Title

Liquid to crystal Si growth simulation using machine learning force field

Journal

The Journal of Chemical Physics, 153(7)

ISSN

0021-9606

Authors

Miao, Ling

Wang, Lin-Wang

Publication Date

2020-08-21

DOI

10.1063/5.0011163

Peer reviewed

The Journal of Chemical Physics

Liquid to crystal Si growth simulation using machine learning force field

Cite as: J. Chem. Phys. **153**, 074501 (2020); https://doi.org/10.1063/5.0011163
Submitted: 18 April 2020 . Accepted: 31 July 2020 . Published Online: 17 August 2020

Ling Miao, and Lin-Wang Wang

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

Liquid to crystal Si growth simulation using machine learning force field

Cite as: J. Chem. Phys. 153, 074501 (2020); doi: 10.1063/5.0011163
Submitted: 18 April 2020 • Accepted: 31 July 2020 •
Published Online: 17 August 2020

Ling Miao[1] and Lin-Wang Wang[2,a)]

AFFILIATIONS

[1]School of Optical and Electronic Information, Huazhong University of Science and Technology, Wuhan 430074, People's Republic of China
[2]Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

Note: This paper is part of the JCP Special Topic on Machine Learning Meets Chemical Physics.
[a)]**Author to whom correspondence should be addressed:** lwwang@lbl.gov

ABSTRACT

Machine learning force field (ML-FF) has emerged as a potential promising approach to simulate various material phenomena for large systems with *ab initio* accuracy. However, most ML-FFs have been used to study the phenomena relatively close to the equilibrium ground states. In this work, we have studied a far from equilibrium system of liquid to crystal Si growth using ML-FF. We found that our ML-FF based on *ab initio* decomposed atomic energy can reproduce all the aspects of *ab initio* simulated growth, from local energy fluctuations to transition temperatures, to diffusion constant, and growth rates. We have also compared the growth simulation with the Stillinger–Weber classical force field and found significant differences. A procedure is also provided to correct a systematic fitting bias in the ML-FF training process, which exists in all training models, otherwise critical results like transition temperature will be wrong.

INTRODUCTION

Machine learning force field (ML-FF) trained based on *ab initio* quantum mechanics calculations has become a very active field of research in recent years. It has potential to replace the classical empirical force field to simulate large scale material problems with accuracy similar to that of *ab initio* calculations, as shown in many types of developed ML-FFs, including water,[1] TiO_2,[2] ZnO,[3] Li_3PO_4,[4] compounds,[5] binary, ternary alloys,[6] and metal–organic frameworks.[7] So far, ML-FFs have mostly been used to study the phenomena relatively close to the equilibrium ground states, e.g., the phonon spectrum, the anharmonic properties, and the solid phase transition with relatively small distortions. One important beyond equilibrium phenomenon is the crystal growth, e.g., from liquid phase. During the growth, the local configuration {**R**} can change dramatically, especially for covalent bonding systems like Si. It thus provides a stringent test for ML-FF.

Not only melt Si growth has tremendous economic significance, the process involving covalent bond breaking and formation, and the phenomenon of step flow growth is also of great physical interest. In the literature, Si crystal growth has been extensively studied using empirical potentials like the Stillinger–Weber (SW) potential[8] and Tersoff potential.[9] While such potential fits well some macroscopic properties such as the phase transition temperature Tc, they could also misrepresent the microscopic bond forming and breaking processes since there is no direct experimental data to fit with. On the other hand, it is too expensive to use density functional theory (DFT) to simulate the crystal growth directly for large systems. Si is one of the most studied systems in terms of ML-FF development.[10,11] For example, Csányi *et al.* has developed a rather comprehensive Si GAP potential for different ground state phases including crystal and amorphous structures.[12] Our previous single atom neural network potential (SANNP) is also an ML-FF for amorphous Si.[13] However, such ML-FF has not been used for crystal growth simulations, and such simulations pose unique challenges. For example, we found that it is essential to correct a systematic bias in almost all the training models, otherwise critical results like transition temperature will be wrong.

To study extended systems, most ML-FFs assume that the total energy of the whole system is a sum of the local atomic energy Ei of each atom "i," which depends only on the local atomic configuration {**R**} within a cut-off distance Rc surrounding this atom "i." To

build a ML-FF model, in most cases, the first step is to extract a set of features from the atomic configuration {**R**}. Such features need to be rotational and permutation invariant and also to be continuous if an atom coming into or going out of the Rc.[10] There are different ways to extract such features or distant kernels, including the atom-centered symmetry functions (ACSF),[14] the smooth-overlap of atomic positions (SOAP),[15] the bispectrum of atomic neighbor density functions,[16] and the contracted moment tensors.[17] After the features are extracted, a ML approach is used to fit the *ab initio* calculated data point. Different ML models can be used, such as all connected dense neural network (DNN),[14] Gaussian process regression,[15] and linear regression.[16,17] Theoretically, different features' set and ML models can be mutually combined, forming a large number of models.

In a previous work,[13] we have introduced a single atom neural network potential (SANNP), with a new set of features based on the localized function. One advantage of this II-III-body feature set is the easiness to systematically increase the number of features by increasing the number of localized functions in the pair and triangle feature terms. We have used a DNN as our ML model. What distinguishes the SANNP from all the other models is that we directly fitted the density functional theory (DFT) atomic energy E_i from our *ab initio* calculations.[18] This is achieved by representing the DFT total energy as a spatial integration of a positive energy density, followed by a Hirshfeld partitioning using isolated atom charges to assign E_i for each atom. For the positive energy density, the kinetic energy is written as an integral of $\frac{1}{2}\sum_i |\nabla\psi_i(r)|^2$, where ψ_i is the Kohn-Sham wave function, and the electrostatic energy is written as a spatial sum of the square of the electric field. More details are given in Ref. 18. The availability of E_i for each atom can significantly simplify the ML fitting process, and it requires much less *ab initio* calculations. Although atomic forces are also used in the fitting of most ML-FF's (including this work), but as it has been pointed out in a recent study,[19] relying only on the atomic forces and the total energy of the supercell can cause major transferability issues, especially for heterogeneous systems where different parts of the system can be allocated erroneous amounts of the total energy. This issue is particularly relevant for the current study, where crystal Si and liquid Si coexist in a supercell. Force is a derivative of the energy, not the energy itself. Inferring atomic energy from its derivative is obviously not as direct as using the atomic energy itself. For phenomena like phase transitions, the correct energy assignment is extremely important.

Obtaining the relevant and sufficient DFT training data, especially for problems far from equilibrium, is not trivial. In general, there are two strategies. One is to use intelligent ways to predict the approximate trajectories of the problems to be studied, and then to generate DFT data along those trajectories.[20] Another way is to generate the data in an iterative fashion. For example, to check the validity of the trained force field during its molecular dynamics (MD) simulation either by checking the variation of multiple models[21,22] or by checking the variance in a Gaussian progress regression model. Thus, either the needed data can be generated again in iterations (active learning) or can be generated on-the-fly.[23] All these are good research topics. In the current study, we have adopted a simple approach, to prepare sufficient *ab initio* data covering the physical environment by simulating sufficient numbers of temperatures for solid Si. As we will show later, this approach provides enough data to generate a good ML-FF for our problem. *Ab initio* molecular dynamics (MD) for Si supercell is carried out by using PWmat package.[24,25] We found that it is essential to correct a systematic bias in almost all the training models, otherwise critical results like transition temperature will be wrong. Our resulting ML-FF is able to reproduce all the aspects of *ab initio* simulated growth, from local energy fluctuations to transition temperatures, to diffusion constant, and growth rates. We have also compared the growth simulation with the Stillinger–Weber classical force field. Our work has demonstrated for the first time that crystal growth can be simulated using ML-FF.

RESULTS AND DISCUSSION

In the current study, we have carried out *ab initio* NPT molecular dynamics (MD) calculations for Si at 500 K, 600 K, 700 K, 800 K, 850 K, 900 K, 950 K, 1 000 K, and 2000 K with 204 atom supercell including solid, liquid, and solid–liquid coexist systems. For each MD step, we have 204 atomic energies, so the dataset is large. In total, we have about 300 000 cases (each case is defined as one atomic energy, three forces, with its corresponding local atomic configuration). A 300 000 case set for the individual atoms only corresponds to about 3000 MD atomic configurations (more details in the Methods section). This is indeed rather small. As far as we know, some of the traditional force field fitting involved hundreds of thousands of MD steps. Figure S1 shows the energy distribution of Ei for different temperatures within a range of 2 eV.

A three layer DNN model (shown in Fig. 1) is used, with the number of nodes at these layers being 252, 80, and 40 respectively. The DNN training is done with the tensor flow package using the stochastic derivative and steepest decent optimization approach. A Rc of 6.2 A is used to cast out a local atomic environment for each atom, and 36 pair features, 216 three particle features are generated. The formalism of piecewise cosine functions to construct features (see details in the supplementary material) is almost as same as in Ref. 13, except formula (6),

$$\varphi_\alpha^{(k)}(R_{ml}) = \begin{cases} \left[\cos\left(\frac{R_{ml}-R_\alpha^k}{h_\alpha^k}\pi\right)+1\right]/2, & \left|R_{ml}-R_\alpha^k\right| < h_k \\ 0, & Otherwise, \end{cases}$$

where

$$R_\alpha^k = 10^{\left[\log R_{inner} + (\alpha-1)\frac{\log R_{outer}-\log R_{inner}}{M_k-1}\right]},$$
$$h_\alpha^k = 2\left(R_\alpha^k - R_{\alpha-1}^k\right).$$

The so trained atomic energy compared with the original DFT result is shown in Fig. 2(a) (note that the same trend is observed for a validation dataset). The trained MAE of Ei is about 40 meV. It should be mentioned that what we reported is for each individual atom energy Ei. If there is no systematic error (e.g., after the bias correction as following), then if one uses the total energy error to calculate the error per atom (like most literature in this field), then the MAE will be 40 meV/sqrt(N_atom) ~ =3 meV, as discussed in our previous publication.[13] This 3 meV error per atom is on par with all the other force fields in this field,[10–12] especially considering that our data contain a larger temperature range and different phases.

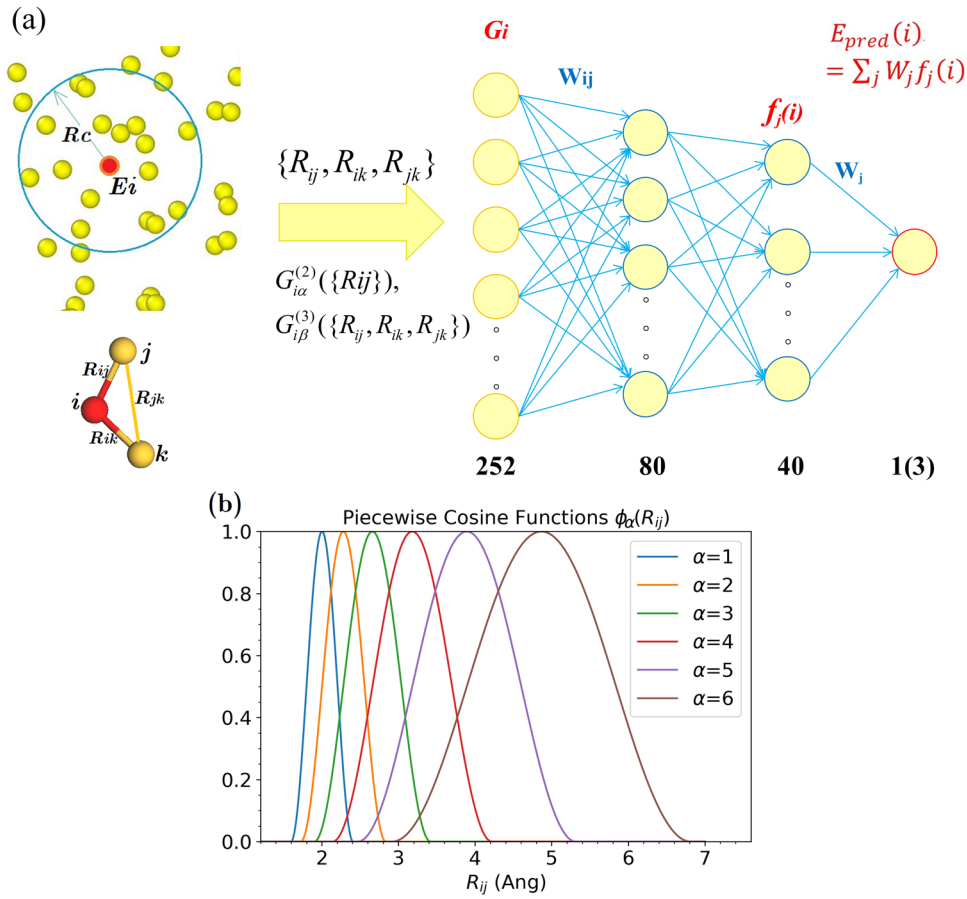

FIG. 1. (a) Schematic representation of the neural network model used in fitting Ei. (b) Piecewise cosine functions as basis functions to construct the symmetry functions.

However, the $Ei(DNN)$-$Ei(DFT)$ as shown in Fig. 2(c) reveals a systematic error. The $Ei(DNN)$ at the higher DFT energy ends tends to be too small, while the points at the lower DFT energy end tend to be too large. As far as we know, this problem has not been discussed in previous literature perhaps because no individual atomic energy E_i was used. Surprisingly, this turns out to be a common feature of most of the fitting procedures, as long as the final result will be linearly scaled by one fitting parameter (e.g., the last step is a linear fitting). The last part of the fitting procedure in the linear regression, Kernel Ridge regression, or DNN can all be expressed as a minimization of the loss function,

$$T = \sum_i [E_{DFT}(i) - \sum_j W_j f_j(i)]^2. \quad (1)$$

Here, index "i" denotes different cases (the atoms), while $f_j(i)$ is either the feature itself in the linear regression or the reference kernel values in the Kernel Ridge regression or the output node value in the last layer of a DNN model (where W_j is the weight to connect the last layer output and the final energy node). Although, in DNN, the $f_j(i)$ will be optimized by other parameters, but independently, the T will also be optimized by the parameter W_j. Thus, taking the minimization condition $\partial T/\partial W_j = 0$, we can yield the following results (see the supplementary material for details):

$$\sum_i \Delta E(i) E_{DFT}(i) = -\sum_i \Delta E(i)^2. \quad (2)$$

Here, $\Delta E(i) = E_{pred}(i) - E_{DFT}(i)$ is the error of the prediction, and $E_{pred}(i) = \sum_j W_j f_j(i)$. Since the right hand side of Eq. (2) is always negative, this means the statistical average of $\Delta E(i)$ at different $E_{DFT}(i)$ energy range cannot be zero at all energy range, and its average will be negative at larger $E_{DFT}(i)$ values, and positive at smaller $E_{DFT}(i)$ regions, as shown in Figs. 2(a) and 2(c). This effect is further amplified when the total energy of a system is obtained by the summation of the atomic energy $E_{sys} = \sum_i E_{pred}(i)$. As shown in the supplementary material, for $E_{sys}(j)$ and $\Delta E_{sys}(j)$, we will have $\sum_j \Delta E_{sys}(j) E_{DFT,sys}(j) = -N_a \sum_j \Delta E_{sys}(j)^2$, where N_a is the number of atoms. Note that if the total energy of the system is directly used in the fitting, like in most of the other methods, then Eq. (2) will hold with ΔE being the total energy ΔE_{sys}; thus, the N_a will not be there. Thus, using atomic energy fitting exacerbates this bias problem for the total energy of a system. This is demonstrated in Fig. 3(a), where a 2 eV bias exists for the total energy of a 204 atom system. In this fitting, the MAE for each atomic energy is 40 meV. Note $\Delta E_{sys}{}^2$ equals $\sum_i \Delta E(i)^2 = 204 \times 0.04^2$ eV^2, thus the right hand side of the above

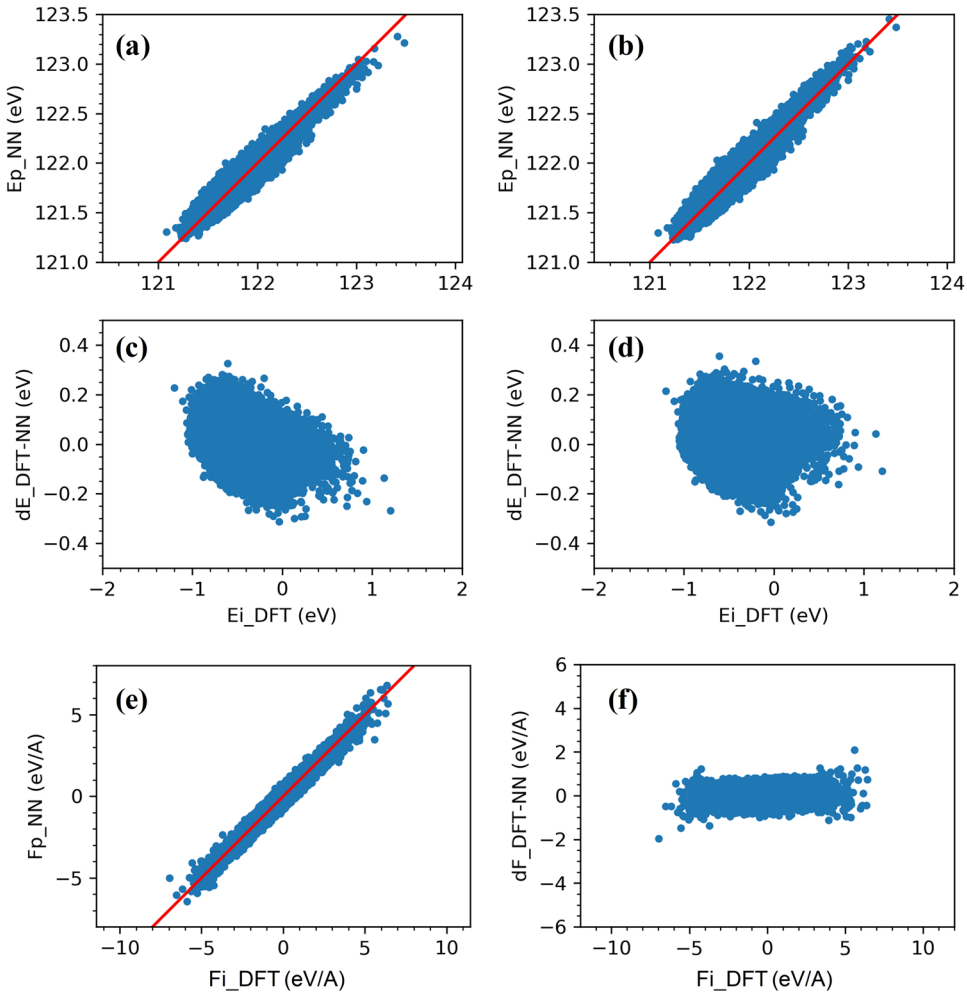

FIG. 2. The trained (without bias) atomic energies (a) compared with the original DFT result and (c) Ei(DNN)-Ei(DFT). The trained (with bias) atomic energies (b) compared with the original DFT result and (d) Ei(DNN)-Ei(DFT). The DNN predicted forces (e) and their errors (f).

equation is $-204^2 \times 0.04^2 = -66$ eV^2 (the extra 204 comes from the N_a factor shown above). According to Fig. 3(a), when $E_{DFT,sys}(j)$ changes 40 eV, we have a bias of about −2 eV; so, the left hand side of the above equation equals to $-2 \times 40 = -80$ eV^2, roughly equal to the right hand side, confirming our formula. We thus conclude that this bias problem is a bigger problem in atomic energy fitting, than the conventional total energy fitting, which has to be corrected here.

To overcome such systematic bias, it is necessary to introduce a parameter α in the loss function

$$T = \sum_i [(1+\alpha)E_{DFT}(i) - \sum_j W_j f_j(i)]^2. \quad (3)$$

One can first set an α (e.g., 0 at the beginning) to carry out a fitting, and then from the fitted result, recalculate $\alpha = -\sum_i E(i)E_{DFT}(i)/\sum_i |E_{DFT}(i)|^2$; then, add this $\Delta\alpha$ on top of the original α, and repeat this until $\Delta\alpha = 0$. Usually one or two iteration will be enough to yield $\sum_i \Delta E(i)E_{DFT}(i) = 0$. The bias corrected results are shown in Figs. 2(b) and 2(d), where $\alpha = 0.1$ is obtained. Note that, since the average atomic force is zero, there is no such systematic error in force. The predicted forces and their errors are shown in Figs. 2(e) and 2(f), and the MAE is about 0.1 eV/A.

It should be noted that this bias correction is not the same issue of the regularization in the linear fitting.[26] The regularization in linear fitting is to avoid over fitting when the variables $f_j(i)$ in Eq. (1) are linearly dependent, or the number of independent sample size i is too small. In our case, due to the large number of samples used in Eq. (1), we do not have an over fitting problem. What we have is a systematic bias problem.

This systematic error for $Epot = \sum_{in\ cell} E(i)$ of unit cell is shown more dramatically in Figs. 3(a) and 3(b), where the low temperature DNN energy is too large. This induces a systematic bias for different phases in Fig. 3(a). Since the higher energy points are at the liquid phase, an under estimation of their energies will lead to a lower transition temperature Tc. The bias corrected results in Fig. 3(b) can follow the DFT Epot values closely. We then proceeded by simulating the crystal phase at 950 K and the liquid phase at 1500 K using both the SANNP and the DFT method for a 256 atom supercell. Their resulting pair distribution functions are shown in

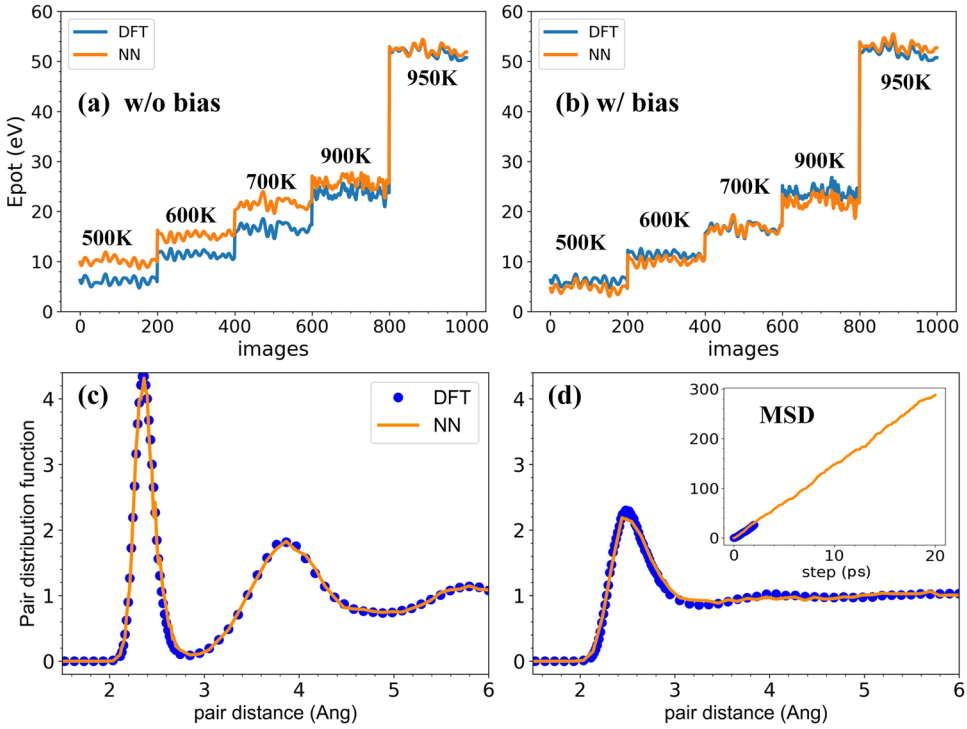

FIG. 3. Comparison of NN and DFT total energies, (a) without bias and (b) with bias. The x-axis corresponds to the images from different temperature MD simulations of crystal growth. Pair distribution functions of (c) crystal phase at 950 K and (d) liquid phase at 1500 K; the inset of (b) is MSD of the liquid phase at 1500 K.

Figs. 3(c) and 3(d), respectively. One can see that there is practically no difference between the SANNP result and DFT results. In the inset of Fig. 3(d), we also show the Si diffusion mean square displacement as a function of time. A linear relationship is found, and the curves for SANNP and DFT are also very close, indicating the same diffusion constants.

For the purpose of studying crystal growth, it is crucial to have accurate modeling of the atomic energies associated with liquid/solid interfaces. Since it is difficult to pinpoint individual energy barrier in such a complex situation, the energy profile changing with time can be used as a good test for different models. We have constructed a 204 atom supercell, with a crystal Si slab exposing two (111) surfaces. Each (111) surface has a step, facilitating a step-flow growth kinetics.[15] We have chosen three Si atoms [19th, 114th, and 137th atom in Fig. 4(a)], which represent solid, liquid, and transition from liquid to solid situations. We first carry out a DFT MD simulation. We then calculate the atomic energies surrounding this atom defined as $E_{loc}(t) = \sum_i E_i(t) \exp\left(-|R_i(t)-R_{i0}(t)|^2/a^2\right) / \sum_i \exp\left(-|R_i(t)-R_{i0}(t)|^2/a^2\right)$, where $E_i(t)$ is the atomic energy either calculated from DFT or from the model prediction, and i_0 is the center atom index (the 19th, 114th, and 137th atoms). We have chosen a = 4.2 A to include the first nearest neighbor in the above summation. The DFT atomic trajectory $R(t)$ is used to evaluate the local energy for different models. For clarity, a 200 fs width Gaussian convolution is first applied to $E_{loc}(t)$ to yield a smoother local energy curve $E'_{loc}(t)$, which is shown in Figs. 4(b) and 4(c) (the original energy curves without Gaussian broadening are shown in Fig. S2). In both Fig. 4(b) and Fig. S2, the SANNP result follows the DFT values closely, with almost no difference between each other. On the other hand, the results of SW potential show significant difference from the DFT results. Such a different energy profile could lead to different microscopic behavior, barrier height, and growth mechanism.

To check the validity of the ML-FF, another more challenging way is to carry out a ML-FF MD, and then check the DFT result on the ML-FF trajectory. This is mostly to check whether the ML-FF MD remains in the region of its original fitting validity. The test result along a segment of the ML_FF NPT MD trajectory with 2000 steps is shown in Fig. S3. We can see that the DFT agrees with the ML-FF results well, indicating that the ML-FF MD is good.

To evaluate the growth rate, we first need to calculate the solid to liquid transition temperature Tc. While the experimental transition temperature is at 1600 K, it is well known that the local density approximation (LDA) and generalized gradient approximation (GGA) predict much lower Tc.[27–29] The Tc can be derived from Gibbs free energy calculation,[27,29] or can be simulated directly from a solid/liquid coexistent equilibrium system using an NVE ensemble.[30] We have first used a 204 atom supercell to carry out such NVE calculations. The resulting temperature variation with time is shown in Figs. 5(a) and 5(b) for DFT and SANNP, respectively. The equilibrium system have both crystal Si and liquid Si and the pressure is close to zero. Both DFT and SANNP yield a Tc around 1090 K ± 20 K. It is surprising that this Tc value is lower than the previously calculated Si crystal/liquid transition temperature using LDA.[27] This value is about 1350 K, estimated based on crystal and liquid free energy calculations.[27] We note that the solid/liquid coexistence calculation is a simpler procedure compared to free energy calculations (e.g., using thermodynamic integration).[27] One

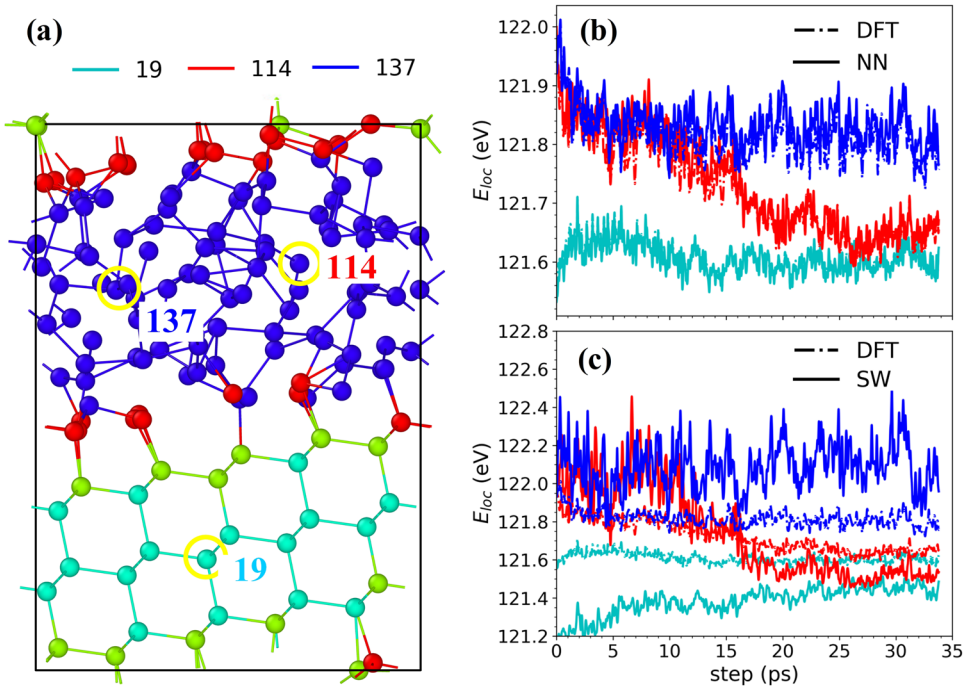

FIG. 4. (a) Supercell structure of a crystal Si slab with (111) surfaces, local energy Eloc(t) of DFT vs (b) NN and (c) SW potential.

possible cause is the *k*-points used. In our simulation, we have used Gamma point for both liquid and solid phases. This might lead to relatively under estimation of the liquid phase energy, which leads to smaller transition temperature. Future investigations are needed to clarify this point. However, here, we are more interested in knowing whether the ML-FF can reproduce the DFT results, regardless of the *k*-points used. So, the possible small error caused by the use of Gamma point should not affect our evaluation for the quality of the ML-FF model. We have used the same procedure for DFT and SANNP methods, which yields the same result, and the so obtained transition temperature can be readily used in crystal growth simulations since the same structure setup is used.

Furthermore, we found that the Tc sensitively depends on the super cell size (see the supplementary material for details) and k-point sampling. When the calculation is repeated with a 1088 atom cell, the Tc further drops to about 930 K ± 10 K [shown in Fig. 5(c)]. This indicates that the small cell size has imposed a periodic constraint on the liquid phase (before the long range Si–Si order has fully subsided), which has increased the liquid phase energy (hence higher Tc).

We note that there are many previous theoretical work for Si solid–liquid transition temperature Tc.[27–29,31,32] For example, recently, Jinnouchi *et al.*[31] have used on-the-fly machine learning force field to carry out a long DFT accuracy MD to calculate the

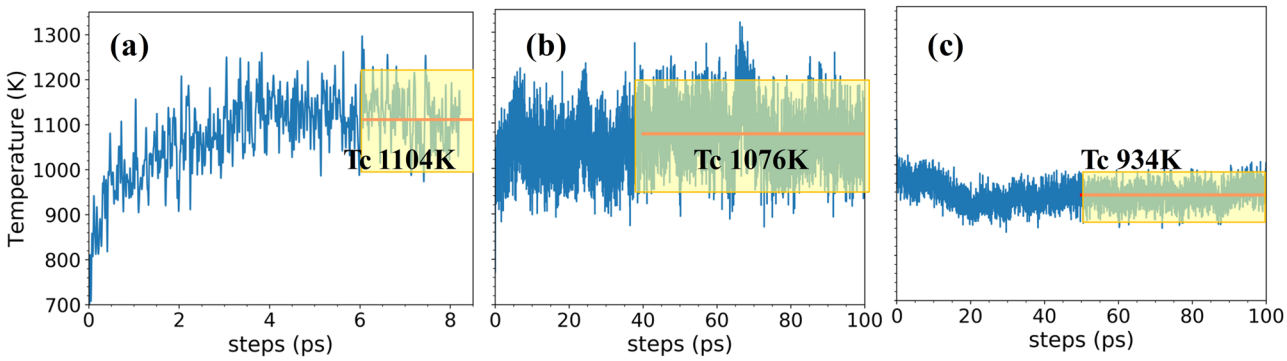

FIG. 5. Temperature variation with time for (a) DFT, (b) DNN simulations of 204 atoms/cell, and (c) DNN simulations of 1088 atoms/cell.

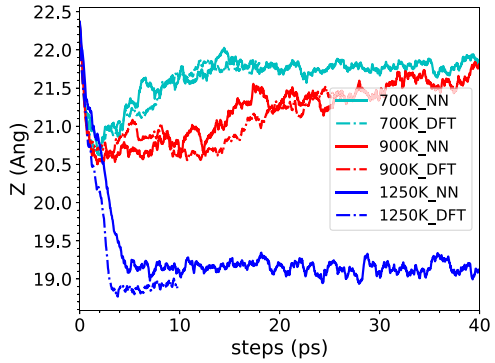

FIG. 6. The growth of Z length with time using both the DFT and SANPP models.

free energy difference between the solid and liquid phase through a thermodynamic perturbation theory integration scheme. Their calculated Tc under LDA XC functional is 1207 K, about 100 K higher than our result of the 208 atom supercell. Their result is based on a 128 atom supercell. Thus, such a Tc difference could be caused by the supercell size difference besides the difference in k-point sampling. In the literature, different groups also reported different Tc with a scatter of about 100 K.[27–29,31,32] Overall, we believe that the use of a larger supercell with ML-FF like the SANNP is important.

It is worth pointing out that correcting such bias is important, otherwise, one cannot reproduce the same DFT phase transition temperature, for the systematic error of Ei. As a simple mathematical proof provided above, the bias correction is simply a fitting issue between the DFT data and the fitted data point, so this has nothing to do with the DFT calculation method and accuracy. A good fitting procedure should regenerate the DFT data points without bias. We believe this is an important issue the community needs to know, because the previous studies do not fit heavily to the energy (since only total energy is used), and previously not many phase transitions are studied using ML-FF. Such systematic bias more apparently appears in such phase transition studies.

To compare the growth rate at different temperatures between DFT and SANPP, we have used the 204 atom cell, carried out MD simulation at constant pressure and temperature (NPT) with T being 700 K, 900 K and 1250 K, and close to zero pressure. Figure 6 shows the growth with time using both the DFT and SANPP models. Here, we have used the cell z-direction length Z in the NPT simulation as an indicator of the amount of crystal Si. For Si, the liquid phase has a smaller volume; so, an increase of z-direction length means more crystal Si. We see that the crystal growth curves (700 K and 900 K) for DFT and SANNP are similar, as well as the melting curve (1250 K).

Finally, we have simulated the growth of a Si crystal using the large 1088 atom supercell under the SANNP model. Two (111) surfaces are exposed, and both have a step on the surface. Fig. S4 shows its structures during growth. The growth curves for different $\Delta T = T - Tc$ are shown in Fig. 7(a). We have used the extended common neighbor analysis method[33] in Ovito software package[34] to classify each Si atom as belonging to liquid or crystal. Simulations up to 100 ps are executed. We see that the crystal begins to grow when ΔT is below −50 K. The same supercell setup is simulated with SW force field, and the results are shown in Fig. 7(b). Although our current simulation does not have sufficient statistics to provide accurate growth rate vs ΔT dependence, we can see that in both SANNP and SW models, the growth rate is roughly proportional to ΔT when $|\Delta T|$ is smaller than 200 K. We do note that the trends for NN and SW are different. At -50 K, the crystal growth in NN is much (about 4 times) faster than that of SW. While in SW, the growth rate increases from −50 K to −300 K, while in NN, the maximum growth rate happens around −150 K, and then the growth decreases from −150 K to −300 K, possibly due to kinetic hindrance. This indicates that the balance between kinetics (the diffusion and reaction barrier) and thermodynamics (the Gibbs free energy drive to form crystals) for NN and SW is different. This also shows why it is useful to have a more accurate NN potential, especially to represent the kinetic process more accurately. To give an order of magnitude, the step growth coefficient defined as $\beta = G/|\Delta T|$ (where G is the growth rate) is 0.16 m/(sK) for the SANNP result, and 0.09 m/(sK) for the SW result. The SW result is slightly smaller than the values reported in Ref. 35. However, as shown in Ref. 35, this rate sensitively depends on the supercell size, which might explain the difference. Interestingly, our SANNP growth rate coefficient is about 1/2 larger than the result of SW. Also interestingly, when ΔT is smaller than −200 K, the growth rate seems to decrease again. This is probably a consequence of the competition between the free energy driving

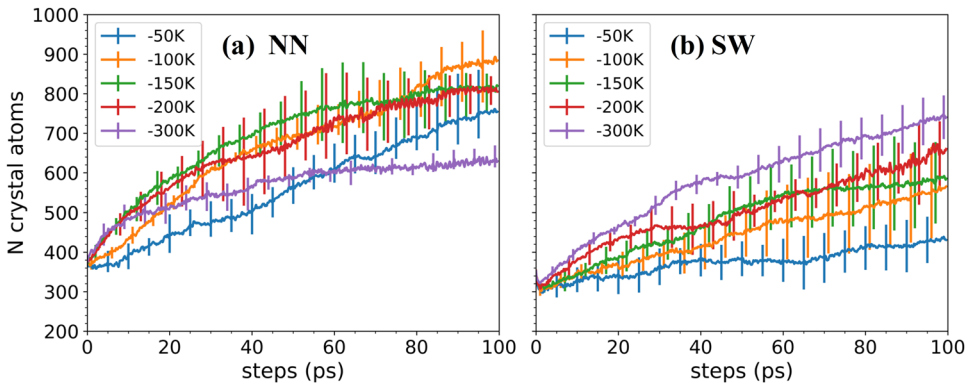

FIG. 7. Growth curve (with error bars) of number of crystal atoms for different ΔT, (a) NN, and (b) SW potential.

force for crystal growth and the Si mobility to overcome the kinetic barriers.

It is worth noting that, in the above, we have compared our SANNP simulation with SW simulation, rather than other machine learning force fields (ML_FF). There are many ML_FFs for Si.[11,12,36,37] However, different ML_FFs have been developed for different purposes, with their own selected training datasets for those special purposes. We thus believe it is difficult to compare them directly here. Indeed, one use of ML_FF is to develop it on-the-flight for a given task at hand, instead of a general purpose force field for all situations.

In summary, we have shown that SANNP can be used to simulate the melt Si growth with accuracy and behavior similar to those of direct DFT simulations. We found that the SANNP can reproduce DFT pair distribution function, Si diffusion constant, local atomic energy fluctuation, the transition temperature, and the growth rates. SANNP results also show significant difference from the SW results. We have also proposed a way to correct the systematic bias error in all the ML-FF trainings. Such a correction is important in order to reproduce the DFT results when there is a phase transition.

METHODS

DFT calculations to get Ei/Fi dataset

All DFT calculations are performed by using the PWmat package[20,21] with SG15 non-conserving pseudo-potential.[38] The exchange-correlation energy is described by local density approximation (LDA).[39,40] The kinetic energy cutoff of electron plane wave functions is 50 Ryd. The structural relaxations are performed by using the conjugated gradient method, and the convergence threshold is set to be 10^{-4} eV in energy and 0.05 eV $Å^{-1}$ in force. The Brillouin zone is sampled by only Gamma point for both liquid and solid phases in the AIMD calculations, since supercells including 204 Si atoms are constructed with relative lager (>10 Ang) cell constants.

The *ab initio* NPT molecular dynamics (MD) calculations are carried out for Si at 500 K, 600 K, 700 K, 800 K, 850 K, 900 K, 950 K, 1000 K, and 2000 K with 204 atoms supercell including solid, liquid, and solid–liquid coexist systems. For the system we calculated, the atomic energy decomposition step to get *Ei* takes about 1/3 of the computational time. These 300 000 DFT *Ei* and *Fi* cases are obtained from above 3000 NPT molecular dynamics snapshots (each snapshot is taken every 10 or 20 MD steps and also called a frame). These have 1000 frames from 800 K to 2000 K including all 204 atoms in the cell for Ei and Fi training, and 2000 frames from 500 K, 600 K, 700 K, 850 K, 900 K, 950 K, and 1000 K including only about 50 interfacial atoms to improve *Ei* training performance.

Training process of ANNs

We first trained the DNN model with the atomic energy E_i only. After the energy loss function is stabilized, the atomic forces are then included in the training (the force training is 20 times more expensive). A weight ratio of 0.9/0.1 is used on the atomic energy vs atom force loss function (when energy is in unit of eV and force is in unit of eV/Ang).

SUPPLEMENTARY MATERIAL

See the supplementary material for some detail math processing and data about atomic energy Ei.

PES.zip: program used to generate the ML potential and also its forces for use in MD calculations.

AUTHORS' CONTRIBUTIONS

L.-W.W. proposed the concept and L.M. carried out the calculations and analysis. All authors contributed to the discussions and writing of the manuscript.

ACKNOWLEDGMENTS

This work was supported by the Director, Office of Science (SC), Basic Energy Science (BES), Materials Science and Engineering Division (MSED), of the U.S. Department of Energy (DOE) under Contract No. DE-AC02-05CH11231 through the Materials Theory program (No. KC2301).

The authors declare no competing financial or non-financial interests.

DATA AVAILABILITY

The data that support the findings of this study are openly available in https://gitlab.com/LingMiao/SANNP_ml.

REFERENCES

[1] T. Morawietz, A. Singraber, C. Dellago, and J. Behler, "How van der Waals interactions determine the unique properties of water," Proc. Natl. Acad. Sci. U. S. A. **113**, 8368–8373 (2016).

[2] N. Artrith and A. Urban, "An implementation of artificial neural-network potentials for atomistic materials simulations: Performance for TiO_2," Comput. Mat. Sci. **114**, 135–150 (2016).

[3] N. Artrith, T. Morawietz, and J. Behler, "High-dimensional neural-network potentials for multicomponent systems: Applications to zinc oxide," Phys. Rev. B **83**, 153101 (2011).

[4] W. Li, Y. Ando, E. Minamitani, and S. Watanabe, "Study of Li atom diffusion in amorphous Li_3PO_4 with neural network potential," J. Chem. Phys. **147**, 214106 (2017).

[5] F. C. Mocanu, K. Konstantinou, T. H. Lee, N. Bernstein, V. L. Deringer, G. Csányi, and S. R. Elliott, "Modeling the phase-change memory material, $Ge_2Sb_2Te_5$, with a machine-learned interatomic potential," J. Phys. Chem. B **122**, 8998–9006 (2018).

[6] K. Gubaev, E. V. Podryabinkin, G. L. W. Hart, and A. V. Shapeev, "Accelerating high throughput searches for new alloys with active learning of interatomic potentials," Comput. Mater. Sci. **156**, 148–156 (2019).

[7] M. Eckhoff and J. Behler, "From molecular fragments to the bulk: Development of a neural network potential for MOF-5," J. Chem. Theor. Comput. **15**, 3793 (2019).

[8] F. H. Stillinger and T. A. Weber, "Computer simulation of local order in condensed phases of silicon," Phys. Rev. B **31**(8), 5262 (1985).

[9] T. Motooka, K. Nisihira, S. Munetoh *et al.*, "Molecular-dynamics simulations of solid-phase epitaxy of Si: Growth mechanisms," Phys. Rev. B **61**(12), 8537 (2000).

[10] J. Behler and M. Parrinello, "Generalized neural-network representation of high-dimensional potential-energy surfaces," Phys. Rev. Lett. **98**(14), 146401 (2007).

[11] J. Behler, R. Martoňák, D. Donadio, and M. Parrinello, "Metadynamics simulations of the high-pressure phases of silicon employing a high-dimensional neural network potential," Phys. Rev. Lett. **100**(18), 185501 (2008).

[12]A. P. Bartók, J. Kermode, N. Bernstein, and G. Csanyl, "Machine learning a general-purpose interatomic potential for silicon," Phys. Rev. X **8**, 041048 (2018).
[13]Y. Huang, J. Kang, W. A. Goddard III, and L. W. Wang, "Density functional theory based neural network force fields from energy decompositions," Phys. Rev. B **99**(6), 064103 (2019).
[14]J. Behler, "Atom-centered symmetry functions for constructing high-dimensional neural network potentials," J. Chem. Phys. **134**, 074106 (2011).
[15]A. P. Bartók, R. Kondor, and G. Csányi, "On representing chemical environments," Phys. Rev. B **87**, 184115 (2013).
[16]A. P. Thompson, L. P. Swiler, C. R. Trott, S. M. Foiles, and G. J. Tucker, "Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials," J. Comput. Phys. **285**, 316–330 (2015).
[17]A. V. Shapeev, "Moment tensor potentials: A class of systematically improvable interatomic potentials," Multiscale Model. Simul. **14**(3), 1153–1173 (2016).
[18]J. Kang and L. W. Wang, "First-principles Green-Kubo method for thermal conductivity calculation," Phys. Rev. B **96**, 020302(R) (2017).
[19]D. Yoo, K. Lee, W. Jeong, D. Lee, S. Watanabe, and S. Han, "Atomic energy mapping of neural network potential," Phys. Rev. Mater. **3**, 093802 (2019).
[20]L. Bonati and M. Parrinello, "Silicon liquid structure and crystal nucleation from *ab initio* deep metadynamics," Phys. Rev. Lett. **121**, 265701 (2018).
[21]J. Behler, "Constructing high-dimensional neural network potentials: A tutorial review," Int. J. Quantum Chem. **115**, 1032–1050 (2015).
[22]L. Zhang, D. Y. Lin, H. Wang, R. Car, and W. E, "Active learning of uniformly accurate interatomic potentials for materials simulation," Phys. Rev. Mater. **3**, 023804 (2019).
[23]R. Jinnouchi, J. Lahnsteiner, F. Karsai, G. Kresse, and M. Bokdam, "Phase transitions of hybrid perovskites simulated by machine-learning force fields trained on the fly with Bayesian inference," Phys. Rev. Lett. **122**, 225701 (2019).
[24]W. Jia, Z. Cao, L. Wang, J. Fu, X. Chi, W. Gao, and L.-W. Wang, "The analysis of a plane wave pseudopotential density functional theory code on a GPU machine," Comput. Phys. Commun. **184**(1), 9–18 (2013).
[25]W. Jia, J. Fu, Z. Cao *et al.*, "Fast plane wave density functional theory molecular dynamics calculations on multi-GPU machines," J. Comput. Phys. **251**, 102–115 (2013).
[26]C. van der Oord, G. Dusson, G. Csányi, and C. Ortner, "Regularised atomic body-ordered permutation-invariant polynomials for the construction of interatomic potentials," Mach. Learn.: Sci. Technol. **1**, 015004 (2020).
[27]O. Sugino and R. Car, "*Ab-initio* molecular dynamics study of first-order phase transitions: Melting of silicon," Phys. Rev. Lett. **74**, 1823 (1995).
[28]D. Alfe and M. J. Gillan, "Exchange-correlation energy and the phase diagram of Si," Phys. Rev. B **68**, 205212 (2003).
[29]F. Dorner, Z. Sukrma, C. Dellago, and G. Kresse, "Melting Si: Beyond density functional theory," Phys. Rev. Lett. **121**, 195701 (2018).
[30]A. Siavosh-Haghighi and D. L. Thompson, "Melting point determination from solid-liquid coexistence initiated by surface melting," J. Phys. Chem. C **111**, 7980 (2007).
[31]R. Jinnouchi, F. Karsai, and G. Kresse, "On-the-fly machine learning force field generation: Application to melting points," Phys. Rev. B **100**, 014105 (2019).
[32]U. R. Pedersen, F. Hummel, G. Kresse, G. Kahl, and C. Dellago, "Computing Gibbs free energy differences by interface pinning," Phys. Rev. B **88**, 094101 (2013).
[33]E. Maras, O. Trushin, A. Stukowski *et al.*, "Global transition path search for dislocation formation in Ge on Si(001)," Comput. Phys. Commun. **205**, 13–21 (2016).
[34]A. Stukowski, "Visualization and analysis of atomistic simulation data with OVITO—The Open Visualization Tool," Modell. Simul. Mater. Sci. Eng. **18**, 015012 (2010).
[35]D. Buta, M. Asta, and J. J. Hoyt, "Kinetic coefficient of steps at the Si(111) crystal-melt interface from molecular dynamics simulations," J. Chem. Phys. **127**, 074703 (2007).
[36]E. Sanville, A. Bholoa, R. Smith, and S. D. Kenny, "Silicon potentials investigated using density functional theory fitted neural networks," J. Phys.: Condens. Matter **20**, 285219 (2008).
[37]V. L. Deringer *et al.*, "Realistic atomistic structure of amorphous silicon from machine-learning-driven molecular dynamics," J. Phys. Chem. Lett. **9**, 2879–2885 (2018).
[38]M. Schlipf and F. Gygi, "Optimization algorithm for the generation of ONCV pseudopotentials," Comput. Phys. Commun. **196**, 36 (2015).
[39]D. M. Ceperley and B. J. Alder, "Ground state of the electron gas by a stochastic method," Phys. Rev. Lett. **45**(7), 566 (1980).
[40]J. P. Perdew and A. Zunger, "Self-interaction correction to density-functional approximations for many-electron systems," Phys. Rev. B **23**(10), 5048 (1981).