

# Potfit: effective potentials from ab-initio data

Peter Brommer and Franz Gähler

Institut für Theoretische und Angewandte Physik (ITAP), Universität Stuttgart,  
Pfaffenwaldring 57, 70550 Stuttgart, Germany

E-mail: [p.brommer@itap.physik.uni-stuttgart.de](mailto:p.brommer@itap.physik.uni-stuttgart.de)

**Abstract.** We present a program called *potfit* which generates an effective atomic interaction potential by matching it to a set of reference data computed in first-principles calculations. It thus allows to perform large-scale atomistic simulations of materials with physically justified potentials. We describe the fundamental principles behind the program, emphasizing its flexibility in adapting to different systems and potential models, while also discussing its limitations. The program has been used successfully in creating effective potentials for a number of complex intermetallic alloys, notably quasicrystals.

*Modelling Simulation Mater. Sci. Eng.* **15** (2007), 295–304  
online at <http://stacks.iop.org/ms/15/295>  
doi: [10.1088/0965-0393/15/3/008](https://doi.org/10.1088/0965-0393/15/3/008)

PACS numbers: 02.60.Pn, 02.70.Ns, 07.05.Tp, 61.44.Br

## 1. Introduction

Classical effective potentials reduce the quantum-mechanical interactions of electrons and nuclei in a solid to an effective interaction between atom cores. This greatly reduces the computational effort in molecular dynamics (MD) simulations. Whereas first principles simulations are limited to a few hundred atoms at most, classical MD calculations with many millions of atoms are routinely performed. Such system sizes are possible, because molecular dynamics with short-range interactions scales linearly with the number of atoms. Moreover, it can easily be parallelized using a geometrical domain decomposition scheme [1, 2], thereby achieving linear scaling also in the number of CPUs.

The study of many problems in materials science and nanotechnology indeed requires simulations of systems with millions of atoms. Quite generally, this is the case whenever long-range mechanical stresses are involved. Examples of such problems are the study of fracture propagation [3], nano-indentation, or the motion and pinning of dislocations. Other problems may be simulated with more moderate numbers of atoms, but require very long simulated times, of the order of nanoseconds, an example of which is the study of atomic diffusion [4]. In either case, if large systems and/or long time scales are required, classical effective potentials are the only way to make molecular dynamics simulations possible.

The reliability and predictive power of classical MD simulations depend crucially on the quality of the effective potentials employed. In the case of elementary solids, such potentials are usually obtained by adjusting a few potential parameters to

optimally reproduce a set of reference data, which typically includes a number of experimental values like lattice constants, cohesive energies, or elastic constants, sometimes supplemented with ab-initio cohesive energies and stresses [5, 6]. In the case of more complex systems with a large variety of local environments and many potential parameters to be determined, such an approach cannot help, however; there is simply not enough reference data available.

The force matching method [7] provides a way to construct physically justified potentials even under such circumstances. The idea is to compute forces and energies from first principles for a suitable selection of small reference systems and to adjust the parameters of the potential to optimally reproduce them.

For that purpose, we developed a program called *potfit*<sup>‡</sup>. By separating the process of optimization from the form of the potential, *potfit* allows for maximal flexibility in the choice of potential model and parametrization.

The underlying algorithms are described in section 2. Section 3 focuses on the implementation of the algorithms, followed by details on employing *potfit* in section 4. We discuss advantages and limitations of the force matching method and our implementation in section 5, and present our conclusions in the final section 6.

## 2. Algorithms

As mentioned above, *potfit* consists of two separate parts. The first one implements a particular parametrized potential model and calculates from a set of potential parameters  $\xi_i$  the target function that quantifies the deviations of the forces, stresses and energies from the reference values. Wrapped around is a second, potential independent part which implements a least squares minimization module. As this part is completely independent of the potential model and just deals with the list of parameters  $\xi_i$ , it is fairly straightforward to change the parametrization of the potential (tabulated or analytic), or even to switch to a different potential model.

### 2.1. Optimization

From a mathematical point of view, force matching is a basic optimization problem: There is a set of parameters  $\xi_i$ , a set of values  $b_k(\xi_i)$  depending on them, and a set of reference values  $b_{0,k}$  which the  $b_k$  have to match. This leads to the well-known method of least squares, where one tries to minimize the sum of squares of the deviations between the  $b_k$  and the  $b_{0,k}$ . In our case, the reference values can either be the components of the force vector  $\vec{f}_{0,j}$  acting on each individual atom  $j$ , or global data  $A_{0,k}$  like stresses, energies, or certain external constraints. We found it helpful to measure the relative rather than the absolute deviations from the reference data, except for very small reference values. The least squares target function thus becomes

$$Z = Z_F + Z_C, \quad (1)$$

$$\text{with } Z_F = \sum_{j=1}^{N_A} \sum_{\alpha=x,y,z} W_j \frac{(f_{j\alpha} - f_{0,j\alpha})^2}{\vec{f}_{0,j}^2 + \varepsilon_j}, \quad (2)$$

$$\text{and } Z_C = \sum_{k=1}^{N_c} W_k \frac{(A_k - A_{0,k})^2}{A_{0,k}^2 + \varepsilon_k}, \quad (3)$$

<sup>‡</sup> <http://www.itap.physik.uni-stuttgart.de/%7Eimd/potfit>

where  $Z_F$  represents the contributions of the forces, and  $Z_C$  that of the global data. The (small and positive)  $\varepsilon_\ell$  impose a lower bound on the denominators, thereby avoiding a too accurate fitting of small quantities which are actually not known to such a precision. The  $W_\ell$  are the weights of the different terms. It proves useful for the fitting to give the total stresses and the cohesion energies an increased weight, although in principle they should be reproduced correctly already from the forces. Even if all forces are matched with a small deviation only, those deviations can add up in an unfortunate way when determining stresses, thus leading to potentials giving wrong elastic constants. Including global quantities in the fit with a sufficiently high weight suppresses such undesired behaviour of the fitting process.

As the evaluation of the highly nonlinear target function (1) is computationally rather expensive, a careful choice of the minimization method has to be made. We chose a combination of a conjugate-gradient-like deterministic algorithm [8] and a stochastic simulated annealing algorithm [9].

For the deterministic algorithm we take the one described by Powell [8], which takes advantage of the form of the target function (which is a sum of squares). By re-using data obtained in previous function calls it arrives at the minimum faster than standard least squares algorithms. It also does not require any knowledge of the gradient of the target function. The algorithm first determines the gradient matrix at the starting point in the high-dimensional parameter space by finite differences. The gradient matrix is assumed to be slowly varying around the starting point. A new optimal search direction towards the minimum is determined by the method of conjugate gradients. Then, the target function is minimized along this direction. This operation is called line minimization. When the minimum is found, the direction unit vector replaces one of the basis vectors spanning the parameter space. The gradient matrix is updated only with respect to this new direction, using the finite differences calculated in the line minimization. In this way, no finite differences have to be calculated explicitly except in the very first step. The line minimization is performed by Brent's algorithm [10] in an implementation taken from the GNU Scientific Library [11].

The algorithm is restarted (including a calculation of the full gradient matrix) when either a step has been too large to maintain the assumption of a constant gradient matrix, the basis vectors spanning the parameter space become almost linearly dependent, or the linear equation involved in Powell's algorithm cannot be solved with satisfactory numerical precision.

The other minimization method implemented is a simulated annealing [12] algorithm proposed by Corana [9]. While the deterministic algorithm mentioned above will always find the closest *local* minimum, simulated annealing samples a larger part of the parameter space and thus has a chance to end up in a better minimum. The price to pay is a computational burden which can be several orders of magnitude larger.

For the basic Monte Carlo move, we chose adding Gaussian-shaped bumps to the potential functions. The bump heights are normally distributed around zero, with a standard deviation adjusted so that on average half of the Monte Carlo steps are accepted. This assures optimal progress: Neither are too many calculations wasted because the changes are too large to be accepted, nor are the steps too small to make rapid progress.

## 2.2. Potential models and parametrizations

The simplest effective potential is a pair potential, which only depends on interatomic distances. It takes the form

$$V = \sum_{i,j < i}^N \phi_{s_i s_j}(r_{ij}), \quad (4)$$

where  $r_{ij}$  is the distance between atoms  $i$  and  $j$ , and  $\phi_{s_i s_j}$  is a potential function depending on the two atom types  $s_i$  and  $s_j$ . This function can either be given in analytic form, using a small number of free parameters, like for a Lennard-Jones potential, or in tabulated form together with an interpolation scheme for distances between the tabulation points. Whereas the parameters of an analytic potential can often be given a physical meaning, such an interpretation is usually not possible for tabulated potentials. On the other hand, an inappropriate form of an analytic potential may severely constrain the optimization, leading to a poor fit. For this reason, we chose the functions  $\phi$  to be defined by tabulated values and spline interpolation, thus avoiding any bias introduced by an analytic potential. This choice results in a relatively high number of potential parameters, compared to an analytic description of the potentials. This is not too big a problem, however. Force matching provides enough reference data to fit even a large number of parameters. The potential functions  $\phi$  only need to be defined at pair distances  $r$  between a minimal distance  $r_{\min}$  and a cutoff radius  $r_{\text{cut}}$ , where the function should go to zero smoothly.

We found pair potentials to be insufficient for the simulation of complex metallic alloys. More suited are EAM (Embedded Atom Method [13, 5]) potentials, also known as glue potentials [14], which have many advantages over pair potentials in the description of metals [14].

EAM potentials include a many-body term depending on a local density  $n_i$ :

$$V = \sum_{i,j < i} \phi_{s_i s_j}(r_{ij}) + \sum_i U_{s_i}(n_i) \quad \text{with} \quad n_i = \sum_{j \neq i} \rho_{s_j}(r_{ij}). \quad (5)$$

$n_i$  is a sum of contributions from the neighbours through a transfer function  $\rho_{s_j}$ , and  $U_{s_i}$  is the embedding function that yields the energy associated with placing atom  $i$  at a density  $n_i$ . Again, all functions are specified by their values at a number of sampling points.

The parameters  $\xi_i$  specifying a tabulated potential are naturally the values at the sampling points. Due to the nature of spline interpolation, either the gradient or the curvature at the exterior sampling points of each function can also be chosen freely. Depending on the type of potential one can keep the gradients fixed, or adapt them dynamically by adding them to the set of parameters  $\xi_i$ .

The EAM potential described by (5) has two gauge degrees of freedom, i.e., two sets of parameter changes which do not alter the physics of the potential:

$$\begin{aligned} \rho_s(r) &\rightarrow \kappa \rho_s(r), \\ U_{s_i}(n_i) &\rightarrow U_{s_i}\left(\frac{n_i}{\kappa}\right), \end{aligned} \quad (6)$$

and

$$\begin{aligned} \phi_{s_i s_j}(r) &\rightarrow \phi_{s_i s_j}(r) + \lambda_{s_i} \rho_{s_j}(r) + \lambda_{s_j} \rho_{s_i}(r), \\ U_{s_i}(n_i) &\rightarrow U_{s_i}(n_i) - \lambda_{s_i} n_i. \end{aligned} \quad (7)$$

According to (6), the units of the density  $n_i$  can be chosen arbitrarily. We use this degree of freedom to set the units such that the densities  $n_i$  computed for the reference

configurations are contained in the interval  $(-1; 1]$ , but not in any significantly smaller interval. The transformation (7) states that certain energy contributions can be moved freely between the pair and the embedding term. An embedding function  $U$  which is linear in the density  $n$  can be gauged away completely. This also makes any separate interpretation of the pair potential part and the embedding term void; the two must only be judged together. The latter degeneracy is usually lifted by choosing the gradients of the  $U_i(n_i)$  to vanish at the average density for each atom type. *potfit* also uses this convention when exporting potentials for plotting and MD simulation. As the average density might change during minimization, *potfit* internally uses a slightly different gauge: It requires that the gradient vanishes at the center of the domain of the respective embedding function.

*potfit* can perform the transformations (6,7) periodically on its own, thus eliminating the need to fix the gauge by an additional term in the target function (1). Unfortunately, for tabulated functions the transformations cannot be performed exactly due to the nature of spline interpolation. A change of gauge therefore can lead to an increase of the target function, which is why we suppress such gauge transformations in the very late stages of a minimization.

### 3. Implementation

*potfit* is implemented in ANSI C. While the user may specify most options in a parameter file read when running the program, some fundamental choices must be made at compile time, like for example the potential model used, or whether to allow for automatic gauge transformations in EAM potentials. This is a compromise between convenience and computation speed. Compile time options can be selected by passing them to the make command, and thus do not require any changes of the source files. For solving the linear equations in Powell's minimization algorithm, *potfit* makes use of routines from the LAPACK library [15], which must be installed separately, probably together with the BLAS library [16] LAPACK is based on.

#### 3.1. Parallelization and optimization

The program spends almost all CPU time in calculating the forces for a given potential; finding a new potential to be tested against the reference data takes only a tiny fraction of that time. Thus, the only way to improve performance is to reduce the total time needed for the force computations, either by minimizing their number, or by making each force computation faster. Powell's algorithm leaves only little room for further reduction of the number of force evaluations. One could for instance adjust the precision required in a line minimization. If the tolerance is too small, time is wasted in refining a minimum beyond need, whereas an insufficient precision may stop too far from the minimum, thus requiring more steps in total. The choice of this tolerance was made empirically.

Much more time can be saved by parallelizing the calculation of forces, energies, and stresses for a given potential. This is done in a straightforward way: As the forces, energies, and stresses of the different reference configurations can be computed independently, we simply distribute the reference configurations on several processes. Before the force computation, the potential parameters are distributed to all processes, and afterwards the computed forces, energies and stresses are collected. The communication is performed using the standard Message Passing Interface (MPI [17]).

This simple parallelization scheme works well as long as the number of configurations per process does not drop below 10 to 15. Otherwise, the communication overhead starts to show up, and load balancing problems may appear. A shared memory OpenMP parallelization also exists, but produces inferior results.

In force matching, the reference configurations stay fixed. Therefore, all distances between atoms remain fixed, and *potfit* can use neighbour lists, which need to be computed only once at startup. In fact, for each neighbour pair all data required for spline interpolation are pre-computed, allowing for a fast lookup of the tabulated functions. This data needs to be recomputed only when the tabulation points of a function are changed.

### 3.2. Input and output files

Tabulated potential functions can be specified with equidistant or with arbitrary tabulation points. For equidistant tabulation points, the boundaries of the domain and the number of sampling points of each function are read from the potential file, followed by a list of function values at the sampling points and the gradients at the domain boundaries. In the case of free tabulation points, only their number is specified at the beginning of the potential file, followed by a list of argument-value pairs and again the gradients of the potential functions at the domain boundaries.

Reference configuration files contain the number of atoms, the box vectors, the cohesive energy, and the stresses on the unit cell, followed by a list of atoms, with atom species, position and reference force for each atom. Such reference configuration files can simply be concatenated.

*potfit* was designed to cooperate closely with the first-principles code VASP [18, 19] and with IMD [20], our own classical MD code. VASP, which is a plane wave code implementing ultrasoft pseudopotentials and the Projector-Augmented Wave (PAW) method [21, 22], is used to compute the reference data for the force matching, whereas the resulting potentials are intended to be used with IMD. For this reason, *potfit* provides import and export filters for potentials and configurations to communicate with these programs. These filters are implemented as scripts, which can easily be modified to interface with other programs.

## 4. Results and validation

As a first test, *potfit* should be able to recover a classical potential from reference data computed with that potential. For this test, we used snapshots from several molecular dynamics runs as reference structures, first for a Lennard-Jones fcc solid, then for a complex Ni-Al alloy simulated with EAM potentials [23]. In order to ensure that all reference data presented to *potfit* is consistent, the potentials were approximated by cubic spline polynomials, in the same way as *potfit* represents the potentials. With such reference data and starting with vanishing potential functions, *potfit* could in both cases perfectly recover the potentials. This test therefore demonstrates the correctness of the program. One should keep in mind, however, that reference data from ab-initio computations often cannot be reproduced perfectly by any classical potential.

Our primary research interest are quasicrystals [24] and other complex metal alloys, for which good potentials are hardly available. *potfit* has been developed in order to generate effective potentials for such complex metal alloys, which feature large (or even infinite) unit cells, several atom species, and a wide variety of different local

environments. So far, force matching had been used mainly to determine potentials for monoatomic metals and a small selection of relatively simple binary alloys.

As a first application beyond simple alloys, we have developed potentials for the quasicrystalline and nearby crystalline phases in the systems Al-Ni-Co, Ca-Cd, and Mg-Zn. Due to the complexity of the structures and also due to the choice of tabulated potential functions, a relatively large number of potential parameters is required. This is especially true for ternary EAM potentials, which comprise 12 tabulated functions, with 10–15 tabulation points each. Correspondingly, a relatively large amount of reference data is required. A computationally efficient implementation of the force matching method is therefore essential. It turned out that *potfit* scales well under those circumstances and is up to its task.

Although the potentials to be generated are intended for (aperiodic) quasicrystals and crystals with large unit cells, all reference structures have to be periodic crystals with unit cell sizes suitable for the ab-initio computation of the reference data. On the other hand, the reference structures should approximate the quasicrystal in the sense, that all their unit cells together accommodate all relevant structural motifs. To do so, they must be large enough. For instance, the quasicrystalline and related crystalline phases of Ca-Cd and Mg-Zn consist of packings of large icosahedral clusters in different arrangements. Reference structures must be able to accommodate such clusters. A further constraint is, that the unit cell diameter must be larger than the range of potentials. We found that reference structures with 80–200 atoms represent a good compromise between these requirements.

Starting from a selection of basic reference structures, further ones were obtained by taking snapshots of MD simulations with model potentials at various temperatures and pressures. Also samples which were strained in different ways were included. For all these reference structures, the ab-initio forces, stresses and energies were determined with VASP, and a potential was fitted to reproduce these data. As reference energy, the cohesive energy was used, i.e., the energies of the constituent atoms was subtracted from the VASP energies. Instead of absolute cohesive energies one can also use the energy relative to some reference structure. Once a first version of the fitted potential was available, the MD snapshots were replaced or complemented with better ones obtained with the new potential, and the procedure was iterated.

As expected, no potential could be found which would reproduce the reference data exactly. During the optimization, the target function (1) does not converge to zero, which indicates that quantum mechanical reality (taking density functional theory as reality) is not represented perfectly by the potential model used. The forces computed from the optimal potential typically differ by about 10% from the reference forces, which seems acceptable. For the energies and stresses a much higher agreement could be reached. Cohesion energy differences for instance can be reproduced with an accuracy better than 1%.

The generated potentials were then used in molecular dynamics simulations to determine various material properties, such as the melting temperature and the elastic constants, for which values consistent with experiment were obtained. The Ca-Cd potentials were especially tuned towards ground-state like structures, whose energies are reproduced with high accuracy, in agreement with ab-initio results. Details of these applications can be found in [4, 25, 26, 27]. Probably the best tested EAM potential constructed with *potfit* was obtained by Rösch, Trebin and Gumbsch [28]. This potential is intended for the simulation of crack propagation in the C15 Laves Phase of NbCr<sub>2</sub>, and has undergone a broad validation. These authors calculated the

lattice constant, the elastic constants and the melting temperature and compared these values to experimental and ab-initio results with reasonable success. They also studied relaxation of surface atoms, surface energy and the crack propagation in NbCr<sub>2</sub>. According to the authors [28], the force-matched potentials created with *potfit* clearly outperform previously published potentials. But this example also shows [28] that a large number of fitting-validation cycles are usually required, before a usable and satisfactory potential is obtained. This makes force matching a time-consuming and tedious process.

## 5. Discussion

### 5.1. Transferability

It should be kept in mind that force-matched potentials will only work well in situations they have been trained to. Therefore, all local environments that might occur in the simulation should also be present in the set of reference configurations. Otherwise the results may not be reliable. Using a very broad selection of reference configurations will make the potential more *transferable*, making it usable for many different situations, e.g. for different phases of a given alloy. On the other hand, giving up some transferability may lead to a higher precision in special situations. By carefully constraining the variety of reference structures one may generate a potential that is much more precise in a specific situation than a general purpose potential, which was trained on a broader set of reference structures. The latter potential, on the other hand, will be more versatile, but less accurate on average. Finding sufficiently many suitable reference structures might not always be trivial. For certain complex structure like quasicrystals, there may be only very few (if any) approximating periodic structures with small enough unit cells.

### 5.2. Optimal number and location of sampling points

Each reference database has an optimum number of parameters it can support. Using too few parameters, the potential functions lack flexibility. On the other hand, exceeding this number may lead to overfitting beyond the limit of the potential model. *potfit* cannot determine that optimal number automatically, but there is a simple strategy the user can employ. The set of reference configurations is split in two subsets, one of which is used for fitting and the other for testing the potential. If the root-mean-square (rms) deviation of the test set significantly exceeds that of the fitting set, the database is probably overfitted [29]. By starting with a relatively low number of parameters, that is increased as long as the rms of the *testing* stage decreases, one can arrive at the optimal number of parameters [30].

This strategy also helps in dealing with oscillatory artefacts of the spline interpolation: If the sampling points are not spaced too densely, and there is enough data to support each tabulation point, artificial wiggles are suppressed. *potfit* provides the frequency with which each tabulation interval is accessed during an evaluation of the target function (1). With this information, sampling point density can be reduced for distances that do not appear frequently enough in the reference configurations.



### 5.3. Number of atom types and choice of reference structures

The most obvious impact of an increasing number of atom types is the corresponding increase in the number of potential parameters. For instance, an EAM potential for  $n$  atom types requires  $n(n+1)/2 + 2n$  tabulated functions, each with 10 to 15 tabulation points. For a ternary system, this already amounts to the order of 150 potential parameters. Whereas such a number of parameters can still be handled, an increasing number of atom types leads to yet another problem, which is more serious. To see this, it must be kept in mind that any potential function depending on the interatomic distance must be determined for the entire argument range between  $r_{\min}$  and  $r_{\text{cut}}$ . If tabulated functions are used, for each tabulation interval there must be distances actually occurring in the reference structures, for otherwise there are potential parameters which do not affect the target function, and which consequently cannot be determined in the fit. The requirement that all distances for all combinations of atom types actually occur in the reference structures becomes especially problematic if the atoms of one type form only a small minority, in which case some distances between such atoms might be completely absent in all reasonable reference structures. If the number of atom types is large, there is unfortunately always at least one element which is a minority constituent. In such situations it might be unavoidable to use a much broader selection of reference structures with varying stoichiometry, instead of a fixed stoichiometry with a minority constituent. It might even be necessary to include energetically less favourable configurations to provide a complete set of reference data.

Another solution would be to use a non-local (or less local) parametrisation of the potential functions, like a superposition of broad gaussians or functions given by analytic formulae. Changing one parameter can then affect the function over a broader range of arguments, making it again possible to fit the function even if only sparse information on it is provided by the reference data. Potentials represented in this way would also not suffer from the wiggle artefacts of spline interpolation described above.

### 5.4. Experimental values as reference data

*potfit* does currently not use experimental data during force matching. The potentials are determined exclusively from ab-initio data, which means they cannot exceed the accuracy of the first principles calculations. While it is possible, in principle, to support also the comparison to experimental values, we decided against such an addition. For once, available experimental values can often not be calculated directly from the potentials, so determining them would considerably slow down the target function (1) evaluation. Secondly, experimental values often also depend on the exact structure of the system, which in most cases is not completely known beforehand for complex structures, for instance due to fractional occupancies in the experimentally determined structure model. A better way to use experimental data is to test whether the newly generated potentials lead to structures that under MD simulation show the behaviour known from experiment.

## 6. Conclusion

Large scale molecular dynamics simulations are possible only with classical effective potentials, but for many complex systems physically justified potentials do not exist so far. Our program *potfit* allows the generation of effective potentials even for complex

binary and ternary intermetallics, adjusting them to ab-initio determined reference data using the force matching method. Potentials for several complex intermetallic compounds have been generated, and were successfully used in molecular dynamics studies of various properties [4, 25, 26, 28]. It should be emphasized, however, that constructing potentials is still tedious and time-consuming. Potentials have to be thoroughly tested against quantities not included in the fit. In this process, candidate potentials often need to be rejected or refined. Many iterations of the fitting-validation cycle are usually required. It takes experience and skill to decide when a potential is finished and ready to be used for production, and for which conditions and systems it is suitable. *potfit* is only a tool that assists in this process. Flexibility and easy extensibility was one of the main design goals of *potfit*. While at present only pair and EAM potentials with tabulated potential functions are implemented in *potfit*, it would be easy to complement these by other potential models, or to add support for differently represented potential functions.

## Acknowledgments

This work was funded by the Deutsche Forschungsgemeinschaft through Collaborative Research Centre (SFB) 382, project C14. Special thanks go to Stephen Hocker and Frohmüt Rösch for fruitful discussion and feedback, and to Hans-Rainer Trebin for supervising the thesis work of the first author.

## References

- [1] Allen M P and Tildesley D J 1987 *Computer Simulation of Liquids*, Oxford Science Publications, (Oxford: Clarendon)
- [2] Beazley D M, Lohmdahl P S, Grønbech-Jensen N, Giles R and Tamayou P 1995 *Parallel algorithms for short-range molecular dynamics* vol III of *Annual Reviews of Computational Physics* (Singapore: World Scientific) pp 119–175 ISBN 981-02-2427-3
- [3] Rösch F, Rudhart C, Roth J, Trebin H R and Gumbsch P 2005 *Phys. Rev. B* **72** 014128
- [4] Hocker S, Gähler F and Brommer P 2006 *Phil. Mag.* **86**(6–8) 1051–1057
- [5] Daw M S, Foiles S M and Baskes M I 1993 *Mater. Sci. Rep.* **9**(7–8) 251–310
- [6] Chantasiriwan S and Milstein F 1996 *Phys. Rev. B* **53**(21) 14080–14088
- [7] Ercolessi F and Adams J B 1994 *Europhys. Lett.* **26**(8) 583–588
- [8] Powell M J D 1965 *Comp. J.* **7**(4) 303–307
- [9] Corana A, Marchesi M, Martini C and Ridella S 1987 *ACM Trans. Math. Soft.* **13**(3) 262–280
- [10] Brent R P 1973 *Algorithms for minimization without derivatives* Prentice-Hall series in automatic computation (Englewood Cliffs, NJ: Prentice-Hall) ISBN 0-13-022335-2
- [11] Galassi M, Davies J, Theiler J, Gough B, Jungman G, Booth M and Rossi F 2005 *GNU Scientific Library Reference Manual - Revised Second Edition* (Bristol: Network Theory Ltd) ISBN 0-9541617-3-4
- [12] Kirkpatrick S, Gelatt C D and Vecchi M P 1983 *Science* **220**(4598) 671–680
- [13] Daw M S and Baskes M I 1984 *Phys. Rev. B* **29**(12) 6443–6453
- [14] Ercolessi F, Parrinello M and Tosatti E 1988 *Phil. Mag. A* **58**(1) 213–226
- [15] Anderson E, Bai Z, Bischof C, Blackford S, Demmel J, Dongarra J, Du Croz J, Greenbaum A, Hammarling S, McKenney A and Sorensen D 1999 *LAPACK Users's Guide, Third Edition* (Philadelphia, PA: Society for Industrial and Applied Mathematics) ISBN 0-89871-447-8
- [16] Lawson C L, Hanson R J, Kincaid D R and Krogh F T 1979 *ACM Trans. Math. Soft.* **5**(3) 308–323
- [17] Gropp W, Lusk E and Skjellum A 1999 *Using MPI - 2nd Edition* (Cambridge, MA: MIT Press) ISBN 0-262-57132-3
- [18] Kresse G and Hafner J 1993 *Phys. Rev. B* **47**(1) 558–561
- [19] Kresse G and Furthmüller J 1996 *Phys. Rev. B* **54**(16) 11169–11186
- [20] Stadler J, Mikulla R and Trebin H R 1997 *Int. J. Mod. Phys. C* **8**(5) 1131–1140
- [21] Blöchl P E 1994 *Phys. Rev. B* **50**(24) 17953–17979

- [22] Kresse G and Joubert D 1999 *Phys. Rev. B* **59**(3) 1758–1775
- [23] Ludwig M and Gumbsch P 1995 *Modelling Simul. Mater. Sci. Eng.* **3**(4) 533–542
- [24] Trebin H R, ed 2003 *Quasicrystals. Structure and Physical Properties* (Weinheim: Wiley-VCH)
- [25] Brommer P and Gähler F 2006 *Phil. Mag.* **86**(6–8) 753–758
- [26] Mihalkovič M and Widom M 2006 *Phil. Mag.* **86**(3–5) 519–527
- [27] Brommer P, Gähler F and Mihalkovič M 2007 *Phil. Mag.* (at press)
- [28] Rösch F, Trebin H R and Gumbsch P 2006 *Int. Journal Fracture* **139**(3–4) 517–526
- [29] Robertson I J, Heine V and Payne M C 1993 *Phys. Rev. Lett.* **70**(13) 1944–1947
- [30] Mishin Y, Farkas D, Mehl M J and Papaconstantopoulos D A 1999 *Phys. Rev. B* **59**(5) 3393–3407