

# Molecular Dynamics for Two-Body Potential from Unobserved Gaussian Regression

Maharavo Randrianarivony<sup>1</sup>

<sup>1</sup> Virtual Material Design, Schloss Birlinghoven, Sankt Augustin, Germany

Correspondence: Maharavo Randrianarivony, Virtual Material Design (Fraunhofer Inst.), Institute for Numerical Simulation, Bonn 53115, Germany. Tel: 49-228-733-845. E-mail: maharavo.randrianarivony@scai.fraunhofer.de

Received: July 31, 2014 Accepted: August 17, 2014 Online Published: August 31, 2014

doi:10.5539/jmsr.v3n4p38 URL: <http://dx.doi.org/10.5539/jmsr.v3n4p38>

## Abstract

A molecular dynamic simulation requires the local potential energy per atom because it provides the force applied to each atom. On the other hand, an electronic structure computation provides only the global potential energy of an atomic system. We propose a stochastic fitting for deducing a potential from a total energy which can be incorporated inside a molecular dynamic program. The objective of that fitting process is twofold. First, the total energy can be reproduced with a sufficient accuracy. In addition, we want to separate the potential energy into atomic contributions. The only inputs in the stochastic regression are the total energies of the atomic systems. As for the applications, we examine the performance of the method with the help of a mathematical model. Afterward, we use it for quantum applications in which we compare the direct method and the unobserved fitting with respect to the energy conservations. For a molecular dynamic application, we examine the atomic cluster formations during freezing when the proposed potential is used. Although the method is generally applicable, we restrict in this paper to total energies which are reproduced from two-body potentials for the molecular dynamic simulation.

**Keywords:** Unobserved fitting, Gaussian, molecular dynamics, potential energy

## 1. Introduction

Multiscale computation (Tadmor & Miller, 2011) is the sequence of simulations from nanoscale quantum mechanics till macroscopic simulation. This sequence usually consists of electronic structure, molecular dynamics (MD), mesoscopic simulation, quasi-continuum, till FEM/CAD (Finite Element Method/Computer Aided Design). The most important step in standard MD packages (Allen & Tildesley, 1989; Ciesla et al., 2007; Kress et al., 1998) is the determination of the local forces which are derived from the potential energy. There are numerous empirical potentials including Lennard-Jones, Finnis-Sinclair, ReaxFF(Reactive Force Field), REBO(Reactive Empirical Bond Order) and BOP(Bond Order Potential) among others. In Figure 1(b) we illustrate a molecular dynamic simulation of a graphene fracture using REBO potentials where the colors indicate the potential energy per atom. In this document, we would like to consider the part of multiscale computation between electronic structure such as DFT and molecular dynamics which is graphically illustrated in Figure 1(a). The main problem in bridging that gap is the separation of atomic contributions of the potential energy. In fact, a quantum packet provides only the total energy of a given atomic configuration whereas a molecular dynamic package requires the atomic contribution in order that the force applied to each atom can be computed at every molecular dynamic step. We are interested in molecular dynamics where the potential is obtained from a prefiting process which consists in splitting the potential energy into atomic local energies. We propose a stochastic method whose only observed inputs are the total energy while the objective is to reconstruct the local energy per atom.

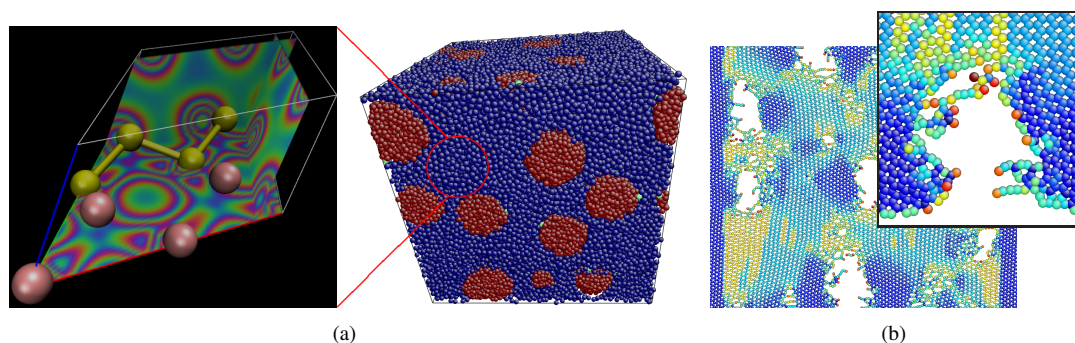


Figure 1. Motivation: (a)Local energy computation and molecular dynamic simulation (b)Potential energy in a graphene fracture

The method presented here is based upon machine learning which uses the concept of learning on the fly. It means in general that when an atom or a configuration passes in one state, a certain computation is performed. Upon passing on that state for the next times, the computation is not repeated any more but experiences from previous computation is used. The machine learning process generates a structure (Seeger et al., 2003) which enables the efficient and fast access of such a methodology. Our motivation is to generate a system which is both accurate and fairly inexpensive to evaluate. The proposed method is based on stochastic method where the expectancy and the variance of the prediction need to be computed. One utilizes the Matern correlation function for the assembly of the covariance matrices. The stochastic posterior depends only on the input observed total energies. As a consequence, it can be computed once for all and it is looked up during future computations. For an MD simulation, what is really required is not the energy but the force derived from that energy. Hence, the derivative of the prediction will be computed analytically. That is important because using Finite Difference is known to diverge the energy conservation in MD. During the simulation, one needs to solve a nonlinear optimization pertaining to the logarithm marginal likelihood. We present a method for avoiding the singularity at the diagonal of the covariance matrix. Our method can be applied to all sorts of quantum information but we mainly focus on the atomic energy in this paper.

The practical implementation is divided into two applications. The first one is the investigation of the method by using some mathematical system. In that case, we consider some exact solution and we reconstruct the internal local energy function by using our theoretical method. The purpose of that test is to examine the performance of the method by comparing the exact solution with the solution provided by the fitting. The local energies are expressed in term of the Coulomb matrix. As for the second application which is based on quantum systems, this paper presents only a preliminary work because evaluating DFT is a very time-consuming process. As a consequence, we compare an MD using Lennard-Jones and an MD where the potential is obtained from unobserved Gaussian regression. In order to gauge the simulation quality, we investigate the radial distribution function (RDF) for various values of the temperature by using the potential which was obtained from the unobserved fitting. The general tendency of the RDF curve is to start from zero values followed by a sudden peak which in turn is followed by a slow approximation of the unit value. In general, the height of the peak depends on the temperature. For very cold temperature, the peak is remarkably high while it is barely observable when the temperature is very warm. To validate the use of the atomic separation approach, we investigate the results of the MD simulation at freezing temperatures. The formation of clusters of atoms which features the phenomenon of freezing temperatures as described in (Chen, 2013; Yamaguchi & Maruyama, 1998) will be practically reconstructed. The adjustment of the temperature value is regulated by using the method of kinetic energy scaling. By using the unobserved Gaussian regression, we achieve total energy conservation as one observes the energy values at many MD steps. In addition, we will compare the energy values by using the direct method and the proposed atomic separation method.

## 2. Theoretical Methodology

### 2.1 Potential Energy w.r.t. Nuclei Coordinates

Consider a system composed of  $n$  atoms  $\mathcal{R} = \{\mathbf{r}_i\}_{i=1}^n$  and  $m$  electrons  $\mathcal{X} = \{\mathbf{x}_i\}_{i=1}^m$ . The main problem in electronic structure computation involves the general Hamiltonian (Tadmor & Miller, 2011; Hill et al., 2005; Saad et al.,

2010) operator

$$\mathcal{H} = - \sum_{k=1}^n \frac{\hbar^2}{2M_k} \Delta_{\mathbf{r}_k} - \frac{\hbar^2}{2m_e} \sum_{i=1}^m \Delta_{\mathbf{x}_i} - \sum_{i=1}^m \sum_{j=1}^n \frac{q_e Z_j}{\|\mathbf{x}_i - \mathbf{r}_j\|} + \frac{1}{2} \sum_{i=1}^n \sum_{i \neq j}^n \frac{Z_i Z_j}{\|\mathbf{r}_i - \mathbf{r}_j\|} + \frac{1}{2} \sum_{i=1}^m \sum_{j \neq i}^m \frac{q_e^2}{\|\mathbf{x}_i - \mathbf{x}_j\|} \quad (1)$$

where  $\hbar$  is the Planck constant divided by  $2\pi$ ,  $q_e$  is the electronic charge,  $m_e$  and  $M_k$  represent the electron mass and the mass of the  $k$ -th nucleus while the charge of the  $j$ -th nucleus is  $Z_j = |q_e|n_j$  in which  $n_j$  is the atomic number. The wave function  $\Psi$  is the eigenfunction corresponding to the smallest eigenvalue of the eigenvalue problem

$$\mathcal{H}\Psi = E_{\min}\Psi, \quad \Psi = \Psi(\mathbf{r}_1, \dots, \mathbf{r}_n, \mathbf{x}_1, \dots, \mathbf{x}_m). \quad (2)$$

The kinetic energies of the nucleus and electron are

$$\langle \Psi | \mathcal{T}_A | \Psi \rangle, \quad \langle \Psi | \mathcal{T}_E | \Psi \rangle \quad \text{where} \quad \mathcal{T}_A := - \sum_{k=1}^n \frac{\hbar^2}{2M_k} \nabla_{\mathbf{r}_k}^2, \quad \mathcal{T}_E := - \frac{\hbar^2}{2m_e} \sum_{i=1}^m \nabla_{\mathbf{x}_i}^2. \quad (3)$$

The third summations of (1) provide the nucleon-electron Coulomb attraction energy. The two last summations are the inter-nuclei and inter-electron repulsion energies. We shall define the atom-electron interaction, the inter-nucleon operator and the inter-electron interaction operators  $\mathcal{V}_{AE}$ ,  $\mathcal{V}_{AA}$  and  $\mathcal{V}_{EE}$  as follows

$$\mathcal{V}_{AE}(\mathbf{x}_1, \dots, \mathbf{x}_m, \mathbf{r}_1, \dots, \mathbf{r}_n) := - \sum_{i=1}^m v_{AE}(\mathbf{x}_i), \quad (4)$$

$$\mathcal{V}_{AA}(\mathbf{r}_1, \dots, \mathbf{r}_n) := \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n Z_i Z_j v(\mathbf{r}_i, \mathbf{r}_j), \quad (5)$$

$$\mathcal{V}_{EE}(\mathbf{x}_1, \dots, \mathbf{x}_m) := \frac{1}{2} \sum_{j=1}^m \sum_{i=1}^{j-1} q_e^2 v(\mathbf{x}_i, \mathbf{x}_j), \quad (6)$$

where

$$v_{AE}(\mathbf{x}) := \sum_{j=1}^n \frac{q_e Z_j}{\|\mathbf{x} - \mathbf{r}_j\|}, \quad v(\mathbf{u}, \mathbf{u}') := \frac{1}{\|\mathbf{u} - \mathbf{u}'\|}. \quad (7)$$

Thus, one has

$$\mathcal{H} = \mathcal{T}_A + \mathcal{T}_E + \mathcal{V}_{AE} + \mathcal{V}_{AA} + \mathcal{V}_{EE}. \quad (8)$$

We use the *Born-Oppenheimer* or adiabatic approximation assumption stating that the mass and the volume of the atoms are very large in comparison to those of the electrons. In fact, an electron is 2000 times lighter than any nucleus such that the atoms move comparatively slower than the electrons. Thus, electrons are supposed to follow the movements of the atoms in the sense that

$$\Psi(\mathbf{r}_1, \dots, \mathbf{r}_n, \mathbf{x}_1, \dots, \mathbf{x}_m) = \Phi(\mathbf{r}_1, \dots, \mathbf{r}_n) \psi_{\mathcal{R}}(\mathbf{x}_1, \dots, \mathbf{x}_m) \quad \text{for each} \quad \mathcal{R} = (\mathbf{r}_1, \dots, \mathbf{r}_n). \quad (9)$$

As a consequence, we treat the time-independent Hamiltonian operator with respect to the a set of nuclei  $\mathbf{r}_i$  which are supposed to be stationary. That is to say, for each given  $\mathcal{R} = (\mathbf{r}_1, \dots, \mathbf{r}_n)$ , the initial electronic structure is reduced to the expression

$$\mathcal{H}_{\text{B.O.}} = - \frac{\hbar^2}{2m_e} \sum_{i=1}^m \Delta_{\mathbf{x}_i} - \sum_{i=1}^m \sum_{j=1}^n \frac{q_e Z_j}{\|\mathbf{x}_i - \mathbf{r}_j\|} + \frac{1}{2} \sum_{j=1}^m \sum_{i=1}^{j-1} \frac{q_e^2}{\|\mathbf{x}_i - \mathbf{x}_j\|}. \quad (10)$$

Thus, the eigenproblem corresponding to the above Born-Oppenheimer Hamiltonian operator becomes

$$(\mathcal{T}_E + \mathcal{V}_{AE} + \mathcal{V}_{EE})\psi_{\mathcal{R}}(\mathbf{r}_1, \dots, \mathbf{r}_n) = E_{\mathcal{R}}\psi_{\mathcal{R}}(\mathbf{r}_1, \dots, \mathbf{r}_n) \quad (11)$$

in which the ground state energy  $E_{\mathcal{R}}$  and the reduced wave function  $\psi_{\mathcal{R}}$  depend on the given nuclear coordinates  $\mathcal{R} = (\mathbf{r}_1, \dots, \mathbf{r}_n)$ . Due to the independence of  $\Phi$  on  $\mathcal{X} = \{\mathbf{x}_i\}$  in (9), one deduces from (11) that

$$\mathcal{H}\Phi(\mathcal{R})\psi_{\mathcal{R}}(X) = (\mathcal{T}_{\mathbf{A}} + \mathcal{T}_{\mathbf{E}} + \mathcal{V}_{\mathbf{AE}} + \mathcal{V}_{\mathbf{AA}} + \mathcal{V}_{\mathbf{EE}})\Phi(\mathcal{R})\psi_{\mathcal{R}}(X) \quad (12)$$

$$= (\mathcal{T}_{\mathbf{A}} + \mathcal{V}_{\mathbf{AA}})\Phi(\mathcal{R})\psi_{\mathcal{R}}(X) + E_{\mathcal{R}}\Phi(\mathcal{R})\psi_{\mathcal{R}}(X) \quad (13)$$

$$= (\mathcal{T}_{\mathbf{A}} + \mathcal{V}_{\mathbf{AA}} + E_{\mathcal{R}})\Phi(\mathcal{R})\psi_{\mathcal{R}}(X). \quad (14)$$

By using (2) and (9) in the last equality, one obtains

$$(\mathcal{T}_{\mathbf{A}} + \mathcal{V}_{\mathbf{AA}} + E_{\mathcal{R}})\Phi(\mathcal{R})\psi_{\mathcal{R}}(X) = E_{\min}\Phi(\mathcal{R})\psi_{\mathcal{R}}(X). \quad (15)$$

By assuming the independence of  $\psi_{\mathcal{R}}(X)$  on the atomic kinetic term such as

$$\mathcal{T}_{\mathbf{A}}(\Phi(\mathcal{R})\psi_{\mathcal{R}}(X)) = \psi_{\mathcal{R}}(X)\mathcal{T}_{\mathbf{A}}\Phi(\mathcal{R}) \quad (16)$$

one obtains from (15)

$$\psi_{\mathcal{R}}(X)[(\mathcal{T}_{\mathbf{A}} + \mathcal{V}_{\mathbf{AA}} + E_{\mathcal{R}})\Phi(\mathcal{R})] = E_{\min}\Phi(\mathcal{R})\psi_{\mathcal{R}}(X). \quad (17)$$

Thus, one obtains the Schrödinger equation with respect to the nuclear coordinates only

$$(\mathcal{T}_{\mathbf{A}} + \mathcal{V}_{\text{pot}})\Phi(\mathbf{r}_1, \dots, \mathbf{r}_n) = E_{\min}\Phi(\mathbf{r}_1, \dots, \mathbf{r}_n), \quad (18)$$

in which

$$\mathcal{V}_{\text{pot}} := \mathcal{V}_{\mathbf{AA}} + E_{\mathcal{R}} \quad \text{such that} \quad E_{\text{pot}} := \langle \Psi | \mathcal{V}_{\text{pot}} | \Psi \rangle. \quad (19)$$

For the molecular dynamic applications, one uses  $\mathcal{V}_{\text{pot}}$  which is the effective inter-atomic potential or the potential energy surface. With the help of the quantum assumptions above, that last equation takes only the nuclei coordinates  $\mathbf{r}_1, \dots, \mathbf{r}_n$  of the atoms as variables. Since it is generally very expensive to evaluate the function

$$(\mathbf{r}_1, \dots, \mathbf{r}_n) \mapsto E_{\text{pot}}(\mathbf{r}_1, \dots, \mathbf{r}_n), \quad (20)$$

one uses a simplification or an approximation of  $E_{\text{pot}}$ . It is impossible to express that function analytically for large MD systems which consist of multi-millions atoms. Efficient empirical approximations of (20) for extremely large systems include: Finnis-Sinclair, EAM (Embedded Atom Method) and BOP (Bond Order Potential). Exact solutions for the above equation are only known for very few special cases. Hence, numerical methods must be used for the general cases. The most direct method is to consider the domain of computation  $\Omega \subset \mathbb{R}^3$  which is supposed to be sufficiently larger than the nuclei cloud  $\{\mathbf{r}_j\}_{j=1}^n$  such that the above operator has neglecting influence beyond it. Every electron  $\mathbf{x}_i \in \Omega$  admits a spin  $\sigma_i$  which is up ( $\sigma_i := +1/2$ ) or down ( $\sigma_i := -1/2$ ) so that one has as variables  $((\mathbf{x}_1, \sigma_1), \dots, (\mathbf{x}_m, \sigma_m)) \in (\Omega \times \{-1/2, +1/2\})^m$ . It is possible to have all spins up or all spins down such as  $((\mathbf{x}_1, \sigma), \dots, (\mathbf{x}_m, \sigma))$ . Usually, the electron spins are used implicitly and they are omitted in the notations. It is possible to consider the spins as part of the variables or to fix them beforehand. It is also possible to have spinless computations which have their limitations (Ciarlet et al., 2003). For the eigenvalues, one has

$$\mathcal{H}_{\mathbf{B.O.}}\Psi(\mathbf{x}_1, \dots, \mathbf{x}_m) = E\Psi(\mathbf{x}_1, \dots, \mathbf{x}_m) \quad (21)$$

such that  $\Psi$  is antisymmetric  $\Psi(\dots, \mathbf{x}_i, \dots, \mathbf{x}_j, \dots) = -\Psi(\dots, \mathbf{x}_j, \dots, \mathbf{x}_i, \dots)$  for all distinct  $i, j = 1, \dots, m$ . The antisymmetrization operator  $\mathcal{A}$  applied to any  $m$ -variate function  $f$  is defined by  $\mathcal{A}(f) := \frac{1}{m!} \sum_{P \in \Pi_m} \text{sgn}(P)f \circ P$  in which  $\Pi_m$  is the set of permutations over  $\{1, \dots, m\}$  and  $\text{sgn}(P)$  designates the signature of a permutation  $P$ . The antisymmetrization operator  $\mathcal{A}$  has the properties that it commutes with the Hamiltonian operator and that for an  $m \times m$  matrix  $\mathbf{M}$  one has  $\mathcal{A}(\mathbf{M}\Phi) = \det(\mathbf{M})\mathcal{A}(\Phi)$ . The Hartree-Fock approach is the variational formulation on the Hamiltonian operator (10) where the trial functions are antisymmetric functions. Some simplifications of the stationary Hamilton operators have already been proposed. For the DFT(Density Functional Theory), one solves a set of equations for each electron. The similarity of the solutions is then derived from the theory of Kohn-Sham (Kohn & Sham, 1965). The Kohn-Sham formalism consists in replacing the complicated initial problem into several ones. For each  $i = 1, \dots, m$

$$\left(-\frac{1}{2}\Delta + V_{\text{eff}}(\mathbf{x})\right)\psi_i(\mathbf{x}) = E_i\psi_i(\mathbf{x}) \quad (22)$$

where  $V_{\text{eff}}$  is the effective potential energy which depends implicitly on the total electron density  $\rho(\mathbf{x}) = \sum_{i=1}^m |\psi_i(\mathbf{x})|^2$  such that  $V_{\text{eff}}(\mathbf{x}) = V_{\text{eff}}[\rho(\mathbf{x})](\mathbf{x})$ . The problem is then reduced from dimensions  $3m$  to  $m$  sets of  $3D$  smaller problems. The influence of one electron with respect to the other electrons is measured by the total electron density.

These approaches enable the treatment of Hamiltonian problem even for an electronic structure admitting a large number of particles on a single desktop. The eigenvalue problem in (22) is nonlinear because its variational operator  $\langle \Psi | -\frac{1}{2}\nabla^2 + V_{\text{eff}}(\mathbf{x}) | \Psi \rangle$  depends on  $\rho$  which in turn depends on  $\psi_i$ . It is solved by using a sequence of the linear eigenvalue problems SCF (Self Consistent Field). The effective potential is constituted of the Hartree potential  $V_H$ , the exchange correlation potential  $V_{XC}$  and the external electrostatic field such as

$$V_{\text{eff}}[\rho(\mathbf{x})](\mathbf{x}) = V_H[\rho(\mathbf{x})](\mathbf{x}) + V_{XC}[\rho(\mathbf{x})](\mathbf{x}) + V_{\text{ext}}[\rho(\mathbf{x})](\mathbf{x}) \quad (23)$$

in which the Hartree potential is the inverse of the Poisson operator such as  $\Delta V_H(\mathbf{x}) = -4\pi\rho(\mathbf{x})$ . The main feature of DFT is that one has to approximate the potential by using some correction terms known as exchange-correlation potential (Perdew & Wang, 1992; Perdew & Zunger, 1981). That is usually done by LDA (Local Density Approximation) or GGA (Generalized Gradient Approximation). Those expressions contain some parameters which are optimized or obtained from some experimental measurements in term of the Weigner-Seitz radius. Analytic expressions of the correlation energy are only known in a few special cases which mainly consist of the high and low density limits. The external electrostatic field potential  $V_{\text{ext}}$  is provided by the kernel  $\sum_{i=1}^n z_i/||\mathbf{x} - \mathbf{x}_i||$ . The above exchange-correlation potential is related to the exchange-correlation energy by  $V_{XC} = \delta E_{XC}/\delta\rho$  where one expresses  $E_{XC} = E_X + E_C$  as the exchange and the correlation parts. In term of the exchange-correlation energy density  $\varepsilon_{XC}$  one has

$$E_{XC}[\rho] = \int \varepsilon_{XC}[\rho](\mathbf{x})\rho(\mathbf{x})d\mathbf{x} \quad \text{where} \quad \varepsilon_{XC} = \varepsilon_X + \varepsilon_C. \quad (24)$$

For the local density approximation (LDA), the exchange energy density is expressed as  $\varepsilon_X^{LDA}(\rho) = -0.75(3\rho(\mathbf{x})/\pi)^{1/3}$  so that  $E_X[\rho] = -0.75(\frac{3}{\pi})^{1/3} \int \rho(\mathbf{x})^{4/3}d\mathbf{x}$ . Once the solution  $E_i$  to (22) for all  $i = 1, \dots, m$  becomes known, the Kohn-Sham approach uses the approximation to  $E$  of (1) by

$$E_{\text{KS}} = \sum_{i=1}^m E_i - \frac{1}{2} \int \rho(\mathbf{x})V_H(\mathbf{x}) + \int \rho(\mathbf{x})(E_{XC}[\rho(\mathbf{x})] - V_{XC}[\rho(\mathbf{x})])d\mathbf{x}. \quad (25)$$

The main improvement from LDA to GGA is that the exchange-correlation energy does not depend only on the total electron density but also on its gradient such as  $E_{XC}^{\text{GGA}} = E_{XC}^{\text{GGA}}[\rho, \nabla\rho](\mathbf{x})$ . In the computations, one represents  $\Psi$  as a set of basis which are usually plane waves (Wimmer et al., 1981), Finite Element Method (Toda et al., 2010; Suryanarayana et al., 2010) or wavelets (Genovese et al., 2011).

## 2.2 Unobserved Information and Gaussian Regression

In this section, we will describe the main points about the unobserved Gaussian regression. We are going beyond the usual reconstruction of a function  $f$  by using the observations  $\{y_i\}$  such that  $y_i = f(\mathbf{x}_i) + \varepsilon_i$  because  $y_i$  is unobserved in our situation. During the data fitting, we do not reconstruct a global function with respect to the atom coordinates. Instead, we use implicit variables  $\mathbf{q}$  of dimension  $D$ . That is to say, for every atomic system  $\mathcal{R}$  having  $n$  atoms we have the new variables  $\mathbf{q}^\alpha(\mathcal{R}) = [q_1^\alpha(\mathcal{R}), \dots, q_D^\alpha(\mathcal{R})] \in \mathbb{R}^D$  for each atom  $\mathbf{x}_\alpha \in \mathcal{R}$ . In this theoretical description,  $\mathbf{q}^\alpha(\mathcal{R})$  is supposed to be a general mapping depending on the atom coordinates. But for the applications, they might have some physical significations such as the Coulomb matrix which we will encounter toward the end of this paper. We consider many observations involving atomic systems  $\{\mathcal{R}_i\}$  of number  $N$ . For each observation, we have a corresponding total energy  $E_{\text{tot}}^i$  such that

$$E_{\text{tot}}^i = \sum_{\alpha=1}^{n_i} E_{\text{loc}}(q_1^\alpha(\mathcal{R}_i), \dots, q_D^\alpha(\mathcal{R}_i)) + \varepsilon_i, \quad \forall i = 1, \dots, N \quad (26)$$

in which  $\varepsilon_i$  are measurement imperfections. The observed data are  $\{E_{\text{tot}}^i\}$  which represent the total energy for quantum application while the unknown is the function  $E_{\text{loc}}$ . We sometimes use the term total energy to refer to the sum of all local energies. It is not to be confused with the usual total energy which is the sum of the kinetic energy and the potential energy. The quantum motivation for this kind of reconstruction is that we are interested in the local energies  $E_{\text{loc}}$  in applications including molecular dynamics, whereas it is impossible to obtain some samples  $\{E_{\text{loc}}^i\}$  from an electronic structure packet. The only values that a quantum simulator provides are a set of total energies  $\{E_{\text{tot}}^i\}$ . It is possible to provide gradient values  $\nabla E_{\text{tot}}$  and Hessian values  $\partial^2 E_{\text{tot}}/\partial x_p \partial x_q$  but we do not treat that yet in this current article because using the gradients and Hessians is computationally very intensive. Fitting using gradient and higher order derivatives will be reported in a subsequent paper. For the next stochastic

deduction, we will use the following Matern correlation function

$$k_{\text{MATERN}}(r) := \frac{2^{1-\nu}}{\Gamma(\nu)} \left( \frac{\sqrt{2\nu}r}{\ell} \right)^\nu K_\nu \left( \frac{\sqrt{2\nu}r}{\ell} \right) \quad (27)$$

where  $(\nu, \ell)$  are positive hyper-parameters and  $K_\nu$  is the modified Bessel function. In dimension  $D$ , the spectral density of the Matern function is

$$S(s) = \frac{2^D \pi^{D/2} \Gamma(\nu + D/2) (2\nu)^\nu}{\Gamma(\nu) \ell^{2\nu}} \left( \frac{2\nu}{\ell^2} + 4\pi^2 s^2 \right)^{(-\nu+D/2)}. \quad (28)$$

If  $\nu$  tends to infinity, one obtains the squared exponential case. For the special case of half-integers such as  $\nu = p + 1/2$ , one obtains a product of an exponential and a polynomial of order  $p$  such as

$$k_\nu(r) = \exp \left( -\frac{\sqrt{2\nu}r}{\ell} \right) \frac{\Gamma(p+1)}{\Gamma(2p+1)} \sum_{i=0}^p \frac{(p+i)!}{i!(p-i)!} \left( \frac{\sqrt{8\nu}r}{\ell} \right)^{p-i}. \quad (29)$$

In the particular cases where  $\nu = 3/2$  and  $\nu = 5/2$ , one has

$$k_{3/2}(r) = \exp \left( -\frac{\sqrt{3}r}{\ell} \right) \left( 1 + \frac{\sqrt{3}r}{\ell} \right), \quad (30)$$

$$k_{5/2}(r) = \exp \left( -\frac{\sqrt{5}r}{\ell} \right) \left( 1 + \frac{\sqrt{5}r}{\ell} + \frac{5r^2}{3\ell^2} \right). \quad (31)$$

The number of atoms in the  $i$ -th observation is  $n_i$  for  $i = 1, \dots, N$  and we denote  $M := \sum_{i=1}^N n_i$ . Let us denote by  $Q$  the following of implicit variable

$$Q := \left[ \begin{array}{cccc} q_1^1(\mathcal{R}_1) & q_2^1(\mathcal{R}_1) & \cdots & q_D^1(\mathcal{R}_1) \\ \vdots & \vdots & \vdots & \vdots \\ q_1^{n_1}(\mathcal{R}_1) & q_2^{n_1}(\mathcal{R}_1) & \cdots & q_D^{n_1}(\mathcal{R}_1) \\ \hline \vdots & \vdots & \vdots & \vdots \\ \hline q_1^1(\mathcal{R}_N) & q_2^1(\mathcal{R}_N) & \cdots & q_D^1(\mathcal{R}_N) \\ \vdots & \vdots & \vdots & \vdots \\ q_1^{n_N}(\mathcal{R}_N) & q_2^{n_N}(\mathcal{R}_N) & \cdots & q_D^{n_N}(\mathcal{R}_N) \end{array} \right] \begin{array}{l} \left. \vphantom{\begin{array}{c} q_1^1(\mathcal{R}_1) \\ \vdots \\ q_1^{n_1}(\mathcal{R}_1) \end{array}} \right\} \text{1-st observation} \\ \left. \vphantom{\begin{array}{c} \vdots \\ \vdots \end{array}} \right\} \text{Next observations} \\ \left. \vphantom{\begin{array}{c} q_1^1(\mathcal{R}_N) \\ \vdots \\ q_1^{n_N}(\mathcal{R}_N) \end{array}} \right\} \text{N-th observation} \end{array}$$

which is a matrix admitting the size  $(M \times D)$  where  $D$  is the dimension representing the number of implicit variables  $\mathbf{q}$  per atom. Let  $L$  be a binary matrix of size  $(M \times N)$  having entries 0 or 1 such that only the Coulomb entries in the  $i$ -th observation are relevant for the  $i$ -th row of  $L^T$ . Denote  $Q = [\mathbf{q}_i]_{i=1}^M$  so that one has the relation  $E_{\text{loc}}^i := E_{\text{loc}}(\mathbf{q}_i)$  where  $\{E_{\text{loc}}^i\}_{i=1}^M$  are unobserved samples. As a consequence, one has the following covariance relation for noise-free observations

$$K(\mathcal{R}_p, \mathcal{R}_s) = \text{COV}(E_{\text{tot}}(\mathcal{R}_p), E_{\text{tot}}(\mathcal{R}_s)) = \text{COV}(E_{\text{tot}}^p, E_{\text{tot}}^s) \quad (32)$$

$$= \text{COV} \left( \sum_{i=1}^M L^T[p, i] E_{\text{loc}}^i, \sum_{j=1}^M L^T[s, j] E_{\text{loc}}^j \right) \quad (33)$$

$$= \sum_{i,j=1}^M L^T[p, i] L^T[s, j] \text{COV}(E_{\text{loc}}^i, E_{\text{loc}}^j) \quad (34)$$

$$= \sum_{i,j=1}^M L^T[p, i] L[j, s] K(\mathbf{q}_i, \mathbf{q}_j) = (L^T K(Q, Q) L)[p, s]. \quad (35)$$

For a noisy data  $E_{\text{tot}}^i = E_{\text{tot}}(\mathcal{R}_i) + \varepsilon$  where the additional noise follows the Gaussian distribution  $\varepsilon \sim \mathcal{N}(0, \sigma_n^2)$ , one has the covariance

$$\text{COV}(E_{\text{tot}}^p, E_{\text{tot}}^s) = (L^T K(Q, Q)L)[p, s] + \sigma_n^2 \delta_{ps}. \quad (36)$$

Introduce the next vectors of observed and unobserved inputs

$$\mathbb{E}_{\text{tot}} = [E_{\text{tot}}^1, \dots, E_{\text{tot}}^N]^T, \quad \mathbb{E}_{\text{loc}} = [E_{\text{loc}}^1, \dots, E_{\text{loc}}^M]^T \quad \text{such that} \quad \mathbb{E}_{\text{tot}} = L^T \mathbb{E}_{\text{loc}}. \quad (37)$$

In the presence of noise, the marginal likelihood is the integral of the likelihood and the prior such as

$$p(\mathbf{y}|\mathbf{X}) = \int p(\mathbf{y}|\mathbf{f}) p(\mathbf{f}|\mathbf{X}) d\mathbf{f}. \quad (38)$$

If  $\mathbb{E}_{\text{tot}}|Q$  follows the Gaussian distribution  $\mathcal{N}(\mathbf{0}, L^T K(Q, Q)L)$  and one has no noise, then the marginal likelihood is

$$p(\mathbb{E}_{\text{tot}}|Q) = (2\pi)^{-N/2} |L^T K(Q, Q)L|^{-1/2} \exp \left[ -\frac{1}{2} \mathbb{E}_{\text{tot}}^T (L^T K(Q, Q)L)^{-1} \mathbb{E}_{\text{tot}} \right] \quad (39)$$

which implies by taking the logarithm marginal likelihood

$$\begin{aligned} p(\mathbb{E}_{\text{tot}}|Q) |L^T K(Q, Q)L|^{1/2} (2\pi)^{N/2} &= \exp \left[ -\frac{1}{2} \mathbb{E}_{\text{tot}}^T (L^T K(Q, Q)L)^{-1} \mathbb{E}_{\text{tot}} \right] \\ \log [p(\mathbb{E}_{\text{tot}}|Q)] + \frac{1}{2} \log |L^T K(Q, Q)L| + \frac{N}{2} \log(2\pi) &= -\frac{1}{2} \mathbb{E}_{\text{tot}}^T (L^T K(Q, Q)L)^{-1} \mathbb{E}_{\text{tot}} \end{aligned}$$

from which one obtains the marginal likelihood

$$\log [p(\mathbb{E}_{\text{tot}}|Q)] = -\frac{1}{2} \mathbb{E}_{\text{tot}}^T (L^T K(Q, Q)L)^{-1} \mathbb{E}_{\text{tot}} - \frac{1}{2} \log |L^T K(Q, Q)L| - \frac{N}{2} \log(2\pi). \quad (40)$$

In the presence of noise  $\varepsilon \sim \mathcal{N}(0, \sigma_n^2)$  by using (36), the log marginal likelihood becomes

$$\begin{aligned} \log [p(\mathbb{E}_{\text{tot}}|Q)] &= -\frac{1}{2} \mathbb{E}_{\text{tot}}^T (L^T K(Q, Q)L + \sigma_n^2 I_{N,N})^{-1} \mathbb{E}_{\text{tot}} \\ &\quad -\frac{1}{2} \log |L^T K(Q, Q)L + \sigma_n^2 I_{N,N}| - \frac{N}{2} \log(2\pi). \end{aligned} \quad (41)$$

The first term  $-\frac{1}{2} \mathbb{E}_{\text{tot}}^T (L^T K(Q, Q)L + \sigma_n^2 I_{N,N})^{-1} \mathbb{E}_{\text{tot}}$  corresponds to the *data-fit term*. The second term  $-\frac{1}{2} \log |L^T K(Q, Q)L + \sigma_n^2 I_{N,N}|$  is the *complexity term*. The last term  $-\frac{N}{2} \log(2\pi)$  is the *normalization term*.

For a given test value  $Q^*$ , the predicted estimation of the local energy follows the Gaussian distribution as

$$\begin{bmatrix} \mathbb{E}_{\text{loc}} \\ \mathbb{E}_{\text{loc}}^* \end{bmatrix} \sim \mathcal{N} \left( \mathbf{0}, \begin{bmatrix} L^T K(Q, Q)L + \sigma_n^2 I_{N,N} & L^T K(Q, Q^*) \\ K(Q^*, Q)L & K(Q^*, Q^*) \end{bmatrix} \right). \quad (42)$$

That is to say, for  $Q$  and  $\mathbb{E}_{\text{tot}}$ , the predictive distribution for a covariate vector  $Q^*$  is Gaussian having the following expectancy and variance

$$\begin{aligned} \text{EXP}[\mathbb{E}_{\text{loc}}^*|Q^*, Q] &= K(Q^*, Q)L [L^T K(Q, Q)L + \sigma_n^2 I_{N,N}]^{-1} \mathbb{E}_{\text{tot}} \\ \text{VAR}[\mathbb{E}_{\text{loc}}^*|Q^*, \mathbb{E}_{\text{tot}}, Q] &= K(Q^*, Q^*) - K(Q^*, Q)L [L^T K(Q, Q)L + \sigma_n^2 I_{N,N}]^{-1} L^T K(Q, Q^*) \end{aligned}$$

where those expressions are dependent on some set of hyper-parameters. The main objective is to theoretically optimize the marginal likelihood. In practice, using the log marginal likelihood is more efficient to implement. Now, we would like to investigate the derivative of the log marginal likelihood with respect to a hyper-parameter  $\theta$ . Let the value of  $r$  for given  $\mathbf{q}_p \in \mathbb{R}^D$  and  $\mathbf{q}_s \in \mathbb{R}^D$  be

$$r^2 = (\mathbf{q}_p - \mathbf{q}_s)^T \begin{bmatrix} 1/\lambda_1^2 & & \\ & \ddots & \\ & & 1/\lambda_D^2 \end{bmatrix} (\mathbf{q}_p - \mathbf{q}_s) \quad (43)$$

in which  $\lambda_1, \dots, \lambda_D$  are hyper-parameters which should be determined by using the log marginal likelihood. One has  $\theta = \log(\ell)$  or  $\theta = \log(\lambda_i)$  where  $i = 1, \dots, D$ . In order to apply an efficient nonlinear optimization to the former functional, we need the partial derivative of the inverse  $[L^T K_\theta(Q, Q)L]^{-1}$  with respect to a hyper-parameter  $\theta$ . From the fact that

$$\frac{\partial}{\partial \theta} [L^T K_\theta(Q, Q)L]^{-1} [L^T K_\theta(Q, Q)L] = \frac{\partial I_{N,N}}{\partial \theta} = 0, \quad (44)$$

one deduces

$$\begin{aligned} \frac{\partial}{\partial \theta} [L^T K_\theta(Q, Q)L]^{-1} &= -[L^T K_\theta(Q, Q)L]^{-1} \frac{\partial}{\partial \theta} [L^T K_\theta(Q, Q)L] [L^T K_\theta(Q, Q)L]^{-1} \\ &= -[L^T K_\theta(Q, Q)L]^{-1} L^T \frac{\partial K_\theta(Q, Q)}{\partial \theta} L [L^T K_\theta(Q, Q)L]^{-1} \end{aligned}$$

which cannot be simplified any further as  $L$  is not invertible. Similarly, one has the relation

$$\frac{\partial}{\partial \theta} \log |L^T K_\theta(Q, Q)L| = \text{trace} \left[ [L^T K_\theta(Q, Q)L]^{-1} \frac{\partial}{\partial \theta} [L^T K_\theta(Q, Q)L] \right]. \quad (45)$$

By using the Matern function  $k$  from (29), the partial derivative of a covariance matrix entry with respect to a hyperparameter  $\theta$  is

$$\partial_\theta [k(r(\mathbf{q}_p, \mathbf{q}_s))] = k'(r) \partial_\theta [r(\mathbf{q}_p, \mathbf{q}_s)], \quad \text{where} \quad r(\mathbf{q}_p, \mathbf{q}_s) = \sqrt{\sum_{i=1}^D \frac{1}{\lambda_i} (q_{p,i} - q_{s,i})^2}. \quad (46)$$

This is singular in the situation where  $\mathbf{q}_p = \mathbf{q}_s$  which occurs on the diagonal entries of the covariance matrix. In fact, one has

$$\partial_{\lambda_\mu} r(\mathbf{q}_p, \mathbf{q}_s) = -\frac{1}{r(\mathbf{q}_p, \mathbf{q}_s)} \left( \frac{q_{p,\mu} - q_{s,\mu}}{\lambda_\mu} \right)^2 \quad \text{for} \quad \mu = 1, \dots, D. \quad (47)$$

In order to suppress the singularity, one utilizes  $R(\mathbf{q}_p, \mathbf{q}_s) := r^2(\mathbf{q}_p, \mathbf{q}_s)$  such that  $\partial_{\lambda_\mu} R(\mathbf{q}_p, \mathbf{q}_s)$  is regular everywhere. In addition,  $\tilde{k}(R) := k(\sqrt{R})$  is smooth at the origin. Indeed, one has  $\tilde{k}(R) = p_{\text{MATERN}}(\sqrt{2\nu}r/\ell)$  such that

$$p_{\text{MATERN}}(t) = f_{\text{MATERN}}(t) \exp(-t) \quad (48)$$

$$p'_{\text{MATERN}}(t) = [f'_{\text{MATERN}}(t) - f_{\text{MATERN}}(t)] \exp(-t) \quad (49)$$

where  $f_{\nu=3/2} = 1 + t$  and  $f_{\nu=5/2} = 1 + t + (1/3)t^2$  so that

$$[f'_{\text{MATERN}}(t) - f_{\text{MATERN}}(t)] = tq_{\text{MATERN}}(t) \quad (50)$$

where  $q_{\text{MATERN}}$  is a polynomial. Therefore, one has

$$\begin{aligned} \frac{d\tilde{k}}{dR} &= \frac{\sqrt{2\nu}}{2\ell\sqrt{R}} \left[ f'_{\text{MATERN}}\left(\frac{\sqrt{2\nu}\sqrt{R}}{\ell}\right) - f_{\text{MATERN}}\left(\frac{\sqrt{2\nu}\sqrt{R}}{\ell}\right) \right] \exp\left(-\frac{\sqrt{2\nu}\sqrt{R}}{\ell}\right) \\ &= \frac{\sqrt{2\nu}}{2\ell} q_{\text{MATERN}}\left(\frac{\sqrt{2\nu}\sqrt{R}}{\ell}\right) \exp\left(-\frac{\sqrt{2\nu}\sqrt{R}}{\ell}\right) \end{aligned}$$

which is regular for all values of  $r(\mathbf{q}_p, \mathbf{q}_s)$ . The partial derivatives of the above estimation with respect to components of  $\mathbf{q}$  can be expressed analytically because we have used linear entities.

### 2.3 Molecular Dynamics of Large Systems

In general, by considering a real valued Lagrange function  $\mathcal{L}(t, x_1, \dots, x_n, v_1, \dots, v_n)$  and several continuously differentiable functions  $q_1, \dots, q_n$ , one obtains the following Euler-Lagrange equation for each  $k = 1, \dots, n$

$$\frac{\partial}{\partial x_k} \mathcal{L}(t, q_1(t), \dots, q_n(t), \dot{q}_1(t), \dots, \dot{q}_n(t)) - \frac{\partial}{\partial t} \frac{\partial}{\partial v_k} \mathcal{L}(t, q_1(t), \dots, q_n(t), \dot{q}_1(t), \dots, \dot{q}_n(t)) = 0. \quad (51)$$

We consider a time-dependent system consisting of  $n$  atoms where the  $k$ -th atom admits the Cartesian coordinates  $\mathbf{r}_k = \mathbf{r}_k(t)$  depending on time. By considering the MD Lagrange function

$$\mathcal{L}(t, \mathbf{r}_1, \dots, \mathbf{r}_n, \dot{\mathbf{r}}_1, \dots, \dot{\mathbf{r}}_n) := \frac{1}{2} \sum_{k=1}^n M_k \|\dot{\mathbf{r}}_k\|^2 - E_{\text{pot}}(\mathbf{r}_1, \dots, \mathbf{r}_n) \quad (52)$$



where  $M_k$  is the mass of the  $k$ -th ion, the Euler-Lagrange equation (51) yields the following equation of motion which is described as an ODE (ordinary differential equation)

$$M_k \ddot{\mathbf{r}}_{k,j}(t) = \frac{\partial E_{\text{pot}}}{\partial \mathbf{r}_{k,j}} =: -F_{k,j}(\mathbf{r}_1, \dots, \mathbf{r}_k, \dots, \mathbf{r}_n) \quad \forall k = 1, \dots, n \quad (53)$$

for each component  $j = 1, 2, 3$ . That expression is the second law of motion where  $[\ddot{r}_{k,1}(t), \ddot{r}_{k,2}(t), \ddot{r}_{k,3}(t)]$  denotes the acceleration while the force  $\mathbf{F}_k = [F_{k,1}, F_{k,2}, F_{k,3}]$  is applied on the  $k$ -th atom. By considering the kernel  $\mathcal{V}_{\text{pot}}$  of the potential energy, one should normally obtain the relation

$$\frac{\partial}{\partial \mathbf{r}_{k,j}} \langle \Psi | \mathcal{V}_{\text{pot}} | \Psi \rangle = 2 \left\langle \frac{\partial}{\partial \mathbf{r}_{k,j}} \Psi \middle| \mathcal{V}_{\text{pot}} \middle| \Psi \right\rangle + \left\langle \Psi \middle| \frac{\partial}{\partial \mathbf{r}_{k,j}} \mathcal{V}_{\text{pot}} \middle| \Psi \right\rangle \quad (54)$$

where  $\Psi$  corresponds to the ground state energy. But the Hellman-Feynmann rule states the independence of the force on the derivative of the ground state energy such that one eventually obtains

$$\frac{\partial}{\partial \mathbf{r}_{k,j}} \langle \Psi | \mathcal{V}_{\text{pot}} | \Psi \rangle = \left\langle \Psi \middle| \frac{\partial}{\partial \mathbf{r}_{k,j}} \mathcal{V}_{\text{pot}} \middle| \Psi \right\rangle. \quad (55)$$

Every MD iteration consists of the following steps. First, one determines the atomic force  $\mathbf{F}_k$ . Then, one has to solve the above equation of motion with the help of any ODE solver. The usual methods for the ODE solver are based on variants of Runge-Kutta or Verlet algorithm. Finally, one updates the atomic position  $\mathbf{r}_k$ . For an expression  $A = A(t)$  depending on  $t$ , we designate the time average by

$$\langle A \rangle = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t A(\theta) d\theta. \quad (56)$$

Since the MD trajectory involves only a large number of discrete values, we utilize the following discrete counterpart for  $L$  number of MD iterations

$$\langle A \rangle = \frac{1}{L} \sum_{k=1}^L A(\theta_k). \quad (57)$$

In the case of two-body potentials where the potential function  $V$  depends only on the inter-atomic distances  $r_{ij} = \|\mathbf{r}_i - \mathbf{r}_j\|$ , let us review some MD thermodynamic expressions. The pressure and the temperature are respectively defined by the expressions

$$P := \frac{1}{3\Omega} \left\langle \sum_{i=1}^n (M_i \mathbf{v}_i \cdot \mathbf{v}_i - \sum_{j>i} \frac{\partial V(r_{ij})}{\partial r_{ij}} r_{ij}) \right\rangle, \quad (58)$$

$$T := \frac{1}{3nk_B} \left\langle \sum_{i=1}^n M_i \mathbf{v}_i \cdot \mathbf{v}_i \right\rangle \quad (59)$$

where  $\Omega$ ,  $n$  and  $k_B$  are respectively the volume, the number of particles and the Boltzmann constant. Equation (59) relates the regulation of the temperature with the kinetic energy. The radial distribution function for nonzero radius  $r$  is

$$g(r) := \frac{1}{\rho 4\pi r^2 n} \left\langle \sum_{i=1}^n \sum_{j \neq i} \delta(r - \|\mathbf{r}_i - \mathbf{r}_j\|) \right\rangle \quad (60)$$

which quantifies the general distribution of the atoms with respect to their corresponding neighbors. We have in addition

$$\sum_{i<j} V(r_{ij}) = \int \sum_{i<j} V(r) \delta(r - r_{ij}) dr \quad (61)$$

$$= \rho 4\pi n \int \frac{r^2}{2} \frac{1}{\rho 4\pi r^2 n} \sum_{i=1}^n \sum_{j \neq i} V(r) \delta(r - r_{ij}) dr. \quad (62)$$

One obtains finally the total potential energy  $PE = 2\rho\pi n \int r^2 g(r) V(r) dr$  which can be used to verify the total energy conservation when the kinetic energy is added to it.

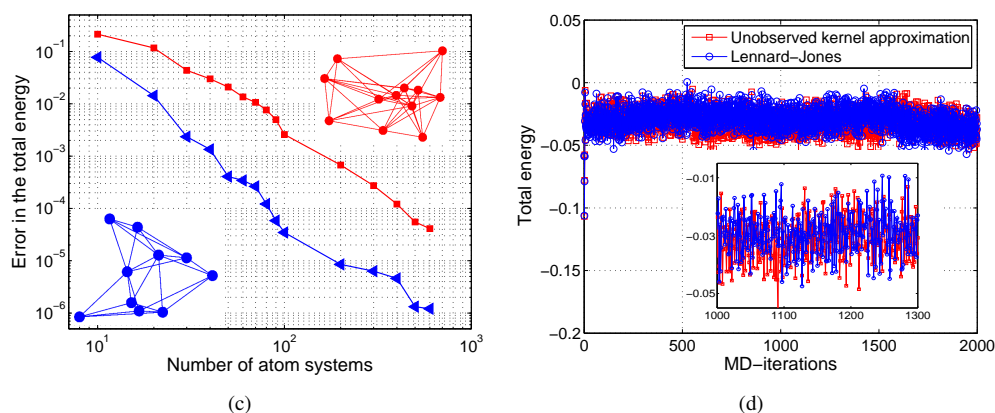


Figure 2. (a)Error in the total energy (b)Energy conservation: Lennard-Jones vs. unobserved fitting potential

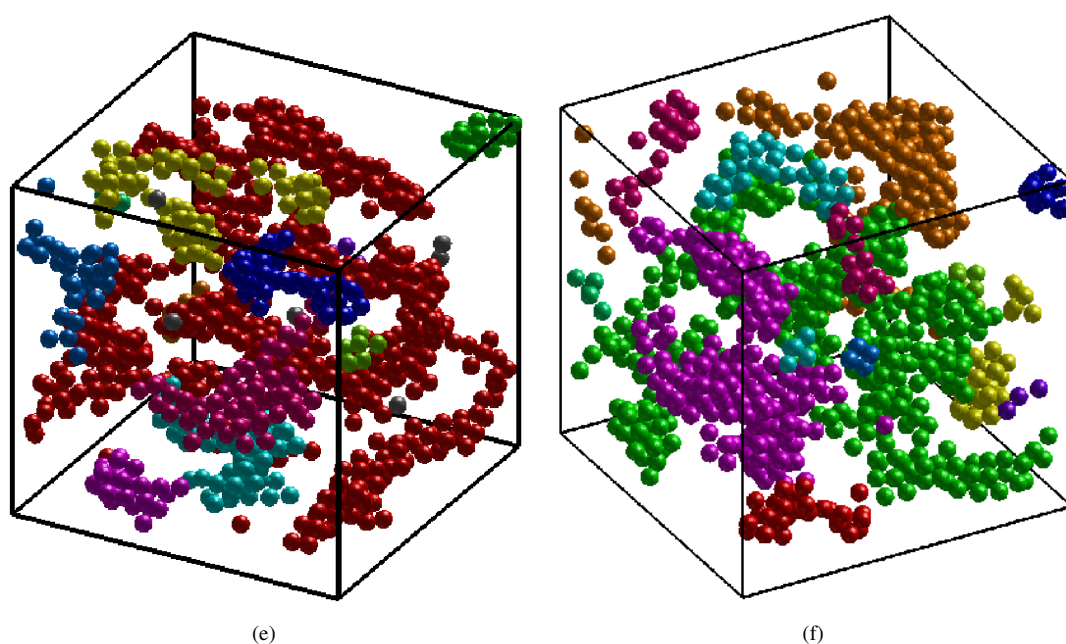


Figure 3. Atom cluster formations after freezing by using the potential from Gaussian unobserved fitting

### 3. Practical Results

In this section, we would like to give some details about the results of the practical implementation of the formerly described theoretical method. We will present results related to mathematical models as well as to quantum application for molecular dynamics. The implementation uses C/C++, BLAS/LAPACK and NLOPT. The BLAS packet is used for the fast vector operations. We use LAPACK for the linear operations such as Cholesky factorization and dense matrix solvers. The code is a very developed version of the matlab implementation provided in (Rasmussen & Williams, 2006). The new additional enhancements from the matlab version consist of the following features. First, using NLOPT provides a lot of improvements as compared to the original matlab nonlinear conjugate gradients. That can be observed when both the number of points and the dimension become large in the nonlinear optimization of the hyper-parameters. Furthermore, the gradient of the kernel-based approximation can be evaluated by using analytical expressions instead of using finite difference. The initial guess of the hyper-parameters is provided by the users. One can consider the determination of the final hyper-parameters as an unconstrained nonlinear optimization. We use NLOPT for the nonlinear operations (Nocedal, 1999) for the optimal hyper-parameters in the log marginal likelihood (41). NLOPT supports diverse nonlinear optimization operations (Johnson, 2010) in which local optimizers are involved. A local one searches only inside a neighborhood of a

certain provided starting initial guess. The optimizations are performed by using derivative-free or gradient-based algorithm which are available in NLOPT. Derivative-free algorithms include BOBYQA (Bond Optimization BY Quadratic Approximation), COBYLA (Constrained Optimization BY Linear Approximation), NEWUOA (NEW Unconstrained Optimization Algorithm). Gradient-based methods include MMA (Method of Moving Asymptotes) and LBFGB (Limited memory Broyden-Fletcher-Goldfarb-Shanno).

As a first application, we consider the reconstruction of a mathematical model where the exact solution is explicitly known. In an atomic system  $\mathcal{R}$  constituting of  $n$  atoms ( $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n$ ), the Coulomb matrix  $[q_{\alpha\beta}]$  admits the following entries for  $\alpha, \beta = 1, \dots, n$ :

$$\begin{cases} q_{\alpha\beta} = Z_\alpha Z_\beta / \|\mathbf{r}_\alpha - \mathbf{r}_\beta\| & \text{if } \alpha \neq \beta \\ q_{\alpha,\alpha} = 0.5 Z_\alpha^{2.4} \end{cases} \quad (63)$$

in which  $Z_\alpha$  is the nuclear charge of the  $\alpha$ -th atom. In our situation, we consider only homogeneous systems consisting of one element type such that  $Z_\alpha \equiv Z$  is a constant. We fix any positive number  $D$  which will then be the implicit dimension. That is to say, for each atom  $\mathbf{r}_\alpha$ , the implicit variables  $[q_1^\alpha(\mathcal{R}), \dots, q_D^\alpha(\mathcal{R})]$  will be the largest Coulomb entries involving the atom  $\mathbf{r}_\alpha$ . Except for the Coulomb entry  $q_i = 0.5 Z_i^{2.4}$ , the largest Coulomb values correspond to the closest atom neighbors. The purpose of the first application is to reconstruct the local energy  $E_{\text{loc}}$  which is a  $D$ -variate function. In order to be able to compare the reconstructed function with the original one, we consider an exact function which maps an atomic system  $\mathcal{R} = (\mathbf{r}_1, \dots, \mathbf{r}_n)$  to

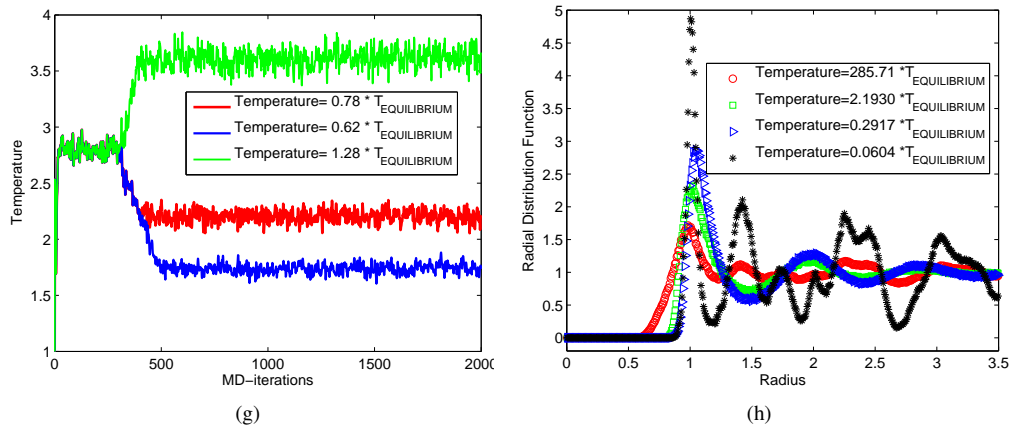


Figure 4. Using the unobserved Gauss reconstruction in reduced units (dimensionless): (a) Temperature adjustment and fluctuation (b) Radial distributed functions for different temperatures

$$E_{\text{tot}}(\mathcal{R}) = \sum_{\alpha=\text{Atom}} E_{\text{loc}}(q_1^\alpha(\mathcal{R}), \dots, q_D^\alpha(\mathcal{R})), \quad \text{where} \quad (64)$$

$$E_{\text{loc}}(\mathbf{q}) := \sum_{i=1}^D \sin(q_i), \quad \mathbf{q} = (q_1, \dots, q_D) \in \mathbb{R}^D. \quad (65)$$

It is possible to evaluate the errors with respect to the local energy or the global one. As in the real situation, the above exact function is unchanged by reordering the sequence of the atoms and the Coulomb matrix. The reconstruction of the implicit function  $E_{\text{loc}}$  from the total energy values follows the theoretical prediction from section 2.2. The inputs are the set of total energies  $\{E_i\}$  in which each total energy  $E_i$  corresponds to one atomic system  $\mathcal{R}_i$ . We do not put any restriction on the size of the atomic systems for that two different atomic systems  $\mathcal{R}_i$  and  $\mathcal{R}_j$  may contain different numbers of atoms and that  $D$  is not necessarily the whole  $n_i$ . On the other hand, the size  $D$  of the Coulomb values per atom has to remain constant during the whole simulation. Otherwise, it would be impossible to assemble the covariance matrices. For our simulation, each coordinate of the atom  $\mathbf{x}_\alpha \in \mathcal{R}$  is randomly defined.

Table 1. Errors in term of number of atomic systems with four Coulomb coordinates per atom

Nb. atomic systems	Implicit size	Size of atomic systems (tests)			
		12	13	14	15
10	138	2.1257e-01	2.0297e-01	1.9596e-01	1.9369e-01
20	272	3.8405e-02	3.5413e-02	3.3224e-02	3.2212e-02
50	682	1.2911e-03	1.0181e-03	9.0925e-04	8.5913e-04
100	1333	3.5975e-04	2.1320e-04	1.7760e-04	1.2839e-04
200	2700	5.7095e-05	2.7146e-05	1.1649e-05	9.6627e-06
300	4032	4.3027e-05	1.5299e-05	5.1034e-06	3.5984e-06
500	6717	1.6109e-05	5.5015e-06	2.5408e-06	1.5625e-06

As a first test for examining the accuracy, we consider several training atomic systems where the number of atoms in each system  $\mathcal{R}_i$  varies from  $n_i = 12$  to  $n_i = 15$  which is chosen by some random number generator. To test the accuracy, we consider additional 5000 atomic systems whose sizes also vary from 12 to 15. The results of the computations are collected in Table which contains the accuracy of the local energy  $E_{loc}$ . It shows the influence of the number  $N$  of training atomic systems and the implicit size which is the size of the involved covariance matrices. Note that the implicit size is much larger than the number of atomic systems so that much larger capacity of the computer memories are required than in the usual fitting (Rasmussen & Williams, 2006). We examine the accuracies for each group of atomic systems categorized according to their sizes  $n_i = 12, \dots, 15$ . In our case, the implicit dimension and the atomic number are  $D = 4$  and  $Z = 5$  respectively. We observe that we have generally a good accuracy improvement when the number  $N$  of the atomic systems increases irrespective of the size  $n_i$  of each atomic system because the value of closest neighbors  $D$  is fixed.

We would like now to investigate the influence of the implicit dimension  $D$  on the accuracy of the local energies. As in the former test from Table , we take again 5000 testing atomic systems. We vary the value of the dimension in the range  $D = 4$  till  $D = 7$  which controls the number of Coulomb values that are considered for each atom. The results of the test are collected in Table where it can be observed that the dimension affects the accuracy. A higher dimension is somewhat more difficult to approximate than a lower dimension. Nonetheless, it can be observed that the error reductions are somehow comparable for all dimensions. As the number of the atomic systems grows, the improvements of the accuracy are relatively of the same size for all considered dimensions. In order to observe the error reduction more clearly, we would like to analyze the accuracy of the total energy. In Figure 1(c), we display the error plots in function of the number of training atom systems. Both the axis for the number of atomic systems and the axis for the error in the total energy are logarithmically scaled. We examine two different cases admitting dimensions  $D = 4$  and  $D = 6$  which correspond respectively to the blue curve with triangular markers and the red curve with square markers. The results exhibit that the performance is efficient for both test cases. The case  $D = 6$  is somewhat more difficult to approximate than the case  $D = 4$  but both situations show that a good accuracy is possible to be reached as more and more atomic systems are added into the training configurations.

Table 2. Influence of the implicit dimension (Number of the closest neighbors)

Nb. atomic systems	Implicit size	Coulomb closest neighbors			
		$D = 4$	$D = 5$	$D = 6$	$D = 7$
10	150	2.0280e-01	3.8312e-01	6.2124e-01	1.0541e+00
20	300	3.8980e-02	1.6537e-01	3.4794e-01	8.0643e-01
50	750	1.0663e-03	7.9429e-03	5.8230e-02	1.3724e-01
80	1200	2.1809e-04	2.1177e-03	1.8412e-02	5.0782e-02
100	1500	7.0613e-05	1.0199e-03	6.2647e-03	3.6338e-02
400	6000	5.9077e-06	2.1401e-05	2.0990e-04	1.6834e-03
500	7500	2.2626e-06	9.4676e-06	9.9216e-05	8.4464e-04
600	9000	2.0130e-06	6.9972e-06	6.6619e-05	5.7867e-04

The purpose of the following test is to apply the potential energies obtained from the fitting process into a molecular dynamic model. We focus only on a two-body potential for the computation of the potential energy. In fact, we utilize the following Lennard-Jones potential energy (Tadmor & Miller, 2011) for the molecular dynamics:

$$E = \sum_{\beta \neq \alpha \in \mathcal{R}} E_{\text{LJ}}(\|\mathbf{r}_\alpha - \mathbf{r}_\beta\|) \quad \text{where} \quad E_{\text{LJ}}(r) := 4\epsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right]. \quad (66)$$

We employ reduced units so that all involved entities are dimensionless. For example, the temperature  $T$  is factored into  $Tk_B/\epsilon$ . In addition,  $\epsilon$  and  $\sigma$  admit unit values in our case while the method is still applicable to more general parameters. This corresponds to the general theory in (26) where the unknown function is  $E_{\text{LJ}}$ . We replace only the sum in (26) by a sum with two indices which can be organized lexicographically in order to obtain a sum with a single index. The implicit variables in this case are  $q_{\alpha,\beta} := \|\mathbf{r}_\alpha - \mathbf{r}_\beta\|$  so that we have the dimension  $D = 1$  for the two-body potential. The atomic forces at each MD step are obtained by taking the analytical partial derivatives of the Gaussian fitting which can be computed explicitly. Taking finite difference from the potential energy is known to possess a diverging energy conservation. The purpose of the test is to compare the MD simulation using the direct potential energy on the one hand and the one using the Gaussian fitting on the other. The total energy which is the sum of the kinetic energy and the potential energy should be kept approximately constant during the whole molecular dynamic steps. We want to compare the results pertaining to the energy conservations. The curves for the total energies are shown in Figure 1(d) for several MD steps. One can observe that the total energies for the two methods align well. The closeup shows that at each MD steps, the two results differ somewhat but the general energy conservations remain practically unchanged.

We want to examine the influence of the temperature on the whole molecular dynamic simulation. The adjustment of the temperatures is regulated by the scaling method of the kinetic energies which is a well-known method for regulating temperatures. In fact, one executes the molecular dynamic simulation for sufficient number of steps until the equilibrium is reached before one applies a constant scaling factor to the kinetic energy. An illustration of that situation is depicted on Figure 1(g) where 200 MD steps are executed to reach the equilibrium. Three different temperature adjustments are displayed there. One of them corresponds to a heating adjustment while the other two are used for cooling. These temperatures are respectively  $T = 0.78T_{\text{equil}}$ ,  $T = 0.62T_{\text{equil}}$  and  $T = 1.28T_{\text{equil}}$ . When the desired cooling or heating temperature values are attained, one applies again a normal MD simulation. In our case, the whole simulation consists of 2000 molecular dynamics steps. We want now to investigate the radial distribution function which was met in (60). We use the local potential energies which have been separated on atoms by using the former method. The radial distribution function quantifies the general distribution of each atom with respect to its neighbors. The values have been averaged with respect to the whole molecular dynamic trajectories. The results of the RDF tests are depicted on Figure 1(h) where the horizontal axis describes the average radius from an atom. We consider four situations which correspond to four different thermal values. In general for all values of the temperature, the RDF value is zero until the ideal radius is reached at the position where the RDF attains its peak. That is followed by some values which alternate in the neighborhood of the unity. The heights of the peak are very much affected by the temperature. The peak values are 1.7, 2.3, 2.8 and 4.9 for the temperature  $T = 285.71T_{\text{equil}}$ ,  $T = 2.1930T_{\text{equil}}$ ,  $T = 0.2917T_{\text{equil}}$  and  $T = 0.0604T_{\text{equil}}$  respectively. By using the proposed potential, one obtains the phenomenon of atom clustering in the case of freezing temperatures. By executing sufficiently many molecular dynamic steps, the atoms automatically form clusters by accumulating close to one another. In Figure 3, we can observe some instances of such cluster formations. Atoms which are close enough to one another are depicted with the same colors. That is another validation that the proposed Gaussian potential energies can efficiently simulate molecular dynamics. This clustering phenomenon can also be confirmed by the radial distribution functions in Figure 1(h) because a high peak averagely reflects the phenomenon of dense neighbors next to each atom.

#### 4. Conclusion

We presented an approach which approximates the total energy and which decomposes it into atomic contributions. The approach is based on stochastic method where the expectancy and the variance are computed. The main purpose is to determine the local energy function while the only observed inputs are the total energy. The method has been tested by using both a mathematical model and a molecular dynamic simulation. For the mathematical model, the accuracy between the exact solution and the fitting results has been investigated while we use two-body potentials for the molecular dynamic model. With the help of the new method, we simulated atom clustering in case of freezing temperatures. In addition, the radial distribution function has been examined for various values of the temperatures.

## Acknowledgments

This work was partially supported by Eurostars Project E!6935 funded by German Federal Ministry of Education and Research. In addition, the author thanks S. Mathias for valuable discussions.

## References

- Allen, M., & Tildesley, D. (1989). *Computer simulation of liquids*. Oxford University Press.
- Chen, L., Shibuta, Y., Kambara, M., & Yoshida, T. (2013). Molecular dynamics simulation of the role of hydrogenated *Si* clusters for fast rate mesoplasma epitaxy, *Journal of Physics D: Applied Physics*, 46(42), 425302. <http://dx.doi.org/10.1088/0022-3727/46/42/425302>
- Ciarlet, P., Lions, J., & Le Bris, C. (2003). *Handbook of numerical analysis: special volume: computational chemistry*. (Vol. X), Oxford University Press.
- Ciesla, M., Pawlowicz, J., & Longa, L. (2007). Molecular dynamics simulation of the Lennard-Jones polymers in a good solvent, *Acta Physica Polonica B*, 38(5), 1727-1736.
- Genovese, L., Videau, B., Ospici, M., Deutsch, T., Goedecker, S., & Méhaut, J. (2011). Daubechies wavelets for high performance electronic structure calculations: The BigDFT project Ondelettes de Daubechies pour le calculs de structure électronique à haute performance : Le code BigDFT. *Comptes Rendus Mécanique*, 339, 149-164. <http://dx.doi.org/10.1016/j.crme.2010.12.003>
- Hill, J., Subramanian, L., & Maiti, A. (2005). *Molecular modeling techniques in material sciences*. Taylor & Francis Group.
- Johnson, S. (2010). *The NLOpt nonlinear-optimization package*. Retrieved from <http://ab-initio.mit.edu/nlopt>
- Kohn, W., & Sham, L. (1965). Self-consistent equations including exchange and correlation effects, *Phys. Rev.*, 140(4A), A1133-A1138. <http://dx.doi.org/10.1103/PhysRev.140.A1133>
- Kress, J., Goedecker, S., Hoisie, A., Wasserman, H., Lubeck, O., Collins, L., & Holian, B. (1998). Parallel  $O(N)$  tight-binding molecular dynamics of polyethylene and compressed methane, *Journal of Computer-Aided Materials Design*, 5(2-3), 295-316. <http://dx.doi.org/10.1023/A:1008637113920>
- Nocedal, J., & Wright, S. (1999). *Numerical optimization*. New York: Springer Series in Operations Research.
- Perdew, J., & Wang, Y. (1992). Accurate and simple analytic representation of the electron-gas correlation energy, *Phys. Rev. B*, 45(23). <http://dx.doi.org/10.1103/PhysRevB.45.13244>
- Perdew, J., & Zunger, A. (1981). Self-interaction correction to density-functional approximation for many-electron systems, *Phys. Rev. B*, 23, 5048-5079. <http://dx.doi.org/10.1103/PhysRevB.23.5048>
- Rasmussen, C., & Williams, C. (2006). *Gaussian processes for machine learning*, the MIT Press.
- Saad, Y., Chelikowsky, J., & Shontz, S. (2010). Numerical methods for electronic structure calculations of materials, *SIAM Review*, 52 (1), 3-54. <http://dx.doi.org/10.1137/060651653>
- Seeger, M., Williams, C., & Lawrence, N. (2003). Fast forward selection to speed up sparse Gaussian process regression. In C.M. Bishop and B. J. Frey, (Eds.), *Proceeding of the 9th Workshop on Artificial Intelligence and Statistics*. Florida, USA.
- Suryanarayana, P., Gavini, V., Blesgen, T., Bhattacharya, K., & Ortiz, M. (2010). Non-periodic finite-element formulation of Kohn-Sham density functional theory, *Journal of the Mechanics and Physics of Solids*, 58, 256-280. <http://dx.doi.org/10.1016/j.jmps.2009.10.002>
- Tadmor, E., & Miller, R. (2011). *Modeling materials, continuum, atomistic and multiscale techniques*, Cambridge University Press, Cambridge.
- Toda, N., Kimizuka, H., & Ogata, S. (2010). DFT-based FEM analysis of nonlinear effects on indentation process in diamond crystal, *International Journal of Mechanical Sciences*, 52(2), 303-308. <http://dx.doi.org/10.1016/j.ijmecsci.2009.09.035>
- Wimmer, E., Krakauer, H., Weinert, M., & Freeman, A. (1981). Full-potential self-consistent linearized-augmented-plane-wave method for calculating the electronic structure of molecules and surfaces:  $O_2$  molecule, *Phys. Rev. B*, 24, 864. <http://dx.doi.org/10.1103/PhysRevB.24.864>

Yamaguchi, Y., & Maruyama, S. (1998). A molecular dynamics simulation of the fullerene formation process, *Chem. Phys. Letters*, 286(3-4), 336–342. [http://dx.doi.org/10.1016/S0009-2614\(98\)00102-X](http://dx.doi.org/10.1016/S0009-2614(98)00102-X)

### **Copyrights**

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).