0 = \pi/a(1, 1, 0)$ (the M point of the BZ) with transverse polarization $\mathbf{e} = (1, \bar{1}, 0)$. The shear constant is related to the sound velocity along the $\Gamma - M$ direction as $\mathbf{k} \rightarrow 0$, for waves polarized along \mathbf{e} . This suggests that a phonon instability at the BZ boundary point M could precede the onset of an elastic instability at $\mathbf{k} \rightarrow 0$.

Table 1: Phonon wavevectors, \mathbf{k} , and polarizations, \mathbf{e} , of the six components of the order parameter, \mathbf{u} .

	u_1	u_2	u_3	u_4	u_5	u_6
\mathbf{k}	$(0\frac{1}{2}\frac{1}{2})$	$(0\frac{1}{2}\frac{1}{2})$	$(\frac{1}{2}0\frac{1}{2})$	$(\frac{1}{2}0\frac{1}{2})$	$(\frac{1}{2}\frac{1}{2}0)$	$(\frac{1}{2}\frac{1}{2}0)$
\mathbf{e}	(100)	$(0\frac{1}{\sqrt{2}}\frac{1}{\sqrt{2}})$	(010)	$(\frac{1}{\sqrt{2}}0\frac{1}{\sqrt{2}})$	(001)	$(\frac{1}{\sqrt{2}}\frac{1}{\sqrt{2}}0)$

The calculations show that actually the doubly degenerate acoustic phonons at M (irrep M_5^- of the group D_{4h}) are becoming soft at $V/V_0 = 0.63$ (V_0 is the equilibrium volume at zero pressure) corresponding to $P = 23$ GPa. There are three inequivalent vectors in the star of M (the star would have 12 points, but they are all equivalent by a lattice translation to one of the three) and two possible polarizations per \mathbf{k} -point: the order parameter, reported in the table, is six-dimensional.

The Landau analysis of the transitions requires the calculation of 3rd- and 4th-invariants formed with the order parameter. For the cubic group of CsI, having only symmorphic operations (with no glide plane or screw axis), the basis functions of the irreps of the space group are simply obtained by multiplying the basis functions of the irreps of the point group (in this case, M_5^-) by a factor $\exp(i\mathbf{k}_i \cdot \mathbf{r})$, for all the \mathbf{k} vectors in the star. The fact that M_5^- is odd wrt inversion is sufficient to guarantee that the $[\Gamma^3]$ and the $\{\Gamma^2\}$ conditions are satisfied: Γ^2 is even wrt inversion and cannot contain the irrep according to which vectors transform (it has to be odd: Γ_4^- for cubic groups); $[\Gamma^3]$ is odd and as a consequence it is not an invariant.

The 2nd-order invariant is $u_1^2 + u_2^2 + u_3^2 + u_4^2 + u_5^2 + u_6^2$. The group theory guarantees that there are no other 2nd-order invariants. The construction of 4th-order terms is less straightforward but not really difficult. They can be found either by brute-force application of group theory or by a more targeted approach. Let us try the latter.

A term of the type $u_i u_j u_l u_m$ will contain a phase factor $\exp(i(\mathbf{k}_i + \mathbf{k}_j + \mathbf{k}_l + \mathbf{k}_m) \cdot \mathbf{r})$. This will be translationally invariant only if $\mathbf{k}_i + \mathbf{k}_j + \mathbf{k}_l + \mathbf{k}_m = \mathbf{G}$. Only combinations in which a given wavevector appears twice (because $\mathbf{k}_i + \mathbf{k}_j \neq \mathbf{G}$ unless $\mathbf{k}_i = \mathbf{k}_j$) or four times are allowed. This conditions is all we need in order to satisfy translational invariance.

One 4th order invariant term is necessarily $(u_1^2 + u_2^2 + u_3^2 + u_4^2 + u_5^2 + u_6^2)^2$. The totally symmetric representation Γ_1^+ of the cubic point group is contained four more times into $[(M_5^-)^4]$, thus four more invariants have to be found. They are generated by symmetrizing (applying all the operations of the point symmetry group) the four terms $u_1^4, u_1^2 u_2^2, u_1 u_2 u_3^2, u_1 u_2 u_3 u_4$.

Finally we arrive at a free energy expansion up to fourth-order:

$$G = G_0 + \omega^2 \left(\sum_{i=1}^6 u_i^2 \right) + \sum_{j=1}^4 b_j f_j^{(4)}(\{u_i\}). \quad (70)$$

The above model has been studied in the literature and all possible minima with definite (and acceptable) symmetry are tabulated. The parameters b_j can be calculated by fitting to a few frozen-phonon calculations, for a few different configuration, at a volume close to the transition volume. ω is the harmonic frequency of the M_5^- phonon. Our hope is to find an orthorhombic structure as minimum energy structure. This corresponds to the minimum along a line labeled as P_1 , in which a single phonon mode is active. We find instead that the lowest-energy structure is along the P_{11} line, corresponding to a distortion in which all six phonons are active. Such a structure has tetrahedral symmetry. No luck!

It turns out that it is the phonon-strain coupling (see previous section) that brings the lowest energy solution along line P_1 . We need to find all invariant terms of the type $u^2 \epsilon$. u^2 terms must be such that $\mathbf{k}_i + \mathbf{k}_j = \mathbf{G}$. There are 9 such symmetric combinations, forming a 3-d reducible representation $u_1 u_2, u_3 u_4, u_5 u_6$, and a 6-d reducible representation $u_1^2, u_2^2, u_3^2, u_4^2, u_5^2, u_6^2$. The former reduces into $\Gamma_2 + \Gamma_3$. The latter reduces into $\Gamma_1 + \Gamma_3 + \Gamma_5$. Invariants can be built by combining an irrep of u^2 with the corresponding irrep of ϵ (see section on elastic instabilities).

The coefficient are again evaluated from frozen-phonon and stress calculations. The picture changes dramatically: the strain has a different effect on different minima. On P_1 , the fourth-order coefficient B becomes large and *negative*, thus implying that sixth-order terms in the expansion are needed. This would be really too complicated. Performing the minimization along the P_1 line

only, one finally finds a slightly first-order transition at a pressure that is somewhat lower than the softening pressure.

The final step is to compare the relative stability of the orthorhombic phase with competing tetragonal phase produced by an elastic instability. For CsI, the former is favoured, while for CsBr the situation is reversed.

Summary of elasticity theory

A long-wavelength distortion (long with respect to the dimensions of the unit cell) is defined by a *strain* $e_{\alpha\beta}$:

$$e_{\alpha\beta} = \frac{\partial r'_\alpha}{\partial r_\beta} \quad (71)$$

or alternatively: strained coordinates \mathbf{r}' are given by

$$r'_\alpha = (\delta_{\alpha\beta} + e_{\alpha\beta})r_\beta. \quad (72)$$

Only symmetric components have a physical meaning: $e_{\alpha\beta} = e_{\beta\alpha}$ (antisymmetric components are equivalent to a rotation). The energy as a function of the strain is, to second order:

$$E = \frac{V}{2} \sum_{\alpha\beta,\gamma\delta} c_{\alpha\beta,\gamma\delta} e_{\alpha\beta} e_{\gamma\delta} \quad (73)$$

where $c_{\alpha\beta,\gamma\delta}$ is the *elastic constants* tensor and the stress (a force per unit surface) is

$$\sigma_{\alpha\beta} = \frac{1}{V} \frac{\partial E}{\partial e_{\alpha\beta}} = \sum_{\gamma\delta} c_{\alpha\beta,\gamma\delta} e_{\gamma\delta} \quad (74)$$

The elastic constant tensor has in general $9 \times 9 = 81$ components, that reduce to $6 \times 6 = 36$ because stress and strain are symmetric tensors (with 6 independent components). It can be demonstrated that the elastic constant tensor is symmetric as well, bringing a further reduction to no more than $6 \times 7/2 = 21$ independent components. The Voigt notation is traditionally used:

$$\epsilon_{\alpha\beta} = e_{\alpha\beta} \text{ se } \alpha = \beta, \quad \epsilon_{\alpha\beta} = 2e_{\alpha\beta} \text{ se } \alpha \neq \beta \quad (75)$$

and for indexes: $xx = 1, yy = 2, zz = 3, xy = 4, xz = 5, yz = 6$.

Symmetry allows a large reduction in the number of independent nonzero elastic constants. For instance, in an hexagonal crystal, only 5 elastic constants $C_{11}, C_{12}, C_{13}, C_{33}, C_{44}$, in Voigt notation, are independent ($C_{66} = (C_{11} - C_{12})/2$ is not). In a cubic crystal there are only 12 nonzero elastic constants and only 3 are independent: C_{11}, C_{12}, C_{44} .

The elastic energy (per unit volume) in the cubic case becomes

$$\frac{E}{V} = \frac{1}{2} C_{11} (\epsilon_{xx}^2 + \epsilon_{yy}^2 + \epsilon_{zz}^2) + C_{12} (\epsilon_{xx}\epsilon_{yy} + \epsilon_{xx}\epsilon_{zz} + \epsilon_{yy}\epsilon_{zz}) + C_{44} (\epsilon_{xy}^2 + \epsilon_{xz}^2 + \epsilon_{yz}^2) \quad (76)$$

For many simple materials the *Cauchy relation* $C_{44} = C_{12}$ approximately holds. It holds exactly if atoms interact with central forces, if all atoms are inversion symmetry centers, in absence of external stress. For an isotropic material the additional condition $C_{44} = (C_{11} - C_{12})/2$ holds. An isotropic material for which the Cauchy relation holds has $C_{11} = 3C_{12}$.

The elastic constants are related to the dispersion of acoustic phonons for $\mathbf{k} \rightarrow 0$ by the general rule (in tensor, not Voigt, notation):

$$\rho \omega^2(\mathbf{k}) u_\alpha = \sum_{\beta,\gamma,\delta} c_{\alpha\beta,\gamma\delta} k_\beta k_\gamma u_\delta \quad (77)$$

for a phonon with polarization \mathbf{u} in a material whose density is ρ . For a cubic crystal the relation between elastic constants and the sound velocity $v_s = \omega(\mathbf{k})/k$ becomes in some cases really simple:

$$v_s^2 = \frac{C_{11}}{\rho} \text{ se } \mathbf{k} = (100), \quad \mathbf{u} = (100) \quad (78)$$

$$v_s^2 = \frac{C_{11} - C_{12}}{2\rho} \quad \text{se } \mathbf{k} = (110), \quad \mathbf{u} = (1\bar{1}0) \quad (79)$$

$$v_s^2 = \frac{C_{44}}{\rho} \quad \text{se } \mathbf{k} = (110), \quad \mathbf{u} = (001) \quad (80)$$

Note that under applied pressure the definition of elastic constants is no longer unique: they can be defined from the stress-strain relation or from the energy, and the result is not the same. The elastic constants that are usually measured, in Brillouin and ultrasonic experiments, are obtained from the stress-strain relation:

$$c_{\alpha\beta,\gamma\delta} = \frac{\partial \sigma_{\alpha\beta}}{\partial \epsilon_{\gamma\delta}}. \quad (81)$$

They are also known as “effective” elastic constants. These can be expressed in terms of the energy and of the applied stress as

$$c_{\alpha\beta,\gamma\delta} = c'_{\alpha\beta,\gamma\delta} + \frac{1}{2} [(\sigma_{\alpha\delta}\delta_{\beta\gamma} + \sigma_{\beta\delta}\delta_{\alpha\gamma} + \sigma_{\alpha\gamma}\delta_{\beta\delta} + \sigma_{\beta\gamma}\delta_{\alpha\delta}) - 2\sigma_{\alpha\delta}\delta_{\beta\gamma}] \quad (82)$$

where $c'_{\alpha\beta,\gamma\delta}$ are the “isothermal” (or “adiabatic”, the difference is negligible) elastic constants:

$$c'_{\alpha\beta,\gamma\delta} = \frac{1}{V} \frac{\partial^2 E}{\partial \epsilon_{\alpha\beta} \partial \epsilon_{\gamma\delta}}. \quad (83)$$

The effective elastic constants are instead obtained from the enthalpy $H = E + PV$:

$$c_{\alpha\beta,\gamma\delta} = \frac{1}{V} \frac{\partial^2 H}{\partial \epsilon_{\alpha\beta} \partial \epsilon_{\gamma\delta}}. \quad (84)$$

The difference between the two sets comes from the dependence of the volume on the strain. In a cubic crystal under hydrostatic pressure, $\sigma_{\alpha\beta} = -P\delta_{\alpha\beta}$ and we obtain in Voigt notation

$$c_{11} = c'_{11} - P; \quad c_{12} = c'_{12} + P; \quad c_{44} = c'_{44} - P. \quad (85)$$

The Cauchy relations hold unchanged for the isothermal, not the effective, elastic constants. For the latter the Cauchy relations in the cubic case becomes $c_{12} = c_{44} + 2P$. See the comment of Brazhkin and Lyapin, Phys.Rev.Lett. **78**, 2493 (1997).