APPLICATION OF THE NEURAL NETWORKS FOR DEVELOPING NEW PARAMETERIZATION OF THE TERSOFF POTENTIAL FOR CARBON

ANTHONY CHUKWUEMEKA NWACHUKWU
AND SZYMON WINCZEWSKI

Faculty of Applied Physics and Mathematics, Gdansk University of Technology, Gabriela Narutowicza 11/12, 80-233 Gdansk, Poland

(received: 4 July 2020; revised: 10 September 2020;
accepted: 20 September 2020; published online: 1 October 2020)

Abstract: Penta-graphene (PG) is a 2D carbon allotrope composed of a layer of pentagons having sp^2- and sp^3- bonded carbon atoms. A study carried out in 2018 has shown that the parameterization of the Tersoff potential proposed in 2005 by Ehrhart and Able (T05 potential) performs better than other potentials available for carbon, being able to reproduce structural and mechanical properties of the PG. In this work, we tried to improve the T05 potential by searching for its parameters giving a better reproduction of the structural and mechanical properties of the PG known from the *ab initio* calculations. We did this using Molecular Statics (MS) simulations and Neural Network (NN). Our test set consisted of the following structural properties: the lattice parameter a; the interlayer spacing h; two lengths of C-C bonds, d_1 and d_2 respectively; two valence angles, θ_1 and θ_2, respectively. We also examined the mechanical properties by calculating three elastic constants, C_{11}, C_{12} and C_{66}, and two elastic moduli, the Young's modulus E and the Poisson's ratio ν. We used MS technique to compute the structural and mechanical properties of PG at $T = 0$ K. The Neural Network used is composed of 2 hidden layers, with 20 and 10 nodes for the first and second layer, respectively. We used an Adams optimizer for the NN optimization and the Mean Squared Error as the loss function. We obtained inputs (about 80 000 different sets of potential parameters) for the Molecular Statics simulation by using randomly generated numbers. The outputs from these simulations became the inputs to our Neural Network. The Molecular Statics simulations were done with LAMMPS while the Neural Network and other computations were done with Python, Pytorch, Numpy, Pandas, GNUPLOT and Bash scripts. We obtained a parameterization which has a slightly better accuracy (lower relative errors of the calculated structural and mechanical properties) than the original parameterization.

Keywords: penta-graphene, mechanical properties, molecular dynamics

DOI: https://doi.org/10.17466/tq2020/24.4/a

1. Introduction

The discovery of graphene in 2004 [1, 2] and the creation of carbon nanotubes in 1992 [3, 4] have initiated a new search field for new carbon allotropes. These two findings have given rise to the suggestion of many new carbon forms based on conceptual considerations. A great number of these structures are yet to be created but the existing ones show the importance of conceptual studies.

Most carbon allotropes have a hexagonal structure which has led to a search for a new carbon allotrope with a pentagonal structure. In 2014 [5] Zhang *et al.* proposed a new two-dimensional carbon allotrope, called penta-graphene (PG) as an exfoliated equivalent of T12-carbon [6], the existence of which was also predicted based on the *ab initio* calculations.

Penta-graphene (PG) can be seen as a layer of pentagons constructed from a mixture of sp^2- and sp^3-bonded carbon atoms. Zhang *et al.* reported that due to its unusual atomic structure, closely resembling the well-known Cairo pentagonal tiling, PG possessed many unique properties. It was found to be mechanically, dynamically and thermally stable, and able to withstand temperatures as high as $T = 1000$ K. It was also reported that PG displayed ultra-high mechanical strength, with the strain at a maximum stress being as high as 21%. The stiffness of PG was also found to be very high, with the corresponding (in-plane) Young's modulus as high as $E = 263$ GPa nm, being more than two-thirds of that of graphene ($E = 345$ GPa nm). PG was also found to exhibit the auxetic behavior, *i.e.*, the anomalous property of becoming wider rather than thinner when stretched. Zhang *et al.* also reported that PG has indirect band gap, as large as $E_g = 3.25$ eV. This feature renders PG as a better material than graphene for two-dimensional transistors, which require the presence of a large band gap to obtain a good switch-off.

1.1. Literature Review

Since its discovery, PG has been studied intensively by several groups [7–20]. The stability of PG and its experimental reachability was questioned by Ewels *et al.* in [8]. Based on the *ab initio* calculations they concluded that PG should be difficult to isolate, also pointing out that PG should rapidly restructure toward graphene in the presence of even few catalytic impurities. A similar observation about the potential instability of PG was made by Cranford [9], who studied finite, hydrogen-terminated sheets of PG. By using the MD simulation he concluded that bond breaking should be observed even at relatively small deformations (*ca.* 5%), leading to transformations of pentagons into hexagons and heptagons. According to him, a similar effect should be observed at elevated temperatures (*ca.* 600 K), resulting in transformation of PG into (defective) graphene.

Sun *et al.* [7] presented the results of extensive studies on the mechanical properties of PG. By combining density functional theory (DFT) calculations with the fourth order continuum elasticity theory they calculated a complete set of (fifteen) anisotropic nonlinear elastic constants of monolayer PG, showing

that the applied continuum formalism (originally proposed by Wei *et al.* for graphene [21]) was able to accurately describe the non-linear elastic behavior of PG in a wide range of strains, even as large as 30%. Sun *et al.* also investigated the fracture of PG. They concluded that due to a longer bond length and lower charge density the sp^3 bonds were more vulnerable to failure than the sp^2 bonds. Sun *et al.* also studied the mechanism of PG deformation, demonstrating that the negative Poisson's ratio of PG originated from de-wrinkling of its structure.

The result of doping on the properties of PG has been also analyzed. Berdiyorov *et al.* [10] discovered that the electronic properties of PG could be altered to obtain a desired result by substituting C atoms with Si, B and N. They showed that the band gap size could be greatly reduced, to 0.2 eV. According to them the greatest reduction in the band gap was obtained for Si substitutions on the top (or bottom) plane of PG. They also showed that surface termination with fluorine and hydroxyl groups resulted in an increase in the band gap. The addition of new functions, features, capabilities, or properties to PG sheets was also investigated by Li *et al.* [11] who found that an addition of hydrogen and fluorine could tune the electronic and mechanical properties of PG, changing the Poisson's ratio from negative to positive, and reducing the Young's modulus.

The thermal conductivity of PG has been also studied. Using classical equilibrium molecular dynamics (MD), Xu *et al.* [12] found that the thermal conductivity of PG at the room temperature was about 170 W/(m K), which is much lower than that of graphene, which is 2000-4000 W/(m K) [22]. They also identified the main mechanism of thermal conduction. By analyzing phonon frequencies and phonon mean free paths they found that the acoustic phonons made a contribution of about 90% to the thermal conductivity, also showing that phonons with mean free paths larger than 100 nm made a contribution over 50%. They also demonstrated that the remarkably lower thermal conductivity of PG (compared with graphene) resulted from lower phonon group velocities and fewer collective phonon excitations.

The influence of functionalization on the thermal properties of PG has been also studied. Using DFT calculations combined with an iterative solution of the phonon Boltzmann transport equation, Wu *et al.* [13] found that hydrogenation of PG led to large (76% increase) improvement in thermal conductivity. Other spatial forms of PG have been also investigated. Yu and Zhang studied the electronic properties of layered PG [14]. They showed that there was no direct-to-indirect band gap transition in PG by varying the strain, layer number, and stacking misalignment. Owing to its characteristics, few-layer PG was recognized by Yu and Zhang [14] as a very promising material for optoelectronic and photovoltaic applications. Recently, Rajbanshi *et al.* [15] studied penta-graphene nanoribbons (PGNRs) by using DFT, concluding that PGNRs were thermodynamically meta-stable with respect to graphene nanoribbons. They also found that on application of uniaxial strain the band gap of PGNRs was decreasing continuously, yielding a straintunable optoelectronic material.

Another potential application of PG has been recently highlighted by Xiao *et al.* [16]. By using DFT calculations they found that PG provided very high ion storage capacity and fast ion diffusivity, and therefore was a promising anode material for Li/Na-ion batteries.

Winczewski *et al.* [23] considered 14 different empirical potentials available for carbon and based on molecular statics/molecular dynamics (MS/MD) simulations tested if any of them were able to describe (at least in a satisfactory manner) the interactions in this unusual system, possessing exotic, sp^2-sp^3 mixed hybridization, not seen in typical carbon forms. Using molecular statics and molecular dynamics simulations they showed that there was only one potential - namely the Tersoff-type potential, proposed by Erhart and Albe in 2005 - which is able to correctly describe all the important properties of penta-graphene.

1.2. Problem Statement

We can see clearly from the above analysis that PG has a great potential for the future. We can also infer that the atomistic modeling techniques are becoming very welcoming. So far, the main method of studying PG has been the DFT method and according to Winczewski *et al.* [23], as at 2018, there were only three papers to report the application of empirical potentials to model PG.

In the first article Cranford [9] employed the reactive force field [24] (REAX) to study the mechanical properties and the chemical stability of PG, while in the second Ebrahimi [17] used the reactive empirical bond order (REBO) potential [25] to study the effect of hydrogen coverage on the buckling of PG. It is important to note that no preliminary validation of the chosen description method was carried out in both of the aforementioned works.

Contrary to this, Xu *et al.* [12] performed such validation by testing four different interatomic potentials: the original Tersoff potential [26, 27], the optimized Tersoff potential [28, 29], the REBO potential [25] and the environment-dependent interatomic potential [30] (EDIP). By comparing the calculated structural and mechanical parameters with the results of the *ab initio* calculations they concluded that it was the original Tersoff potential that most closely reproduced the properties of PG among the tested interaction models. Therefore, it was used in [12] to study the transport phenomena in PG. However, it must be noted that the Poisson's ratio of PG calculated with the use of the original Tersoff potential was found in [12] to be equal to -0.174, which strongly (more than 2.5-times) differed from the *ab initio* result of Zhang *et al.* [5], which was -0.068. Such a significant difference suggests that the original Tersoff potential may not be the best choice when modeling PG at the empirical level.

It is well known that the credibility of the results obtained from any atomistic simulation depends first and foremost on the quality of the model employed to describe interatomic interactions. The last three decades have brought significant advances in the development of empirical potentials. This is especially true in the case of carbon. Due to the significance of this element for the nanotechnology revolution, many new potentials have been proposed for carbon [24–44],

in order to describe its various - and often very different - forms. Since all potentials for carbon have been parameterized without accounting for the properties of PG - which was unknown when the potentials were developed - before starting modeling PG at the empirical level it is important to first test how well, or whether at all, the existing potentials reproduce the properties of this new - and very exotic - form of carbon.

The system is a bit challenging for the empirical potentials because the form of PG is uncommon. This is a result of the fact that there are two types of bonding schemes found in PG, with the entire structure being a "mixture" of sp^2- and sp^3- hybridized atoms. This is a reason why the basic building blocks in PG differ significantly from their counterparts seen in diamond and graphite/graphene (*e.g.* the carbon-carbon bond lengths characteristic for diamond and graphite/graphene are 1.54 Å and 1.42 Å, respectively, the corresponding bonds in PG have the lengths of 1.55 Å and 1.34 Å, respectively, also the valence angles typical for PG - *i.e.* 98.6°, 112.2°, 113.5° and 134.4° - differ significantly from those typical for diamond - 109.5° - and graphite/graphene - 120°). Therefore, it is not clear if even the potentials, which are known as being able to (simultaneously) describe purely sp^2- and purely sp^3-systems, will be able to correctly capture the characteristics of PG.

1.3. Objective and Aims

In this work we decided to search for a set of parameters of T05 that would give a better reproduction of the fundamental properties of PG. Our aim was to get a result better than that of Winczewski *et al.* [23] obtained by employing the parameterization of the Tersoff potential proposed by Erhart and Albe in 2005 (T05 potential).

This work is organized as follows. In Section 2 we give a theoretical background of our work. We describe the structure of PG and recap the results of the *ab initio* calculations of Zhang *et al.* [5]. We also briefly present the Tersoff potentials and finally give a general overview of the Neural Network (NN).

In Section 3 we discuss the procedure which we employed in generating the data used for our simulations, how we calculated the structural and mechanical properties of PG, and describe the NN architecture used for the NN training and the evaluation technique used.

In Section 4 we present the outcome of the NN model and the corresponding calculations of the structural, energetic and mechanical properties of PG and compare them with the results of Ref. [23].

We conclude with a summary in Section 5.

2. Theoretical Foundation

The proper understanding of the total energy of a system of atoms as a function of the atomic coordinates is a requirement for obtaining the solutions of many problems encountered in chemistry, physics and material science. Some examples of these are diffusion paths and barriers, determination of surface reconstructions,

the dispersion and interactions of the phonon, and even the thermal and mechanical properties of materials.

Calculations using quantum-mechanical methods have brought great progress in the bid to address the problem [45–49]. However, these quantum-mechanical calculations at the present time are not feasible for addressing problems that involve large systems and those that require statistical averages. This is due to their numerical intensity.

A solution to solve this problem could be to create an empirical inter-atomic potential $E(\{r\})$. This gives the total energy E of a set of particles, as an explicit mathematical function of the set $\{r\}$ of the particle coordinates. Peradventure, this function is reasonably easy to find, and if it is able to describe a real system of interest with a reasonable accuracy, then realistic calculations of the properties of a larger system can be performed. This approach is expected to give a significant loss in accuracy when compared with the *ab initio* calculations.

Prior to the present time, the majority of the empirical inter-atomic potentials were divided into two groups, one group including the pair potentials with the very popular examples of the Lennard-Jones and exponential Morse potentials [50, 51, 45–49]. These potentials can be used directly on completely arbitrary atom configurations, however, they describe accurately only simplest closed-shell systems. To be specific, pair potentials fail to describe strongly covalent systems, semiconductors, for example.

The other group of potentials is constructed to accurately describe small distortions from the ground state in more complex systems such as diamond-structure semiconductors [52]. Perhaps the most famous of these is the Keating model [53]. Such potentials are useful for describing phonons and elastic deformations, but they cannot describe the energy of states which differ qualitatively from the tetrahedral ground state.

It is of great interest to know that these two approaches are in accordance with the leading term found in a mathematical expansion of the energy seen as a function of the atomic positions. The Keating model, and similar methods, resemble the Taylor expansions of the energy about its minimum. They have a high accuracy of descriptions for small displacements, but progressively reduce the accuracy for large displacements. A recent review of such approaches has been given by Kane [52]. The energy of N interacting particles may be written as

$$\sum_i V_1(r_i) + \sum_{i<j} V_2(r_i, r_j) + \sum_{i<j<k} V_3(r_i, r_j, r_k) + \ldots \tag{1}$$

Here, V_1, V_2, V_3 represent one body, two body and three body potentials, respectively. The first term of the expansion relates to an external force, and hence, disappears, if we decide to consider the inter-atomic forces only. It is only the pair potential (V_2) that may be appropriate for rather closed packed structures (liquefied noble gases Ar, Kr, Xe are typical for such systems) but not suitable for describing systems that are strongly covalent with a more open structure. Due

to this, the next term in the expansion (the three body potential V_3) will have to be taken into account. Adding these to the two body term can result in a sufficiently accurate description of the real physical system of interest [54, 55].

The specific form of the V_2 term (in fact, only V_2 and V_3 are significant since no external interactions are usually taken into account, and adding more terms in the expansion will make the computation impracticable) varies from the $1/r^n$ interaction ('Lennard-Jones' type), to the $e^{-\alpha r}$ interaction (Morse type) [55, 56], or a combination of these [54]. In such cases, a cutoff function is added to limit the range of the potential and permit a reduction in the computational time.

2.1. Tersoff Potential

Tersoff [56] abandoned the use of the N-body potential form and proposed a new approach by effectively coupling two body and higher multi-atom correlations into the model. The central idea comes from simple quantum-mechanical arguments [51]: the more neighbors an atom has, the weaker the bond to each neighbor will be. In general, the bond strength, or the bond order, depends, in a complex way, on the geometry.

A pair potential, the strength of which depends on the environment, was then developed by Tersoff. It was first calibrated for silicon [56] and later for carbon [57]. The Morse form is adopted for the Biswas and Hamann potential [55], which relates to the exponential decay dependence of the electronic density. Its form is as follows:

$$E = \sum_i E_i = \frac{1}{2} \sum_{i \neq j} V_{ij}, \tag{2}$$

$$V_{ij} = f_C(r_{ij}) \left[a_{ij} f_R(r_{ij}) + b_{ij} f_A(r_{ij}) \right] \tag{3}$$

In the above expression, E represents the total energy of the system, for convenience, this is decomposed into bond energy V_{ij} and site energy E_i. The indices i and j run over the atoms of the system, and r_{ij} is the distance from atom i to atom j. The function f_A stands for an attractive pair potential associated with bonding, while f_R stands for a repulsive pair potential. The additional term f_C is just a smooth cutoff function, which limits the range of the potential. Short-ranged functions allow a great reduction in the computational effort for most applications.

$$\begin{aligned} f_A(r) &= -Be^{-\alpha_2 r} \\ f_R(r) &= Ae^{-\alpha_1 r} \end{aligned} \tag{4}$$

$$f_C(r) = \begin{cases} 1, & r < R - D \\ \frac{1}{2} - \frac{1}{2} \sin\left[\frac{\pi}{2} (r - R)/D \right], & R - D < r < R + D \\ 0, & r > R + D \end{cases} \tag{5}$$

It has to be noticed that the parameters R and D are not systematically optimized but are chosen to include the first-neighbor shell only. The function f_C decreases from 1 to 0 in the range $R-D<r<R+D$.

The function b_{ij} represents the bond order and is assumed to be a monotonically decreasing function of the coordination of atoms i and j. The basic idea is that the strength of each bond depends upon the local environment and is lowered when the number of neighbors is relatively high. This dependence is expressed by b_{ij}, which can accentuate or diminish the attractive force relative to the repulsive force, according to the environment.

$$b_{ij} = \frac{1}{\left(1+\beta^n \zeta_{ij}^n\right)^{1/2n}}$$

$$\zeta_{ij} = \sum_{k\neq i,j} f_c(r_{ij}) g(\theta_{ijk}) e^{\left[\lambda_3^3 (r_{ij}-r_{ik})^3\right]} \tag{6}$$

$$g(\theta) = 1 + \frac{c^2}{d^2} - \frac{c^2}{\left[d^2 + (h-\cos\theta)^2\right]}$$

The term ζ_{ij} defines the effective coordination number of atom i, *i.e.* the number of nearest neighbors, taking into account the relative distance of two neighbors $r_{ij}-r_{ik}$ and the bond-angle θ. The function $g(\theta)$ has a minimum for $h=\cos(\theta)$, the parameter d determines how sharp the dependence on the angle is, and c expresses the strength of the angular effect. θ_{ijk} is the bond angle between bonds ij and ik. While $b_{ij} \neq b_{ji}$, this fact has no significance other than for the division of the total energy into a sum of site energies in (2). If, for aesthetic reasons, a more symmetric form is desired, the sum over pairs of atoms in (2) can be replaced with the sum over bonds ($i>j$), and then b_{ij} can be replaced with the symmetrized function $b_{ij}=(b_{ij}+b_{ji})/2$. The form proposed for a_{ij} is

$$a_{ij} = \left(1+\alpha^n \eta_{ij}^n\right)^{-1/2n}$$

$$\eta_{ij} = \sum_{k\neq i,j} f_c\left(r_{ik} \exp^{\left[\lambda_3^3 (r_{ij}-r_{ik})^3\right]}\right) \tag{7}$$

However, in the present work, a_{ij} is not actually used, *i.e.*, $a=0$, so $a_{ij}=1$. Equation (7) is included for completeness, because the potential can probably be further improved with such a term.

The parameters provided in Table 1 are from the paper [31], *i.e.* they correspond to parameterization proposed by Erhart and Albe.

2.2. Structure of penta-graphene

Penta-graphene is an allotrope of carbon completely made up of carbon pentagons and having a resemblance to the Cairo pentagonal tiling [58]. PG was first proposed in 2014 for the purpose of analyses and simulations [58]. Calculations showed that it was not stable in its pure form [59], but could be made

Table 1. Parameters of the Erhart and Albe parameterization of the Tersoff potential

Parameter	Value
A (eV)	2019.8449
B (eV)	175.42665
λ_1 (Å^{-1})	4.18426
λ_2 (Å^{-1})	1.9309
λ_3 (Å^{-1})	0
β	1
n	1
c	181.91
d	6.28433
$\cos\theta_0$	-0.5556
R (Å)	2
D (Å)	0.15
m	1
γ	0.11233

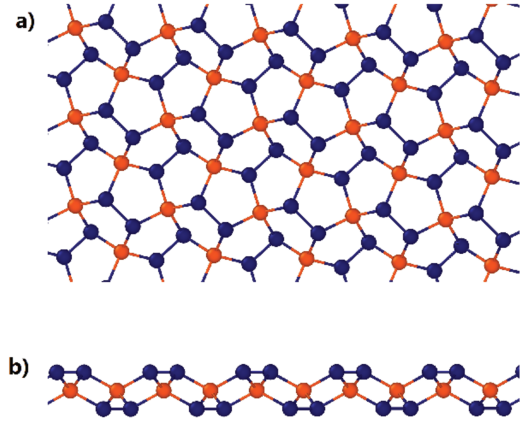

Figure 1. Structure of penta-graphene

stable by hydrogenation [60]. Due to its atomic configuration, it has an unusually negative Poisson's ratio and very high ideal strength which is believed to exceed that of graphene, a material with similar properties [58].

Figure 1 shows the structure of PG as described by Zhang *et al.* in [5]. PG has a tetragonal lattice with $P-42_1m$ symmetry having a space group number of 113. Its unit cell has two types of carbon atoms: sp^3- (later on denoted as C1 atoms) and sp^2- bonded (C2 atoms), and its characterized by lattice parameters $aQQ = b = 3.64$ Å. Hence, a unit cell is made up of 2 C1 atoms and 4 C2 atoms.

The structure of PG is layered. As can be seen in Figure 1b, three of these layers are present. The C2 atoms form two outside layers having $z = \pm h$, while the C1 atoms lie on the middle layer having $z = 0$. By h, we mean the interlayer spacing which is equal to 0.6 Å, as shown in [5]. The C1 atoms can be seen as connectors, with each being connected to four C2 atoms, two from each outer layer. For C2, each of them is connected to two C1 atoms and one C2 atom coming from the same layer.

PG has two types of bonds, C1-C2 and C2-C2. All C1-C2 bonds have the same length of $d_1 = 1.55$ Å while all C2-C2 bonds have the same length of $d_2 = 1.34$ Å [5]. The pentagons are the same as the interior angles equal to 113.5 (the C1-C2-C2 angle, occurring twice), 112.4 (the C1-C2-C1 angle) and $\theta_1 = 98.6$ degrees (the C2-C1-C2 angle, occurring twice). There exists another C2-C1-C2 angle that measures the PG crumpling. This angle is equal to $\theta_2 = 134.4$ degrees.

There is a slight difference between the geometry of PG and Cairo pentagonal tiling. The Cairo pentagonal tiling has the interior angles of 90 degrees (this occurs twice) and 120 (this occurs three times) and the shorter edge to longer edge ratio of its pentagons is $\sqrt{3}-1 \approx 0.73$, while for PG it is $d_2/d_1 \approx 0.86$.

Since PG has a tetragonal symmetry, its in-plane mechanical properties can be described with three independent elastic constants: C_{11}, C_{12} and C_{66}. Here, we have employed the Voigt notation which will be used throughout this work. Knowing the elastic constants one may easily calculate the Young's modulus E and the Poisson's ratio v, using the following formulas:

$$E = \frac{C_{11}^2 - C_{12}^2}{C_{11}} \tag{8}$$

and

$$v = \frac{C_{12}}{C_{11}}. \tag{9}$$

In the case of the tetragonal lattice both of the above moduli depend on the direction of measurement. The presented formulas correspond to the <1,0> family of directions. The elastic constant C_{66} defines the shear modulus $\mu = C_{66}$. Ref. [5] reported that $C_{11} = 265$ GPa nm, $C_{12} = -18$ GPa nm and $C_{66} = 152$ GPa nm.

We now describe the processes involved in the calculations of the mechanical properties of our penta-graphene. We used the molecular statics and the direct method together to calculate the mechanical properties of penta-graphene at zero temperature.

2.2.1. Overview of Mechanical Properties

Hook's law can be used to describe the mechanical properties of a linear elastic material as seen below:

$$\sigma = C\varepsilon \tag{10}$$

where σ is the stress tensor, ϵ is the strain tensor while C is the stiffness tensor which describes the mechanical properties. Both ϵ and σ tensors have 6 elements

which do not depend on themselves. The stiffness tensor C can have up to 21 independent elements. Eq. (10) can be expanded in this way:

$$\begin{bmatrix}\sigma_1\\ \sigma_2\\ \sigma_3\\ \sigma_4\\ \sigma_5\\ \sigma_6\end{bmatrix} = \begin{pmatrix} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16}\\ C_{21} & C_{22} & C_{23} & C_{24} & C_{25} & C_{26}\\ C_{31} & C_{32} & C_{33} & C_{34} & C_{35} & C_{36}\\ C_{41} & C_{42} & C_{43} & C_{44} & C_{45} & C_{46}\\ C_{51} & C_{52} & C_{53} & C_{54} & C_{55} & C_{56}\\ C_{61} & C_{62} & C_{63} & C_{64} & C_{65} & C_{66}\end{pmatrix} \begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \varepsilon_3\\ \varepsilon_4\\ \varepsilon_5\\ \varepsilon_6\end{bmatrix} \tag{11}$$

Since we are considering a 2D system, (11) reduces to:

$$\begin{bmatrix}\sigma_1\\ \sigma_2\\ \sigma_6\end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{16}\\ C_{12} & C_{22} & C_{26}\\ C_{16} & C_{26} & C_{66}\end{bmatrix} \begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \varepsilon_6\end{bmatrix} \tag{12}$$

Given a two dimensional structure, we need to find the values of six elastic constants to characterize the mechanical properties of the structure. Due to the symmetry, the stiffness tensor of a two dimensional square lattice becomes:

$$C = \begin{bmatrix} C_{11} & C_{12} & 0\\ C_{12} & C_{11} & 0\\ 0 & 0 & C_{66}\end{bmatrix} \tag{13}$$

Knowing the elastic constants C_{11}, C_{12} and C_{66}, the mechanical moduli Y, ν and μ can be calculated from the following relations:

$$Y = \left(C_{11}^2 - C_{12}^2\right)/C_{11} \tag{14},$$

$$\nu = C_{12}/C_{11} \tag{15}$$

and

$$\mu = C_{66} \tag{16}$$

2.2.2. Direct Method

We used the direct method to investigate the mechanical properties. This method involves the application of deformation (strain ϵ) and the observation of how this applied deformation influences the energy of the system E. Then, we study the dependence of the stress σ on the strain ϵ in order to determine the mechanical properties.

The application of deformation on two-dimensional systems causes the total energy E to change. The dependence of energy E on strain ϵ can be described with:

$$E = E_0 + \frac{1}{2}A_0 \sum_i \sum_j C_{ij}\varepsilon_i\varepsilon_j + O(\varepsilon^3) \tag{17}$$

Indices i and j run over 1, 2, and 6. Symbols E and E_0 represent energies of the deformed and undeformed system, respectively, while A_0 represents the surface area of the undeformed system. $O(\epsilon^3)$ represents the higher order contributions of order 3 or greater.

The elastic energy density ρ_{el} can be expressed as:

$$\rho_{el} = \frac{E_{el}}{A_0} = \frac{E - E_0}{A_0} \approx \frac{1}{2} \sum_i \sum_j C_{ij} \varepsilon_i \varepsilon_j \tag{18}$$

In the above equation, we supposed that the deformation was small and this made the higher order terms vanish. It is easy to find C_{ij} since we know the dependence of ρ_{el} on ϵ.

By studying the dependence of stress σ on the applied strain ϵ, we can also find the elastic constants. This relation has the form:

$$\sigma_i = \sum_j C_{ij} \varepsilon_j + O(\varepsilon^2) \tag{19}$$

For low deformation (for linear regime, we apply Hook's law), (19) becomes:

$$\sigma_i = \sum_j C_{ij} \varepsilon_j \tag{20}$$

While doing the computations in the simulation box described with the vectors $A = [L_x, 0]$ and $B = [X_y, L_y]$, the strains ϵ_1, ϵ_2 and ϵ_6 can be obtained using:

$$\varepsilon_1 = \varepsilon_{xx} = \frac{L_x}{L_{x,0}} - 1, \tag{21}$$

$$\varepsilon_2 = \varepsilon_{yy} = \frac{L_y}{L_{y,0}} - 1, \tag{22}$$

and

$$\varepsilon_6 = \gamma_{xy} = 2\varepsilon_{xy} = \frac{X_y}{L_{y,0}}. \tag{23}$$

The symbols $L_{x,0}$ and $L_{y,0}$ in equations (21), (22) and (23) represent the sizes of the undeformed system. It is also assumed that $X_{y,0} = 0$

To find the mechanical properties, the following three deformations (for each deformation we give the form of the strain tensor ϵ) are formulated:

$$D_1 = \begin{pmatrix} \delta \\ \delta \\ 0 \end{pmatrix}, \quad D_2 = \begin{pmatrix} \delta \\ -\delta \\ 0 \end{pmatrix}, \quad D_3 = \begin{pmatrix} 0 \\ 0 \\ \delta \end{pmatrix} \tag{24}$$

When we plug these deformations into equations (18) and (20), the result is an expression which shows how the elastic energy density e_{el} and the stress σ depend on the deformation magnitude δ. The expression for the elastic energy density for D_1 is:

$$e_{el}(\delta) = (C_{11} + C_{12})\,\delta^2, \tag{25}$$

while for the deformation D_2 it reads:

$$e_{el}(\delta) = (C_{11} - C_{12})\,\delta^2, \tag{26}$$

and for D_3 it reads:

$$e_{el}(\delta) = \frac{1}{2} C_{66} \delta^2. \tag{27}$$

For the deformation D_1, the expression for the stress is:

$$\frac{1}{2}(\sigma_1+\sigma_2)=(C_{11}+C_{12})\delta, \tag{28}$$

with

$$\sigma_6=0. \tag{29}$$

For the deformation D_2,we have:

$$\frac{1}{2}(\sigma_1-\sigma_2)=(C_{11}-C_{12})\delta, \tag{30}$$

with

$$\sigma_6=0. \tag{31}$$

Finally, for the Deformation D_3 we obtain:

$$\sigma_1=\sigma_2=0 \tag{32}$$

and

$$\sigma_6=C_{66}\delta. \tag{33}$$

2.3. Neural Network

An artificial neural network [61], popularly referred as the Neural Network (NN), is a mathematical model for predicting the system performance (*i.e.*, the system output) inspired by the structure and function of human biological neural networks. The ANN is developed and derived to have a function similar to the human brain by memorizing and learning various tasks and behaving accordingly. It is trained to predict specific behavior and remember such behavior in the future like the human brain does. Also its architecture is similar to human neuron layers in the brain as far as the functionality and the inter-neuron connection are considered [62–65].

The NN has been successful at predicting inter-atomic potentials of materials. Bukkapatnam *et al.* [66] showed that the performance of the genetic algorithm for fitting empirical interatomic potentials could be improved by using the neural network. They found a reduction in the computational time by over two orders of magnitude. Also the potentials which they estimated from the functions (the Tersoff potential in their work) were within 0.1% of the actual potential. Purja *et al.* [65] showed that great improvement could be seen in the transferability of ML potentials by combining a general physics-based model (*i.e.* the analytical bond-order potential) with a neural network regression. Hobday *et al.* [63] showed that "(many body) interatomic potential functions for multi-component systems can be derived by training a specially constructed NN on a variety of structural data."

2.3.1. Brief History

McCulloch and Pitts [67] (1943) created a computational model for neural networks based on algorithms called threshold logic. This model resulted in the research being split into two approaches. One of the approaches was based

on biological processes while the other on the application of neural networks to artificial intelligence. This work gave birth to the research on nerve networks and their connection to finite automata [68]. This was just the beginning of the Neural Networks (NN).

Hebb [69] in the late 1940s invented a learning hypothesis which was based on the mechanism of neural plasticity and was later known as the Hebbian learning. The Hebbian learning is an unsupervised learning and it developed into models for long-term potentiation. In 1948, using Turing's B-type machines, researchers began to apply these ideas to computational models. Computational machines, that were known then as calculators, were first used by Farley and Clark [70] (1954) to simulate a Hebbian network. Rochester, Holland, Habit and Duda (1956) created other neural network computational machines [71].

An algorithm used for the recognition of patterns known as the perceptron was created by Rosenblatt [72] (1958). Circuitry such as the exclusive-or circuit not in the basic perceptron that could not be processed by neural networks at the time was described by Rosenblatt using the mathematical notation [73]. Based on their discovery of two types of cells in the primary visual cortex: simple cells and complex cells, Nobel laureates, Hubel and Wiesel proposed a biological model in 1959 [74]. In 1965, the first functional networks with multiple layers were published by Ivakhnenko and Lapa, as the Group Method of Data Handling [75–77].

There was a stagnation of research after the machine learning studies by Minsky and Papert [78], who made a discovery of two major issues with the computational machines that processed neural networks. The first discovery was the incapability of basic perceptrons in the processing of exclusive-or circuit. The second finding was the inability of the computer processing power to effectively handle the work from large neural networks. Until computers were able to achieve much greater processing power, there was reduced research in the neural network. For a long time, the principal focus of the artificial intelligence was on high-level (symbolic) models processed by explicit algorithms, and characterized by expert systems with knowledge embodied in if-then rules. Until the late 1980s there had been no expansion of research to low-level (sub-symbolic) machine learning, characterized by the embodiment of knowledge in the parameters of a cognitive model.

2.3.2. Basic Unit

A basic unit of computation in a neural network is the neuron. It is also known as a unit or a node. The neuron receives input from an external source or from some other nodes, makes computations and produces an output. There is an associated weight (w) with each input, which has its assignment based on its relative importance to the other inputs. As can be seen in Figure 2 below, a function f (defined below) is applied by the node to the weighted sum of its inputs.

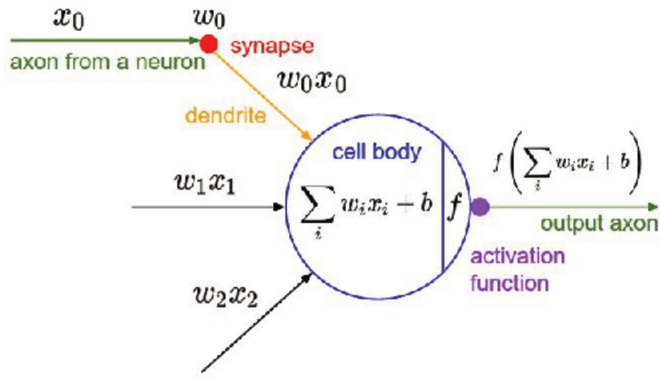

Figure 2. Single neuron

In Figure 2, we can see how the output from the neuron is computed. The function f is non-linear. It is called the Activation Function. The activation function serves the purpose of introduction of non-linearity into the output of a neuron which is linear on its own. This is of great necessity as the majority of data from the real world is non-linear and we wish that neurons should learn these non-linear representations.

Every activation function (which is non-linear) accepts a single number and performs some fixed mathematical operation on it. There are many activation functions to come across in practice. Some of these activation functions are shown below:

Sigmoid: accepts a real-valued input and shrinks it to the range of 0 and 1

$$\alpha(x) = \frac{1}{1+\exp(-x)} \tag{34}$$

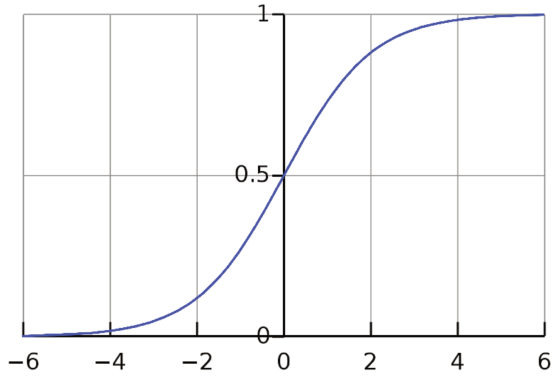

Figure 3. Sigmoid activation function

tanh: accepts a real-valued input and shrinks it to the range of −1 and 1 both inclusive.

$$\tanh(x) = 2\alpha(2x) - 1 \tag{35}$$

ReLU: ReLU is an acronym for the Rectified Linear Unit. It accepts a real-valued input and thresholds it at zero (replacing the negative values with zero).

$$f(x) = \max(0, x) \tag{36}$$

Reason for Bias: The major reason why we introduce bias is to provide every node with a constant value that is trainable (an addition to the normal inputs accepted by each node).

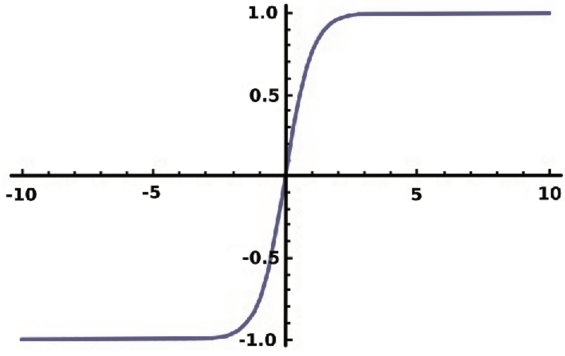

Figure 4. Tanh activation function

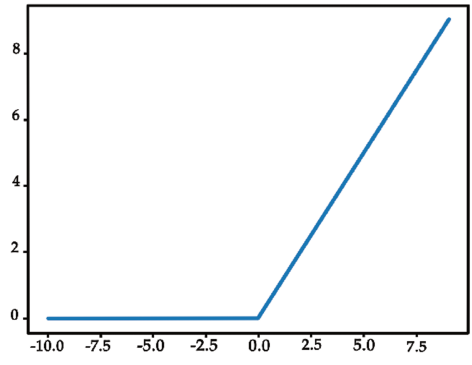

Figure 5. ReLu activation function

2.3.3. Feedforward Neural Network

The simplest and first type of an artificial neural network formed was the feedforward neural network [79]. The feedforward neural network contains many neurons (or nodes) that are packed in layers. Adjacent layered nodes have

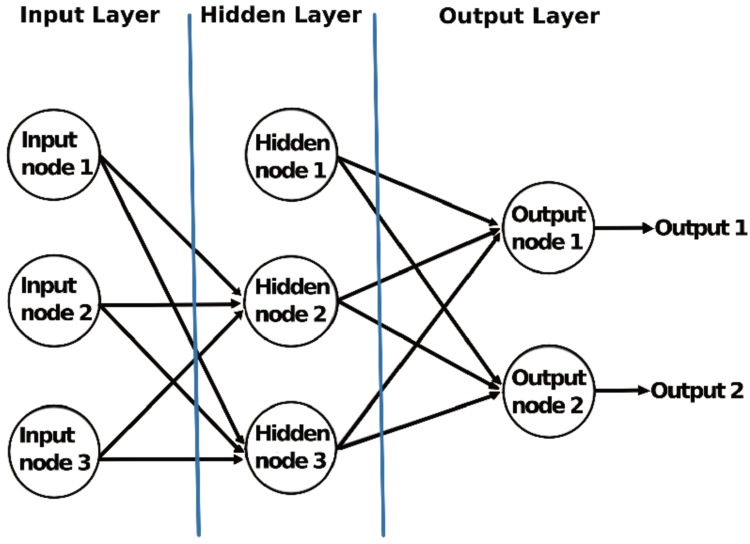

Figure 6. Architecture of feedforward neural network

edges or connections between them. Each of these connections has associated weights.

Figure 6 is an example of a feedforward neural network. A typical example of a feedforward neural network consists of three main types of nodes:

Input Nodes - The input nodes are all together known as the *Input Layer*. They get information from the outside world and introduce it to the network. There is no computation in the input nodes - they only transfer the information to the hidden nodes.

Hidden Nodes - The hidden nodes as the name suggests do not have any direct linkage with the external world. They only do computations and the passage of information from the input to the output nodes. The combination of hidden nodes gives a *Hidden Layer*. A feedforward network has only a single input layer and a single output layer, but the number of Hidden Layers can start from zero but has no upper limit.

Output Nodes - The output nodes are collectively called the *Output Layer*. Their responsibility is to make computations and transfer information from the network (hidden layers and/or input layers) to the external world.

A feedforward neural network allows the movement of information only in one direction - forward direction - starting from the input nodes, via the hidden nodes (if any) and finally to the output nodes. It does not contain any cycles or loops in the network [79] (this property differentiates the Feedforward Neural Networks from the Recurrent Neural Networks which form a cycle by allowing connections between nodes). Two examples of feedforward networks are given below:

Single Layer Perceptron - The simplest feedforward neural network is a single layer perceptron [79]. It does not have any hidden layers. More more details about Perceptrons can be found in [80].

Multi Layer Perceptron - There are one or more hidden layers for the Multi Layer Perceptron. Since multi layer perceptrons are more useful than single layer perceptrons for practical applications today, we will discuss them in detail.

2.3.4. Multi-Layer Perceptron

Unlike a single layer perceptron, a Multi-Layer Perceptron (MLP) has one or more hidden layers (in addition to one input and one output layers). A single layer perceptron has the ability to learn only linear functions, while a multi-layer perceptron has the ability to learn both linear and non-linear functions.

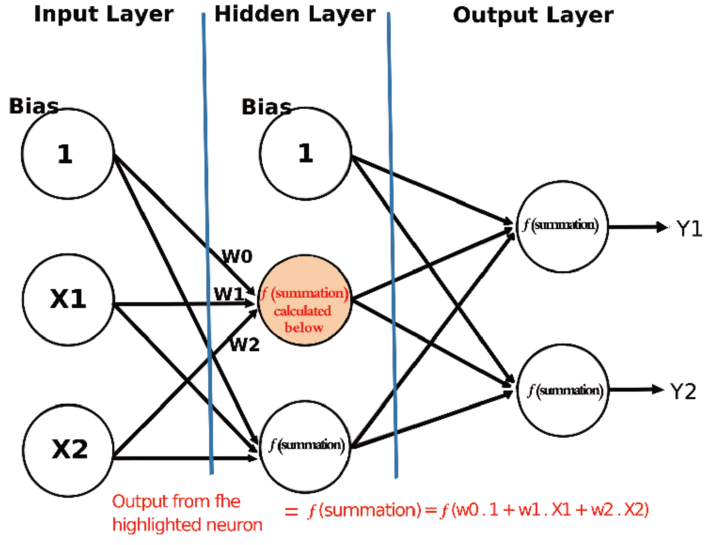

Figure 7. A multi-layer perceptron having one hidden layer

Figure 7 presents a multi-layer perceptron having a single hidden layer. It is noteworthy that all connections have associated weights, although we showed only three weights $(w0, w1, w2)$ in the figure for simplicity.

Input Layer: We can see that the input layer has three nodes. The value of 1 is assigned to the bias node. The other two nodes accept $X1$ and $X2$ as external inputs (these are numerical values which are dependent on the input dataset). As we have already stated, no computation is allowed in the Input Layer, hence, the outputs from nodes in the input layer can be seen as 1, $X1$ and $X2$, respectively, which are fed to the hidden layer.

Hidden Layer: There are also three nodes with a bias node having an output of 1 in the hidden layer. The output of the two other nodes in the hidden layer depends on the outputs of the input layer $(1, X1, X2)$ and also on weights associated with the connections (or edges). The output of the calculation for one of the hidden

nodes (highlighted) can be seen in Figure 7. In the same manner, we can as well calculate the output from other hidden node. f is known as the activation function. We then feed these outputs to the nodes in the output layer.

Output Layer: There are two nodes in the output layer which take inputs from the hidden layer and do similar computations as shown in the highlighted hidden node. $Y1$ and $Y2$ are the outputs of the Multi-Layer Perceptron; they are the calculated values from the computations.

A Multi-Layer Perceptron can learn the relationship between features, say $X = (x1, x2, ...)$, and the associated target values, say y, for either regression or classification problems.

2.3.5. Training the NN: the Back-Propagation Algorithm

The back-propagation algorithm involves the process of making the Multi Layer Perceptron learn. Essentially, what we do is to propagate the errors backward and it is often abbreviated as BackProp. There are several other ways of training the NN, and this is one of them. This method is a supervised training scheme, and it means that it learns from labeled training data (the labels act as a supervisor, to guide the learning). To put it simply, BackProp is like *learning from mistakes*. The duty of the supervisor is to correct the NN whenever it makes mistakes.

As discussed before, a NN is made up of nodes in different layers; the input layer, the intermediate or hidden layer(s) and the output layer. The interconnections between the nodes of adjacent layers have associated *weights*. The aim of the learning process is to make correct assignment of weights to these edges. When presented with an input vector, these weights will have to determine the output vector. Supervised learning involves a labeled training set. This implies that for some given inputs, we have the knowledge of the associated desired/expected output (label).

3. Methodology

In order to compute the structural and mechanical properties of PG at $T = 0$ K the Molecular Statics (MS) technique was used in this work. In order to obtain the best parameters for the Tersoff potential, the obtained results from our calculations were compared with the *ab initio* calculations of Zhang *et al.* [5] which were used as reference. Our test set consisted of the following structural properties: the lattice parameter a; the interlayer spacing h; two lengths of C-C bonds, d_1 and d_2 respectively; two valence angles, θ_1 and θ_2, respectively. We also examined the mechanical properties by calculating three elastic constants, C_{11}, C_{12} and C_{66}, and two elastic moduli, the Young's modulus E and the Poisson's ratio ν.

3.1. Description of the Processes Involved

As seen in the flowchart 8, first the Tersoff parameters are generated with a random number generator. Inputs for the NN are needed (the inputs

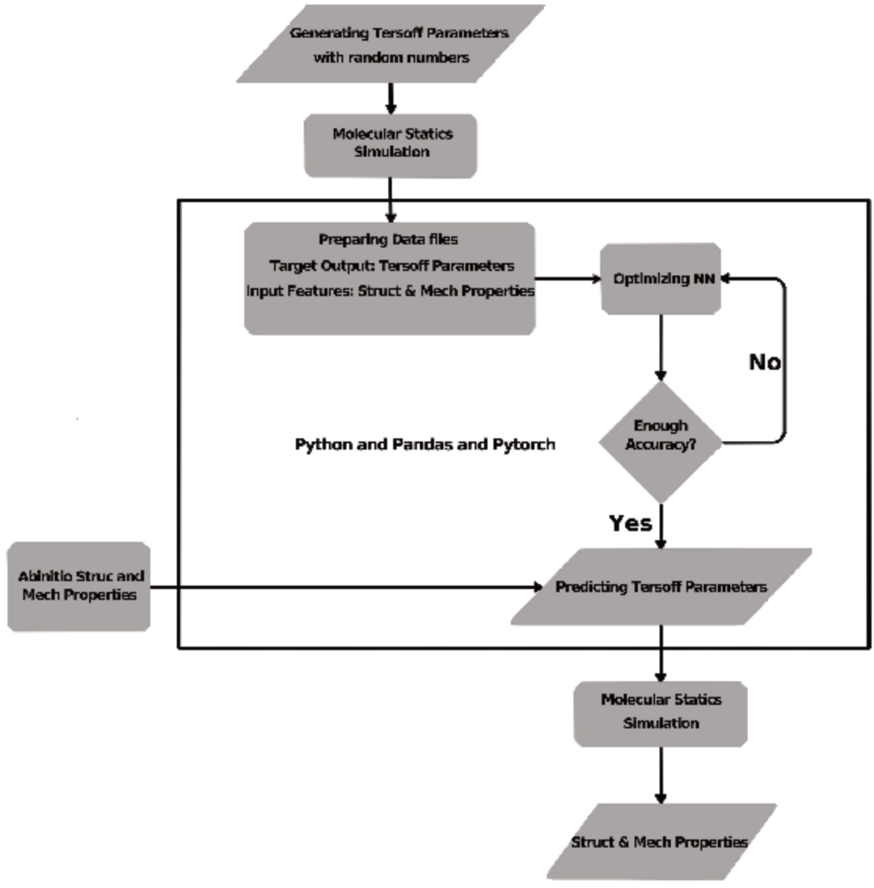

Figure 8. Flowchart of the procedure involved in the work

will be the structural and mechanical properties of the PG). To achieve this, simulations need to be run taking Tersoff parameters as the input and generate the structural and mechanical properties. Then, first, the sets of parameters are generated for the Tersoff potential using a python random number generator.

Next, these sets of parameters are sent into LAMMPS for MS simulation. The results of the simulations (the structural and mechanical properties of PG) will now be the inputs for our NN while the corresponding sets of parameters that yield the PG properties will be the target values of the NN. Python and Pandas are employed to prepare this data (here the data is normalized using a standard scalar) before being sent into the NN.

Basically what the neural network does is to keep adjusting its parameters until a good fit is obtained. When we are satisfied with the accuracy of the NN, we stop the iterations and what we have is a model. The *ab initio* structural

and mechanical properties are fed into the model and it produces a set of numbers which are the predicted - optimized - parameters of the Tersoff potential.

These Tersoff parameters from our model are sent to LAMMPS for MS simulation to test their accuracy. The structural and mechanical properties are obtained from the simulation. Finally, the relative error of these predicted structural and mechanical properties and that of the *ab initio* are computed.

3.2. Calculating the Structural Properties

To start with, the structure was optimized to determine the equilibrium structure of PG at $T = 0$ K. The minimization was started with the *ab initio* result of [5] as an initial configuration and involved repeated alternating optimization of the box and the atomic coordinates to obtain a structure that would correspond to the minimum of energy, a zero (negligible) stress tensor and zero (negligible) forces acting on individual atoms. The strict convergence criteria were used. The termination criterion was to be below 10^{-8} eV Å^{-1} for all the atomic force components, with the final positions accurate to no less than 10^{-6}Å. The minimization was carried out using the Polak-Ribiere [81] formulation of the conjugate gradient method.

The entire minimization procedure was then repeated for different initial structures that differed in the initial values of the lattice parameter a and the interlayer spacing h. This was done to check if there were any other local minima of potentially lower energy which were not accessible by means of local minimization when starting the optimization with the *ab initio* structure. 31×17 equi-spaced points were tested in the (a,h) search space, with $3.34\text{Å} \leq aQQ \leq 3.94\text{Å}$ and $0.43\text{Å} \leq h \leq 2.16\text{Å}$.

All the calculations described here (and hereafter) were performed using the LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator) simulation package developed by Plimpton *et al.* at Sandia [82]. The calculations were performed for a system consisting of 10×10 repetitions of the six-atom elementary cell, with the total number of atoms $N_{pg} = 600$. The periodic boundary conditions (PBC) were applied along the x and y directions. The non-periodic boundary condition was used along the z direction. The structural parameters a, h, d_1, d_2, θ_1 and θ_2 were obtained from the simulations results using Ovito [83] and with the help of Python [84] codes developed by the authors. The processes were automated with python codes which made the whole procedure less cumbersome.

3.3. Calculating the Mechanical Properties

Once the equilibrium structure had been determined, elastic constants were calculated. For this reason, the dependence of the elastic energy $\Delta E = E - E_0$ on the strain was computed. Here, E_0 and E represent the potential energies of equilibrium (unstrained) and strained systems, respectively. Three different deformations were considered: the equibiaxial strain with $\varepsilon_1 = \varepsilon_2 = x$ (later denoted by D_1), the volume preserving biaxial strain with $\varepsilon_1 = [(1+x)/(1-x)]^{1/2} - 1$

and $\varepsilon_2 = [(1-x)/(1+x)]^{1/2} - 1$ (later denoted by D_2), and the shear strain with $\varepsilon_6 = x$ (later denoted by D_3).

The equilibrated (*i.e.* corresponding to the zero strain) structure was deformed by incrementally dilating the simulation box along the loading direction and applying an equal affine transformation to the atomic positions. This was followed by minimizing the energy of the atomic coordinates. The elastic energy was calculated for 41 different values of x, equally spaced in the interval $[-0.005, 0.005]$ for each deformation type. This approach allowed us to obtain well-behaving (*i.e.* clearly quadratic) strain-energy dependencies.

The first two deformations were used to calculate the C_{11} and C_{12} elastic constants. The following expressions for the elastic energy density e (written in terms of the elastic constants C_{11} and C_{12}, and the magnitude of the applied strain x) were used:

$$e = \Delta E/A_0 \approx (C_{11} + C_{12})x^2 \tag{37}$$

for deformation D_1, and

$$e = \Delta E/A_0 \approx (C_{11} - C_{12})x^2 \tag{38}$$

for deformation D_2. Here, A_0 denotes the surface area of the unstrained system. The last deformation (D_3) allowed us to calculate C_{66}, since the following equation is obtained for it:

$$e = \Delta E/A_0 \approx \frac{1}{2} C_{66} x^2 \tag{39}$$

Having all the elastic constants C_{11}, C_{12} and C_{66} the Young's modulus E and the Poisson's ratio ν were calculated using the equations (8) and (9).

3.4. Training the Neural Network

Neural Networks seek to imitate the ability to adapt, intelligent decision making and the ability to properly process information of the brain. We designed a simple NN considering the time of training and the accuracy of the model. The input to the network were the structural and mechanical properties while the output were the Tersoff potential parameters.

3.4.1. Input to Neural Network

There are six structural properties from the *ab initio* [5] calculations which we seek to reproduce and three mechanical properties making it nine properties in total. These are: the lattice parameter a; the interlayer spacing h; the two lengths of C-C bonds, d_1 and d_2, respectively; two valence angles, θ_1 and θ_2, respectively and the three elastic constants, C_{11}, C_{12} and C_{66}. Hence our input has the shape of $m \times 9$. The output of the NN will be the Tersoff parameters; m, γ, λ_3, c, d, $\cos\theta_0$, n, β, λ_2, B, R, D, λ_1 and A. These are fourteen in number, but since some of these are chosen as constants, $m = 1$, $\lambda_3 = 0$, $n = 1$, $\beta = 1$, $R = 2$, $D = 0.15$, we will want to exclude them from the network. Thus, finally we have to following parameters left, γ, c, d, $\cos\theta_0$, λ_2, B, λ_1 and A. This implies that the output of the NN will be $m \times 8$. m is the number of observations, *i.e.*, the total number of rows in the dataset.

We needed a reasonable amount of training data to help our network learn well. To achieve this, we used random numbers (with the normal distribution). We generated a huge dataset containing sets of data ranging from 0.75 to 1.25 for each element (structural/mechanical property) of the Tersoff parameters of 2005 [31] with the mean being the parameters and a standard deviation of 0.5 of the corresponding parameter. The reason for such choice of ranges was that there were some number inputs that would cause the LAMMPS software to crash; we knew from the formula of the Tersoff potential that some parameters would appear in the denominator, hence a very small number in the denominator would result in a very large distortion which the system would be incapable of handling, hence, we could not just make totally random guesses. Also, the cosine naturally would yield values from -1 to 1, hence, we had to ensure that the generated values satisfy $-1 \leq \cos\theta_0 \leq 1$. Also, since there was a known result as shown by [23] which did well, it was wise to work around these parameters.

	a	h	d1	d2	theta_1	theta_2	c11	c22	c66	gamma	c	d	costheta0	lambda2	B	lambda1	A
0	4.146	0.838	1.791	1.738	102.651	124.193	106.2	0.7	65.9	0.110	171.982	6.632	-0.551	1.892	161.799	3.806	1985.688
1	4.076	0.869	1.779	1.731	103.816	121.493	107.3	1.8	66.5	0.120	183.996	6.805	-0.503	2.013	176.318	3.819	1872.268
2	3.506	0.691	1.508	1.458	102.123	125.449	255.0	-4.8	160.2	0.122	169.972	6.723	-0.567	1.849	164.490	4.283	2177.977
3	3.282	0.648	1.414	1.357	102.133	125.426	362.7	-31.2	240.4	0.117	186.029	6.405	-0.552	1.864	158.118	4.462	1985.200
4	3.506	0.689	1.512	1.434	102.006	125.730	330.0	-52.1	238.2	0.115	183.637	5.753	-0.535	1.894	183.856	4.177	1938.768
5	3.238	0.646	1.400	1.328	102.298	125.031	458.8	-72.3	329.6	0.115	193.159	5.883	-0.528	1.928	187.517	4.369	1874.340
6	3.395	0.623	1.446	1.377	100.684	128.993	343.9	-40.7	232.5	0.106	196.544	6.554	-0.595	1.816	173.850	4.262	1903.388
7	3.993	0.749	1.704	1.641	101.135	127.862	132.6	-6.2	85.2	0.110	180.698	6.682	-0.596	1.923	173.148	4.005	2218.558
8	4.243	0.886	1.844	7.331	103.329	122.610	84.5	-1.4	53.8	0.103	168.956	6.084	-0.519	2.070	174.686	3.833	1964.835
9	3.544	0.645	1.506	1.444	100.558	129.311	267.1	-21.4	174.3	0.107	176.285	6.522	-0.611	1.818	170.962	4.243	2141.723

Figure 9. Example of data used for training the NN

About eighty thousand (80000) datasets were generated. These were fed to the LAMMPS [82] software which was used to carry out the MS simulations as discussed in the previous sections and with the help of Python [84] and GNUPLOT [85] we were able to calculate the structural and mechanical properties for each dataset row. The output of the simulations (structural and mechanical properties) becomes the input to our NN while the input to the simulation software becomes the expected output of our NN as shown in Figure 9.

3.4.2. Neural Network Architecture

In Figure 10 we present architecture of the NN, which we found to work well with our data. There are 9 nodes for the input layer $I_1 ... I_9$ as expected since our input parameters (the structural and mechanical properties) are nine in number. We discovered that beyond 2 hidden layers the NN did not perform any better, hence we used 2 hidden layers $H^{(1)}, H^{(2)}$ with the first having 20 nodes $H_1^{(1)} ... H_{20}^{(1)}$ while the second had 10 nodes $H_1^{(2)} ... H_{10}^{(2)}$.

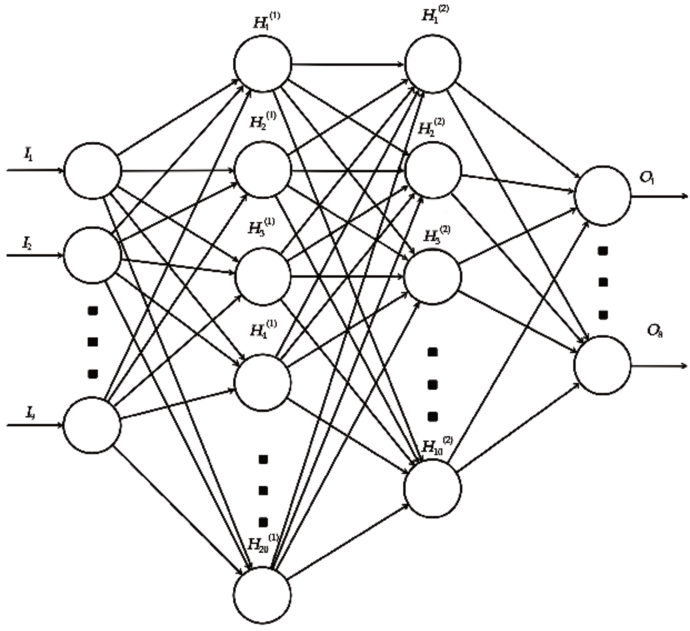

Figure 10. Architecture of the used NN

Each node of each hidden layer has the ReLu activation function. This is to create a non-linear behavior in the system. Since our output is a regression function, we do not add any activation function to it. As expected, the output layer O has 8 nodes $O_1..O_8$, these are the Tersoff parameters, our target.

Essentially, our goal was to find the Tersoff parameters that would reproduce the structural and mechanical properties of penta-graphene. The Is (9 in total) in the diagram are the structural and mechanical properties that we obtained from the simulations while the Os (8 in total) are the corresponding Tersoff parameters that yielded them.

3.4.3. Optimizer - Adam

Adam [86] (short for Adaptive Moment Estimation) is an updated version of the RMSProp [87] optimizer. In this optimization algorithm, running averages of both gradients and second moments of gradients are used. Given the parameters $w^{(t)}$ and a loss function $L^{(t)}$, where t indexes the current training iteration (indexed at 0), the Adam parameter update is given by:

$$m_w^{(t+1)} \leftarrow \beta_1 m_w^{(t)} + (1-\beta_1)\nabla_w L^{(t)}$$

$$v_w^{(t+1)} \leftarrow \beta_2 v_w^{(t)} + (1-\beta_2)(\nabla_w L^{(t)})^2$$

$$\widehat{m}_w = \frac{m_w^{(t+1)}}{1-\beta_1}$$

$$\hat{v}_w = \frac{v_w^{(t+1)}}{1-\beta_2}$$

$$w^{(t+1)} \leftarrow w^{(t)} - \eta \frac{\hat{m}_w}{\sqrt{\hat{v}_w} + \epsilon}$$

where ϵ is a small scalar (*e.g.* 10^{-8}) used to prevent division by 0, and β_1 (*e.g.* 0.9) and β_2 (*e.g.* 0.999) are the forgetting factors for gradients and second moments of gradients, respectively. Squaring and square-rooting is done element-wise.

3.4.4. Loss Function - Mean Squared Error (MSE)

We used the mean squared error due to its great performance on regression problems since the output of our network is regression. The mean squared error also known as the mean squared deviation (MSD) of an estimator measures the average of the squares of errors (*i.e.*, the average squared difference between the estimated values and the actual value). MSE is a risk function, corresponding to the expected value of the squared error loss. The MSE is a measure of the quality of the estimator, it is always non-negative, and values closer to zero are better.

MSE is computed in the following way:

$$MSE = \frac{1}{m} \sum_{i=1}^{m} \left(Y_i - \hat{Y}_i \right)^2 \tag{40}$$

where m is the number of rows in our dataset, Y_i is vector containing the Tersoff parameters, while $\hat{Y}_i$ is a vector of the Tersoff parameters predicted by the NN.

3.5. Computing the Structural and Mechanical Properties from NN results

When we have trained the NN and obtained a minimal MSE, then we feed the structural and mechanical properties $a = 3.64$ Å, $h = 0.6$ Å, $d_1 = 1.55$ Å, $d_2 = 1.34$ Å, $\theta_1 = 98.6^\circ$, $\theta_2 = 134.2^\circ$, $C_{11} = 265$ GPa, $C_{12} = -18$ GPa and $C_{66} = 152$ GPa from the *ab initio* calculations into the trained model to obtain the predicted results which represent the target Tersoff parameters. Next, we feed these parameters into LAMMPS and perform MS simulations to get our desired penta-graphene properties.

4. Results

4.1. Structural Properties

We now present the parameters obtained together with their structural and mechanical properties. First, we present the results of the structural properties. We observed that the Tersoff parameters found by us performed better than the T05 parameters in describing the structural parameters of penta-graphene as seen in Table 2.

Table 2. Structural parameters of penta-graphene as predicted by T05 Tersoff parameters and parameters obtained in this work. For the convenience of comparison, the results of the *ab initio* calculations [5] are also presented. The value in the parentheses represents the signed relative error (in percents)

Potential	a (Å)	h (Å)	d_1 (Å)	d_2 (Å)	θ_1 (°)	θ_2 (°)
ab initio	3.64	0.6	1.55	1.34	98.6	134.2
T05	3.5923 (-1.309)	0.7057 (17.62)	1.5461 (-0.25)	1.4831 (10.68)	102.0251 (3.474)	125.6843 (-6.346)
new parameters	0.6058 (-0.941)	0.6938 (15.627)	1.5463 (-0.237)	1.4827 (10.648)	101.612 (3.055)	126.6855 (-5.6)

The relative errors were computed from Table 2 and Table 3 to give

$$Relative\_error = \frac{Result - ab\_initio}{ab\_initio} \times 100 \tag{41}$$

Then we computed the absolute difference in the relative error, ADRE of T05 obtaining

$$ADRE = \frac{Relative\_error\_T05 - Relative\_error\_Result}{Relative\_error\_T05} \times 100 \tag{42}$$

The absolute difference in the relative error of T05 and our result calculated for the lattice parameter a is $-1.309-(-0.941)=0.368$, which is a significant improvement of about 28% from the T05 result; for the interlayer spacing h we have $17.62-15.627=1.993$, and this also is a significant improvement of about 11% from the T05 result; two lengths of the C-C bonds, d_1 and d_2, respectively yield $-0.25-(-0.237)=0.013$ and $10.68-10.648=0.032$ with improvement of 5.2% and 0.3%, respectively. While the improvement for d_1 is significant, it is very for d_2; two valence angles, $\theta_1 = 3.474-3.055=0.121$ and $\theta_2 = -6.346-(-5.6)=0.746$, both with improvement of 12%, which is significant. These are clear when viewed in the bar-chart in Figures 11 and 12.

4.2. Mechanical Properties

Now, the results of the mechanical properties are presented. We observed that the Tersoff parameters found by us also performed better than the T05 parameters in describing the mechanical properties of penta-graphene as seen in Table 3.

Table 3. Mechanical properties of penta-graphene as predicted by T05 Tersoff parameters and parameters obtained in this work. The elastic constants C11, C12, C66 and Young's modulus E are expressed in units of GPa nm. For the convenience of comparison, the results of the *ab initio* calculations [5] are also presented. The value in the parentheses represents the signed relative error (in percents)

Potential	C_{11}	C_{12}	$C_{66}=\mu$	E	ν
ab initio	265	-18.0	152	264	-0.068
T05	244.1(-7.89)	-20.6(14.4)	162.3(6.78)	242.36(-8.20)	0.0844(24.1)
new parameters	244.3(-7.81)	-19.5(8.33)	160.8(5.79)	242.74(-8.05)	0.0798(17.4)

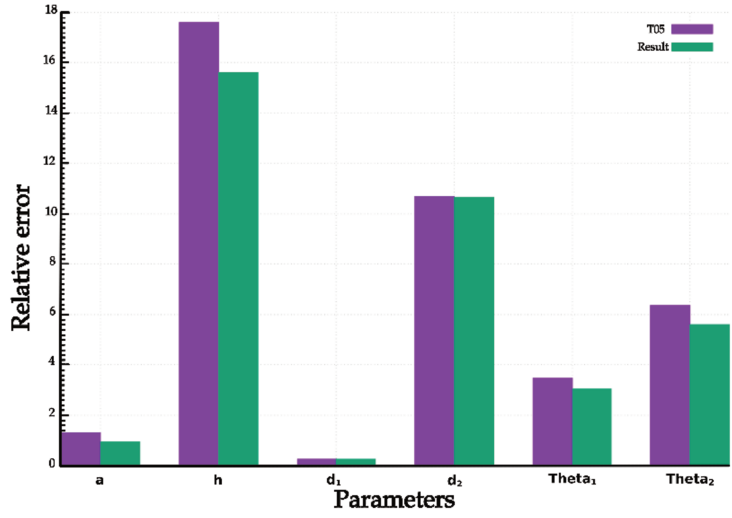

Figure 11. Relative error of calculated structural properties compared with *ab initio* structural properties

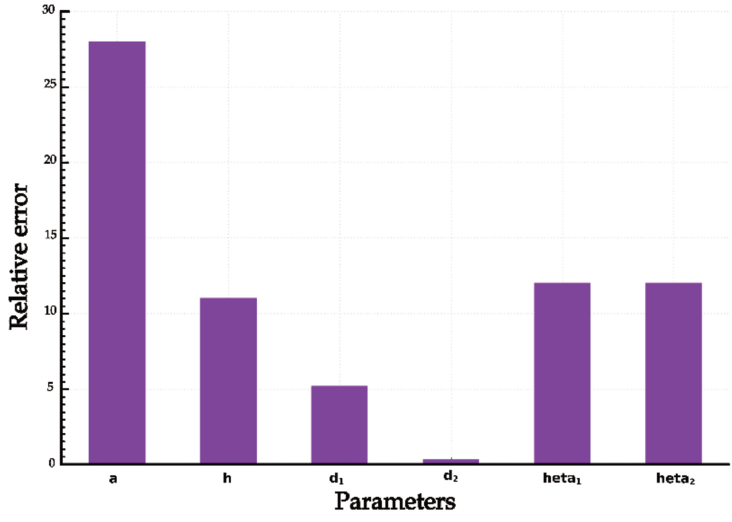

Figure 12. Difference in error of T05 structural properties and structural properties calculated by the authors

The absolute difference in the relative error of T05 and our result calculated for the elastic constant C_{11} is $-7.887-(-7.811)=0.076$, which is an improvement of about 0.96% from the T05 result which is small; for the elastic constant C_{12} we have $14.444-8.333=6.111$, which is a significant improvement of about 42.3% from the T05 result; while the elastic constant C_{66} yields $6.776-5.789=0.987$ with improvement of about 15% which is significant; for Young's modulus E we have $-8.197-(-8.053)=0.144$, this is a slight improvement of about 1.76% compared

with the T05 result; while the Poisson's ratio ν yields $24.1 - 17.4 = 6.7$ with improvement of about 27.8% which is significant. These are clear when viewed in the bar-chart in Figure 13 and 14.

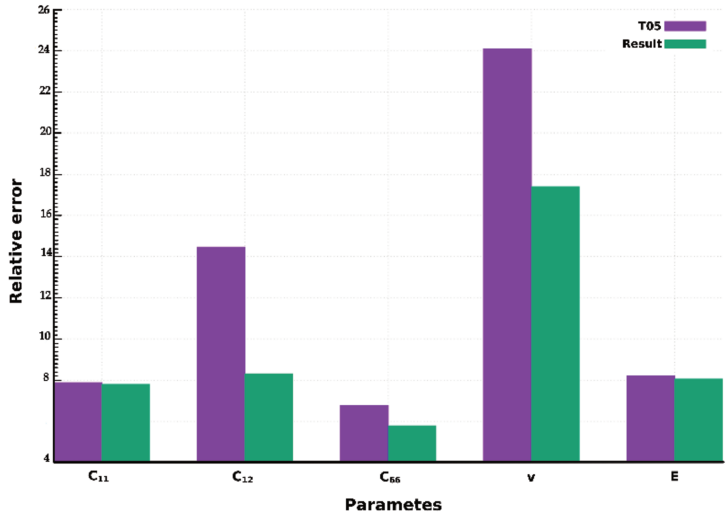

Figure 13. Relative error of calculated mechanical properties compared with *ab initio* mechanical properties

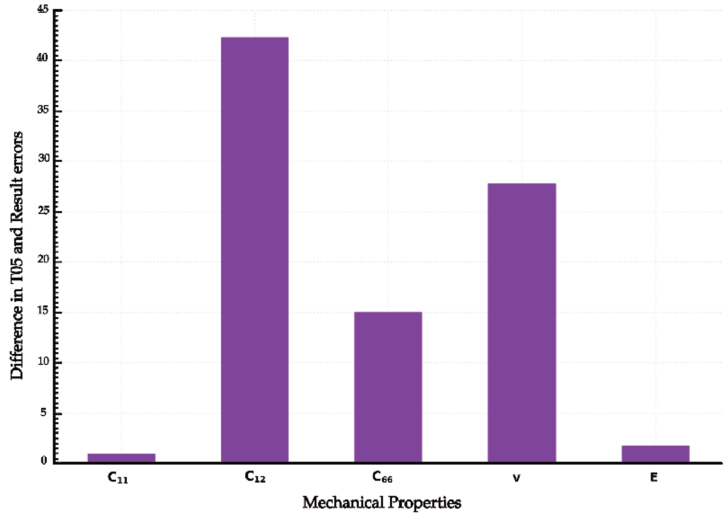

Figure 14. Difference in error of T05 mechanical properties and mechanical properties calculated by the authors

Therefore we conclude that the Tersoff parameters found by us and shown in Table 4 are the choice parameters for the Tersoff potential when modeling PG

Table 4. Parameters of the optimized Tersoff potential

Parameter	**Value**
A (eV)	1978.514
B (eV)	185.782
λ_1 (Å^{-1})	4.112
λ_2(Å^{-1})	1.922
λ_3(Å^{-1})	0
β	1
n	1
c	181.91
d	66.37
$\cos\theta_0$	-0.574
R (Å)	2
D (Å)	0.15
m	1
γ	0.104

and taking the structural and mechanical properties as the main indicator of the quality of the potential.

5. Summary

We produced about hundred thousand inputs to LAMMPS with the use of randomly generated numbers. We ran simulations within a couple of weeks to generate inputs to our neural networks using the computational power of Tryton. We built a neural network and trained the network using the outputs from LAMMPS employing the functionality of Python with such libraries as Pandas, Numpy and Pytorch. We finally obtained the results - new parameterization of the Tersoff potential - after series of trainings and twisting of parameter values. We did it with the goal to find suitable parameters for the Tersoff potential appropriate for modeling penta-graphene.

We calculated the properties of PG at 0 K using the Molecular Statics method. Comparing the obtained results with the available *ab initio* data we demonstrated that the parameters for the Tersoff potential found by us are able to describe PG more accurately than the original parameters proposed by Erhart and Albe in 2005 [31]. We showed that these parameters gave structural and mechanical properties complying with the *ab initio* studies.

References

[1] Berger C, Song Z, Li T, Li X, Ogbazghi A Y, Feng R, Dai Z, Marchenkov A N, Conrad E H, First P N and de Heer W A 2004 *Ultrathin Epitaxial Graphite: 2D Electron Gas Properties and a Route toward Graphene-based Nanoelectronics*, *The Journal of Physical Chemistry*, American Chemical Society (ACS), **108** (52) 19912

[2] Novoselov K S, Geim A K, Morozov S V, Jiang D, Zhang Y, Dubonos S V, Grigorieva I V and Firsov A A 2004 *Electric Field Effect in Atomically Thin Carbon Films*, American Association for the Advancement of Sciences, **306** (5696) 666

[3] Iijima S 1991 *Helical microtubules of graphitic carbon*, *Nature*, Springer Science and Business Media LLC, **354** (6348) 56

[4] Ebbesen T W and Ajayan P M 1992 *Large-scale synthesis of carbon nanotubes*, *Nature*, Springer Science and Business Media LLC, **358** (6383) 220

[5] Zhang S, Zhou J, Wang Q, Chen X, Kawazoe Y and Jena P 2015 *Penta-graphene: A new carbon allotrope*, *Proceedings of the National Academy of Sciences*, Proceedings of the National Academy of Sciences, **112** (8) 2372

[6] Zhao Z, Tian F, Dong X, Li Q, Wang Q, Wang H, Zhong X, Xu B, Yu D, He J, Wang H, Ma Y and Tian Y 2012 *Tetragonal Allotrope of Group 14 Elements*, *Journal of the American Chemical Society*, American Chemical Society (ACS), **134** (30) 12362

[7] Sun H, Mukherjee S and Singh C V 2016 *Mechanical properties of monolayer penta-graphene and phagraphene: a first-principles study*, *Physical Chemistry Chemical Physics*, Royal Society of Chemistry (RSC), **18** (38) 26736

[8] Ewels C P, Rocquefelte X, Kroto H W, Rayson M J, Briddon P R and Heggie M I 2015 *Predicting experimentally stable allotropes: Instability of penta-graphene*, *Proceedings of the National Academy of Sciences*, Proceedings of the National Academy of Sciences, **112** (51) 15609

[9] Cranford S W 2016 *When is 6 less than 5? Penta- to hexa-graphene transition*, *Carbon*, Elsevier BV, **96** 421

[10] Berdiyorov G R, Dixit G and Madjet M E 2016 *Band gap engineering in penta-graphene by substitutional doping: first-principles calculations*, *Journal of Physics: Condensed Matter*, IOP Publishing, **28** (47) 475001

[11] Li X, Zhang S, Wang F Q, Guo Y, Liu J L and Wang Q 2016 *Tuning the electronic and mechanical properties of penta-graphene via hydrogenation and fluorination*, *Physical Chemistry Chemical Physics*, Royal Society of Chemistry (RSC), **18** (21) 14191

[12] Xu W, Zhang G and Li B 2015 *Thermal conductivity of penta-graphene from molecular dynamics study*, *The Journal of Chemical Physics*, AIP Publishing, **143** (15) 154703

[13] Xufei Wu X, Varshney V, Lee J, Zhang T, Wohlwend J L, Roy A K and Luo T 2016 *Hydrogenation of Penta-Graphene Leads to Unexpected Large Improvement in Thermal Conductivity*, *Nano Letters*, American Chemical Society (ACS), **16** (6) 3925

[14] Yu Z G and Zhang Y 2015 *A comparative density functional study on electrical properties of layered penta-graphene*, *Journal of Applied Physics*, AIP Publishing, **118** (16) 165706

[15] Rajbanshi B, Sarkar S, Mandal B, Sarkar P 2016 *Energetic and electronic structure of penta-graphene nanoribbons*, *Carbon*, Elsevier BV, **100** 118

[16] Xiao B, Li Y, Yu X and Cheng J 2016 *Penta-graphene: A Promising Anode Material as the Li/Na-Ion Battery with Both Extremely High Theoretical Capacity and Fast Charge/Discharge Rate*, *ACS Applied Materials & Interfaces*, American Chemical Society (ACS), **8** (51) 35342

[17] Ebrahimi S 2016 *Effect of hydrogen coverage on the buckling of penta-graphene by molecular dynamics simulation*, *Molecular Simulation*, Informa UK Limited, **42** (17) 1485

[18] Avramov P, Demin V, Luo M, Choi C H, Sorokin P B, Yakobson B and Chernozatonskii L 2015 *Translation Symmetry Breakdown in Low-Dimensional Lattices of Pentagonal Rings*, *The Journal of Physical Chemistry Letters*, American Chemical Society (ACS), **6** (22) 4525

[19] Stauber T, Beltrán J I and Schliemann J 2016 *Tight-binding approach to penta-graphene*, *Scientific Reports*, Springer Science and Business Media LLC, **6** (1) 2045

[20] Rahaman O, Mortazavi B, Dianat A, Cuniberti G and Rabczuk T 2017 *Metamorphosis in carbon network: From penta-graphene to biphenylene under uniaxial tension*, *FlatChem*, Elsevier BV, **1** 65

[21] Wei X, Fragneaud B, Marianetti C A and Kysar J W 2009 *Nonlinear elastic behavior of graphene: Ab initio calculations to continuum description*, *Physical Review B*, American Physical Society (APS), **80** (20) 205407

[22] Pop E, Varshney V and Roy A K 2012 *Thermal properties of graphene: Fundamentals and applications*, *MRS Bulletin*, Cambridge University Press (CUP), **37** (12) 1273

[23] Winczewski S, Shaheen M Y and Rybicki J 2018 *Interatomic potential suitable for the modeling of penta-graphene: Molecular statics/molecular dynamics studies*, *Carbon*, Elsevier BV, **126** 165

[24] Chenoweth K, van Duin A C T and Goddard W A 2008 *ReaxFF Reactive Force Field for Molecular Dynamics Simulations of Hydrocarbon Oxidation*, *The Journal of Physical Chemistry*, American Chemical Society (ACS), **112** (5) 1040

[25] Brenner D W, Shenderova O A, Harrison J A, Stuart S J, Ni B and Sinnott S B 2002 *A second-generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons*, *Journal of Physics: Condensed Matter*, IOP Publishing, **14** (4) 783

[26] Tersoff J 1989 *Modeling solid-state chemistry: Interatomic potentials for multicomponent systems*, *Physical Review B*, American Physical Society (APS), **39** (8) 5566

[27] Tersoff J 1990 *Erratum: Modeling solid-state chemistry: Interatomic potentials for multicomponent systems*, *Physical Review B*, American Physical Society (APS), **41** (5) 3248

[28] Lindsay L and Broido D A 2010 *Optimized Tersoff and Brenner empirical potential parameters for lattice dynamics and phonon thermal transport in carbon nanotubes and graphene*, *Physical Review B*, American Physical Society (APS), **81** (20) 205441

[29] Lindsay L and Broido D A 2010 *Erratum: Optimized Tersoff and Brenner empirical potential parameters for lattice dynamics and phonon thermal transport in carbon nanotubes and graphene*, *Physical Review B*, American Physical Society (APS), **82** (20)

[30] Marks N 2002 *Modelling diamond-like carbon with the environment-dependent interaction potential*, *Journal of Physics: Condensed Matter*, IOP Publishing, **14** (11) 2901

[31] Erhart P and Albe K 2005 *Analytical potential for atomistic simulations of silicon, carbon, and silicon carbide*, *Physical Review B*, American Physical Society (APS), **71** (3) 35211

[32] P. Mahon P, Pailthorpe B A and Bacskay G B 1991 *A quantum mechanical calculation of interatomic interactions in diamond*, *Philosophical Magazine B*, Informa UK Limited, **63** (6) 1419

[33] Barnard A S and Russo S P 2002 *Development of an improved Stillinger-Weber potential for tetrahedral carbon using ab initio (Hartree-Fock and MP2) methods*, *Molecular Physics*, Informa UK Limited, **100** (10) 1517

[34] Brenner D W 1990 *Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamond films*, *Physical Review B*, American Physical Society (APS), **42** (15) 9458

[35] Brenner D W 1992 *Erratum: Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamond films*, *Physical Review B*, American Physical Society (APS), **46** (3) 1948

[36] Stuart S J, Tutein A B and Harrison J A 2000 *A reactive potential for hydrocarbons with intermolecular interactions*, *The Journal of Chemical Physics*, AIP Publishing, **112** (14) 6472

[37] O'Connor T C, Jan Andzelm J and Robbins M O 2015 *AIREBO-M: A reactive model for hydrocarbons at extreme pressures*, *The Journal of Chemical Physics*, AIP Publishing, **142** (2) 24903

[38] van Duin A C T, Dasgupta S, Lorant F and Goddard W A 2001 *ReaxFF: A Reactive Force Field for Hydrocarbons*, *The Journal of Physical Chemistry*, American Chemical Society (ACS), **105** (41) 9396

[39] Nielson K D, van Duin A C T, Oxgaard J, Deng W and Goddard W A 2005 *Development of the ReaxFF Reactive Force Field for Describing Transition Metal Catalyzed Reactions, with Application to the Initial Stages of the Catalytic Formation of Carbon Nanotubes*, *The Journal of Physical Chemistry*, American Chemical Society (ACS), **109** (3) 493

[40] Srinivasan S G, van Duin A C T and Ganesh P 2015 *Development of a ReaxFF Potential for Carbon Condensed Phases and Its Application to the Thermal Fragmentation of a Large Fullerene*, *The Journal of Physical Chemistry*, American Chemical Society (ACS), **119** (4) 571

[41] M. I. Baskes 1992 *Modified embedded-atom potentials for cubic materials and impurities*, *Physical Review B*, American Physical Society (APS), **46** (5) 2727

[42] Los J H and Fasolino A 2003 *Intrinsic long-range bond-order potential for carbon: Performance in Monte Carlo simulations of graphitization*, *Physical Review B*, American Physical Society (APS), **68** (2) 24107

[43] Zhou X W, Ward D K and Foster M E 2015 *An analytical bond-order potential for carbon*, *Journal of Computational Chemistry*, Wiley, **36** (23) 1719

[44] Marks N A 2000 *Generalizing the environment-dependent interaction potential for carbon*, *Physical Review B*, American Physical Society (APS), **63** (3) 35401

[45] Yin M T and Cohen M L 1980 *Microscopic Theory of the Phase Transformation and Lattice Dynamics of Si*, *Physical Review Letters*, American Physical Society (APS), **45** (12) 1004

[46] Yin M T and Cohen M L 1982 *Theory of static structural properties, crystal stability, and phase transformations: Application to Si and Ge*, *Physical Review B*, American Physical Society (APS), **26** (10) 5668

[47] Pandey K C 1981 *New*Π*-Bonded Chain Model for Si(111)-(2×1) Surface*, *Physical Review Letters*, American Physical Society (APS), **47** (26) 1913

[48] Northrup J E 1986 *Origin of surface states on Si(111)(7×7)*, *Physical Review Letters*, American Physical Society (APS), **57** (1) 154

[49] Car R, Kelly P J, Oshiyama A and Pantelides S T 1984 *Microscopic Theory of Atomic Diffusion Mechanisms in Silicon*, *Physical Review Letters*, American Physical Society (APS), **52** (20) 1814

[50] Dodson B W 1987 *Development of a many-body Tersoff-type potential for silicon*, *Physical Review B*, American Physical Society (APS), **35** (6) 2795

[51] Abell G C 1985 *Empirical chemical pseudopotential theory of molecular and metallic bonding*, *Physical Review B*, American Physical Society (APS), **31** (10) 6184

[52] Kane E O 1985 *Phonon spectra of diamond and zinc-blende semiconductors*, *Physical Review B*, American Physical Society (APS), **31** (12) 7865

[53] Keating P N 1966 *Effect of Invariance Requirements on the Elastic Strain Energy of Crystals with Application to the Diamond Structure*, *Physical Review B*, American Physical Society (APS), **145** (2) 637

[54] Stillinger F H and Weber t A 1985 *Computer simulation of local order in condensed phases of silicon*, *Phys. Rev. B*, American Physical Society, **31** 5262

[55] Biswas R and Hamann D R 1987 *New classical models for silicon structural energies*, *Phys. Rev. B*, American Physical Society, **36** 6434

[56] Tersoff J 1988 *Empirical interatomic potential for silicon with improved elastic properties*, *Phys. Rev. B*, American Physical Society, **38** 9902

[57] Tersoff J 1988 *Empirical Interatomic Potential for Carbon, with Applications to Amorphous Carbon*, *Phys. Rev. Lett.*, American Physical Society, **61** 2879

[58] Zhang S, Zhou J, Wang Q, Chen X, Kawazoe Y and Jena P 2015 *Penta-graphene: A new carbon allotrope*, *Proceedings of the National Academy of Sciences*, Proceedings of the National Academy of Sciences, **112** (8) 2372

[59] Ewels C P, Rocquefelte X, Kroto H W, Rayson M J, Briddon P R and Heggie M I 2015 *Predicting experimentally stable allotropes: Instability of penta-graphene*, *Proceedings of the National Academy of Sciences*, Proceedings of the National Academy of Sciences, **112** (51) 15609

[60] Einollahzadeh H, Fazeli S M and Dariani R S 2016 *Studying the electronic and phononic structure of penta-graphane*, *Science and Technology of Advanced Materials*, Informa UK Limited, **17** (1) 610

[61] Bataineh M H 2012 *Artificial neural network for studying human performance*, The University of Iowa

[62] Shimamura K, Fukushima S, Koura A, Shimojo F, Misawa M, Kalia R K, Nakano A, Vashishta P, Matsubara T and Tanaka S 2019 *Guidelines for creating artificial neural network empirical interatomic potential from first-principles molecular dynamics data under specific conditions and its application to -Ag2Se*, *The Journal of Chemical Physics*, AIP Publishing, **151** (12) 124303

[63] Hobday S, Smith R and Belbruno J 1999 *Applications of neural networks to fitting interatomic potential functions*, *Modelling and Simulation in Materials Science and Engineering*, IOP Publishing, **7** (3) 397

[64] Lee K, Yoo D, Jeong W and Han S 2019 *SIMPLE-NN: An efficient package for training and executing neural-network interatomic potentials*, *Computer Physics Communications*, Elsevier BV, **242** 95

[65] Purja Pun G P, Batra R, Ramprasad R and Mishin Y 2019 *Physically informed artificial neural networks for atomistic modeling of materials*, *Nature Communications*, Springer Science and Business Media LLC, **10** (1) 2339

[66] Bukkapatnam S, Malshe M, Agrawal P M, Raff L M and Komanduri R 2006 *Parametrization of interatomic potential functions using a genetic algorithm accelerated with a neural network*, *Physical Review B*, American Physical Society (APS), **74** (22) 224102

[67] McCulloch W S and Pitts W 1943 *A logical calculus of the ideas immanent in nervous activity*, *The Bulletin of Mathematical Biophysics*, Springer Science and Business Media LLC, **5** (4) 115

[68] Kleene S C 1956 *Representation of events in nerve nets and finite automata*, Princeton University Press

[69] Morris R G M 1999 *D.O. Hebb: The Organization of Behavior, Wiley: New York; 1949*, *Brain Research Bulletin*, Elsevier BV, **50** (5-6) 437

[70] Farley B and Clark W 1954 *Simulation of self-organizing systems by digital computer*, *Transactions of the IRE Professional Group on Information Theory*, Institute of Electrical and Electronics Engineers (IEEE), **4** (4) 76

[71] Rochester N, Holland J, Haibt L and Duda W 1956 *Tests on a cell assembly theory of the action of the brain, using a large digital computer*, *IEEE Transactions on Information Theory*, Institute of Electrical and Electronics Engineers (IEEE), **2** (3) 80

[72] Rosenblatt F 1958 *The perceptron: A probabilistic model for information storage and organization in the brain*, *Psychological Review*, American Psychological Association (APA), **65** (6) 386

[73] Werbos P J 1974 *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*

[74] Hubel D H and Wiesel T N 2004 *Brain and Visual Perception*, Oxford University Press

[75] Schmidhuber J 2015 *Deep learning in neural networks: An overview*, *Neural Networks*, Elsevier BV, **61** 85

[76] Ivakhnenko A G and Lapa V G 1973 *Cybernetic Predicting Devices*, CCM Information Corporation
[77] Ivakhnenko A G, Lapa V G and McDonough R N 1967 *Cybernetics and Forecasting Techniques*, American Elsevier Publishing Company
[78] Minsky M and Seymour P 1969 *Perceptrons: An introduction to computational geometry*, MIT Press
[79] Schmidhuber J 2015 *Deep learning in neural networks: An overview*, *Neural Networks*, Elsevier BV, **61** 85
[80] Rosenblatt F 1958 *The perceptron: A probabilistic model for information storage and organization in the brain*, *Psychological Review*, American Psychological Association (APA), **65** (6) 386
[81] Polak E and Ribiere G 1969 *Note sur la convergence de methodes de directions conjuguees*, *ESAIM: Mathematical Modelling and Numerical Analysis - Modelisation Mathematique et Analyse Numerique*, Dunod, **3** (R1) 35
[82] Plimpton S 1995 *Fast Parallel Algorithms for Short-Range Molecular Dynamics*, *Journal of Computational PhysicS*, Elsevier BV, **117** (1) 1
[83] Stukowski A *Visualization and analysis of atomistic simulation data with OVITO - the Open Visualization Tool*
[84] Van Rossum G and Drake F 2009 *Python 3 Reference Manual*, Scotts Valley, CA: CreateSpace
[85] Williams T and Kelley C *Gnuplot 4.6: an interactive plotting program*
[86] Kingma D P and Ba J 2014 *Adam: A Method for Stochastic Optimization*
[87] Hinton G 2012 *Lecture 6e rmsprop: Divide the gradient by a running average of its recent magnitude*, University of Toronto
[88] Golda A 2004 *Principles of training multi-layer neural network using backpropagation*, *http://galaxy.agh.edu.pl/∼vlsi/AI/backp_t_en/backprop.html*, AGH University of Science and Technology
[89] Car R and Parrinello M 1985 *Unified Approach for Molecular Dynamics and Density-Functional Theory*, *Physical Review Letters*, American Physical Society (APS), **55** (22) 2471
[90] Payne M C, Bristowe P D and Joannopoulos J D 1987 *Ab initio determination of the structure of a grain boundary by simulated quenching*, *Physical Review Letters*, American Physical Society (APS), **58** (13) 1348
[91] Quijano-Briones J J, Fernández-Escamilla H N and Tlahuice-Flores A 2017 *Chiral penta-graphene nanotubes: Structure, bonding and electronic properties*, *Computational and Theoretical Chemistry*, Elsevier BV, **1108** 70
[92] Avramov P, Demin V, Luo M, Choi C H, Sorokin P B, Yakobson B and Chernozatonskii L 2015 *Translation Symmetry Breakdown in Low-Dimensional Lattices of Pentagonal Rings*, *The Journal of Physical Chemistry Letters*, American Chemical Society (ACS), **6** (22) 4525
[93] Su C, Jiang H and Feng J 2013 *Two-dimensional carbon allotrope with strong electronic anisotropy*, *Physical Review B*, American Physical Society (APS), **87** (7) 75453
[94] Yin W, Xie Y, Liu L, Wang R, Wei X, Lau L, Zhong J and Chen Y 2013 *R-graphyne: a new two-dimensional carbon allotrope with versatile Dirac-like point in nanoribbons*, *Journal of Materials Chemistry*, Royal Society of Chemistry (RSC), **1** (17) 5341
[95] Sharma B R, Manjanath A and Singh A K 2014 *pentahexoctite: A new two-dimensional allotrope of carbon*, *Scientific Reports*, Springer Science and Business Media LLC, **4** (1)
[96] Wang Z, Zhou X F, Zhang X, Zhu Q, Dong H, Zhao M and Oganov A R 2015 *Phagraphene: A Low-Energy Graphene Allotrope Composed of 5–6–7 Carbon Rings with Distorted Dirac Cones*, *Nano Letters*, American Chemical Society (ACS), **15** (9) 6182

[97] Zhang X, Wei L, Tan J and Zhao M 2016 *Prediction of an ultrasoft graphene allotrope with Dirac cones*, *Carbon*, Elsevier BV, **105** 323

[98] Jiang X, Arhammar C, Liu P, Zhao J and Ahuja R 2013 *The R3-carbon allotrope: a pathway towards glassy carbon under high pressure*, *Scientific Reports*, Springer Science and Business Media LLC, **3** (1) 1877

[99] Zhanpeisov N U 2012 *Theoretical DFT study on structure and chemical activity of new carbon K4 clusters*, *Research on Chemical Intermediates*, Springer Science and Business Media LLC, **39** (5) 2141

[100] Zhang S, Wang Q, Chen X and Jena P 2013 *Stable three-dimensional metallic carbon with interlocking hexagons*, *Proceedings of the National Academy of Sciences*, Proceedings of the National Academy of Sciences, **110** (47) 18809

[101] Burchfield L A, Al Fahim M, Wittman R S, Delodovici F and Manini N 2017 *Novamene: A new class of carbon allotropes*, *Heliyon*, Elsevier BV, **3** (2) 242

[102] Stukowski A 2009 *Visualization and analysis of atomistic simulation data with OVITO — the Open Visualization Tool*, *Modelling and Simulation in Materials Science and Engineering*, IOP Publishing, **18** (1) 15012

[103] Tersoff J 1988 *Empirical interatomic potential for silicon with improved elastic properties*, *Physical Review B*, American Physical Society (APS), **38** (14) 9902

[104] Murani A P, Mattens W C M, de Boer F R and Lander G H 1985 *Neutron-inelastic-scattering study of the compound YbCuAl*, *Physical Review B*, American Physical Society (APS), **31** (1) 52

[105] Plimpton S 1995 *Fast parallel algorithms for short-range molecular dynamics*, *J. Comput. Phys.* **11** 1