Improved cutoff functions for short-range potentials and the Wolf summation

Martin H. Müser

Taylor & Francis
Taylor & Francis Group

Improved cutoff functions for short-range potentials and the Wolf summation

Martin H. Müser

Department of Materials Science and Engineering, Saarland University, Saarbrücken, Germany

ABSTRACT

A class of radial, polynomial cutoff functions $f_{cn}(r)$ for short-ranged pair potentials or related expressions is proposed. Their derivatives up to order n and $n+1$ vanish at the outer cutoff r_c and an inner radius r_i, respectively. Moreover, $f_{cn}(r \le r_i) = 1$ and $f_{cn}(r \ge r_c) = 0$. It is shown that the used order n can qualitatively affect results: stress and bulk moduli of ideal crystals are unavoidably discontinuous with density for $n = 0$ and $n = 1$, respectively. Systematic errors on energies and computing times decrease by 20–50% for Lennard-Jones with $n = 2$ or $n = 3$ compared to standard cutting procedures. Another cutoff function turns out beneficial to compute Coulomb interactions using the Wolf summation, which is shown to not properly converge when local charge neutrality is obeyed only in a stochastic sense. However, for all investigated homogeneous systems with thermal noise (ionic crystals and liquids), the modified Wolf summation, despite being infinitely differentiable at r_c, converges similarly quickly as the original summation. Finally, it is discussed how to reduce the computational burden of numerically exact Monte Carlo simulations using the Wolf summation even when it does not properly converge.

1. Introduction

The efficiency of molecular simulations hinges on the truncation of interaction potentials [1,2]. One possibility to achieve that is to multiply the interaction potential, or, functions entering their calculation, with cutoff functions [3–5]. However, there are two major, mutually exclusive requirements on them. The cutoff radius r_c should be as large as possible to reduce systematic errors [6] but also be as small as possible to boost computational efficiency. Similarly, given a particular value for r_c, $f_c(r)$ should be close to unity for as long as possible to reduce discrepancies from the real energy. At the same time, $f_c(r)$ should be decreased to zero as smoothly as possible to avoid abnormally large forces or curvatures at large distances, which generally induce undesired behaviour [7,8]. A compromise is certainly needed, but it does not seem that a generally applicable one has been identified. Another common strategy to cut off short-range potentials is by making all their derivatives up to n th order vanish continuously at r_c through

$$U_{sn}(r) = \{U(r) - \mathcal{T}[U(r), r_c, n]\}\Theta(r_c - r), \quad (1)$$

where $U(r)$ is a pair-potential or a related local function, $\mathcal{T}[U(r), r_c, n]$ the n th order Taylor expansion of $U(r)$ about $r = r_c$, and $\Theta(r)$ the Heaviside function. A shifted-potential (SP) potential is obtained for $n = 0$, a shifted-force (SF) potential for $n = 1$ [9–11], and a shifted-curvature (SC) potential for $n = 2$. A disadvantage of the shifting procedure is that binding or cohesive energies decrease rather quickly with n at fixed r_c. For this but also for other reasons, it is often desirable to sum up potentials or other local functions so that the contribution from nearest neighbours is exact, for example, when computing the embedding density within a potential based on the

embedded-atom method [12]. To achieve this, the partial densities are multiplied with cutoff functions, $f_c(r)$, which are set to unity up to an inner radius r_i and then swiftly decreased toward zero [4]. However, in order to avoid qualitative cutoff artifacts, $f_c(r)$ has to approach zero in a sufficiently smooth fashion.

Good cutoff functions are central to balancing computational efficiency and systematic errors but are surprisingly little discussed even in stellar textbooks on molecular simulation [1,2]. Unfortunately, there is no unique optimum. It would depend not only on the potential but also on the property of interest. For example, when studying sublimation, reproducing the energies themselves is crucial. However, forces and curvature of potentials determine mechanical properties. In this context, it is useful to keep in mind that an attractive potential that is cut without shifting has a δ-function spike in the force. Its proper implementation will unavoidably lead not only to unphysical dynamics but require non-trivial terms to be averaged when determining static properties like pressure [1]. In a cut-and-shift potential, stress is still discontinuous in density for ideal crystals (and thus potentially for other systems too) so that upon reversion the density is discontinuous in pressure. Each higher-order in a cut-and-shift procedure mitigates artifact to one higher-order in the response function so that $n = 3$ is the lowest cut-and-shift order, which systematically avoids a discontinuity of elastic properties with pressure. In MD simulations based on standard symplectic integrators, $n \ge 2$ is needed to ensure energy conservation [10] so that $n = 2$ can be seen as the lowest order of a generally applicable cutting-off procedure.

In this paper, cutoff functions with beneficial properties are proposed. The cutoff functions are designed to take a value of

unity up to r_i and to approach zero continuously as their argument approaches r_c. Moreover, the function itself and its derivatives up to order n vanish continuously at the outer cutoff r_c. It is demanded to be one order higher at the inner cutoff, because artifacts arise not only when atoms or entire neighbour shells cut through r_c but also through r_i. This choice is made because short-ranged potentials, their forces and curvatures tend to be larger at the inner radius than at the outer cutoff, so more care is required at r_i than at r_c.

Shifting potentials have been discussed in particular in regard to the Wolf summation [13–16]. Wolf et al. showed that cutting and shifting the Coulomb interaction is equivalent to placing a charge-balancing countercharge at r_c, thereby reproducing an important element of the Evjen summation [17]. Applying the shifting procedure to the damped Coulomb interaction arising in the real-space part of the Ewald summation [18] rather than to the original Coulomb interaction, allows the convergence with increasing r_c to be quickly reached, even when neglecting the non-zero-wavenumber contributions to the Fourier portion of the Ewald sum. While systematic errors in the Wolf summation cannot be made arbitrarily small with the same low computational cost as with other Coulomb interaction summation techniques, most notably the particle mesh Ewald method [19], it may yet be interesting for a variety of reasons: It can be used (i) for quick prototyping, (ii) in Monte Carlo simulations, which, unlike molecular dynamics, does not benefit from the simultaneous update or thermalisation of all degrees of freedom, and (iii) in conjunction with multiple-time stepping schemes [20]. This is why cutoff functions in the context of the Wolf summation are also investigated. This includes a discussion of how to effectively use the (modified) Wolf summations when it fails to converge.

2. Background

2.1. Conventional cutoff functions

In principle, any shifted potential can also be obtained with a cutoff function defined implicitly through $f_c(r) \equiv U_{sn}(r)/U(r)$, where n is the largest-order derivative of the potential going continuously to zero at the cutoff. The resulting cutoff function would not be near unity at a typical nearest-neighbour distance unless r_c were very large. This is why shifting procedures should generally be inferior to more general cutoff functions with similar behaviour for $U^{(n)}(r \lesssim r_c)$. We can therefore dismiss simple-shifting procedures as a competitive alternative to well-designed shifting functions.

One of the most frequently used cutoff functions, supposedly proposed by Tersoff [4], is given by

$$f_{cF}(r) = \Theta(r_i - r)$$
$$+ \frac{\Theta(r - r_i)\Theta(r_c - r)}{2}\left\{1 + \cos\left(\pi\frac{r - r_i}{r_c - r_i}\right)\right\}. \quad (2)$$

This function, just like SF potentials, makes the force go linearly to zero as r approaches r_c but has a discontinuity in the curvature. Since $f_{cF}(r)$ is mirror symmetric about $(r_m, 1/2)$,

where $r_m = (r_i + r_c)/2$ can be called the mid-point, it has the same non-analyticity at r_i and r_c.

An improved version of and thus replacement for the SP potential can be generated with the cutoff function

$$f_{cP}(r) = \Theta(r_i - r) + \Theta(r - r_i)\Theta(r_c - r)\sin\left(\frac{\pi}{2}\frac{r_c - r}{r_c - r_i}\right), \quad (3)$$

which obeys the proposed rule of the cross-over function being one order more continuous at r_i than at r_c. This rule is meant to be useful when a potential decays with $1/r^3$ or faster, because the relative number of interactions, inner versus outer radius, where discontinuities in derivatives matters, scales with $(r_i/r_c)^2$ in three spatial dimensions. For an electrostatic monomer-dipole interaction, the same order discontinuity at r_i and r_c would be recommended as this would balance errors at the inner radius and the outer cutoff. In lower dimensions, the exponents have to be corrected appropriately.

Other cutoff functions exist [21], e.g. functions that also have mirror symmetry about $(r_m, 1/2)$ but with higher-order vanishing derivatives at r_i and r_c than f_{cF}. However, they are not considered here, because they violate our mantra that more care needs to be taken at the inner than at the outer radius, and/or, because they have more 'structure' than simple polynomials. Finally, we note that this mantra does not apply to long-range potentials, where it may be beneficial to have smaller errors at r_c than at r_i, in particular if $r_i = 0$ is chosen.

2.2. Polynomial cutoff functions

Cutoff-function proposed in prose in abstract and introduction are given by $f_{cn}(r) = 1 - P_n(x)$ with $x = (r - r_i)/(r_c - r_i)$ and

$$P_1(x) = x^2,$$
$$P_2(x) = 4x^3 - 3x^4, \quad (4)$$
$$P_3(x) = 15x^4 - 24x^5 + 10x^6.$$

The $P_n(x)$ are constructed as the lowest-order polynomials to vanish with order $0, \ldots, n+1$ at $x = 0$ and to assume $P(1) = 1$ while all derivatives up to order n vanish at $x = 1$. The resulting cutoff function are depicted in Figure 1 together
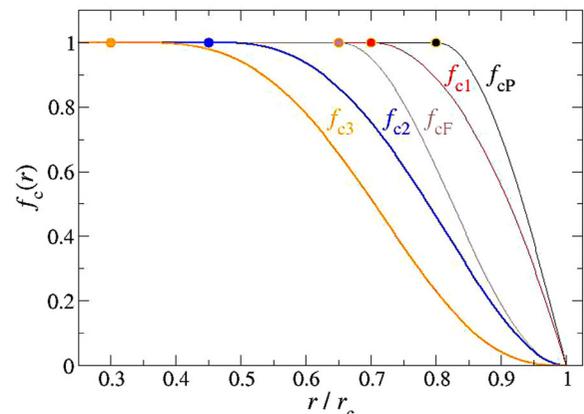

Figure 1. Selected cutoff functions. Circles indicate the location of the inner cutoff radius r_i.

with f_{cP} and f_{c2}. Inner cutoffs were chosen for aesthetic reasons so that different functions do not cross.

We are confronted with the task of determining 'optimal' values for r_i, which depends on the cutoff function in addition to the (pair) potential and the property of interest. Since the parametrisation of potential is done w.r.t. the cutoff function, it seems in place to suggest a generic choice in the *one-size-fits-all* spirit. Choosing $r_i = 0$ and $r_i = r_c$ are certainly lower and upper bounds. However, they are obviously anything but helpful. One way of proceeding would be to demand that forces or derivatives at $r \geq r_i$ should not be greater in magnitude than at $r = r_i$. For a dispersive $1/r^6$ attraction, r_i/r_c would then turn out to lie within 0.8049 and 0.8221 for all cutoff functions discussed so far, except for the popular $f_{cF}(r)$ cutting function, which would require $r_i/r_c = 0.7447$. This smaller ratio arises because the non-analyticity of $f_{cF}(r)$ at the inner cutoff is as significant as at the outer cutoff. Thus, larger compromises would have to be made on the cohesive energy using $f_{cF}(r)$ than for the remaining cutoff functions. Unfortunately, the just reported r_i/r_c ratios still turn out too aggressive for the Lennard-Jones potential: the equation of state of an ideal, face-centred cubic (fcc) Lennard-Jones remains discontinuous. This undesired behaviour could be eliminated by reducing the ratio to $r_i/r_c = 2/3$.

2.3. Cutoff functions infinitely often differentiable at r_c

Cutoff functions going to zero such that all their derivatives vanish at r_c can be beneficial, e.g. in the context of generalised embedded-atom-method (EAM) based potentials, in which derivatives of the charge density enter the definition of the potential [22,23]. This can be achieved with a function combining the Stillinger–Weber (SW) [3] cutoff function and the idea of a polynomial expansion pursued in this paper. Specifically,

$$f_{SWn}(r) = \frac{\mathcal{T}[\text{denom}(r), 0, n]}{1 + \exp\left(\frac{\Delta r_c}{r_c - r}\right)} \Theta(r_c - r) \quad (5)$$

is such a cutoff function. Here, $\text{denom}(r)$ is the denominator of the quotient on the r.h.s. of the equation, $\mathcal{T}[\cdots]$ denotes a Taylor series expansion as above, and Δr_c determines, as a
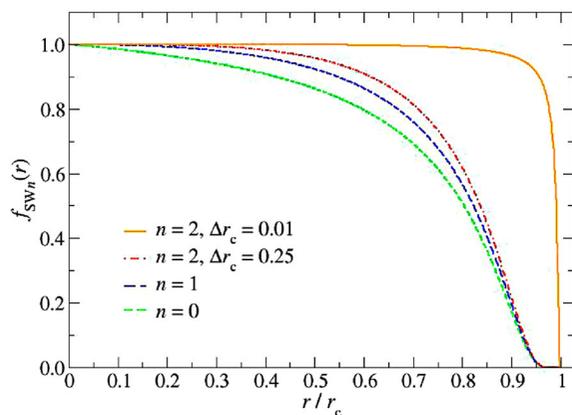

Figure 2. Selected Stillinger–Weber inspired cutoff functions. All dashed lines use $\Delta r_c = 0.25 r_c$.

function of n, how closely to r_c the cutoff function assumes the value 0.5. Some selected SW generalised cutoff functions are shown in Figure 2. The original one proposed by Stillinger and Weber corresponds to $n = 0$.

For $\Delta r_c \ll r_c$ and/or large values of n, the effect of $f_{SWn}(r)$ will be similar to that of a harsh cutoff, as $f_c(r)$ is close to unity up to the immediate vicinity of r_c, in which case the disadvantages from harsh cutoffs are inherited. We see no reason to use $f_{SWn}(r)$ for regular short-range potentials, however, it could benefit, for example, the systematically modified embedded atom method, [23] for which the embedding energy depends on (higher-order) derivatives of the embedding density. Higher-order derivatives place higher demands on the way in which the charge density is brought down to zero at r_c, in particular, when an individual atom breaks its final bond.

Although using f_{SWn} is disadvised for regular short-range potentials, it appears to be a suitable candidate to cut off long-range potentials, since its analyticity is of higher order at r_c than at r_i. In fact, a cutoff function closely related to that proposed by Stillinger and Weber, $f_m(r) = \prod_{q=1}^{m}(1 - x^q)$, was explored recently and found beneficial for the Wolf summation [24]. For $m \to \infty$, $f_m(r)$ closely resembles the SW function with $\Delta r_c/r_c = 1.2$. Using higher-order n in f_{SWn} might further improve convergence. This expectation is explored in the context of the Wolf summation in Section 4.

3. Cutting short-range potentials

The generic (pair) potential used to describe non-bonded interaction is the Lennard-Jones (LJ) potential

$$U(r) = U_0\left\{\left(\frac{r_0}{r}\right)^{12} - 2\left(\frac{r_0}{r}\right)^6\right\}, \quad (6)$$

where U_0 is the binding energy of a dimer and r_0 its equilibrium bond length. The standard cutoff used for LJ is $r_c = 2.5\sigma$, where $\sigma = \sqrt[6]{2} r_0$ is also called the LJ radius. Often, the LJ potential is merely shifted using this default value. This procedure is standard practice and certainly acceptable. Nonetheless, simultaneous improvements on both accuracy and computing time should be possible, which is explored next.

To demonstrate the effect of the various cutting schemes, the fcc LJ crystal will be investigated. It allows artifacts to be highlighted, while keeping computing times and numerical errors minimal. The local structure certainly differs between LJ crystals with well-defined neighbour shells and liquid Lennard–Jonesium, which is close to random-sphere packing. As a consequence, typical bond distances at zero or what-would-be ambient pressure are less than r_0 in the crystal but greater r_0 in the liquid. Next-nearest neighbour distances and associated coordination numbers, to be defined, e.g. through a skew-normal-distribution analysis of peaks in the radial distribution $g(r)$ [25], differ even more between crystal and liquid. Including into the discussion the radial distribution functions arising in (united-atom based) models of polymers makes it even more difficult to identify guidelines for how to pick r_i and r_c so that they both coincide with minima in $g(r)$. Thus, any

final choice should yield robust results no matter how r_i and r_c relate to the maxima and minima in $g(r)$ of any particular system of interest. Any critical situation is included in the analysis when analysing the cohesive energy and the equation of state (EOS) in the range $0.85 \leq a_0/r_0 \leq 1.3$, where a_0 is the (mean) nearest-neighbour bond length. This is because the energy of an individual LJ bond is already positive at $r = 0.85r_0$, which is a situation of a very high compressive stress or force. At the other end at $a_0 = 1.3r_0$, a LJ bond can be considered broken, because this bond length is beyond the inflection point of the LJ potential, i.e. past the point of maximum tensile force.

The goal is to identify parameters for the cutoff function(s) and radii that globally outperform the standard cut-and-shift procedure. To this end, we chose arbitrarily $r_c = 2.5r_0$, which reduces the interaction volume to 70% compared to that of the default cutoff, $r_c = 2.5\sigma$, and thereby the number of force evaluations by a similar percentage. Of course, it would be a simple matter to include mean-field corrections for the cohesive stress [26] so that smaller cutoff radii could be trivially achieved without losing accuracy. However, they would not be useful for heterogeneous systems, e.g. when surfaces are present. Moreover, such corrections are not always available in popular software packages. This is why mean-field corrections are not included in this study either.

For $r_c = 2.5r_0$, an inner cutoff of $r_i = 2r_c/3$ was found beneficial. It makes the cohesive energy of an fcc crystal be just below the default cut-and-shift procedure with $r_c = 2.5\sigma$, at least in the 'interesting range' of $0.85 < a_0/r_0 < 1.3$, which is demonstrated in Figure 3, where the values of the pertinent potential energies are almost within line width in panel (a). Both the default cut-and-shift as well as the $f_{c3}(r)$ cutting yield a similar minimum in the cohesive energy of about $8U_0$, per atom, while the nearest-shell approximation yields exactly $6U_0$. The exact binding energy is about $8.59U_0$. Figure 3(a) also reveals that using a harsh, unshifted cutoff at $2.5r_0$ does not significantly lower the energy compared to a method using the same cutoff radius but the high-order smoothing function $f_{c3}(r)$. However, the discontinuities occurring when using harsh cutoffs generally yield unacceptable behaviour.

Figure 3(b) resolves the error $\Delta U = U(\text{appr.}) - U(\text{exact})$ over a relevant range. Errors are multiplied with a_0^3 to make results approximately constant. Values turn out close to the ones expected from the mean-field correction to the dispersive interaction, i.e.

$$\frac{\Delta U_{\text{disp}}}{U_0} \approx \int_{r>r_c} \mathrm{d}^3 r \rho \left(\frac{r_0}{r}\right)^6 = \frac{\sqrt{32}\pi}{3}\frac{r_0^6}{r_c^3 a_0^3},$$

the numerical prefactor evaluating to approximately 0.38 after having inserted the fcc atomic number density of $\rho = \sqrt{2}/a_0^3$. Since repulsion reduces the binding energies, 0.38 is merely a lower bound for the numbers reported in Figure 3(b).

Since the energies of the various approximations schemes are quite close to each other, so will their EOS. In fact, they turn out to be within line width in Figure 3(c), except for the nearest-neighbour approximation revealing a significantly reduced (theoretical) maximum cohesive stress. However,

zooming into parts of the EOS resolves that the standard cut-and-shift procedure induces a discontinuous EOS. Similar discontinuities also occur under compression, however, their relative effect is negligible. Of course, even minor thermal fluctuations smear out the discontinuities, so one certainly does not need to be concerned when using the standard $r_c = 2.5\sigma$ LJ cut-and-shift procedure. Nonetheless, they can become relevant for other potentials or for smaller cutoffs.

It can be summarised that the $f_{c2}(r)$ and $f_{c3}(r)$ cutoff function lead to smaller errors than the standard cut-and-shift procedure for the cohesive energy and the EOS in the range what we deemed to be interesting. At the same time, they require only about 70% of the force evaluations. However, this latter point is only advantages when look-up tables for interatomic potentials and forces are used. Otherwise, the additionally required floating-point operations needed to evaluate forces from smoothly cut potentials would be prohibitively expensive.

4. Wolf summation

4.1. Background

Ewald [18] demonstrated that Coulomb interactions in periodically repeated systems can be meaningfully summed up by dividing the summation into a real-space and a Fourier or reciprocal-space contribution. The latter contains two terms at zero wave vector, one of which is the so-called self-interaction energy and the other the electrostatic energy resulting from the electrostatic field generated by the mean dielectric polarisation. In detail, given a charge-density distribution of $\rho(\mathbf{r}) = \sum_i q_i \delta(\mathbf{r} - \mathbf{r}_i)$ with zero net charge, the electrostatic energy reads [18,27]

$$U_{\text{C}}(\{\mathbf{r}\}) = \sum_{i,j>i} \frac{q_i q_j}{4\pi\varepsilon_0} \frac{\text{erfc}(k_{\text{E}} r_{ij})}{r_{ij}} - \sum_i \frac{q_i^2}{4\pi\varepsilon_0} \frac{k_{\text{E}}}{\sqrt{\pi}}$$
$$+ \frac{\mathbf{p}_{\text{tot}}^2}{2(2\varepsilon_{\text{ext}}+1)\varepsilon_0 V} + \frac{1}{2\varepsilon_0 V} \sum_{\mathbf{k},k\neq0} \frac{|\tilde{\rho}(\mathbf{k})|^2}{k^2} \, e^{-k^2/4k_{\text{E}}^2},$$

(7)

where V is the volume of the periodically repeated (simulation) cell, $\tilde{\rho}(\mathbf{k})$ is the Fourier transform of $\rho(\mathbf{r})$, and $\mathbf{p}_{\text{tot}}$ the total dipole moment of the simulation cell, i.e. $\mathbf{p}_{\text{tot}} = \sum_i q_i(\mathbf{r}_i - \mathbf{r}_{0i})$ assuming the dipole moment for the set of reference coordinates $\{\mathbf{r}_0\}$ to vanish. Finally, ε_{ext} is the relative permittivity of an external embedding medium. Including its effect matters for finite clusters, which are placed into a simulation cell with a vacuum buffer, in which case $\varepsilon_{\text{ext}} = 1$.

The last summand on the r.h.s. of Equation (7) can become irrelevant for sufficiently small k_{E}. This might have enticed Wolf et al. [13] to ignore that term completely. In order to effectively enforce charge-neutrality within r_c, Wolf et al. [13] used a cut-and-shift potential, and corrected the self-interaction energy to

$$U_{\text{DSP}}\text{self}(i) = -\left(\frac{k_{\text{E}}}{\sqrt{\pi}} + \frac{\text{erfc}(k_{\text{E}} r_c)}{2r_c}\right)\frac{q_i^2}{4\pi\varepsilon_0}.$$
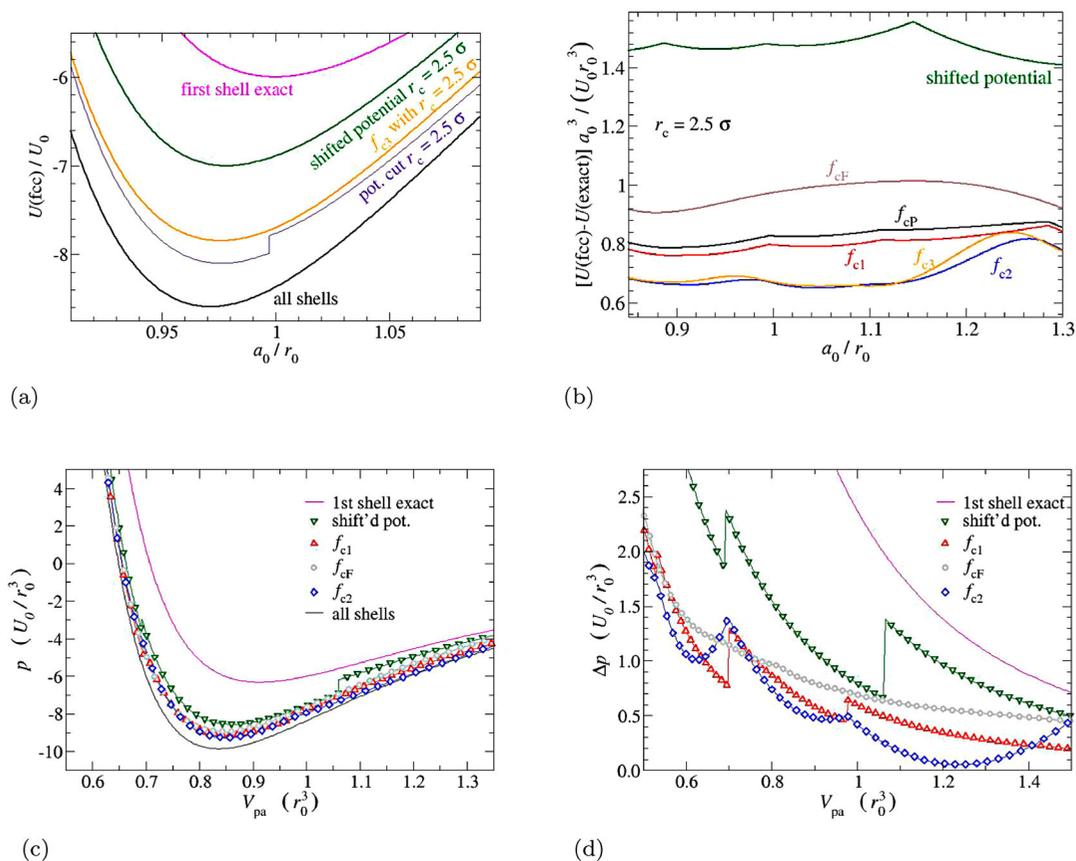
(8)

Figure 3. Effect of various cutting procedures on (a,b) the fcc binding energy U(fcc) as function of nearest-neighbour distance a_0 and on (c,d) the equation of state. (a) Cohesive energies U(fcc) of various approximation schemes for fcc Lennard–Jonesium. (b) Weighted errors of various approximation schemes to the exact cohesive energy. (c) Global equation of state (EOS). (d) Error in the pressure in the EOS.

Although simply cutting and shifting potentials is problematic for reasons discussed above as well as in [6,11,14], this original Wolf summation is taken as the reference for alternative cutting procedures investigated here below.

Picking k_E properly when performing a real Ewald summation is crucial to achieving a good compromise between speed and accuracy. Using a fast Ewald method, k_E can be kept constant irrespective of the system size, or, particle number N. For the conventional Ewald sum, the apparently optimum choice is $k_E \propto 1/\sqrt[6]{N}$, in which case the computational effort to yield results with a target error scales as $N^{3/2}$, both in real as well as in reciprocal space [2,28]. For both fast and conventional Ewald summation, increasing the demand on accuracy by a given factor then only necessitates an increase in computing time scaling sub-logarithmically in this factor.

The large convergence rate of Ewald summations cannot be achieved using the Wolf summation. However, even an algebraic dependence would be desirable, in particular in the context of Monte Carlo simulations, which, unlike molecular-dynamics simulation, does not profit from a parallel update of all degrees of freedom. To ensure convergence using the Wolf summation, k_E must be made a function of r_c. As discussed in more detail here below, the overall best choice when using a Wolf summation turns out to satisfy $k_E \approx 1/\sqrt{a_0 r_c}$, where a_0 is a typical distance of adjacent anions and cations.

4.2. Convergence for ideal and perturbed rock-salt structures

A first convergence analysis for the Wolf summation is presented in Figure 4. Panel (a) shows the Madelung constant α_M, while panel (b) depicts the magnitude of its error. The latter reveals that convergence is not only algebraic but even exponential with r_c for this highly symmetric structure when using $k_E = 1/\sqrt[4]{4 a_0^3 r_c}$. While the original Wolf summation converges the most quickly, kinks in α_M indicate indirectly that the original summation will unavoidably lead to artifacts. Essentially exponential convergence is also obtained for f_{c2} and f_{c3}, for which $r_i = r_c/2$ was used. Although rates are a little less than for the original Wolf summation, the new summations are much better behaved whenever a neighbour sell cuts through the cutoff radius. The f_{SW2}-data was obtained using $r_i = 0$ and $\Delta r = r_c$. The general $n = 2$ expansion of the numerator on the r.h.s. in Equation (5) reads $1 + \exp(\tilde{w})\{1 + \tilde{r} + \tilde{w}(2 + \tilde{w})\tilde{r}^2/2\}$ with $\tilde{w} = \Delta r_c/r_c$ and $\tilde{r} = r/\tilde{r}_c$.

Before proceeding to less idealised cases, some observations will be reported. First, ideal rock salt was the only structure for which choosing $k_E \approx 1/\sqrt[4]{4 a_0^3 r_c}$ was clearly optimal. In all other cases, $k_E \approx 1/\sqrt{a_0 r_c}$ turned out to be the apparently best option for reasons stated further below. Second, rock salt was the only structure for which f_{c2} and f_{c3} 'outperformed' f_{SW2} at large r_c. Third, making the order n at r_i greater than at r_c, i.e. replacing $1 - P_n(x)$ in the definition of $f_{cn}(r)$ with
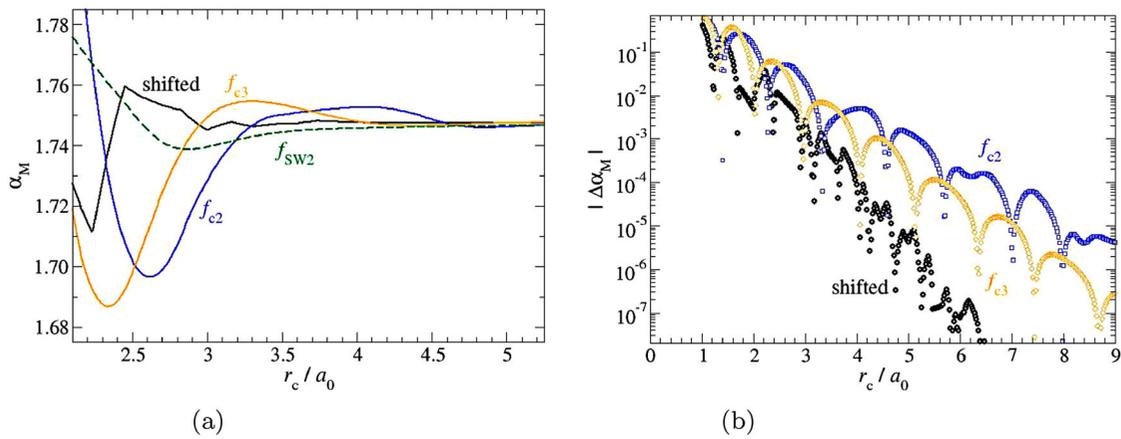
Figure 4. Convergence of Madelung constant α_M in ideal rock salt with increasing cutoff radius r_c in units of the bond length a_0 for various cutting schemes for the damped Coulomb interaction. (a) Absolute values. (b) Magnitude of deviation.

$P_n(1-x)$, did not improve results. Fourth, f_{SW2} outperformed all other f_{SWn}. Observations 3 and 4, whose reasons we do not yet understand, also hold for the other investigated structures.

Deviations from the ideal rock-salt structure were also investigated. First, a small random distance was added to each atomic coordinate so that the far field of an atoms is identical to that of a point charge augmented with a random dipole. Second, the charge of each atom on an ideal lattice was augmented or reduced randomly by half an elementary charge with the constraint that the net charge remains unchanged. The result is an ionic solid solution with positional disorder. The such produced configurations will be called random-dipole and random-charge crystals, respectively.

For sufficiently large systems, the Madelung constants of both random crystals is identical to that of regular rock salt. This is because (a) the field of a random dipole or higher-order multipole has a random direction so that placing another multipole into its field has, on average, zero potential energy and (b) the expectation value of the product $q_i q_j$ satisfies $\langle q_i q_j \rangle = \langle q_i \rangle \langle q_j \rangle$ in the limit of infinite particle numbers. In finite systems, systematic deviations occur because the fluctuation of a given charge is perfectly correlated with that of its

periodic images but slightly anti-correlated with all other charges and their periodic images.

Figure 5(a) shows that the Wolf summation converges to the proper effective or mean Madelung constant α_M for the random-dipole crystal (within statistical fluctuations from one random realisation to the next) if k_E is made an appropriate function of r_c, e.g. $k_E = \kappa/(a_0 r_c)^\beta$ with the prefactor $\kappa = 1.2$ and exponents $\beta = 1/2$ or $\beta = 1/3$. This is not surprising, since the dipole-dipole interaction of oriented dipoles is just no longer integrable in three spatial dimensions, so that sums over randomly oriented dipoles are unconditionally integrable or summable. Figure 5(a) allows the following, additional observations to be made: The Wolf-summation results can be fit quite accurately using $\alpha_M = \alpha_M(\infty) + \beta/r^\gamma$ at large r_c. A smaller exponent β leads to a smaller exponent γ, however, the asymptotic scaling is reached at smaller r_c/a_0 ratios. In the given example, $\gamma = 1$ for $\beta = 1/3$ and $\gamma = 3/2$ for $\beta = 1/2$. Moreover, the modified Wolf summation has the same asymptotic approach to $\alpha_M(\infty)$ as the original summation.

It is also noted that the SW2-modified Wolf summation tends to be closer to the exact result than the original Wolf
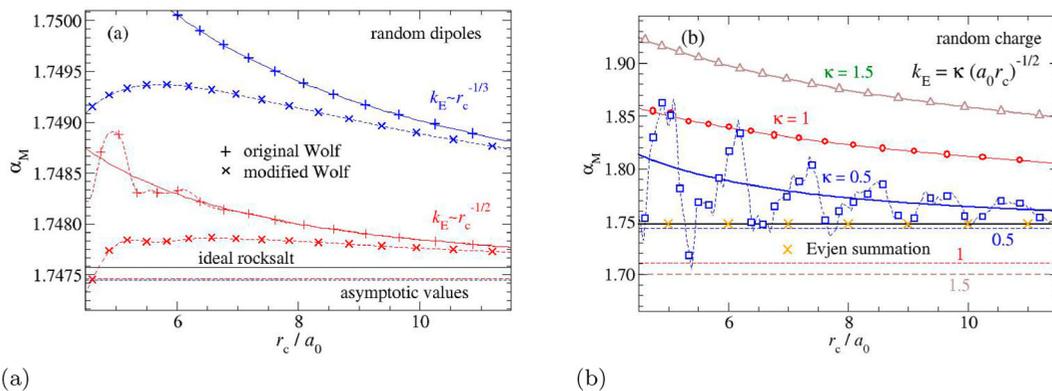

Figure 5. Convergence of the mean Madelung constant α_M for (a) random-dipole and (b) random-charge crystals containing $24 \times 24 \times 24$ atoms in total. For $r_c > 8a_0$, the Wolf sums were fit with $\alpha_M = \alpha_M(\infty) + \beta/r_c^\gamma$ and shown as coloured, solid lines on the entire domain. The asymptotic values, $\alpha_M(\infty)$, are drawn as dashed lines in their respective color. In both panels, the Madelung constant of ideal rock salt is drawn as a solid, black line. (a) Plus symbols and crosses show the original and a f_{SW2}-modified Wolf summation, respectively. The Ewald parameter was chosen as $k_E = 1.2/(a_0^{1-\beta} r_c^\beta)$ with $\beta = 1/3$ and $\beta = 1/2$. (b) Different symbols refer to different prefactors κ used for $k_E = \kappa/\sqrt{a_0 r_c}$, i.e. $\kappa = 1.5$ (triangles), $\kappa = 1$ (circles) and $\kappa = 0.5$ (squares). Crosses show the Evjen summation.

summation, however, asymptotic scaling sets in at larger r_c. Besides producing continuous forces and potential curvatures at r_c, this is one reason why the use of the SW2-modified Wolf summation would be suggested for simulations for which r_c is a fixed quantity. However, extrapolating $\alpha_M(r_c \to \infty)$ is more easily done using the original Wolf summation, or, for example, the f_{c3}-modified Wolf summation. For this reason, most of the subsequent convergence analysis is made on the original Wolf summation.

For the random-charge crystal, the Wolf summation no longer converges to the correct Madelung constant, as is revealed in Figure 5(b), at least as long as r_c is less than half the size of the periodically repeated cell. This time, the prefactor κ to the $k_E = \kappa/\sqrt{a_0 r_c}$ was varied. The exponent γ in the (seemingly) asymptotic $\alpha_M = \alpha_M(\infty) + \beta/r^\gamma$ relation was again not universal but turned $\gamma \lesssim 1/2$ for $\kappa \gtrsim 1$. Thus, being locally charge neutral in a stochastic sense is not a sufficiently strong condition for the Wolf summation to converge. If positive and negative charges separate deterministically, which can be caused by a structural heterogeneity on scales exceeding r_c, the Wolf summation will obviously be even more erroneous than for random charge neutrality.

4.3. Convergence for crystalline and liquid silica

The convergence rate of the modified Wolf summation is also explored on crystals of lower symmetry than rock salt and a corresponding ionic melt, namely silica. As a reference crystal, cristobalite was chosen. It is a polymorph of silica, in which the silicon atoms form a cubic diamond lattice and the bridging oxygen atoms predominantly rotate in a safe distance about their average, crystallographic positions, which are located halfway between two adjacent silicon atoms [29,30]. The (local) symmetry of atoms in this polymorph is lower than in rock salt, because the field gradients on oxygen atoms even in the crystallographic positions are anisotropic, while the anisotropy of fields of atomic positions in the ideal rock salt structure appears first in its third spatial derivative. Since
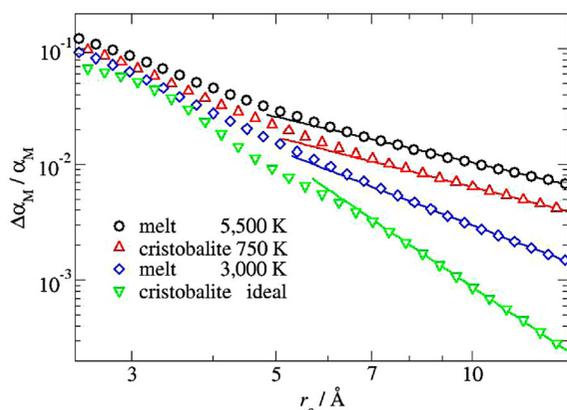
the real positions of oxygen are quite distant from the crystallographic ones, oxygen atoms tend to sit at sites with a relatively rather large electric field. Silica is simulated with the potential proposed by van Beest, Kramer, and van Santen (BKS) [31] using a house-written code described before [30]. Despite some shortcomings, the BKS potential has reproduced various properties of liquid [32] and crystalline [30,33] silica.

Figure 6 shows the relative error in the Coulomb energy, which was obtained for silica melts at two different temperatures as well as for cristobalite, one time with oxygen atoms being constrained to their crystallographic positions and one time at a temperature just above the phase transformation temperature from the high-symmetry β-cristobalite phase to the optically active α-cristobalite [34]. As expected, the Wolf summation converges more quickly for the ideal, crystallographic crystal than for the thermal crystal, for which the Wolf summation converges similarly quickly, or, depending on viewpoint, slowly as the random-dipole crystal considered in Section 4.2.

A surprising result of Figure 6 is the relatively fast convergence of the Wolf summation for the 'low-temperature' ($T = 3000$ K) melt, which is not only faster than at $T = 5500$ K melt but also faster than for the 750 K, thermal crystal. This may have to do with the fact that Madelung sums should actually converge for homogeneous melts, even mixtures [16], since the (partial) density autocorrelation function in dense liquids are damped oscillations at large r not 'suffering' from distant neighbour shells carrying large number of atoms and thereby preventing lattice sums from unconditional convergence. Ultimately, (twice) the electrostatic energy per point charge can be cast as an integral over the charge-density autocorrelation function, $C_{\rho\rho}(r) \equiv \langle \rho(0)\rho(r) \rangle$ via

$$U_C = \frac{1}{4\pi\varepsilon_0} \int_{0^+} d^3 r \frac{C_{\rho\rho}(r)}{r} = \frac{1}{\varepsilon_0} \int_{0^+}^{\infty} dr\, r\, C_{\rho\rho}(r), \qquad (9)$$

where 0^+ is meant to indicate that self-interactions of charges at $r = 0$ are excluded from the integral. The (negative) integrand in the last term of Equation (9) is shown in Figure 7. It reveals that subsequent peaks in the integrand becomes ever smaller in the melt but not necessarily in the crystal. In dense, three-dimensional liquid, this behaviour can be rationalised using the Ornstein–Zernike theory [35], which predicts the density oscillations to obey asymptotically $\cos(r/\lambda + \varphi) \exp(-r/\zeta)$, where λ is a wavelength, φ a phase shift and ζ a correlation length. Maxima and minima in the negative integrand, which could be interpreted as a Madelung constant density, are located near the maxima of the partial dislike and like-ion radial distribution functions, respectively. While the magnitude of the integrand for large r is clearly bound by a simple exponential in the liquid, this is not true for the crystal, where the extrema at $r \gtrsim 9$ Å and $r \lesssim 10$ Å are more pronounced than those in the interval 7–8 Å.

As a small side remark to this article, we wish to note that the computational burden of the regular Ewald summation can be slightly reduced when the Fourier part of the interaction is not evaluated every time step but only every $\mathcal{O}(a_0 k_E)$ time steps. Such a reduction is possible because the long-wavelength dynamics are slower than the ones associated with short

Figure 6. Relative error in the Coulomb energy $\Delta\alpha_M/\alpha_M$ as a function of the cutoff radius r_c in units of Å for a modified Wolf summation. The studied systems were a melt at 5500 K (circles) and at 3000 K (diamonds) as well as a thermal cristobalite crystal at 750 K (red triangles up) and an ideal crystal at the same density, for which, however, all atoms were placed onto their ideal lattice positions (triangles down). Lines represent power laws with exponents $\gamma = 1.31, 1.43, 2.15, 3.8$ (top to bottom).
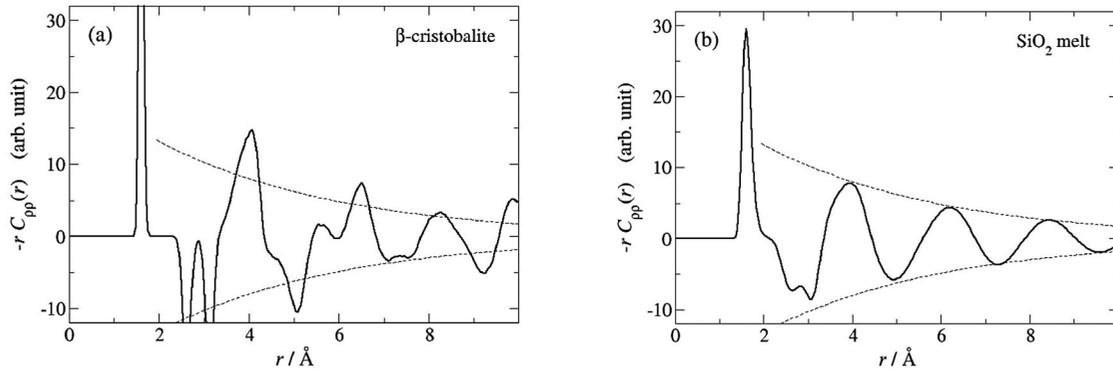
Figure 7. Weighted charge-density autocorrelation function, $-rC_{\rho\rho}(r)$, as a function of distance r for (a) β-cristobalite at 750 K and (b) a SiO$_2$ at 3000 K. The dashed lines are an exponential function proportional to $\pm \exp(-r/\zeta)$ with $\zeta = 4$ Å.

wavelengths. The CPU time needed for the real-space sum would then scale as before with Nr_c^3, while the Fourier part would be reduced from $k_E^3 N^2$ to $k_E^4 N^2$, assuming that, say, a wavenumber cutoff of $k_c \approx 4k_E$ is generally acceptable. For systems with stark (charge) heterogeneity on arbitrary wavelengths, it would be required to chose $k_c \propto 1/r_c$ as to avoid uncontrollable summation errors, which would otherwise arise if a structural heterogeneity existed on a wavelength exceeding simultaneously r_c and $2\pi/k_c$. Minimising the total CPU time through a proper choice of k_E would then lead to a $N^{10/7}$ rather than a $N^{3/2}$ scaling of the numerical effort with particle number N.

5. Discussion and conclusions

The topic of this article is the proper balance between accuracy and efficiency when cutting potentials. This is certainly an important, albeit somewhat neglected issue, Ref. [10] being a notable exception. The need for another discussion of cutting and smoothing potentials was recognised while writing a review on interatomic potentials [21], where it would have been inappropriate to suggest new functions and their properties. Of course, the necessity for good cut-off or smoothing functions is mitigated with advanced methods allowing the long-range tails of interactions to be summed effectively [36]. However, despite their efficiency, such methods are generally limited to pair interactions and are not always available.

This article emphasises that well-designed cutoff functions should generally outperform cut-and-shift potentials and that the discontinuities in cutoff functions at the inner radius deserve at least the same attention as at the outer cutoff radius, in particular for short-range potentials decaying more quickly than $1/r^3$. On the simple Lennard-Jones potential, it shows that the standard cutting procedure can be optimised in that errors on pressure and energy could be reduced by roughly 20 to 50%. While these gains are relatively minor, the incredibly large number of computations assuming Lennard-Jones potentials might make it worthwhile implementing the cutting-off procedure defined in this work.

Although the Wolf summation [13] was scrutinised in excellent earlier works [14,16,37], they remained somewhat vague regarding the asymptotic scaling of errors with the cutoff radius, the exact requirements for the homogeneity of matter for its convergence, and how to best pick the Ewald parameter k_E. Here, we confirmed that it is well behaved for most homogeneous systems [16], but that local charge neutrality should be obeyed more systematically than in a purely stochastic sense. Previous work on random-sphere packing of Na$^+$Cl$^-$ did not identify the marginal, sixth-digit error in the Coulomb energy, or potentially deemed it irrelevant [15]. The failure of the Wolf summation for deterministic charge separation, e.g. due to structural heterogeneity, had already been established earlier [15,37]. Moreover, we found $k_E = 1/\sqrt{a_0 r_c}$ as a kind of optimum choice in that it worked well for all investigated practical situations involving thermal or structural fluctuations. In fact, this choice appears is the 'sweet spot', similar to a critically damped case, where for the more general choice of $k_E = \kappa/(a_0^{1-\beta} r_c^\beta)$, the scaling of α_M with r_c crosses over from an 'overdamped' (convergence with small exponent) to oscillatory behaviour upon either an increasing β or decreasing κ. It shall also be noted that the f_{SW2}-modified Wolf summation allows one to keep intra-molecular potentials (essentially) unmodified without affecting its properties. Assuming a cutoff radius of 8 Å and a bond length of 2 Å, the smoothing function assumes a value of close to 0.97, even within a radius of zero.

Despite being problematic when the charge density is not strictly locally neutral, the Wolf summation can still be useful under such conditions. However, an exact summation of the k-space contribution would have to be made sporadically, in particular in Monte Carlo simulations, which, unlike molecular dynamics, does not benefit from a simultaneous update of all propagated degrees of freedom. The entire simulation between two such k-space evaluations would then constitute one large trial move so that the latest configuration after many steps using only the Wolf summation would be considered a trial configuration. That trial configuration could be accepted or rejected using, for example, the Metropolis algorithm [38], where the energy difference between new and old k-space contribution, ΔU_k, would enter the Boltzmann factor. While the rejection of such a time-intensive trial move is certainly regrettable, a reasonable scaling of the overall numerical effort with particle number N should be achievable. Of course, as is the case with the traditional Ewald summation as used in molecular dynamics [28], r_c would have to increase algebraically with N so that the absolute error induced by local 'Wolf moves' decreases with increasing N.

Acknowledgments

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

[1] Allen MP, Tildesley DJ. Computer simulation of liquids. Oxford: Oxford University Press; 2017.

[2] Frenkel D, Smit B. Understanding molecular simulation. San Diego: Elsevier; 2002.

[3] Stillinger FH, Weber TA. Computer simulation of local order in condensed phases of silicon. Phys Rev B. 1985;31(8):5262–5271.

[4] Tersoff J. New empirical model for the structural properties of silicon. Phys Rev Lett. 1986;56(6):632–635.

[5] Steinbach PJ, Brooks BR. New spherical-cutoff methods for long-range forces in macromolecular simulation. J Comput Chem. 1994;15(7):667–683.

[6] Patra M, Karttunen M, Hyvönen MT, et al. Molecular dynamics simulations of lipid bilayers: major artifacts due to truncating electrostatic interactions. Biophys J. 2003;84(6):3636–3645.

[7] Mattoni A, Colombo L, Cleri F. Atomic scale origin of crack resistance in brittle fracture. Phys Rev Lett. 2005;95(11):115501.

[8] Pastewka L, Pou P, Pérez R, et al. Describing bond-breaking processes by reactive potentials: importance of an environment-dependent interaction range. Phys Rev B. 2008;78(16):161402(R).

[9] Beck DAC, Armen RS, Daggett V. Cutoff size need not strongly influence molecular dynamics results for solvated polypeptides. Biochemistry. 2005 Jan;44(2):609–616.

[10] Toxvaerd S, Dyre JC. Communication: shifted forces in molecular dynamics. J Chem Phys. 2011;134:081102.

[11] Waibel C, Feinler MS, Gross J. A modified shifted force approach to the Wolf summation. J Chem Theory Comput. 2019;15(1):572–583.

[12] Daw MS, Baskes MI. Embedded-atom method: derivation and application to impurities, surfaces, and other defects in metals. Phys Rev B. 1984;29(12):6443–6453.

[13] Wolf D, Keblinski P, Phillpot SR, et al. Exact method for the simulation of coulombic systems by spherically truncated, pairwise r^{-1} summation. J Chem Phys. 1999;110(17):8254–8282.

[14] Fennell CJ, Gezelter JD. Is the Ewald summation still necessary? Pairwise alternatives to the accepted standard for long-range electrostatics. J Chem Phys. 2006;124(23):234104.

[15] Vlugt TJH, García-Pérez E, Dubbeldam D, et al. Computing the heat of adsorption using molecular simulations: the effect of strong coulombic interactions. J Chem Theory Comput. 2008;4:1107–1118.

[16] Rahbari A, Hens R, Jamali SH, et al. Effect of truncating electrostatic interactions on predicting thermodynamic properties of water–methanol systems. Mol Simul. 2018;45:336–350.

[17] Evjen HM. On the stability of certain heteropolar crystals. Phys Rev. 1932;39(4):675–687.

[18] Ewald PP. Die Berechnung optischer und elektrostatischer Gitterpotentiale. Ann Phys. 1921;369(3):253–287.

[19] Essmann U, Perera L, Berkowitz ML, et al. A smooth particle mesh Ewald method. J Chem Phys. 1995;103(19):8577–8593.

[20] Tuckerman M, Berne BJ, Martyna GJ. Reversible multiple time scale molecular dynamics. J Chem Phys. 1992;97(3):1990–2001.

[21] Müser MH, Sukhomlinov SV, Pastewka L. Interatomic potentials: achievements and challenges. 2022. Advances in Physics X (in press, DOI: 10.1080/23746149.2022.2093129)

[22] Baskes MI, Nelson JS, Wright AF. Semiempirical modified embedded-atom potentials for silicon and germanium. Phys Rev B. 1989;40(9):6085–6100.

[23] Jalkanen J, Müser MH. Systematic analysis and modification of embedded-atom potentials: case study of copper. Model Simul Mater Sci Eng. 2015;23(7):074001.

[24] Stenqvist B, Aspelin V, Lund M. Generalized moment correction for long-ranged electrostatics. J Chem Theory Comput. 2020;16:3737–3745.

[25] Sukhomlinov SV, Müser MH. Determination of accurate, mean bond lengths from radial distribution functions. J Chem Phys. 2017;146(2):024506.

[26] in 't Veld PJ, Ismail AE, Grest GS. Application of Ewald summations to long-range dispersion forces. J Chem Phys. 2007;127 (14):144711.

[27] de Leeuw SW, Perram JW, Smith ER. Simulation of electrostatic systems in periodic boundary conditions. I. Lattice sums and dielectric constants. Proc R Soc A: Math Phys Eng Sci. 1980;373 (1752):27–56.

[28] Perram JW, Petersen HG, De Leeuw SW. An algorithm for the simulation of condensed matter which grows as the 3/2 power of the number of particles. Mol Phys. 1988;65(4):875–893.

[29] Dove MT, Heine V, Hammonds KD. Rigid unit modes in framework silicates. Mineral Mag. 1995;59(397):629–639.

[30] Müser MH, Binder K. Molecular dynamics study of the α-β transition in quartz: elastic properties, finite size effects, and hysteresis in the local structure. Phys Chem Min. 2001;28(10):746–755.

[31] van Beest BWH, Kramer GJ, van Santen RA. Force fields for silicas and aluminophosphates based on ab initio calculations. Phys Rev Lett. 1990;64(16):1955–1958.

[32] Vollmayr K, Kob W, Binder K. Cooling-rate effects in amorphous silica: a computer-simulation study. Phys Rev B. 1996;54 (22):15808–15827.

[33] Herzbach D, Binder K, Müser MH. Comparison of model potentials for molecular-dynamics simulations of silica. J Chem Phys. 2005;123(12):124711.

[34] Schmahl WW, Swainson IP, Dove MT, et al. Landau free energy and order parameter behaviour of the α/β phase transition in cristobalite. Z Kristallogr. 1992;201(1-2):125–145.

[35] Baxter RJ. Ornstein–Zernike relation and Percus–Yevick approximation for fluid mixtures. J Chem Phys. 1970;52 (9):4559–4562.

[36] Isele-Holder RE, Mitchell W, Hammond JR, et al. Reconsidering dispersion potentials: reduced cutoffs in mesh-based Ewald solvers can be faster than truncation. J Chem Theory Comput. 2013;9:5412–5420.

[37] Cisneros GA, Karttunen M, Ren P, et al. Classical electrostatics for biomolecular simulations. Chem Rev. 2013;114(1):779–814.

[38] Metropolis N, Rosenbluth AW, Rosenbluth MN, et al. Equation of state calculations by fast computing machines. J Chem Phys. 1953;21(6):1087–1092.