Journal of Computational Physics 300 (2015) 592–604

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

Wavelets as basis functions to represent the coarse-graining potential in multiscale coarse graining approach

M. Maiolo [a,c], A. Vancheri [a], R. Krause [b], A. Danani [a,*]

[a] SUPSI, Department of Innovative Technology, Galleria 2, 6928 Manno, Switzerland
[b] USI, Institute of Computational Science, Via Buffi 13, 6906 Lugano, Switzerland
[c] ZHAW, Institut für Angewandte Simulation, Grüental, CH-8820 Wädenswil, Switzerland

ARTICLE INFO

Article history:
Received 17 February 2014
Received in revised form 13 May 2015
Accepted 24 July 2015
Available online 30 July 2015

Keywords:
Molecular dynamics
Coarse graining
MSCG
Multiresolution analysis
Wavelets
Sparsity

ABSTRACT

In this paper, we apply Multiresolution Analysis (MRA) to develop sparse but accurate representations for the Multiscale Coarse-Graining (MSCG) approximation to the many-body potential of mean force. We rigorously framed the MSCG method into MRA so that all the instruments of this theory become available together with a multitude of new basis functions, namely the wavelets. The coarse-grained (CG) force field is hierarchically decomposed at different resolution levels enabling to choose the most appropriate wavelet family for each physical interaction without requiring an a priori knowledge of the details localization. The representation of the CG potential in this new efficient orthonormal basis leads to a compression of the signal information in few large expansion coefficients. The multiresolution property of the wavelet transform allows to isolate and remove the noise from the CG force-field reconstruction by thresholding the basis function coefficients from each frequency band independently. We discuss the implementation of our wavelet-based MSCG approach and demonstrate its accuracy using two different condensed-phase systems, i.e. liquid water and methanol. Simulations of liquid argon have also been performed using a one-to-one mapping between atomistic and CG sites. The latter model allows to verify the accuracy of the method and to test different choices of wavelet families. Furthermore, the results of the computer simulations show that the efficiency and sparsity of the representation of the CG force field can be traced back to the mathematical properties of the chosen family of wavelets. This result is in agreement with what is known from the theory of multiresolution analysis of signals.

1. Introduction

Despite the tremendous increase in computer power and algorithms efficiency, an adequate investigation of many complex systems remains well beyond the present capabilities of conventional atomistic simulation methods. To overcome this problem, in the recent years, coarse-grained (CG) models have become very popular. In these models, groups of atoms are clustered into new CG sites and averaging over less important details allows to focus on essential features, consequently obtaining a significant jump in the accessible spatial/temporal scales.

* Corresponding author.
E-mail addresses: massimo.maiolo@zhaw.ch (M. Maiolo), alberto.vancheri@supsi.ch (A. Vancheri), rolf.krause@usi.ch (R. Krause), andrea.danani@supsi.ch (A. Danani).

http://dx.doi.org/10.1016/j.jcp.2015.07.039
0021-9991/

In the CG approach, a mapping operator (typically the center of mass or the center of geometry of a set of atoms) is used to define the location of each CG site in an atomistic system as a function of the positions of the atoms. Then the key issue is the choice of the properties used to link the models of different resolutions. Several CG techniques are based on a "bottom-up" approach which employs information from the atomistic model to construct the potential for a CG model of the corresponding system.

A straightforward and popular approach is the Boltzmann inversion [1], together with its extensions like iterative Boltzmann inversion and inverse Monte Carlo method [2,3]. Another class of CG methods uses variational approaches to determine in a systematic way the CG potential within a rigorous statistical mechanical framework, such as the Multiscale Coarse-Graining (MSCG) introduced by Izvekov and Voth [4,5].

The basic assumption of the MSCG method is that the atomistic and the coarse-grained canonical distribution functions are compatible. The compatibility condition between CG and all-atom (AA) models is obtained by requiring that the probability to find the CG system in a given configuration is equal to the probability to find the AA system in the subset of AA configurations corresponding to the CG reduction defined by the mapping operator [6–8].

The MSCG force field is determined by minimizing a defined variational residual which generates a linear least-squares problem. The resulting inverse problem is solved by choosing an appropriate basis set of functions. The numerical implementation of the variational principle in MSCG can become a hard demanding task. There are two main challenges that arise by solving the least-squares problem. First, with a uniform grid, the huge number of equations requires an enormous amount of storage memory and computational power. Secondly, especially for complex systems, a poor sampling can induce statistical noises in the construction of the CG potential or in the worst case a degeneracy of the solution. Lu et al. have discussed in great details different new algorithms aiming at enhancing the efficiency and the stability of the MSCG computations [9]. In particular, they proposed various dense and sparse matrix algorithms to speed up the calculations and to regularize ill-posed MSCG problems.

In the present paper, we introduce a rigorous framework for the use of the wavelet transform in the context of the multiresolution analysis (MRA) for implementing the MSCG variational principle. This approach represents an innovative way to address the problems related to the sparsity of the CG potential representation. In a slightly different context, Ismail et al. introduced a coarse-graining methodology based on the wavelet transform, as a method for sampling polymer chains [10].

Thank to the MRA, the CG force field is decomposed at multiple space-scales simultaneously through a sequence of nested vector spaces to reflect the actual physical processes underlying the observed behavior as closely as possible. Moreover, the wavelet transform produces excellent results in signal denoising [11,12]. By shrinking or suppressing the wavelet coefficients under a predetermined threshold we are able to reduce the noise strongly.

The paper is organized in the following manner. Section 2 is devoted to the MSCG theory. In Section 3, wavelets as new basis functions are introduced. In Section 4, the implementation of the wavelet-based MSCG approach and the demonstration of its accuracy is discussed using two different condensed-phase systems, namely liquid water and methanol. Simulations of liquid argon were performed to test different wavelets properties. Finally, the last two Sections 5 and 6 contain discussion and conclusions, respectively.

2. The MSCG method: background

The first step in the construction of a CG force field is the reduction of the degrees of freedom of the atomistic system by means of a map that groups atoms into CG sites generally called beads. Let $\mathbf{r} = \{\mathbf{r}_1, \ldots, \mathbf{r}_n\}$ be the Cartesian coordinates of an instantaneous configuration of our atomistic model, composed by n atoms, and let $\mathbf{R} = \{\mathbf{R}_1, \ldots, \mathbf{R}_N\}$ be the corresponding state of our CG model composed by N CG sites. The configuration of the CG system is specified by a linear mapping operator $\mathbf{M}(\mathbf{r}) = \{\mathbf{M}_1(\mathbf{r}), \ldots, \mathbf{M}_N(\mathbf{r})\}$ of the form

$$\mathbf{M}_I(\mathbf{r}) = \sum_{i=1}^{n} c_{Ii}\mathbf{r}_i = \mathbf{R}_I \quad \text{for } I = 1, \ldots, N. \tag{1}$$

The CG site I is composed by all the atoms i such that $c_{Ii} \neq 0$. The numerical value of the coefficients is typically chosen so that $\mathbf{R}_I$ is either the center of mass or the geometrical center of the bead I. A basic assumption of the CG approach is that the atomistic and the coarse-grained canonical distribution functions are compatible. Equating AA and CG canonical distribution functions brings to the following effective potential U^{CG} of the CG system:

$$U^{\mathrm{CG}}(\mathbf{R}) = -\beta^{-1} \ln \int e^{-\beta U(\mathbf{r})} \delta\big(\mathbf{R} - \mathbf{M}(\mathbf{r})\big) \mathrm{d}\mathbf{r}\,. \tag{2}$$

It is also called potential of mean force (PMF) and represents the free energy of the atomistic system conditioned to the surface $\mathbf{M}(\mathbf{r}) = \mathbf{R}$. It is defined only up to an arbitrary additive constant. The PMF defines the CG force field through the relation:

$$\mathbf{F}_I(\mathbf{R}) = -\frac{\partial U^{\mathrm{CG}}}{\partial \mathbf{R}_I} \tag{3}$$

where $\mathbf{F}_I(\mathbf{R})$ is the force acting on the CG site I.

Let us associate a force vector $\{\mathbf{G}_I(\mathbf{R})\}$ to each CG configuration $\mathbf{R}$ and each CG site I, $\{I = 1, \ldots, N\}$. The MSCG method determines an optimal CG force field from the PMF by minimizing with respect to $\{\mathbf{G}_I(\mathbf{R})\}_{I=1}^{N}$ the following quadratic functional:

$$\chi^2[\mathbf{G}] = \frac{1}{3N}\left\langle \sum_{I=1}^{N} \left| \mathbf{F}_I^{\mathrm{AA}}(\mathbf{r}) - \mathbf{G}_I\big(\mathbf{M}(\mathbf{r})\big) \right|^2 \right\rangle_{\mathrm{AA}}, \tag{4}$$

where $\mathbf{F}_I^{\mathrm{AA}}(\mathbf{r})$ is the sum of the atomistic forces acting on the atoms which belong to the site I in the AA configuration $\mathbf{r}$. The average in equation (4) is computed with respect to the atomistic canonical distribution. The functional χ^2 measures to what extent the set of force vectors $\mathbf{G}_I$ matches the average of the total atomistic force acting on the CG sites. Hence, the minimization of the functional χ^2 is equivalent to a force matching condition.

In order to make the variational problem (4) numerically manageable, the functions $\mathbf{G}_I(\mathbf{R})$ are first represented as a sum of central, pairwise contributions:

$$\mathbf{G}_I(\mathbf{R}) = \sum_J f(R_{IJ})\hat{\mathbf{R}}_{IJ}, \tag{5}$$

where R_{IJ} is the distance between the CG sites I and J, $\hat{\mathbf{R}}_{IJ}$ is the unit vector directed from J to I and f is the component of the force between I and J at the distance R_{IJ} along the direction $\hat{\mathbf{R}}_{IJ}$. Let us notice that, if all the CG sites are identical, the functions f do not depend on the labels I and J. The unknown function f is then expanded with respect to a suitable set of basis functions $\{f_m\}$:

$$\mathbf{G}_I(\mathbf{R}) = \sum_J \sum_{m=1}^{\infty} \phi_m f_m(R_{IJ})\hat{\mathbf{R}}_{IJ}\,. \tag{6}$$

Defining the vector functions $\mathcal{G}_{m;I}(\mathbf{R}) = \sum_J f_m(R_{IJ})\hat{\mathbf{R}}_{IJ}$ and rearranging the sum in (6), we obtain the following expansion of the functions $\mathbf{G}_I(\mathbf{R})$:

$$\mathbf{G}_I(\mathbf{R}) = \sum_{m=1}^{\infty} \phi_m \mathcal{G}_{m;I}(\mathbf{R}). \tag{7}$$

The vector functions $\mathcal{G}_{m;I}(\mathbf{R})$ describe the contribution to the force $\mathbf{G}_I(\mathbf{R})$ acting on the site I deriving from the spatial arrangement $\mathbf{R}$ of the CG sites. The representation of the forces $\mathbf{G}_I(\mathbf{R})$ as a sum of pairwise contributions is approximated because, by construction, the CG forces contain in principles many-body contributions at all orders. Nevertheless this approximation is sufficiently accurate for the systems treated in this paper.

The expansion (7) is subsequently truncated in such a way that only a finite number M of basis force fields contribute. The functional (4) is minimized with respect to the coefficients $\{\phi_m\}_{m=1}^{M}$ of the truncated expansion. This procedure gives an approximate solution of the variational problem that can be improved by increasing the number of basis force fields. It is possible to prove that the values of the coefficients that minimize the functional χ^2 satisfies the following set of linear algebraic equations:

$$\sum_n A_{mn}\phi_n = b_m \tag{8}$$

where

$$A_{mn} = \frac{1}{3N}\left\langle \sum_{I=1}^{N} \mathcal{G}_{m;I}\big(\mathbf{M}(\mathbf{r})\big)\mathcal{G}_{n;I}\big(\mathbf{M}(\mathbf{r})\big) \right\rangle_{\mathrm{AA}} \tag{9}$$

$$b_m = \frac{1}{3N}\left\langle \sum_{I=1}^{N} \mathcal{G}_{m;I}\big(\mathbf{M}(\mathbf{r})\big)\mathbf{F}_I^{\mathrm{AA}}(\mathbf{r}) \right\rangle_{\mathrm{AA}}. \tag{10}$$

The entries of the matrices A and b measure the correlations between all couples of basis force fields and the correlation of each basis force field with the atomistic force field F^{AA}, respectively. The canonical averages appearing in (9) and (10) are numerically evaluated through a sampling of AA configurations in an atomistic MD simulation.

The system of linear equations (8) is solved using standard methods, such as Gaussian elimination or Singular Value Decomposition (SVD). The CG force field is finally reconstructed by substituting the solutions $\tilde{\phi}_m$ of (8) into the expansion (7) of the force field truncated at the term M.

The dimension and the condition number of the matrix A becomes a crucial point for complicated systems, where the bonded interactions are included and several CG types are present. It is evident that a sparse representation of the interactions is needed.

In this paper, we use wavelets for the functions $\{f_m\}$ in the expansion (6). In the first versions of the MSCG method, the authors proposed piecewise constant, linear and cubic spline basis sets to represent various CG interactions [7]. The drawback of this solution is that the CG force field expansion is carried into a uniform grid and consequently a large number of functions is necessary to approximate the interaction potential with acceptable accuracy. Basis functions with small support are needed especially in the regions with high force-field derivatives, where the reconstruction accuracy is of fundamental importance. As a result, in the cited paper, more than 200 basis functions were required to describe a typical nonbonded short-range interaction. For very complex systems, in presence of many different CG types, many nonbonded short-range interactions should be evaluated and consequently the equation matrix may grow dramatically. By increasing the number of grid points, it may occur that some basis function is supported by poor statistics which provides a noisy CG potential because of uncertainty in the coefficients or, in the worst case, a singular matrix. For examples, for nonbonded potentials, the core repulsive region is the zone with less sampled data and with a very dense grid, a function might be never evaluated and consequently the associated matrix becomes singular.

To optimize the number of basis functions a non-uniform grid with spacing depending on the details of the force field should be used. Das and Andersen proposed piecewise polynomials with on a non-uniform grid [13]. This construction, however, requires a prior knowledge of the CG potential shape in order to thicken the grid in the region where this is needed.

Other solutions were proposed, for example using higher-order polynomials, which was achieved using B-spline basis functions [14] or basis functions for particular requirements [13]. A further improvement was reached by introducing multiscale cubic polynomials [15] combined with a modified version of the elastic net method [16]. With this approach, the authors are able to consistently reduce the number of basis functions. Their construction led to a smoothed Haar wavelet but without the property of the multiresolution analysis.

In our approach, using the wavelets in the framework of the multiresolution analysis, the choice of the grid is automatically solved. In the core repulsive region, a finer grid is adaptively set to represent a function with a high derivative; on the other hand in the cut-off region, where the function is almost flat, a coarser grid gives a sparser representation with sufficient accuracy. This approach represents a further improvement compared to the penalty term in the multiscale construction proposed by Das and Andersen in Ref. [15].

3. Wavelets as new basis functions for the CG potential

In transform theories, problems with time and frequency resolution due to the Heisenberg principle exist regardless of the transform used. While Fourier analysis gives a perfect frequency resolution but loses time information, Multi-Resolution Analysis (MRA) captures both frequency and time information, formalizing the intuitive idea that a function $f(x)$ in $\mathbf{L}^2(\mathbb{R})$ can be analyzed at different scales in such a way that only relevant details are taken into account. Wavelets are one of the main concepts in MRA. The construction of a wavelet orthonormal basis uses *scaling functions* $\varphi(x)$ that have the following fundamental property: they can be written as linear combinations of $\varphi(2x-k)$, the half-scaled and $n/2$ translated versions of $\varphi(x)$. That is:

$$\varphi(x) = \sum_n h_0(n)\varphi(2x-n)\ . \tag{11}$$

This is called the *two-scale relation* for the scaling functions φ and the coefficients $h_0(n)$ are fixed real numbers specific to each different family of wavelets. If we fix that the function $\varphi_0(x)$ spans the space V_0, the functions

$$\varphi_{jn}(x) = 2^{-\frac{j}{2}}\varphi_0(2^{-j}x-n)\ , \quad n = 0\ldots 2^j-1$$

span the spaces V_j. The two-scale relation (11) generates a nested sequence of subspaces:

$$\mathbf{L}^2(\mathbb{R}) \supset \ldots \supset \mathbf{V}_{-2} \supset \mathbf{V}_{-1} \supset \mathbf{V}_0 \supset \mathbf{V}_1 \supset \mathbf{V}_2 \supset \ldots \tag{12}$$

Since $V_{j-1} = V_j \oplus W_j$, where W_j is the orthogonal complement of V_j respect to V_{j-1}, the wavelet $\psi \in W_j$ is defined through the functions $\varphi(2x-n)$:

$$\psi(x) = \sum_n h_1(n)\varphi(2x-n)\ , \tag{13}$$

where $h_1(n) = (-1)^n\ h_0(1-n)$. This is the *two-scale relation* for the wavelets. Equations (11) and (13) hold at any scale. A fixed number of coefficients $h_0(n)$ and $h_1(n)$ relates the scaling functions and wavelets at one resolution to the scaling functions at the next lower resolution. The scaling function $\varphi_0(x)$ is also called *father* wavelet while $\psi(x) \in W_0$ is called the *mother* wavelet. For this reason, the term wavelet includes both scaling and wavelet functions, although scaling functions are not proper wavelets. Equation (14) motivates formally this way of using the term wavelet. The sequence of closed spaces $\mathbf{V}_j$, $j \in \mathbb{Z}$ defines a multiresolution and provides discrete different levels of signal approximation associated with a system of superimposed grids at different space scales. Given that the spaces $\mathbf{V}_j$ and $\mathbf{V}_{j-1}$ contain approximations at scales 2^j and 2^{j-1} respectively, it is meaningful to interpret the functions belonging to the orthogonal complement $\mathbf{W}_j$ of $\mathbf{V}_j$ in $\mathbf{V}_{j-1}$ as

details at the scale 2^{j-1}. The equation $\mathbf{V}_{j-1} = \mathbf{V}_j \oplus \mathbf{W}_j$ can be recursively expanded to connect a more resolute level $j-m$ to a less resolute one j in the following way:

$$\mathbf{V}_{j-m} = \mathbf{V}_j \oplus \mathbf{W}_j \oplus \mathbf{W}_{j-1} \oplus \ldots \oplus \mathbf{W}_{j-m+1}. \tag{14}$$

Equation (14) means that an approximation of a function f at a scale 2^{j-m} can be obtained starting with an approximation at a less resolute scale 2^j by adding details at more and more resolute scales. Multiresolution theory therefore provides techniques to construct a basis for the spaces $\mathbf{W}_j$ as translated and scaled versions of the *mother* wavelet ψ, that is:

$$\psi_{jn}(x) = 2^{-j/2}\psi\left(2^{-j}x - n\right) \tag{15}$$

and we have the following representation of the function f, starting from a reference scale J:

$$f = \sum_{k=-\infty}^{+\infty} a_k \phi_{Jk} + \sum_{j=J}^{\infty} \sum_{k=-\infty}^{+\infty} d_{jk}\psi_{jk} \tag{16}$$

and the $\mathbf{L}^2(\mathbb{R})$ norm of a function f can be written as follows

$$\|f\|^2 = \sum_{k=-\infty}^{+\infty} |a_k|^2 + \sum_{j=J}^{\infty} \sum_{k=-\infty}^{+\infty} |d_{jk}|^2. \tag{17}$$

Thus the squares $|d_{jk}|^2$ of the coefficients d_{jk} can be interpreted as a measure of the contribution of a detail at a scale 2^{j-1} at the position $x = k2^j$. This suggests a procedure of decimation of the basis functions using a chosen threshold suited to obtain a sparse representation of the force field.

Actually large-amplitude wavelet coefficients can detect and measure short high-frequency variations because they have narrow time localization at high frequencies. At low frequencies their time resolution is lower, but they have a better frequency resolution. In this way MRA analysis provides a frame where these opposite behaviors of the force field can be taken into account simultaneously in an effective way.

In molecular dynamic simulations, the nonbonded potentials are cut at a certain distance $r_{\max}$, beyond which the interactions between particles become negligible and can be set to zero. Moreover, due to the high repulsive core for short distances, MSCG formalism introduces a threshold $r_{\min}$ as the smallest value of the distance between CG beads where we can collect enough statistics at all-atom level. For these reasons, the CG force field should be considered as a signal on a bounded interval $[r_{\min}, r_{\max}]$ rather than on $\mathbb{R}$. Working with a force field defined on a limited interval forced us to adapt the multiresolution theory to signals defined on a bounded interval. Applying the wavelet transform, which was designed for an infinite length signal, directly to the CG force field leads to artifacts near the boundaries. To avoid such edge artifacts in the CG force field reconstruction, we implemented so-called edge wavelets [17]. To simplify the construction, the closed interval $I = [r_{\min}, r_{\max}]$ is finally mapped into $[0, 1]$.

4. Results

In this work we tested the wavelet based MSCG approach (WMSCG) on water and methanol bulk systems at the liquid state. We also present a systematic study of the WMSCG approach for a simple model of Lennard-Jones (LJ) particles, focusing on some wavelet family properties for the CG potential approximation. All classical MD simulations were performed using GROMACS [18] simulation package. The CG force field of the LJ system was reconstructed using Daubechies (db) and Symlets [19] (sym) from 2 to 16 on the interval, while liquid water and methanol were analyzed using db6 on the interval.

4.1. Liquid argon

Being a monoatomic system, the liquid argon enable us to map trivially each atom to a single CG site. In this way the AA system coincides with the CG system and the whole MSCG method can be regarded as a reverse engineering procedure for reconstructing the potential from the trajectories. Knowing the solution in advance (the CG potential coincides actually with the LJ potential used for generating the trajectories) we can compute exactly the reconstruction error obtained using different wavelets families in different operating conditions. This allowed to address several issues related to the optimization of the choice of the wavelets.

After equilibration, the system composed by 1000 argon atoms was simulated in the constant NVT ensemble for 5 ns using the leap-frog algorithm [20] with an integration timestep of 1 fs. The interactions between atoms were modeled by a Lennard-Jones potential [21] and a cut-off distance of 0.9 nm was used. The simulation was performed in a cubic box of $V = (3.587\ \text{nm})^3$ under periodic boundary conditions with a target temperature of 90 K maintained by a velocity-rescaling thermostat [22] and a coupling parameter of 0.1 ps. The coordinates and the forces for each atom were recorded every picosecond during the course of the trajectory to obtain 5000 configurations. Fig. 1 illustrates the force field reconstruction using four different wavelet families, namely db1, db2, db9 and db16 with respectively 0, 1, 8 and 15 vanishing moments. As we will see in Section 5, the number of vanishing moments is one of the most important properties that contributes to

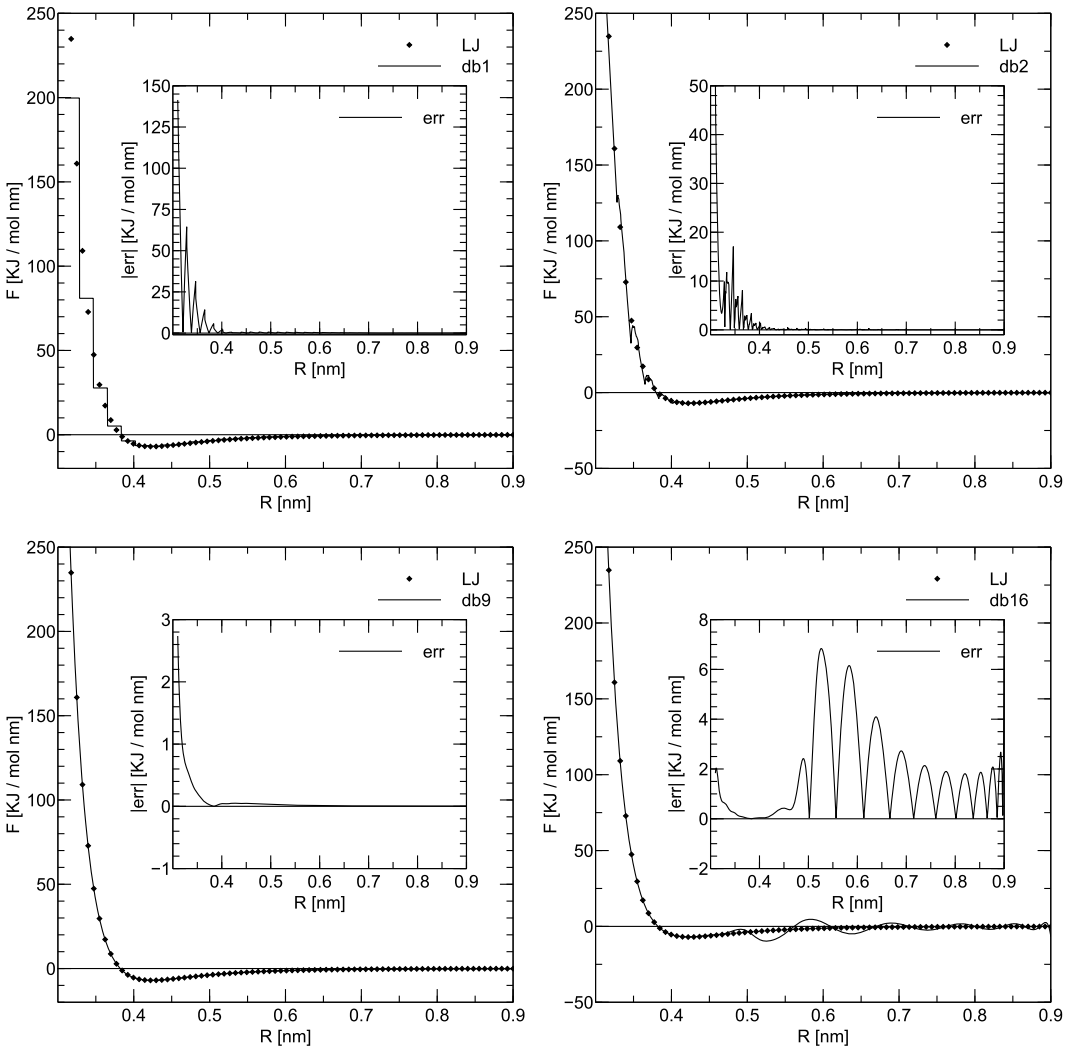

Fig. 1. Force profile reconstruction using scaling functions of the family db1 (top left), db2 (top right), db9 (bottom left) and db16 (bottom right). The solid line is the WMSCG force field approximation with 33 scaling functions at $J = -5$ without the correction provided by wavelets. Dots represent the exact solution. In the insert of each figure, the error between the WMSCG and the exact LJ curves is highlighted.

the sparsity of the CG force field. The force profile was obtained only with the contribution of scaling functions, at the scale $J = -5$. The $\mathbf{L}^2$ norm of the error in the reconstruction is depicted in the insert of each figure. It can be noticed that the accuracy in the reconstruction of the repulsive region improves with the increasing of vanishing moments however, in the region of cut-off the force profile obtained with db16 scaling functions oscillate, differently from the others. This behavior can be explained by the fact that the wavelet support length increases with its number of vanishing moments. Since the frequency content of the nonbonded interaction changes rapidly near to the repulsive core, some basis functions should take into account about opposite behaviors in its support. Daubechies 9 is among the four choices depicted the one that has the best compromise between vanishing moments and support length. The role played by the number of vanishing moments and the accuracy in the reconstruction was investigated in more details and the results are shown in Fig. 2. In this test we isolate the two regions of the interval with different frequency content to avoid opposite behaviors in the wavelet support due to a rapid change in spectral content. Thus, we divided the force field domain $I = [0.31, 0.90]$ nm into two non-overlapping intervals, namely $I = [0.31, 0.37]$ and $I = [0.37, 0.90]$ nm. In the first interval, high-order polynomials are required locally to approximate the potential, so wavelets with many vanishing moments perform better. The force field in the interval $I = [0.37, 0.90]$ nm is characterized by having low frequencies, in this region wavelets with relative few vanishing moments and consequently a shorter support size are more advisable for obtaining a sparser reconstruction. To perform this analysis we tested fifteen different symlets (sym) families (from 2 to 16) with 1 to 15 vanishing moments, respectively. For each family we have measured the approximation error as the $\mathbf{L}^2$ norm of the difference between the atomistic LJ and the wavelet representation of the CG force field. The $\mathbf{L}^2$ norm of the error $\|f^{AA} - f^{CG}\|^2$ as function of the wavelet family is depicted in Fig. 2. In both intervals the reconstruction error decreases as the number of vanishing moments increases.

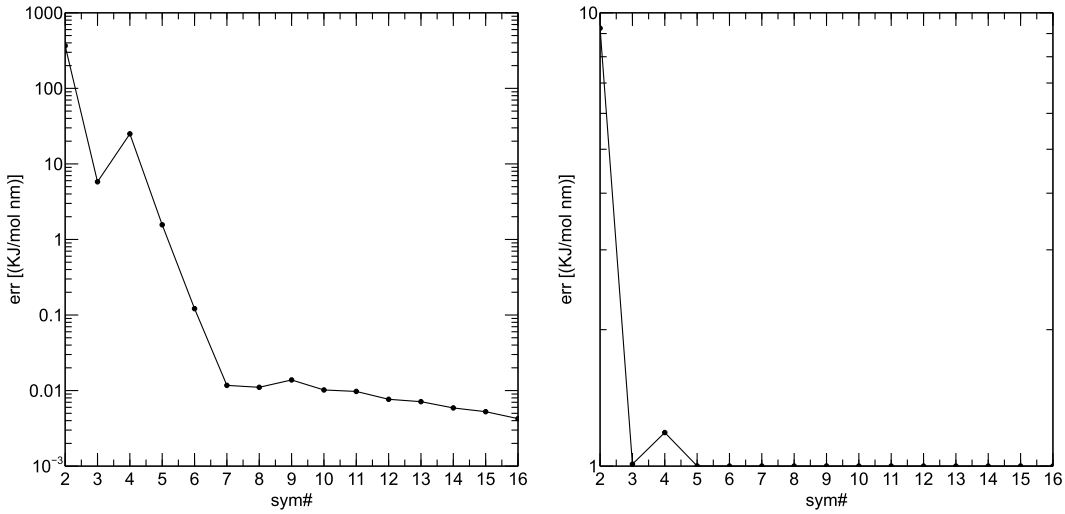

Fig. 2. Reconstruction error with different symlets wavelet functions in two non-overlapping bounded intervals. Left: error $\mathbf{L}^2$ norm in logarithmic scale interval $I = [0.31, 0.37]$ nm obtained with scaling functions and wavelets at $J = -6$. In this interval, it should be preferred basis functions orthogonal to high-order polynomials. Right: error $\mathbf{L}^2$ norm in logarithmic scale interval $I = [0.37, 0.90]$ nm obtained with scaling functions and wavelets at $J = -6$. In this interval the LJ functions is well represented already by low-order polynomials.

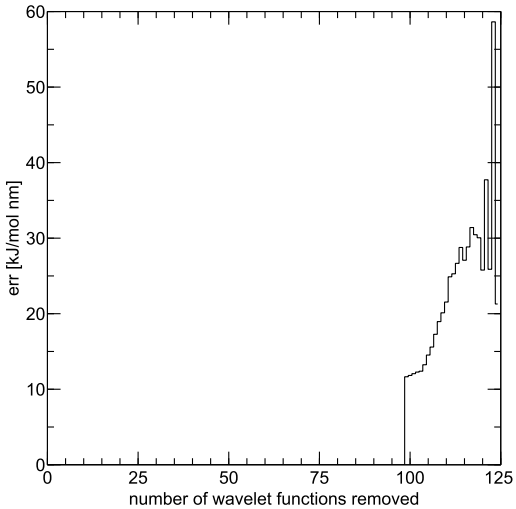

Fig. 3. Approximation error as a function of the number of basis functions. By removing one-by-one the wavelet basis functions sorted by ascending magnitude value the accuracy of the approximation is not affected until the last 25 more important basis functions are removed.

In the interval $I = [0.37, 0.90]$ nm having more than 4 vanishing moments do not improve the solution significantly, so in this interval it is preferable wavelets orthogonal to low-order polynomials because they have a shorter support. On the contrary, in the interval $I = [0.31, 0.37]$ nm, the higher the number of vanishing moments the more accurate the solution. The optimal wavelet family is the one that better satisfies the compromise between accuracy and sparsity. It is worth noticing that the FF in these two intervals has a different spectral content: in the left region very high frequencies are present, on the contrary the right part of the non-bonded potential contains only low frequencies. The best solution could be achieved by mixing wavelets with different features (dictionaries). The choice of a different wavelet family for the two region, i.e. in the left region a wavelet with many vanishing moments, whereas in the right intervals a wavelet with a short support size, would give a more sparse representation.

Another important parameter that should be carefully chosen is the threshold value under which the basis function coefficients are set to zero. To test how this value affects the solution we started representing the force field using db6 wavelets at the scales $J = -4, -5, -6$ obtaining 125 coefficients. Thereafter, we removed one-by-one the wavelet basis functions sorted by ascending magnitude value. After each coefficient suppression we computed the approximation error. The result is depicted in Fig. 3, we can notice that only removing the 25 more important wavelet functions we introduce an appreciable error in the reconstruction. This test suggest also that working with an orthonormal set of basis functions we could implement a stop criterion that, differently from the present approach, inserts one-by-one a new basis function.

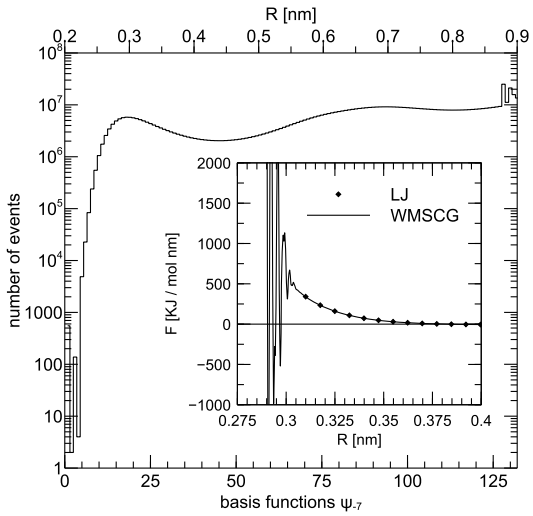

Fig. 4. Number of events caught by wavelet basis functions at $J = -7$ in the interval $I = [0.29, 0.9]$ nm. The region of core potential is rarely sampled, which can be seen as few events (distance between beads) in the most left side of the graph. In the insert: pairwise force profile between argon atoms as a function of distance. Solid line: WMSCG reconstruction using db6 with $J = -4, \ldots, -7$, dots represent the atomistic Lennard-Jones force field.

Moreover, it is important to notice that using orthonormal basis, the computation of the coefficients of a new candidate basis functions do not requires to re-compute all the other coefficients, it is sufficient to project the old solution in the new greater space. In Fig. 4 it is shown how the representation is affected by a poor statistics by setting a too small $r_{\min}$, in this case equal to 0.29 nm. This parameter should be set as small as possible to have an accurate representation in the repulsive region where the potential has a high derivative but great enough to catch in the sampled atomistic configurations sufficient events to stabilize the coefficients magnitude (9) and (10). The figure depicts in the x-axis the different basis functions and in the y-axis their sampled frequency. The result of setting a too small $r_{\min}$ is that the first dozen of wavelet functions are supported by few events which unveils in the uncertainty of the coefficients and consequently a reconstructed force field noisy and unreliable (shown in the insert).

4.2. Methanol

An all atoms model of 800 methanol molecules was simulated for 10 ns with an integration timestep of 1 fs and 2 fs, in AA and CG respectively. Interactions between atoms were modeled using the parameters from the OPLS-AA force field [23]. The simulation, after equilibration, was performed in the constant NVT ensemble in a cubic box of $V = (3.815 \text{ nm})^3$ under periodic boundary conditions with a target temperature of 300 K. Long-ranged and short-ranged non-bonded interactions were truncated at a cut-off distance of 1.00 nm. The coordinates and the forces for each atom were recorded every 2 ps during the course of the trajectory to obtain 5000 configurations. Each methanol molecule is replaced by a bead and the center of mass is used as a mapping operator. The mass of each site was defined as the total mass of the methanol molecule and zero net charge was assigned to each bead. The WMSCG potential was defined by a sum of short-ranged non-bonded central pair potential between identical CG sites. We opted for the wavelet family db6 on the interval $I = [0.29, 1.00]$ nm. From our simulations we see that this wavelet family satisfy the trade-off between sparsity and accuracy of the results globally in the interval. The resolution level of the approximation was $J = -5$ and only one level of wavelets was employed.

Note that, exploiting the MRA properties, the choice of reference V_0 is arbitrary: conceptually we can start at any resolution level but the choice has consequences on the sparsity of the representation. This means that if we analyze a signal using the combination of scaling function and wavelets, the scaling function by itself takes care of the spectrum otherwise covered by all the wavelets up to scale J. A final set of 44 basis functions was selected thresholding the coefficients at $T = 0.3$. The region of the force field in the interval $I =]0, r_{\min}]$ is extrapolated by means of a function like $f_{\text{ext}}(r) = \frac{A}{(r-B)^{13}} + C$. The wavelets at resolutions lower than $J = -5$ were all eliminated during the thresholding procedure. The resulting force field and the related methanol–methanol radial distribution function for both the atomistic and CG systems are depicted in Fig. 5. The RDF of the CG system was computed as the radial distribution function between the center of mass position of the atomistic methanol molecules. The radial distribution function of the CG model agrees excellently with that obtained from atomistic trajectory.

We simulated the CG methanol using the iterative Yvon–Born–Green method (YBG) introduced in [24] by Cho and Chu. The connection between YBG and MSCG have been discussed by Mullinax and Noid in [25] and [26]. This procedure enables to introduce information associated with three body correlations in order to obtain an optimal two-body approximation of the PMF. The results of our simulations showed that the corrections provided by the iterative YBG procedure was not

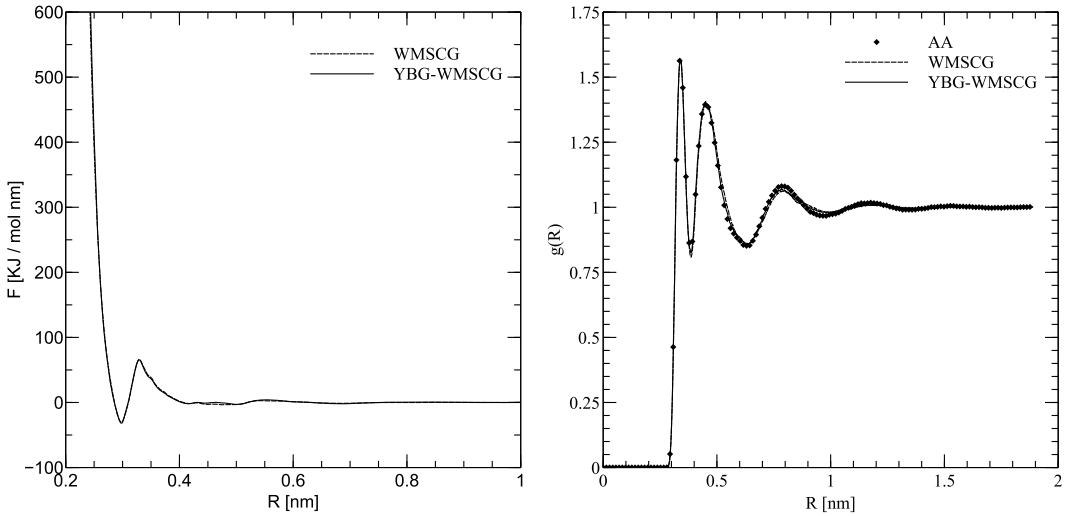

Fig. 5. Left: Pairwise force between CG methanol molecules as a function of interdistance. Dotted line: force profile applying the WMSCG approach. Solid line: force profile applying one loop of YBG method. Right: RDFs between the center of mass of methanol molecules. Dots represent the atomistic RDF, dashed line the RDF obtained with WMSCG approach and solid line applying one loop of YBG.

relevant for this system (see Fig. 5). This indicates that the many-body PMF for the one-site CG model is well represented only by a sum of pair-additive interactions between sites, without the corrections provided by the YBG method.

4.3. Water

An atomistic bulk system of 895 water molecules at 298 K was simulated using a TIP3P [27] water model to mimic the atomic interactions. The system, after equilibration, was simulated for 10 ns with an integration timestep of 2 fs and 5 fs, in the AA and CG respectively. The simulation was performed in the constant NVT ensemble in a cubic box of $V = (2.998\ \text{nm})^3$ under periodic boundary conditions. Long-ranged electrostatic interactions and short-ranged non-bonded interactions were truncated at 1.2 nm. The coordinates and the forces for each atom were recorded every 2 ps during the course of the trajectory to obtain 5000 configurations. The CG system is constructed replacing a single water molecule by a bead and both the center of mass (CoM) and the geometrical center (CoG) were tested as mapping operators. We observed that the CoM gives better results in agreement with previous work [28]. The mass of each site was defined as the total mass of the water molecule and zero net charge was assigned to each bead. The wavelet based MSCG interaction potential is defined by a sum of short-ranged non-bonded central pair potential. For the reconstruction of the CG force field the wavelet family db6 on the interval $I = [0.24, 1.20]$ nm was employed. We opted for such interval to have enough statistics that support each basis functions. The starting level of approximation (resolution) was $J = -5$ and only one level of wavelets was required to have a sufficient accurate reconstruction. To cut out the contribution of unnecessary wavelets a threshold value $T = 0.3$ was set and after the filtering of the coefficients we ended up with 44 basis functions. In the interval $I =]0, r_{\text{min}}]$ the CG force field is extrapolated using the function $f_{\text{ext}}(r) = \frac{A}{(r-B)^{13}} + C$, where the values of the parameters A, B and C were chosen in order to have continuous connection between the analytical function and the WMSCG force field.

To account for three body correlations, after the first CG simulation we ran a single loop of the YBG method previously introduced. The force field and the relative RDFs computed with and without YBG are compared in Fig. 6. From the computed RDFs one evinces that at least one loop of YBG is necessary in order to account for the polar properties of the water molecule. Without the information provided by the YBG method, the CG two-body potential badly reproduce the correct atomistic radial distribution function. This inability is in particular evident in the first solvation shell and may be caused by the angular anisotropy of water. With the contributions provided by one iteration of the YBG procedure, the WMSCG results can better represent the atomistic results, which can be seen in the improvement of the structure of the first solvation shell.

5. Discussion

In this paper we have proposed a new, general and systematic approach to the choice of basis functions for the representation of the CG force field in the frame of MSCG. The method is based on the use of wavelets and related multiresolution systems in order to represent the CG force field. Our method exploits the ability of wavelets and related MR systems to approximate functions efficiently, that is with as few coefficients as possible. A sparse representation of the force field is obtained if the CG potential has few non-negligible wavelet coefficients at the fine-scale (high-resolution). Fig. 3 shows that wavelet bases are particularly suited for their ability to efficiently approximate functions with few non-zero coefficients. The

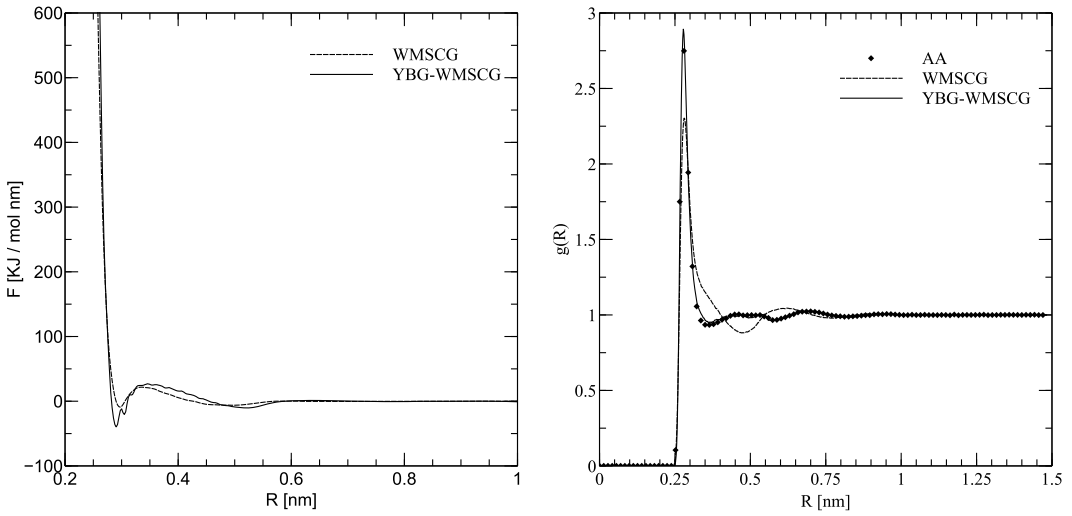

Fig. 6. Left: Pairwise force between CG water molecules as a function of inter-distance. Dotted line: force profile applying the WMSCG approach. Solid line: force profile applying one loop of YBG at the WMSCG approach. Right: RDFs between the center of mass of water molecules. Dots represent the atomistic RDF, dashed line the RDF obtained with WMSCG approach and solid line applying one loop of YBG.

simulations show indeed that with only about 25 basis functions we are able to approximate a nonbonded potential. Scaling functions in V_j synthesize the wavelets representation at resolutions $k > j$ (equation (14)). This MRA property allows to start the signal decomposition/reconstruction at an arbitrary resolution level with the contribution of scaling functions adding the wavelets details at lower scales. In principle, although the resolution level from which we start our FF reconstruction can be chosen arbitrary, in wavelets transform on $]0, 1[$ it is more advisable to avoid starting the decomposition of non-bonded potentials with global scaling and wavelet functions. Indeed the presence in the force field of a repulsive core on one side and a vanishing interaction on the other side, generates some difficulties with global functions. A global function is associated with a large coefficient because its support size extends to whole interval and it is "forced" to adapt to the repulsive core (see Ref. [17]). But this requires a cancellation mechanism in the region of the cut-off, where the force field tends to zero, which is provided by wavelets. For this reason the better results in terms of sparsity were reached with a low value of J, the choice made in the paper. The fact that we obtained a sparse representation could be explained by the fact that scaling functions acted as a low-pass filter in the signal and alone were able to catch 90% of the features of the force field. On the one hand the fact that we could reach a sparse representation even making use of a "too refined" description at the level of scaling functions is a further evidence that the method works well in giving sparse representations of the force field (e.g. box functions and cubic polynomials are not able to perform the same way). On the other hand the performances of the approach could strongly be improved if the problem with the global functions would be solved in a satisfactory way. A first evidence in this sense could be obtained studying systems were bonded interactions without a repulsive core are present. In this systems the CG force field could be optimally represented starting with global functions. In the case of non-bonded interactions the problem could be perhaps circumvented using multiresolutions on the interval based on rational functions that could be especially suited to capture the shape of LJ-like contributions. Another reason to start with scaling functions at a fine scale is that the submatrix A_{vv} with only the cross correlations between scaling functions has a better condition number that the submatrix A_{ww} with the wavelets correlations. This observation can be explained by the fact that first, scaling functions are better supported by the data and second, scaling functions produce an average reconstruction (low-pass filter) and consequently dumping high frequency. This suggest that in order to obtain a numerical more stable solution it is better to start with a resolution at a low scale.

We tested our approach with three simple models, namely argon, water and methanol. In the argon system there is an identification of the CG site with the AA site and no degree of freedom is integrated out. So the result has no added entropic effects and the outcome of the WMSCG is a wavelet representation of the atomistic non-bonded pairwise forces. Therefore the exact CG force field is known and a more reliable comparison between different choices of multiresolution is possible, allowing to test rigorously some property of MRA. Because of the simple nature of the model system, it is possible to make some a priori statements about the expected CG force field, which can be a guide for the interpretation of the numerical results. Despite the simplicity of the model, it provides a nontrivial test of the basis functions and we could investigate the role played by different features of the chosen multiresolution systems. Water and methanol systems were also tested to investigate the efficacy of the wavelets to provide a sparse representation of the CG potential. For both, we ended up with 44 basis functions which correspond to a strong improvement compared to Ref. [15]. We also observed that in the case of water system, the application of YBG method is necessary in order to take into account the three-body correlations. For both systems, CG and AA RDF agree excellently.

One of the most relevant aspects to be taken into account in using wavelets as basis functions is that, the ability of a wavelet to produce a sparse representation of the force field, depends strongly on some of its properties, mostly on the regularity, the number of vanishing moments and the support size. Some of the wavelet properties are competing against the other, so one has to find a trade-off. One of the most important wavelet properties, in the contest of sparse signal representation, is its number of vanishing moments.

The number of vanishing moments measures the rate of decay of a functions toward infinity. The higher the number of vanishing moments of a wavelet, the higher the degree of the polynomials that are orthogonal to the wavelets. Hence, the wavelet expansion of a signal that is locally well approximated by a polynomial of degree $n \leqslant p$ at the scale J will have small wavelets coefficients at resolutions $j > J$. Actually, keeping the approximation error constant, if we choose a wavelet family with more vanishing moments we obtain a sparser representation. Furthermore the number of vanishing moments is related to the smoothness of the wavelet: for important families of multiresolution systems such as Daubechies, the wavelet regularity generally increases with the number of vanishing moments. Smoothness provides numerical stability, so better quality CG potential representations are obtained with wavelets that are continuously differentiable. Hence, increasing the number of vanishing moments, will lead, at least for some important classes of multiresolution systems, to a double advantage: a more sparse representation and, at the same time, a higher quality of the reconstructed CG force field. But let us emphasize that, though the number of vanishing moments and the regularity of orthogonal wavelets are related, it is the number of vanishing moments and not the regularity that affects the amplitude of the wavelet coefficients at fine scales. Consequently, as far as the sparsity of the representation is a concern, we should concentrate our attention in the selection of the optimal wavelet on the number of vanishing moments. We investigated the role played by the vanishing moments of different wavelet families (see Figs. 1 and 2). Given the previous discussion, one could expect to observe that a high number of vanishing moments correlates systematically with a sparser representation of the force field. This behavior is indeed observed in the liquid argon system, specifically in the region of the repulsive core. In fact, db16, which has a greater number of vanishing moments with respect to db1, gave better results. Hence our results support the idea that the nonbonded CG potential needs more vanishing moments in the repulsive region where the potential is approximated locally by high-order polynomials. But in the cut-off region we observe an opposite behavior: less vanishing moments and consequently a shorter support size is preferred to obtain a more compact representation.

This result can be explained observing that the frequency content of the CG force field varies strongly with the distance r. Regions around the core potential with $r \gtrsim r_{\min}$, where the signal varies strongly with the distance r, contains high frequency components, whereas the region near the cut-off contains low frequencies. In general, if a signal $f(t)$ has an isolated singularity at t_0 and if t_0 is inside the support of $\psi_{j,n}(t)$, then $\langle f, \psi_{j,n} \rangle$ may have a large amplitude. If ψ has a compact support of size K, at each scale 2^j there are K wavelets $\psi_{j,n}$ whose support includes t_0. In order to minimize the number of high amplitude coefficients in presence of singularities we should choose wavelets with a short support size of ψ. Unfortunately the requirement to minimize the size of the support is in contrast with the requirement to maximize the number of vanishing moments. The optimal wavelet family is the one that better satisfies the compromise between the length of the support and the number of vanishing moments and the regularity. Daubechies wavelets, among the wavelet families, are optimal in the sense that they have the minimal support for a prescribed number of vanishing moments. Wavelets compactly supported in space domain are not bandlimited in frequency domain, however most of them have a rapid fall-off so that they can be considered bandlimited. For the Heisenberg's uncertainty principle, the better the space resolution the poorer the frequency resolution. This discussion suggests that an optimal choice of basis functions is the result of a trade-off between different conditions imposed by the force field. We observed that db2 seems to perform better than db16 in the force field tail. In $r \cong 0.4$ nm an abrupt change in frequency content occurs that has similar effects as the presence of a singularity. This resembles border effects observed when a signal defined on an interval is periodically repeated. This effect could be enhanced increasing p and therefore the support size of the scaling functions. Actually, the scaling functions sharing their support with the two regions of the force field could induce an oscillatory behavior in the coefficients that needs cancellation through the contribution of wavelets at lower levels. Thus when choosing a particular wavelet, we face a trade-off between the number of vanishing moments and the support size. The support size of a function and the number of vanishing moments are a priori independent. However, the constraints imposed on orthogonal wavelets imply that if ψ has p vanishing moments then its support is at least of size $2p - 1$. If f has few isolated singularities, like the nonbonded CG force field at $r = 0$, and is very regular between singularities, we must choose a wavelet with many vanishing moments to produce a large number of small wavelet coefficients $\langle f, \psi_{j,n} \rangle$. On the other hand, in the region of cut-off a wavelet with a relative short support, and consequently few vanishing moments, is preferable to obtain a more compact representation. The optimal choice would be wavelet that better satisfy the compromise between accuracy and sparsity. These results suggest that a better choice would be the implementation of dictionaries so that the different families of wavelets could be mixed to build an even sparser CG force field representation. Let us finally observe that, from a computation complexity point of view, wavelets with a small support should be preferred.

Another issue in this paper concerns the signal denoising process. The denoising of the reconstructed force field has been carried on by thresholding the coefficients with an amplitude below a predetermined value. Thresholding the wavelets coefficients yields not only a representation with less parameters but also a denoised signal reconstruction. Moreover, the introduction of an energy-connected threshold provides a tool for globally influencing the approximation of the wavelets. The advantage of working with an orthonormal basis set of functions is the relationship between the threshold value and the $\mathbf{L}^2$ norm of the signal removed (this follows directly from the Parseval identity, see equation (17)). The $\mathbf{L}^2$ norm of the

removed coefficients give us a clear idea of the quantity of signal "energy" we are deleting. By increasing the threshold, the basis functions will be removed according to their importance, and a change of regime will be observed. The local support of the basis functions allows us to localize them both in spatial and in frequency domain and rejecting a particular basis will only affect its area of support. This important property enables an elegant control of the local level-of-detail of the approximation. If a system of basis functions is orthonormal, each coefficient is connected with the $\mathbf{L}^2$ norm of the function to be spanned in a simple way: the square of the modulus of the coefficient equals the amount of $\mathbf{L}^2$ norm carried by the corresponding basis function. This makes meaningful the introduction of a threshold below which the contribution of a basis functions is disregarded. In the approach of Das and Andersen, the use of a basis function is associated with a cost and the approach consists in optimizing an objective function by measuring costs and benefits associated with each contribution. We believe that our approach has the double advantage to be easier in the implementation and conceptually more intrinsic because it measures the usefulness of a contribution in terms of a meaningful property of the function, that is $\mathbf{L}^2$ (energy) norm. This represents a completely different approach compared to the penalty term of the objective function in the elastic net method [15]. In the latter approach, the penalty cost has not a physical meaning. It is worth noticing that it is very difficult to compare our approach and the one of Das and Andersen with the information at our disposal about their computational setting. Only a comparison related to the same approximation for the same reconstruction would be reliable.

A third important issue concerns the number of times a basis function is sampled in the evaluation of the coefficients when formulas (9) and (10) are used. The importance of the number of configurations needed by each basis function is depicted in Fig. 4. Poor statistics results in unstable coefficients and an unreliable force field reconstruction. In the previously proposed approaches, to improve the accuracy one has to increase the number of grid points. As a consequence, to avoid unsampled basis functions, one should extend the atomistic simulation trajectory which turns in a bigger matrix. Indeed, as the grid become thicker as higher the probability that a function is never supported by events. In a dense grid, it is sufficient that a single basis function is never sampled for the matrix to become singular. Never or rarely sampled basis functions cannot be just removed because this would open a hole in the reconstructed force field. In our approach, the fact that scaling functions and wavelets overlap each other at different scales, enables to remove wavelets contributions related to unsampled basis function avoiding singularity in the matrix. In the multiresolution analysis, one can start approximating the force field with few scaling functions at a coarse level, and than adding as many layers of wavelets which will add the details as needed. Before the inversion of the matrix, we can check the frequency by which each wavelet was sampled and suppress all those functions seldom sampled.

6. Conclusions

With this work we endow MSCG method with a new theoretical framework, the multiresolution analysis on bounded interval which allows further enhancements in particular in the research of an efficient, accurate and sparse representation of the CG force field. The advantage to frame MSCG in MRA theory is that a plenty of new instruments for the optimization of the basis functions become available. Hundreds of algorithms and ideas already published could be added taking advantage of a sound theory and an ascertain experience of a multitude of scientists in this field. Our results support the idea that the properties of the chosen multiresolution system (like e.g. the number of vanishing moments and the support size) strongly affect the efficiency of the WMSCG approach. Indeed, given a set of criteria, this approach makes possible a systematic investigation of the basis functions optimization using all the instruments available in the multiresolution framework. Daubechies and Symlets wavelets used in this work are only two of the many possible choices. The introduction of the MRA framework allows to choose among a huge number of wavelet classes the one that, thank to its properties, gives the most compact representation. Nevertheless, in order to identify the most effective wavelet family, a further investigation about the properties of different families is required.

The wavelet transform reveals to be effective in approximating the CG potential with few non-zero coefficients. This is useful for data compression (sparse representation) and fast calculations. This is achieved by employing a hierarchical decomposition of the coarse-grained force field in a wavelets basis functions, in which the CG potential is analyzed at various scales and resolutions. This transform uses short windows at high frequencies providing an arbitrarily good time resolution and long windows at low frequencies yielding an arbitrarily good frequency resolution. Therefore it is particular suited for the characterization and manipulation of CG potentials whose frequency components strongly vary in space. Indeed, force fields are characterized by regions, such as the repulsive core, with high frequency components that require strongly localized and high frequency components and regions, for example near the cut-off distance where the force field tend to zero and only a few low frequency components are required. Multiresolution analysis provides a frame where these opposite behaviors of the force field can be taken into account simultaneously in a very effective way. The decomposition of the force field into a coarse approximation, obtained by the scaling functions, and the details, added by wavelets at different resolution levels, represent an innovative approach with respect to state-of-the-art to obtain a sparse CG force field representation. Variations of the thresholds of the wavelet coefficients allow the control of the error of the FF approximation. Moreover, the framework of the wavelet transform enables an elegant focus on the local level-of-detail of any particular region of interest within the data set.

In this paper, we introduced the wavelet-based MSCG approach. After the preliminary results presented here, further work will be devoted to the implementation of bonded CG interactions to extend the method to more complex systems like proteins, DNA and membranes.

Acknowledgements

This research was supported by “Ticino in rete” financed by DECS, Canton Ticino. Allocation for computer time at CSCS-Swiss Center of Scientific Computing are gratefully acknowledged. The authors thank Federica Trudu for the added-value provided with the simulations, Sergio Albeverio, Melissa Gajewski and Marco Deriu for carefully reading the manuscript.

References

[1] D. Reith, M. Puetz, F. Mueller-Plathe, Deriving effective mesoscale potentials from atomistic simulations, J. Comput. Chem. 24 (2003) 1624–1636.
[2] A.P. Lyubartsev, A. Laksonen, Calculation of effective interaction potentials from radial distribution functions: a reverse Monte Carlo approach, Phys. Rev. E 52 (1995) 3730–3737.
[3] A.K. Soper, Empirical potential Monte Carlo simulation of fluid structures, Chem. Phys. 202 (1996) 295–306.
[4] S. Izvekov, G.A. Voth, A multiscale coarse-graining method for biomolecular systems, J. Phys. Chem. Lett. 109 (2005) 2469–2473.
[5] S. Izvekov, G.A. Voth, Multiscale coarse graining of liquid state systems, J. Chem. Phys. 123 (2005) 134105.
[6] W.G. Noid, J.-W. Chu, G.S. Ayton, V. Krishna, S. Izvekov, G.A. Voth, A. Das, H.C. Andersen, The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models, J. Chem. Phys. 128 (2008) 244114.
[7] W.G. Noid, J.-W. Chu, G.S. Ayton, V. Krishna, S. Izvekov, G.A. Voth, A. Das, H.C. Andersen, The multiscale coarse-graining method. II. A rigorous bridge between atomistic and coarse-grained models, J. Chem. Phys. 128 (2008) 244115.
[8] A. Das, H.C. Andersen, The multiscale coarse-graining method. V. Isothermal–isobaric ensemble, J. Chem. Phys. 132 (2010) 164106.
[9] L. Lu, S. Izvekov, A. Das, H. Andersen, G.A. Voth, Efficient, regularized, and scalable algorithms for multiscale coarse-graining, J. Chem. Theory Comput. 6 (2010) 954–965.
[10] A. Ismail, G. Rutledge, G. Stephanopoulos, Topological coarse-graining of polymer chains using wavelet-accelerated Monte Carlo. I. Freely-jointed chains, J. Chem. Phys. 122 (23) (2005) 234901.
[11] D. Donoho, De-noising by soft-thresholding, IEEE Trans. Inf. Theory 41 (3) (1995) 613–627.
[12] D. Donoho, I. Johnstone, Adapting to unknown smoothness via wavelet shrinkage, J. Am. Stat. Assoc. (1995) 1200–1224.
[13] A. Das, H.C. Andersen, The multiscale coarse-graining method. III. A test of pairwise additivity of the coarse-grained potential and of new basis functions, J. Chem. Phys. 131 (2009) 034102.
[14] L. Lu, G.A. Voth, Systematic coarse-graining of a multicomponent lipid bilayer, J. Phys. Chem. B 113 (2009) 1501.
[15] A. Das, H.C. Andersen, The multiscale coarse-graining method. VIII. Multiresolution hierarchical basis functions and basis function selection in the construction of coarse-grained force fields, J. Chem. Phys. 136 (2012) 194113.
[16] J. Friedman, T. Hastie, H. Hölfling, R. Tibshirani, Pathwise coordinate optimization, Ann. Appl. Stat. 1 (2) (2007) 302–332.
[17] S. Bertoluzza, S. Falletta, Building wavelets on]0, 1[at large scales, J. Fourier Anal. Appl. 9 (2003) 261–288.
[18] E. Lindahl, B. Hess, D. Spoel, GROMACS 3.0: a package for molecular simulation and trajectory analysis, J. Mol. Model. 7 (5) (2001) 306–317.
[19] I. Daubechies, Ten Lectures on Wavelets, CBMS/NSF Series in Applied Math., vol. 61, SIAM, 1992.
[20] R.W. Hockney, S.P. Goel, J. Eastwood, Quiet high-resolution computer models of a plasma, J. Comput. Phys. 14 (1974) 148–158.
[21] A. Rahman, Correlations in the motion of atoms in liquid argon, Phys. Rev. 136 (2A) (1964) A405–A411.
[22] G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling, J. Chem. Phys. 126 (2007) 014101.
[23] W.L. Jorgensen, D.S. Maxwell, J. Tirado-Rives, Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, J. Am. Chem. Soc. 118 (1996) 11225–11236.
[24] H.M. Cho, J.W. Chu, Inversion of radial distribution functions to pair forces by solving the Yvon–Born–Green equation iteratively, J. Chem. Phys. 131 (2009) 134107.
[25] J.W. Mullinax, W.G. Noid, Generalized Yvon–Born–Green theory for molecular systems, Phys. Rev. Lett. 103 (2009) 198104.
[26] J.W. Mullinax, W.G. Noid, Generalized Yvon–Born–Green theory for determining coarse-grained interaction potentials, J. Phys. Chem. C 12 (114) (2010) 5661–5674.
[27] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935.
[28] S. Izvekov, M. Parrinello, C.J. Burnham, G.A. Voth, Effective force field for condensed phase systems: a new method for force-matching, J. Chem. Phys. 120 (2004) 10896–10913.