

Relative Resolution: A multipole approximation at appropriate distancesAviel Chaimovich ^{1,*}, Kurt Kremer,² and Christine Peter³¹Max Planck Institute of Colloids and Interfaces, Potsdam 14476, Germany²Max Planck Institute for Polymer Research, Mainz 55128, Germany³Department of Chemistry, University of Konstanz, Konstanz 78457, Germany

(Received 31 May 2019; published 30 September 2019)

Recently, we introduced Relative Resolution (RelRes) as a hybrid formalism for fluid mixtures [Chaimovich *et al.*, *J. Chem. Phys.* **143**, 243107 (2015)]. The essence of this approach is that it switches molecular resolution in terms of relative separation: While nearest neighbors are characterized by a detailed fine-grained model, other neighbors are characterized by a simplified coarse-grained model. Once the two models are analytically connected with each other via energy conservation, RelRes can capture the structural and thermal behavior of various multicomponent and multiphase systems across state space. This current work is a natural continuation of our original communication. Most importantly, we present the comprehensive mathematics of RelRes, casting it as a multipole approximation at appropriate distances; the current set of equations technically applies for any arbitrary system in soft matter (e.g., water). Besides, we continue examining the capability of this multiscale approach in molecular simulations of various (nonpolar) uniform liquids, specifically examining a 2:1 mapping for dumbbell-like molecules, as well as a 6:1 mapping and a 6:2 mapping for butterflylike molecules. In turn, we exhaustively show that RelRes can successfully retrieve for these systems their static and dynamic behavior, given that the fine-grained and coarse-grained potentials are switched at the boundary between the first and second coordination shells, the location at which orientational correlations vanish. We finally conclude by discussing how RelRes is the inherent variant of the “cell-multipole” approach for molecular simulations and, thus, this multiscale framework is especially promising for studying biological systems.

DOI: [10.1103/PhysRevResearch.1.023034](https://doi.org/10.1103/PhysRevResearch.1.023034)**I. INTRODUCTION**

Over the past half of a century, invoking molecular simulations has become one of the most promising routes for studying soft matter. The conventional protocol is proficient for describing systems that just comprise small spatial and short temporal dimensions (i.e., $<10^{-6}$ m and $<10^{-6}$ s, respectively), but it is deficient in describing systems that also involve large spatial and long temporal dimensions (i.e., $>10^{-3}$ m and $>10^{-3}$ s, respectively); resolving such challenges is especially important for biological processes, whose particularities usually span orders of magnitude in scale. For overcoming this dimensionality issue, much attention has been given to enhancing the computational efficiency of molecular simulations, while ensuring that the phenomena of interest are still correctly retrieved. One route involves the intelligent manipulation of statistical mechanics, while designing sophisticated algorithms that target a specific aspect of a molecular system. For example, various successful strategies have been developed for improving the calculation of free energies [1–4], reaction coefficients [5–8], etc.

Our work revolves around another set of algorithms that has received much attention in recent decades: it is commonly called the multiscale approach. Rather than focusing on the calculation of a specific feature of a molecular simulation, multiscale algorithms aim at improving the computational efficiency of the entire system, while ideally capturing all of its static and dynamic behavior. Importantly, the main signature of all multiscale simulations is that they involve two systems, one constructed of detailed fine-grained (FG) models with many degrees of freedom (usually corresponding with atomistic coordinates), and one constructed of simplified coarse-grained (CG) models with few degrees of freedom (usually corresponding with gravitational centers). While many multiscale strategies exist that hybridize quantum-classical or discrete-continuum models, the emphasis of our work is on algorithms that combine FG and CG molecular models of Newtonian particles. We continue by discussing the two main routes for implementing this multiscale approach, which are the in-serial [9–16] and in-parallel [17–24] formats.

The in-serial multiscale methods generally focus on the FG and CG systems in successive order: In the usual case, they begin by postulating a particular set of FG models, and after performing a sufficient amount of molecular simulations of this detailed system, they optimize, based on a certain criterion, a respective set of CG models, continuing the remainder of their investigation only with molecular simulations of this simplified system. While in practice, there are many variants of this type of multiscale strategies [25,26], each of these can be derived via one of two comprehensive frameworks:

*aviel.chaimovich@mpikg.mpg.de

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

one stems in the relative entropy [9], while the other aims at “force matching” [10,11]. The former multiscale approach minimizes a functional which measures the logarithmic ratios between the FG and CG configurational probabilities [9]; it has been shown that the relative entropy underlies the uniqueness theorem of Henderson [12], which is the basis for the multiscale strategies that aim at capturing radial distributions between molecular pairs [13,14,27,28]. The latter multiscale approach minimizes a functional which measures the squared differences between the FG and CG instantaneous forces; it has been notably shown that “force matching” is equivalent with a generalized version of the Yvon-Born-Green formalism [15]. Besides, both of these approaches can retrieve the multibody potential of Kirkwood [16].

While the in-serial multiscale methods have generally achieved much theoretical success, they still have several computational challenges. On a fundamental level, it is agreed upon that no model can optimally transfer across state space (e.g., across temperature and density), while also concurrently representing all structural correlations and thermal properties of a given system [29,30]. This technically means that one must choose a particular aspect of a molecular system for optimization (e.g., a configurational probability or an instantaneous force), while having no guarantee of replicating any of its other behavior, and on top of this, this protocol must be repetitively performed at each state point of interest. Furthermore, on a very practical level, the most significant challenge with all of these multiscale strategies is the fact that a very undesirable computational step is always required: This step involves performing a molecular simulation of a particular FG model before even using a respective CG model; realize that, in general, if we can construct a molecular simulation of the former, we do not really need the latter.

The in-parallel multiscale methods generally contain the FG and CG information simultaneously in a unified system. A single molecular simulation contains both FG and CG models, with vital aspects described by the former, and trivial aspects described by the latter. Diverse strategies have been developed that follow this hybrid computational path, and we categorize them here in four main classes. The most straightforward approach is the “group-based” multiscale method: here, essential molecular groups (e.g., solute molecules) are described via FG interactions, and auxiliary molecular groups (e.g., solvent molecules) are described via CG interactions [17,18,31,32]. The next class of strategies can be thought of as the “time-based” multiscale method: a given molecular simulation spends some of its time as the FG system and some of its time as the CG system, with the results being recorded for the former and discarded for the latter [19,20]. A powerful variation of this approach actually exchanges among FG and CG replicas of the molecular simulation, which in turn enhances the overall efficiency of the computational procedure [21,22]. In either case, this hybrid approach is very convenient for overcoming various timescale barriers in a system of interest.

Perhaps the most successful class of the in-parallel multiscale methods has been the one which is commonly called Adaptive Resolution. With much research done on it for over a decade, this approach is a “one-body distance-based” multiscale method. In a single molecular simulation, adaptive

molecules switch their resolution in terms of absolute position [23,24,33,34]. For example, as molecules move around a system, they adopt a FG model if near to the origin and a CG model if far from the origin. In practice, Adaptive Resolution essentially constructs hybrid molecules that embody both FG and CG models, and in fact, the main difference between all variations of Adaptive Resolution has technically been the mixing rule for the hybrid interaction, which is expressed as a function of absolute position. In the original publication of Adaptive Resolution, Praprotnik *et al.* introduced the hybrid interaction as a linear combination of forces, and this has been the ideal choice for evolving Newtonian trajectories [23]; at the same time, Abrams presented an alternative route that switches between the two models in a stochastic manner [24]. Conversely, Ensing *et al.* formulated the hybrid interaction as a linear combination of energies [33]; in turn, Potestio *et al.* showed that such a Hamiltonian version of Adaptive Resolution naturally corresponds with many fundamental aspects of statistical mechanics [34]. On a practical level, Adaptive Resolution is especially useful if a given system has a specific region of special interest (e.g., a protein at the origin, with everything else being water), and as such, it has been already applied in several biological scenarios [35,36].

Very recently, we evolved another type of multiscale method, which we called Relative Resolution (RelRes) [37]. Inspired by Adaptive Resolution, this approach is a “two-body distance-based” multiscale method: In a single molecular simulation, hybrid molecules switch their resolution in terms of relative separation; in particular, molecules that are near neighbors (i.e., their pairwise distance is small) interact via FG potentials, and molecules that are far neighbors (i.e., their pairwise distance is large) interact via CG potentials [37]. Importantly, RelRes is the sole class of multiscale simulations which describes all molecules, for all times, with both FG and CG models; the resolution of a given molecule is always relative to its observer. While other strategies, reminiscent of RelRes, have been also developed [38,39], our formalism is unique in that it mathematically finds a natural connection between the FG and CG potentials. While Ref. [38] numerically parametrizes between the two models via “force matching,” Ref. [39] does not make a clear connection between the two models. Conversely in our case, we developed an analytical parametrization between the FG and CG models, which is just based on equating energy functions at an infinite limit [37]. Unlike those other strategies [38,39], we were consequently able to correctly retrieve across state space, the structural and thermal behavior of several fluid mixtures [37]. Besides, we showed that our hybrid approach can be considered as a generalized extension of established theories for uniform liquids, which assume a mean field for interactions beyond a certain distance [40–43].

The current work is the comprehensive continuation of our original communication [37]. Foremost, we present mathematically the generalized framework for RelRes. Specifically, in the spirit of the familiar multipole expansion for the Coulombic potential [44,45], we formulate an analogous Taylor series for an inverse law of an arbitrary power. We show that the zero-order term, which is sufficient for nonpolar systems, is identical with the succinct expression we derived in our initial publication; in consideration of polar systems,

the current work also presents the first and second terms of the Taylor series, both of which involve molecular orientation. By use of this multipole approximation, we introduce RelRes as a Hamiltonian that switches between the FG and CG models at an appropriate distance. Building on the computational results of our original work, we continue examining RelRes with molecular simulations of nonpolar systems. Importantly, we exhaustively examine an elementary system of dumbbell-like molecules, in which we map from two FG sites to one CG site. We also proceed with a system of butterflylike molecules, mapping from six to one sites or from six to two sites; this latter study is particularly relevant for oligomeric and polymeric scenarios which one may encounter in soft matter. Through these case studies, we demonstrate that the ideal switching distance between the FG and CG models is about the same as the location at which angular correlations between molecules vanish, and this is roughly equivalent with the transition between the respective primary and secondary coordination shells. Furthermore, we also show here that RelRes can correctly capture not just static behavior but also dynamic behavior. In summary, we reinforce the notion that RelRes can be a powerful multiscale tool for efficiently studying soft matter via molecular simulations.

II. THEORETICAL FOUNDATION

Contrary to our original publication [37], we present here our hybrid formalism in a reversed order. In the previous work, we started by defining RelRes, and we ended by parametrizing between the FG and CG potentials. In the current work, we start by parametrizing between the FG and CG potentials, and we end by defining RelRes. Note also that the current framework is the comprehensive one, applying for all molecular systems (the original publication is only applicable for nonpolar scenarios).

A. Topology and energy of a molecular pair

Most of the ensuing derivation is based on a Taylor series that solely considers a pair of molecules in vacuum (they are not necessarily identical). One can think of them in terms of a time slice in a molecular simulation with no boundaries. These thorough mathematics then readily define the parametrization between the FG and CG models, as well as the RelRes potential. We then conclude the multiscale framework by making an extension to a collection of molecules, which naturally represents a realistic liquid that can be studied in a molecular simulation.

In Fig. 1, we present an arbitrary configuration of a molecular pair. We label the gravitational centers by the Latin indices i and j , and we label the atomistic coordinates by the Greek indices μ and ν ; we assume that the coordinates map on the centers in the conventional way through their masses, but, in principle, other linear mappings may be employed as well. In our notation throughout, we denote n_i or n_j as the number of sites on a molecule i or j , respectively. The relative separation between the centers is \bar{r}_{ij} , and the relative separation between the coordinates is $\bar{r}_{\mu_i\nu_j}$. Furthermore, the distance from a certain atomistic coordinate to a respective gravitational center is $\bar{\Delta}_{\mu_i}$ or $\bar{\Delta}_{\nu_j}$. By Fig. 1, it is clear that

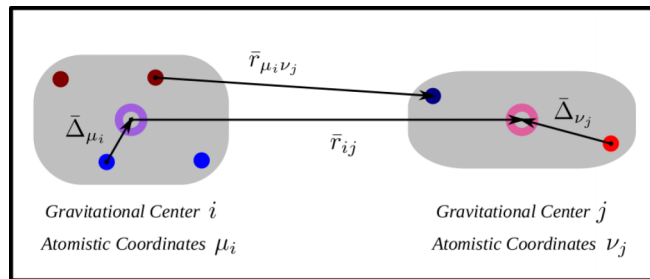


FIG. 1. A pair of molecules in vacuum. The gravitational centers are represented by hollow rings, and the atomistic coordinates are represented by replete disks. The left side is for molecule i with its μ_i sites, and the right side is for molecule j with its ν_j sites. The various colors of these mean that they can all be inherently different in terms of their interaction parameters. The gray shading does not have much of a physical meaning: it just helps us delineate an effective boundary of a molecule. The various arrows are distance vectors. \bar{r}_{ij} is the distance between centers i and j , and $\bar{r}_{\mu_i\nu_j}$ is the distance between coordinates μ_i and ν_j . Within each molecule, $\bar{\Delta}_{\mu_i}$ or $\bar{\Delta}_{\nu_j}$ is a vector from an atomistic coordinate, μ_i or ν_j , to a gravitational center, i or j , respectively.

these distances are related as

$$\bar{\Delta}_{\mu_i\nu_j} = \bar{r}_{ij} - \bar{r}_{\mu_i\nu_j} \quad (1)$$

and we define in this expression the following variable:

$$\bar{\Delta}_{\mu_i\nu_j} = \bar{\Delta}_{\nu_j} - \bar{\Delta}_{\mu_i}. \quad (2)$$

Note that we introduce here all distances in vector form, and their corresponding scalar magnitudes are realized by removing their “bars.” We also define dimensionless variables that will compact all of the ensuing mathematics:

$$\bar{\xi}_{\mu_i\nu_j} = \frac{\bar{\Delta}_{\mu_i\nu_j}}{r_{ij}}, \quad (3)$$

$$\cos \theta_{\mu_i\nu_j} = \frac{\bar{r}_{ij} \cdot \bar{\Delta}_{\mu_i\nu_j}}{r_{ij} \Delta_{\mu_i\nu_j}}. \quad (4)$$

With these definitions, together with some rearrangement, we derive this useful distance ratio through the dot product of Eq. (1) with itself:

$$\frac{r_{\mu_i\nu_j}}{r_{ij}} = \sqrt{1 - 2\bar{\xi}_{\mu_i\nu_j} \cos \theta_{\mu_i\nu_j} + \bar{\xi}_{\mu_i\nu_j}^2}. \quad (5)$$

The governing energy function here is defined as follows:

$$u_{ij}(\bar{r}_{ij}; \{\bar{\Delta}_{\mu_i}\}, \{\bar{\Delta}_{\nu_j}\}) = \sum_{\mu_i\nu_j} u_{\mu_i\nu_j}(r_{\mu_i\nu_j}). \quad (6)$$

Here, $u_{\mu_i\nu_j}$ is the intrinsic potential between atomistic coordinates μ_i and ν_j , and u_{ij} is the resultant potential between gravitational centers i and j ; the latter is obtained by the respective summation of the former. Notice that we are allowing for absolute nonuniformity in these systems, considering the indices on the potentials. Importantly, we assume that the former is isotropic, being exclusively a function of the scalar $r_{\mu_i\nu_j}$. Conversely, we expect that the latter is geometric, being chiefly a function of the vector \bar{r}_{ij} , as well as of the two sets of intramolecular distances $\{\bar{\Delta}_{\mu_i}\}$ and $\{\bar{\Delta}_{\nu_j}\}$. Notice that Eq. (6) limits our multiscale approach to pairwise interactions

between unique FG sites located in distinct CG sites; while an extension for multibody potentials may be possible, we do not explore such an option in this work. Besides, we ignore intramolecular energetics during most of the derivation: these are accounted for once we present the entire Hamiltonian of RelRes.

One of the focal assumptions in our analysis is that the potential can be cast as an inverse law of the pairwise distance with the exponent m :

$$u_{\mu_i v_j}(r) = \frac{c_{\mu_i v_j}}{r^m}. \quad (7)$$

In most scenarios, like with the Coulombic potential ($m = 1$), as well as with the Lennard-Jones (LJ) potential ($m = 12$ and 6), this inverse function is a natural choice. The proportionality coefficient $c_{\mu_i v_j}$ corresponds with the unique parameters involved in the interaction between sites μ_i and v_j , and it is usually known empirically for use in molecular simulations. We will later make an important assumption for $c_{\mu_i v_j}$ (i.e., we will express it as a product of two separate parameters). Besides, our entire derivation in this work can be readily generalized for the case in which the potential is actually a linear combination of several basis functions as the one given in Eq. (7); we briefly discuss such an extension at a much later stage of our mathematics.

The goal, in the remainder of the derivation, is to reformulate u_{ij} of Eq. (6) in terms of r_{ij} , together with variations of $\bar{\xi}_{\mu_i v_j}$ and $\cos \theta_{\mu_i v_j}$, as defined by Eqs. (3) and (4). Such a reformulation holds a promise for the efficient calculation of the energy of the pair in Fig. 1; obviously, computation of a single distance r_{ij} is preferable over the computation of $n_i n_j$ distances $r_{\mu_i v_j}$, which the summation of Eq. (6) requires. Besides, such a reformulation may involve approximations and, thus, we must make valid assumptions for $\bar{\xi}_{\mu_i v_j}$ and $\cos \theta_{\mu_i v_j}$ so that we maintain a fairly correct energy for this pair.

We now substitute Eq. (7) in Eq. (6), which yields the following:

$$u_{ij}(\bar{r}_{ij}; \{\bar{\xi}_{\mu_i v_j}\}, \{\cos \theta_{\mu_i v_j}\}) = \sum_{\mu_i v_j} \frac{c_{\mu_i v_j}}{r_{\mu_i v_j}^m}. \quad (8)$$

For clarity throughout the ensuing analysis, $\{\bar{\xi}_{\mu_i v_j}\}$ and $\{\cos \theta_{\mu_i v_j}\}$ are frequently omitted yet implied in the functionalities involved in Eq. (8); the same also goes for the “bar” on \bar{r}_{ij} . Consider now the radial distance which appears here together with the power of m ; we correspondingly exponentiate Eq. (5), and we present the ensuing reciprocal:

$$\left(\frac{r_{ij}}{r_{\mu_i v_j}}\right)^m = (1 - 2\bar{\xi}_{\mu_i v_j} \cos \theta_{\mu_i v_j} + \bar{\xi}_{\mu_i v_j}^2)^{-m/2}. \quad (9)$$

We can use this expression to substitute the computationally inexpensive r_{ij} for the computationally expensive $r_{\mu_i v_j}$ in Eq. (8). We consequently attain the following expression:

$$u_{ij}(r_{ij}) = \frac{1}{r_{ij}^m} \sum_{\mu_i v_j} [c_{\mu_i v_j} (1 - 2\bar{\xi}_{\mu_i v_j} \cos \theta_{\mu_i v_j} + \bar{\xi}_{\mu_i v_j}^2)^{-m/2}]. \quad (10)$$

B. Definition of the molecular pair in the infinite limit

Since the current consideration is in vacuum, we usually expect that $r_{ij} \gg \Delta_{\mu_i v_j}$ (i.e., $\bar{\xi}_{\mu_i v_j} \ll 1$); in other words, the separation between the molecules essentially approaches infinity. As such, we now begin a derivation for an energy function which we will name the molecular pair in the infinite limit (MPIL). Let us thus invoke the familiar multipole series:

$$(1 - 2\xi \cos \theta + \xi^2)^{-m/2} = \sum_{\aleph} [\xi^{\aleph} \mathbb{P}_{m, \aleph}(\cos \theta)]. \quad (11)$$

The function $\mathbb{P}_{m, \aleph}$ is defined by the appropriate derivative in terms of the Hebrew index \aleph :

$$\mathbb{P}_{m, \aleph}(\cos \theta) = \frac{1}{\aleph!} \left[\frac{\partial^{\aleph}}{\partial \xi^{\aleph}} (1 - 2\xi \cos \theta + \xi^2)^{-m/2} \right]_{\xi=0}. \quad (12)$$

Clearly, $\mathbb{P}_{m, \aleph}$ only depends on $\cos \theta$, with the functionality of ξ appearing as a power in Eq. (11). Notice also that the exponent m of the inverse law is just a parameter in $\mathbb{P}_{m, \aleph}$. Besides, if $m = 1$ (i.e., the Coulombic potential), Eq. (12) simply becomes the common definition of the Legendre polynomials.

Substituting Eq. (11) in Eq. (10) yields the following power series for our energy function:

$$u_{ij}(r_{ij}) = \frac{1}{r_{ij}^m} \sum_{\aleph} \left[\sum_{\mu_i v_j} [c_{\mu_i v_j} \bar{\xi}_{\mu_i v_j}^{\aleph} \mathbb{P}_{m, \aleph}(\cos \theta_{\mu_i v_j})] \right], \quad (13)$$

and it is particularly convenient to cast the above as

$$u_{ij}(r_{ij}) = \frac{1}{r_{ij}^m} \sum_{\aleph} c_{ij, \aleph}(\{\bar{\xi}_{\mu_i v_j}\}, \{\cos \theta_{\mu_i v_j}\}) \quad (14)$$

in which we obviously defined the following:

$$c_{ij, \aleph}(\{\bar{\xi}_{\mu_i v_j}\}, \{\cos \theta_{\mu_i v_j}\}) = \sum_{\mu_i v_j} [c_{\mu_i v_j} \bar{\xi}_{\mu_i v_j}^{\aleph} \mathbb{P}_{m, \aleph}(\cos \theta_{\mu_i v_j})]. \quad (15)$$

One may think of $c_{ij, \aleph}$ as an effective coefficient for a given configuration of Fig. 1; $\{\bar{\xi}_{\mu_i v_j}\}$ and $\{\cos \theta_{\mu_i v_j}\}$ are omitted yet implied in the functionality of $c_{ij, \aleph}$ throughout much of the derivation. In any case, these expressions can be quite useful for computational purposes: By summing over all sites μ_i and v_j , one can separately calculate the energetic coefficient of any \aleph term by Eq. (15), and by including as many \aleph terms in Eq. (14) as desired, one can have a valid approximation for the overall interaction between the molecular pair of Fig. 1. Of particular importance is the leading (nonvanishing) \aleph term: This is the energy function which we alluded to above, the one we call MPIL, and it is in fact one of the focal aspects of our framework.

1. A few terms in the Taylor series

By not even performing any differentiation, the $\aleph = 0$ term is readily determined by Eq. (12) as the following:

$$\mathbb{P}_{m, 0} = 1. \quad (16)$$

Conveniently, this term is always unity, regardless of the value of m . Plugging this in Eq. (15), we attain this expression:

$$c_{ij, 0} = \sum_{\mu_i v_j} c_{\mu_i v_j}. \quad (17)$$

Notice here that this is not dependent at all on the configuration of the molecular pair: this means $c_{ij,0}$ is a constant that can be calculated before any molecular simulation is performed. If this is nonzero, we can approximate the energy in Eq. (14) by the following:

$$u_{ij}(r_{ij}) \approx \frac{c_{ij,0}}{r_{ij}^m},$$

$$u_{ij}(r_{ij}) \approx \frac{\sum_{\mu_i\nu_j} c_{\mu_i\nu_j}}{r_{ij}^m}. \quad (18)$$

In the latter part of this expression, we obviously invoked Eq. (17). Notice the similarity between Eqs. (8) and (18): we start with an m inverse law in the former, and we end with an m inverse law in the latter. In Eq. (8), we sum over the many distances between the atomistic coordinates, with their intrinsic interaction coefficients, and in Eq. (18), we take in the single distance between the gravitational centers, with its resultant interaction coefficient. This is in effect the monopole term; once we introduce another assumption, we show that this is specifically equivalent with the monopole-monopole interaction.

The above is essentially the sole term that we derived in our previous work. We simply attained it by vanishing all bond lengths (i.e., $\Delta_{\mu_i\nu_j} = 0$) in Fig. 1; this is basically equivalent to setting all $\xi_{\mu_i\nu_j} = 0$ in Eq. (10), which in turn eliminates all $\cos\theta_{\mu_i\nu_j}$ from the energy function. The only discrepancy of Eq. (18) with the corresponding one in our original work is that in this one, we have the assumption of the inverse power for the interaction. We can temporarily remove this assumption by invoking Eq. (7) for Eq. (18), and we in turn get the same expression for the MPIL as we had in the original publication [37]:

$$u_{ij}(r_{ij}) \approx \sum_{\mu_i\nu_j} u_{\mu_i\nu_j}(r_{ij}). \quad (19)$$

Realize that this approximation is a special case of the MPIL; as mentioned before, the formal definition of the MPIL is the leading (nonvanishing) \aleph term in Eq. (13).

Importantly in our original work, we showed that if this expression is implemented with RelRes, it can successfully describe nonpolar liquids that are based on the LJ potential. Theoretically, this approximation shall also work with the Coulombic potential for ionic molecules that have a net charge. In fact for RelRes, we expect that Eq. (18) shall be adequate for any m , given that this monopole term does not vanish. Nevertheless, once the above summation for a certain molecular pair contains both positive and negative values for $c_{\mu_i\nu_j}$, there may be some issues; specifically, if all $c_{\mu_i\nu_j}$ in the summation of Eq. (17) cancel each other, $c_{ij,0}$ is strictly zero. The most obvious such case is the zwitterionic scenario, in which the molecules have Coulombic charges, yet they are overall neutral (e.g., water). Of course, the interaction between such polar pairs must be finite and, consequently, one must go beyond the monopole term of Eq. (18) in order to describe the corresponding energy, at least approximately. This is in fact the purpose of the ensuing mathematics. Note that the equations below may be also useful as correction terms for the interaction between the molecular pair, even if their monopole term is nonzero.

Obviously, all the terms of the Taylor series can be evaluated by successive differentiation as presented in Eq. (12). Here is the first-order term:

$$\mathbb{P}_{m,1}(\cos\theta) = [m(\cos\theta - \xi)(1 - 2\xi\cos\theta + \xi^2)^{-(m+2)/2}]_{\xi=0},$$

$$\mathbb{P}_{m,1}(\cos\theta) = m\cos\theta. \quad (20)$$

Consequently, the $\aleph = 1$ term of Eq. (15) is the following:

$$c_{ij,1}(\{\xi_{\mu_i\nu_j}\}, \{\cos\theta_{\mu_i\nu_j}\}) = m \sum_{\mu_i\nu_j} [c_{\mu_i\nu_j} \xi_{\mu_i\nu_j} \cos\theta_{\mu_i\nu_j}]. \quad (21)$$

Here is the second-order term:

$$\mathbb{P}_{m,2}(\cos\theta) = \frac{1}{2}[m(m+2)(\cos\theta - \xi)^2$$

$$\times (1 - 2\xi\cos\theta + \xi^2)^{-(m+4)/2}$$

$$- m(1 - 2\xi\cos\theta + \xi^2)^{-(m+2)/2}]_{\xi=0},$$

$$\mathbb{P}_{m,2}(\cos\theta) = \frac{m}{2}[(m+2)\cos^2\theta - 1]. \quad (22)$$

Consequently, the $\aleph = 2$ term of Eq. (15) is the following:

$$c_{ij,2^+}(\{\xi_{\mu_i\nu_j}\}, \{\cos\theta_{\mu_i\nu_j}\}) = \frac{m}{2} \sum_{\mu_i\nu_j} [c_{\mu_i\nu_j} \xi_{\mu_i\nu_j}^2 \cos^2\theta_{\mu_i\nu_j}], \quad (23)$$

$$c_{ij,2^-}(\{\xi_{\mu_i\nu_j}\}, \{\cos\theta_{\mu_i\nu_j}\}) = \frac{m}{2} \sum_{\mu_i\nu_j} [c_{\mu_i\nu_j} \xi_{\mu_i\nu_j}^2]. \quad (24)$$

For compactness, we introduced here this definition:

$$c_{ij,2}(\{\xi_{\mu_i\nu_j}\}, \{\cos\theta_{\mu_i\nu_j}\}) = (m+2)c_{ij,2^+} - c_{ij,2^-}. \quad (25)$$

Notice that if $m = 1$ (e.g., the Coulombic scenario), $\mathbb{P}_{m,1}$ and $\mathbb{P}_{m,2}$ become the familiar first and second Legendre polynomials, respectively.

Of course, these $c_{ij,\aleph}$ can be readily used for evaluating the energy function of Eq. (14): they may serve as correction terms for Eq. (18), and if $c_{ij,1}$ or $c_{ij,2}$ is actually the leading term of this multipole expansion, it can even become the sole approximation for the energy (i.e., the MPIL). Notice that unlike $c_{ij,0}$, these current expressions are functions of $\cos\theta_{\mu_i\nu_j}$, as well as of $\xi_{\mu_i\nu_j}$. Thus, while Eq. (18) is isotropic, involving just the scalar r_{ij} in its functionality, the energy which corresponds with other $c_{ij,\aleph}$ is geometric, involving also the vector $\bar{\Delta}_{\mu_i\nu_j}$ in its functionality. Therefore, for $\aleph \neq 0$, the coefficient must be computed at each step of a molecular simulation. As such, going beyond the monopole term requires significant computational cost, and it is only recommended if it is necessary (i.e., for polar systems in which these are the leading terms in the Taylor series).

2. An assumption for the proportionality coefficient

Focus now on the proportionality coefficient $c_{\mu_i\nu_j}$ of the inverse law of Eq. (7). Usually, such a parameter is expressed in terms of some mixing rules between two separate coefficients c_{μ_i} and c_{ν_j} of sites μ_i and ν_j , respectively. For example with the Coulombic potential, $c_{\mu_i\nu_j}$ is fundamentally the product of partial charges; an analogous product is also often used for the LJ potential. For the remainder of our work, we assume that the following equation holds:

$$c_{\mu_i\nu_j} = c_{\mu_i}c_{\nu_j}. \quad (26)$$

For completeness, this means that Eq. (8), becomes the following:

$$u_{ij}(\bar{r}_{ij}; \{\bar{\xi}_{\mu_i v_j}\}, \{\cos \theta_{\mu_i v_j}\}) = \sum_{\mu_i v_j} \frac{c_{\mu_i} c_{v_j}}{r_{\mu_i v_j}^m}. \quad (27)$$

It turns out that the assumption of Eq. (26) substantially facilitates the computational algorithm for RelRes, especially since a variant of $c_{ij,\aleph}$, for any \aleph , becomes a constant during a molecular simulation.

We begin with the zero-order term. Substituting Eq. (26) in Eq. (17), we attain the following equation:

$$c_{ij,0} = c_{i,0} c_{j,0} \quad (28)$$

in which we have introduced the definition of the monopole $c_{i,0}$ or $c_{j,0}$, for each molecule i or j , respectively:

$$\begin{aligned} c_{i,0} &= \sum_{\mu_i} c_{\mu_i}, \\ c_{j,0} &= \sum_{v_j} c_{v_j}. \end{aligned} \quad (29)$$

If both of these monopoles are nonzero, the MPIL is the following according to Eq. (14):

$$\begin{aligned} u_{ij}(r_{ij}) &\approx \frac{c_{i,0} c_{j,0}}{r_{ij}^m}, \\ u_{ij}(r_{ij}) &\approx \frac{(\sum_{\mu_i} c_{\mu_i})(\sum_{v_j} c_{v_j})}{r_{ij}^m}. \end{aligned} \quad (30)$$

Analogous with our comparison between Eqs. (8) and (18) above, notice the similarity between Eqs. (27) and (30): performing the approximation, we start with several products of inverse coefficients $c_{\mu_i} c_{v_j}$, and we end with one product of inverse coefficients $c_{i,0} c_{j,0}$.

Note now that the product assumption of Eq. (26) yields some convenience for the monopole energy since Eq. (30) is slightly more computationally efficient than Eq. (18): the former involves two summations over n_i and n_j parameters, while the latter involves one summation over $n_i n_j$ parameters. Regardless, all of these summations can be performed before a molecular simulation and, thus, they have negligible computational cost. The computational superiority introduced via Eq. (26) becomes very apparent once we move to the first and second terms of the multipole expansion.

We thus apply now the assumption of Eq. (26) for evaluating any $c_{ij,\aleph}$ [i.e., Eq. (15)], after the appropriate differentiation of Eq. (12). As an important part of the ensuing mathematics, we introduce the familiar definitions for the dipoles $\bar{p}_{i,1}$ and $\bar{p}_{j,1}$ of molecules i and j , respectively:

$$\begin{aligned} \bar{p}_{i,1} &= \sum_{\mu_i} [c_{\mu_i} \bar{\Delta}_{\mu_i}], \\ \bar{p}_{j,1} &= \sum_{v_j} [c_{v_j} \bar{\Delta}_{v_j}]. \end{aligned} \quad (31)$$

Moreover, we introduce components of the familiar definitions for the quadrupoles $\{\bar{q}_{i,2}, q_{i,0}\}$ and $\{\bar{q}_{j,2}, q_{j,0}\}$ of

molecules i and j , respectively:

$$\begin{aligned} \bar{q}_{i,2} &= \sum_{\mu_i} [c_{\mu_i} \bar{\Delta}_{\mu_i} \bar{\Delta}_{\mu_i}], \\ \bar{q}_{j,2} &= \sum_{v_j} [c_{v_j} \bar{\Delta}_{v_j} \bar{\Delta}_{v_j}], \end{aligned} \quad (32)$$

$$\begin{aligned} q_{i,0} &= \sum_{\mu_i} [c_{\mu_i} \Delta_{\mu_i}^2], \\ q_{j,0} &= \sum_{v_j} [c_{v_j} \Delta_{v_j}^2]. \end{aligned} \quad (33)$$

Casting a linear combination of these, $(m+2)\bar{q}_{i,2} - q_{i,0}\bar{\mathcal{I}}$ and $(m+2)\bar{q}_{j,2} - q_{j,0}\bar{\mathcal{I}}$, we retrieve the conventional definitions of the quadrupoles for molecules i and j , respectively; this is most apparent for $m=1$ (e.g. the Coulombic scenario). As in the usual case, we employ here generalized tensor algebra (e.g., $\bar{\mathcal{I}}$ is the identity tensor); the amount of ‘‘bars’’ on these parameters, together with their last index for extra emphasis, corresponds with the order of the tensor. Importantly, remember that the convenience of these parameters is that they can be calculated by just considering the topology of a single molecule on its own, rather than the entire configuration of the molecular pair of Fig. 1. For compactness, we introduce here a notation for the set of dipoles (i.e., $\{\bar{p}_{ij,1}\} = \{\bar{p}_{i,1}, \bar{p}_{j,1}\}$), as well as for the set of quadrupoles (i.e., $\{\bar{q}_{ij,2}\} = \{\bar{q}_{i,2}, \bar{q}_{j,2}\}$ and $\{q_{ij,0}\} = \{q_{i,0}, q_{j,0}\}$).

By invoking these parameters, we specifically show in Appendix A how the following expressions for $c_{ij,\aleph}$ can be derived. Here is the first-order term:

$$c_{ij,1}(\{\bar{p}_{ij,1}\}) = \frac{m}{r_{ij}} [c_{i,0}(\hat{r}_{ij} \cdot \bar{p}_{j,1}) - c_{j,0}(\hat{r}_{ij} \cdot \bar{p}_{i,1})]. \quad (34)$$

Here is the second-order term:

$$\begin{aligned} c_{ij,2^+}(\{\bar{p}_{ij,1}\}, \{\bar{q}_{ij,2}\}) &= \frac{m}{r_{ij}^2} \left[-(\hat{r}_{ij} \cdot \bar{p}_{i,1})(\hat{r}_{ij} \cdot \bar{p}_{j,1}) \right. \\ &\quad \left. + \frac{1}{2} c_{i,0}(\hat{r}_{ij} : \bar{q}_{j,2}) + \frac{1}{2} c_{j,0}(\hat{r}_{ij} : \bar{q}_{i,2}) \right], \end{aligned} \quad (35)$$

$$\begin{aligned} c_{ij,2^-}(\{\bar{p}_{ij,1}\}, \{q_{ij,0}\}) &= \frac{m}{r_{ij}^2} \left[-(\bar{p}_{i,1} \cdot \bar{p}_{j,1}) + \frac{1}{2} c_{i,0} q_{j,0} + \frac{1}{2} c_{j,0} q_{i,0} \right]. \end{aligned} \quad (36)$$

Of course, keep in mind the linear combination of Eq. (25), which we express as follows here:

$$c_{ij,2}(\{\bar{p}_{ij,1}\}, \{\bar{q}_{ij,2}, q_{ij,0}\}) = (m+2)c_{ij,2^+} - c_{ij,2^-}. \quad (37)$$

Note that the ‘‘hats’’ above r denote a distance tensor of a unity magnitude; their amount is equivalent with the order of the tensor. Moreover, we omit yet imply here the dependence of both $c_{ij,\aleph}$ on the monopoles, as well as on r_{ij} . Besides, notice that for $m=1$, we retrieve the familiar Coulombic case, considering our definitions of the quadrupoles in Eqs. (32) and (33).

Importantly in our general formalism, these coefficients can be employed in Eq. (14) for evaluating the interaction

between the molecular pair of Fig. 1: these can be used as correction terms for Eq. (30), and for cases in which the zero-order term vanishes, $c_{ij,1}$ or $c_{ij,2}$ may even serve as the MPIL. Let us now consider the computational implementation of Eqs. (34) and (37), especially in relation with our corresponding discussion of Eqs. (21) and (25), respectively [i.e., the discussion before we assumed the product rule of Eq. (26)]. Above all, Eqs. (34) and (37) are dependent on parameters of gravitational centers i and j (i.e., $\bar{p}_{ij,1}$, as well as $\bar{q}_{ij,2}$ and $q_{ij,0}$), while Eqs. (21) and (25) are dependent on parameters of atomistic coordinates μ_i and ν_j (i.e., $\xi_{\mu_i\nu_j}$, as well as $\cos\theta_{\mu_i\nu_j}$). At the most basic level, the former involves two summations over n_i and n_j parameters, while the latter involves one summation over n_in_j parameters. Regardless, the chief computational facilitation generated by Eq. (26) is the following: $\bar{p}_{ij,1}$, together with $\bar{q}_{ij,2}$ and $q_{ij,0}$, can be evaluated analytically irrespective of \bar{r}_{ij} , while $\xi_{\mu_i\nu_j}$, together with $\cos\theta_{\mu_i\nu_j}$, must be calculated in terms of \bar{r}_{ij} , at each step of a molecular simulation. Note though that Eqs. (34) and (37) still involve vectors, which is unlike the monopole-monopole term of Eq. (28) that solely depends on scalars. We thus generally recommend invoking $c_{ij,1}$ or $c_{ij,2}$ only if necessary, and this is usually the case with polar systems (e.g., aqueous solutions).

C. Definition of Relative Resolution

Up until now, we have been essentially looking only at the infinite limit associated with Fig. 1. How do we deal with an arbitrary distance between the molecular pair? RelRes in fact resolves this issue, yet we must begin by clearly defining the FG and CG potentials, which mostly amounts to introducing the appropriate labels on the potentials we have been working with above.

1. Hybrid potential in terms of its two components

Considering Fig. 1, the FG potential $u_{\mu_i\nu_j}^{\text{FG}}$, only a function of $r_{\mu_i\nu_j}$, is the fundamental interaction between atomistic coordinates μ_i and ν_j , and the CG potential u_{ij}^{CG} , notably a function of r_{ij} , is the apparent interaction between gravitational centers i and j ; the latter is an approximation of the appropriate summation of the former:

$$u_{ij}^{\text{CG}}(\bar{r}_{ij}; \{\bar{\Delta}_{\mu_i}\}, \{\bar{\Delta}_{\nu_j}\}) \approx \sum_{\mu_i\nu_j} u_{\mu_i\nu_j}^{\text{FG}}(r_{\mu_i\nu_j}). \quad (38)$$

Compare this expression with Eq. (6); importantly, the right sides but not the left sides are identical between the two expressions (i.e., $u_{\mu_i\nu_j}^{\text{FG}} = u_{\mu_i\nu_j}$, but $u_{ij}^{\text{CG}} \neq u_{ij}$), and this is in fact the reason for the approximation in Eq. (38). With an inverse assumption analogous with Eq. (7), we can clearly resolve the approximation here via the multipole expansion we have performed above [i.e., between Eqs. (10) and (13)]. The following then ensues:

$$u_{\mu_i\nu_j}^{\text{FG}}(r_{\mu_i\nu_j}) = \frac{c_{\mu_i\nu_j}}{r_{\mu_i\nu_j}^m}, \quad (39)$$

$$u_{ij}^{\text{CG}}(r_{ij}; \{\xi_{\mu_i\nu_j}\}, \{\cos\theta_{\mu_i\nu_j}\}) = \frac{c_{ij,\aleph^*}(\xi_{\mu_i\nu_j}, \cos\theta_{\mu_i\nu_j})}{r_{ij}^m}. \quad (40)$$

Remember of course the relation between the coefficients which appear here as given by Eq. (15), and the Hebrew index

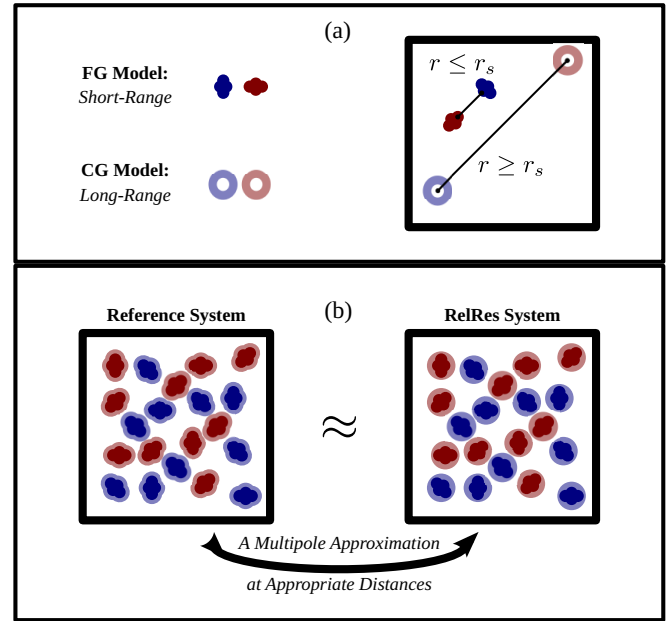


FIG. 2. A schematic representation of our multiscale approach for nonpolar systems. The red and blue colors mean that we can formally have two different molecules. For purposes of generality, the molecules here are intentionally depicted with an arbitrary number of sites (i.e., they do not correspond with any of the molecules we examine in our work). The top panel (a) characterizes the geometric FG and isotropic CG models on its left, and on its right, it illustrates that the FG potential applies between atomistic coordinates, if their relative separation is small, and the CG potential applies between gravitational centers, if their relative separation is large. The bottom panel (b) basically shows two molecular simulations, for the reference system on the left and for the RelRes system on the right. The arrow represents the MPIL parametrization, which makes the two systems approximately equal to each other.

\aleph^* is the leading (nonvanishing) term in the Taylor series of Eq. (12); this in fact corresponds with the MPIL. Of course, if we invoke the product assumption of Eq. (26), we can cast these potentials in terms of the ensuing multipoles [e.g., Eq. (28)].

We now cast the interaction between the molecular pair of Fig. 1 as a single function by using Eqs. (39) and (40). For this purpose, we must again remember the MPIL that we thoroughly discussed above. If the molecules of Fig. 1 are near to each other, the approximation of MPIL is unreasonable, and the FG potential of Eq. (39) is the relevant one. If the molecules of Fig. 1 are far from each other, the approximation of MPIL is legitimate, and the CG potential of Eq. (40) is the relevant one. We illustrate these ideas for nonpolar molecules in the top panel of Fig. 2. This is in fact the main idea of RelRes, and we define its pair potential as follows:

$$\tilde{u}(r_{\mu_i\nu_j}; r_{ij}) = \sum_{\mu_i\nu_j} \tilde{u}_{\mu_i\nu_j}^{\text{FG}}(r_{\mu_i\nu_j}) + \tilde{u}_{ij}^{\text{CG}}(r_{ij}). \quad (41)$$

In essence, RelRes is a linear combination of $\tilde{u}_{\mu_i\nu_j}^{\text{FG}}$ and $\tilde{u}_{ij}^{\text{CG}}$, which are slight modifications of the FG and CG potentials of Eqs. (39) and (40), respectively; we thoroughly discuss these functions below. Besides, note that we emphasize in Eq. (41) that the RelRes potential is a function of various

pairwise distances (i.e., between atomistic coordinates $r_{\mu_i v_j}$ and between gravitational centers r_{ij}); in such a way, RelRes maintains all degrees of freedom.

The modifications of the FG and CG potentials must ensure that for small relative separations, the derivatives of $\tilde{u}_{\mu_i v_j}^{\text{FG}}$ and $u_{\mu_i v_j}^{\text{FG}}$ are identical between each other, while the derivative of $\tilde{u}_{ij}^{\text{CG}}$ is zero, and for large relative separations, the derivatives of $\tilde{u}_{ij}^{\text{CG}}$ and u_{ij}^{CG} are identical between each other, while the derivative of $\tilde{u}_{\mu_i v_j}^{\text{FG}}$ is zero. Their exact functionalities are not unique, and specifically in our current formulation, we choose the piecewise functions of Ref. [42], casting them as follows here:

$$\tilde{u}_{\mu_i v_j}^{\text{FG}}(r) = \begin{cases} u_{\mu_i v_j}^{\text{FG}}(r) - u_{\mu_i v_j}^{\text{FG}}(r_s) & \text{if } r = (0, r_s], \\ 0 & \text{if } r = [r_s, \infty), \end{cases} \quad (42)$$

$$\tilde{u}_{ij}^{\text{CG}}(r) = \begin{cases} u_{ij}^{\text{CG}}(r_s) & \text{if } r = (0, r_s], \\ u_{ij}^{\text{CG}}(r) & \text{if } r = [r_s, \infty). \end{cases} \quad (43)$$

Reminiscent of Ref. [43], the switching distance r_s is a tuning parameter in our framework that represents the relative separation at which the MPIL switches from being unreasonable to being legitimate. Note that the vertical shifts $u_{\mu_i v_j}^{\text{FG}}(r_s)$ and $u_{ij}^{\text{CG}}(r_s)$ merely ensure that the potentials are continuous throughout r ; via the linear combination of RelRes, these vertical shifts approximately cancel each other, with the extent of the cancellation depending on how many nonzero terms the summation of Eq. (41) includes.

If one is interested in Newtonian trajectories, the step functionality of Eqs. (42) and (43) must be modified, so that their respective forces do not have a singularity at r_s . Consider the interval $[r_s - \delta r_s, r_s + \delta r_s]$, with δr_s being a parameter for smoothing the switching between $u_{\mu_i v_j}^{\text{FG}}$ and u_{ij}^{CG} [23,24,33,34,38,39]. We consequently reformulate the RelRes potential by these:

$$\tilde{u}_{\mu_i v_j}^{\text{FG}}(r) = \begin{cases} u_{\mu_i v_j}^{\text{FG}}(r) - u_{\mu_i v_j}^{\text{FG}}(r_s) & \text{if } r = [0, r_s - \delta r_s], \\ f_{\mu_i v_j}^{\text{FG}}(r) & \text{if } r = [r_s - \delta r_s, r_s + \delta r_s], \\ 0 & \text{if } r = [r_s + \delta r_s, \infty), \end{cases} \quad (44)$$

$$\tilde{u}_{ij}^{\text{CG}}(r) = \begin{cases} u_{ij}^{\text{CG}}(r_s) & \text{if } r = [0, r_s - \delta r_s], \\ f_{ij}^{\text{CG}}(r) & \text{if } r = [r_s - \delta r_s, r_s + \delta r_s], \\ u_{ij}^{\text{CG}}(r) & \text{if } r = [r_s + \delta r_s, \infty). \end{cases} \quad (45)$$

Here, $f_{\mu_i v_j}^{\text{FG}}$ and f_{ij}^{CG} can be any functions that respectively ensure the continuity of $\tilde{u}_{\mu_i v_j}^{\text{FG}}$ and $\tilde{u}_{ij}^{\text{CG}}$, together with their corresponding derivatives, at the two boundaries $r_s \pm \delta r_s$. This means for the lower boundary

$$\begin{aligned} f_{\mu_i v_j}^{\text{FG}}(r_s - \delta r_s) &= u_{\mu_i v_j}^{\text{FG}}(r_s - \delta r_s) - u_{\mu_i v_j}^{\text{FG}}(r_s), \\ \partial_r f_{\mu_i v_j}^{\text{FG}}(r_s - \delta r_s) &= \partial_r u_{\mu_i v_j}^{\text{FG}}(r_s - \delta r_s), \end{aligned} \quad (46)$$

$$\begin{aligned} f_{ij}^{\text{CG}}(r_s - \delta r_s) &= u_{ij}^{\text{CG}}(r_s), \\ \partial_r f_{ij}^{\text{CG}}(r_s - \delta r_s) &= 0 \end{aligned} \quad (47)$$

and this means for the higher boundary

$$\begin{aligned} f_{\mu_i v_j}^{\text{FG}}(r_s + \delta r_s) &= 0, \\ \partial_r f_{\mu_i v_j}^{\text{FG}}(r_s + \delta r_s) &= 0, \end{aligned} \quad (48)$$

$$\begin{aligned} f_{ij}^{\text{CG}}(r_s + \delta r_s) &= u_{ij}^{\text{CG}}(r_s + \delta r_s), \\ \partial_r f_{ij}^{\text{CG}}(r_s + \delta r_s) &= \partial_r u_{ij}^{\text{CG}}(r_s + \delta r_s) \end{aligned} \quad (49)$$

with ∂_r being the notation for the distance derivative. In this work, we specifically invoke a four-term polynomial of the inverse distance:

$$f(r) = \frac{\gamma_{m_1}}{r^{m_1}} + \frac{\gamma_{m_2}}{r^{m_2}} + \frac{\gamma_{m_3}}{r^{m_3}} + \frac{\gamma_{m_4}}{r^{m_4}}. \quad (50)$$

The exponents m must be distinct integers, and the values of the respective coefficients γ_m are determined via the appropriate set of four linear equations which satisfy the boundary conditions: For the FG coefficients, these are given by Eqs. (46) and (48), and for the CG coefficients, these are given by Eqs. (47) and (49). Realize that γ_m between $f_{\mu_i v_j}^{\text{FG}}$ and f_{ij}^{CG} are strictly different, while m may be actually the same between the two.

At this point, we remind that a pair potential is usually defined as a linear combination of several basis functions. Because of the linearity of our entire approach [the MPIL equations (34) and (37), the RelRes equations (41) and (50), etc.], one can still use all of the above expressions in a straightforward manner. Specifically here, the set of Eqs. (42) and (43) or the set of Eqs. (44) and (45) can be used separately on each function of the pair potential.

2. Lennard-Jones potential as an example

As mentioned earlier, the inverse potential is very commonly used in molecular simulations, and in this work, we focus on the LJ function. Considering sites μ_i and v_j , we define the corresponding FG potential as the following:

$$u_{\mu_i v_j}^{\text{FG}}(r) = 4\epsilon_{\mu_i v_j} \left[\frac{\sigma_{\mu_i v_j}^{12}}{r^{12}} - \frac{\sigma_{\mu_i v_j}^6}{r^6} \right] \quad (51)$$

with $\sigma_{\mu_i v_j}$ and $\epsilon_{\mu_i v_j}$ being the length and energy parameters, respectively, of the appropriate mixed interaction; for clarity, we technically omit here their FG indices (i.e., $\sigma_{\mu_i v_j} = \sigma_{\mu_i v_j}^{\text{FG}}$ and $\epsilon_{\mu_i v_j} = \epsilon_{\mu_i v_j}^{\text{FG}}$). In the context of Eq. (7), the LJ potential is just a difference between two inverse functions with $m = 12$ and 6 , and for each, the proportionality coefficient is the following:

$$c_{\mu_i v_j} = 4\epsilon_{\mu_i v_j} \sigma_{\mu_i v_j}^m. \quad (52)$$

We can plug this expression in Eq. (17), and we attain this expression:

$$c_{ij,0} = 4 \sum_{\mu_i v_j} \epsilon_{\mu_i v_j} \sigma_{\mu_i v_j}^m. \quad (53)$$

Because all parameters here have the same sign, no terms cancel in this summation and, thus, this nonzero coefficient becomes the basis for the MPIL in the LJ case. Note now

Eq. (38). Considering sites i and j , we assume the corresponding CG potential as the following:

$$u_{ij}^{\text{CG}}(r) = 4\epsilon_{ij} \left[\frac{\sigma_{ij}^{12}}{r^{12}} - \frac{\sigma_{ij}^6}{r^6} \right] \quad (54)$$

with σ_{ij} and ϵ_{ij} being the length and energy parameters, respectively, of the appropriate mixed interaction; for clarity, we essentially omit here their CG indices (i.e., $\sigma_{ij} = \sigma_{ij}^{\text{CG}}$ and $\epsilon_{ij} = \epsilon_{ij}^{\text{CG}}$). If we perform some mathematical manipulations, while keeping Eq. (18) in mind, the following relationships between the FG and CG parameters hold:

$$\sigma_{ij} = \left[\sum_{\mu_i \nu_j} \epsilon_{\mu_i \nu_j} \sigma_{\mu_i \nu_j}^{12} \right]^{1/6} / \left[\sum_{\mu_i \nu_j} \epsilon_{\mu_i \nu_j} \sigma_{\mu_i \nu_j}^6 \right]^{1/6}, \quad (55)$$

$$\epsilon_{ij} = \left[\sum_{\mu_i \nu_j} \epsilon_{\mu_i \nu_j} \sigma_{\mu_i \nu_j}^6 \right]^2 / \left[\sum_{\mu_i \nu_j} \epsilon_{\mu_i \nu_j} \sigma_{\mu_i \nu_j}^{12} \right]. \quad (56)$$

These are very convenient expressions: they mean that the LJ parameters can be determined by MPIL irrespective of a molecular simulation.

As mentioned earlier, one often assumes a product rule for mixing the proportionality coefficients of inverse functions. For the LJ parameters of a FG potential, if two identical sites μ_i have $\{\sigma_{\mu_i}, \epsilon_{\mu_i}\}$, and two identical sites ν_j have $\{\sigma_{\nu_j}, \epsilon_{\nu_j}\}$, the product rule means the following:

$$\sigma_{\mu_i \nu_j} = \sqrt{\sigma_{\mu_i} \sigma_{\nu_j}}, \quad (57)$$

$$\epsilon_{\mu_i \nu_j} = \sqrt{\epsilon_{\mu_i} \epsilon_{\nu_j}}. \quad (58)$$

In the context of Eq. (26), we have the following:

$$\begin{aligned} c_{\mu_i} &= 2\sqrt{\epsilon_{\mu_i} \sigma_{\mu_i}^m}, \\ c_{\nu_j} &= 2\sqrt{\epsilon_{\nu_j} \sigma_{\nu_j}^m}. \end{aligned} \quad (59)$$

In turn, we get the following LJ monopoles by Eq. (29):

$$\begin{aligned} c_{i,0} &= 2 \sum_{\mu_i} \sqrt{\epsilon_{\mu_i} \sigma_{\mu_i}^m}, \\ c_{j,0} &= 2 \sum_{\nu_j} \sqrt{\epsilon_{\nu_j} \sigma_{\nu_j}^m}. \end{aligned} \quad (60)$$

Again, there is no cancellation in the summations here, meaning that these nonzero coefficients can be used for purposes of MPIL in the LJ case. Consequently, we may assume a product rule here as well. For the LJ parameters of a CG potential, if two identical sites i have $\{\sigma_i, \epsilon_i\}$, and two identical sites j have $\{\sigma_j, \epsilon_j\}$, the product rule means the following:

$$\sigma_{ij} = \sqrt{\sigma_i \sigma_j}, \quad (61)$$

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}. \quad (62)$$

By substituting these in Eqs. (55) and (56), while performing some mathematical manipulations, we get the following relationships between the FG and CG parameters:

$$\begin{aligned} \sigma_i &= \left(\sum_{\mu_i} \sqrt{\epsilon_{\mu_i} \sigma_{\mu_i}^{12}} \right)^{1/3} / \left(\sum_{\mu_i} \sqrt{\epsilon_{\mu_i} \sigma_{\mu_i}^6} \right)^{1/3}, \\ \sigma_j &= \left(\sum_{\nu_j} \sqrt{\epsilon_{\nu_j} \sigma_{\nu_j}^{12}} \right)^{1/3} / \left(\sum_{\nu_j} \sqrt{\epsilon_{\nu_j} \sigma_{\nu_j}^6} \right)^{1/3}, \end{aligned} \quad (63)$$

$$\epsilon_i = \left(\sum_{\mu_i} \sqrt{\epsilon_{\mu_i} \sigma_{\mu_i}^6} \right)^4 / \left(\sum_{\mu_i} \sqrt{\epsilon_{\mu_i} \sigma_{\mu_i}^{12}} \right)^2,$$

$$\epsilon_j = \left(\sum_{\nu_j} \sqrt{\epsilon_{\nu_j} \sigma_{\nu_j}^6} \right)^4 / \left(\sum_{\nu_j} \sqrt{\epsilon_{\nu_j} \sigma_{\nu_j}^{12}} \right)^2. \quad (64)$$

Comparing these two expressions with Eqs. (55) and (56), these are even better for use: not only that the LJ parameters can be determined by MPIL with no molecular simulation, but they can be just determined for each molecule separately.

An immense simplification can be further attained if all sites are identical, with their sole parameters being σ and ϵ . This translates to the following for the FG interaction:

$$\sigma_{\mu_i \nu_j} = \sigma, \quad (65)$$

$$\epsilon_{\mu_i \nu_j} = \epsilon. \quad (66)$$

We can then substitute these expressions in Eqs. (55) and (56). This translates to the following for the CG interaction:

$$\sigma_{ij} = \sigma, \quad (67)$$

$$\epsilon_{ij} = n_i n_j \epsilon. \quad (68)$$

Note that while the sites must be identical, the molecules themselves are not necessarily the same: the number of sites on each molecule, as well as their internal location, may be different. Besides, realize that these relations apply even if a product rule does not hold.

For the scenario in which all LJ interactions are identical, we present the corresponding hybrid potential in Fig. 3. The top and bottom panels pertain to the FG and CG potentials, respectively. Note that the latter is scaled by the product of the number of sites between the two molecules. The gray curve is the true LJ potential, and it is the same between both panels. The various colors represent different switching distances in RelRes.

3. Multiscale framework for an arbitrary liquid

The hybrid potential above, Eq. (41), is in fact the pairwise version of RelRes, which only applies for a molecular pair in vacuum. Suppose now that we have a system with many molecules (e.g., a liquid). Let us make the conventional assumption that its governing energy function is defined as a summation of Eq. (6) over all of its distinct molecular pairs; as in the usual case, effective interactions, that account for multibody effects, rather than vacuum interactions, shall be

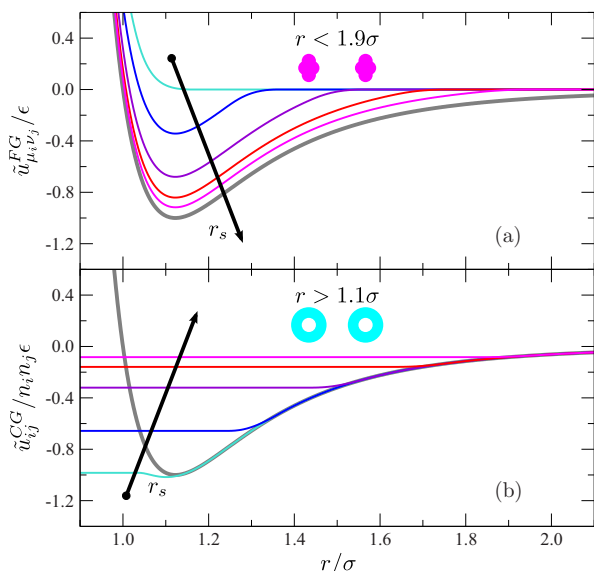


FIG. 3. The LJ potential in the context of RelRes. The top panel (a) is for the modified FG interaction between sites that are near to each other, and the bottom panel (b) is for the modified CG interaction between sites that are far from each other. These functions are specifically for the case in which all the sites are identical, with parameters σ and ϵ . Notice that the latter is scaled by the product of the number of sites on each molecule $n_i n_j$. The various colors correspond with a different switching distance in RelRes (i.e., turquoise for $r_s = 1.1\sigma$, blue for $r_s = 1.3\sigma$, violet for $r_s = 1.5\sigma$, red for $r_s = 1.7\sigma$, magenta for $r_s = 1.9\sigma$, etc.); each arrow indicates an increasing switching distance. In both panels, the gray curve is for the true LJ potential; in (a), one can think of it as a hybrid potential with $r_s \rightarrow \infty$, and in (b), one can think of it as a hybrid potential with $r_s \rightarrow 0$. Note the schematic in each panel, with each one of them representing a pair of molecules separated by $r \approx 1.5\sigma$. In the top panel, the magenta color indicates that the RelRes of the pair has $r_s = 1.9\sigma$, meaning that they interact via the FG interaction, and in the bottom panel, the turquoise color indicates that the RelRes of the pair has $r_s = 1.1\sigma$, meaning that they interact via the CG interaction. Note that we use here the smooth version of RelRes [i.e., Eqs. (44) and (45)].

used in such an expression. Because of the pairwise nature of the energy function, the hybrid formulation above can be applied on each molecular pair separately and, in turn, the definition for the complete version of RelRes ensues:

$$\begin{aligned} \tilde{U} &= \frac{1}{2} \sum_{i \neq j} \tilde{u}(r_{\mu_i \nu_j}; r_{ij}), \\ \tilde{U} &= \frac{1}{2} \sum_{i \neq j} \left[\sum_{\mu_i \nu_j} \tilde{u}_{\mu_i \nu_j}^{\text{FG}}(r_{\mu_i \nu_j}) + \tilde{u}_{ij}^{\text{CG}}(r_{ij}) \right]. \end{aligned} \quad (69)$$

We omit the functionality of this complete version of RelRes throughout our work for clarity. Besides, realize that while a generalized variant of RelRes, that formally incorporates multibody terms, may be possible, we do not explore such a possibility in this work.

This Hamiltonian introduces a notable conceptual complexity, as compared with Eq. (41) of the pairwise case. For

a molecular pair in vacuum, the entire system is basically either a pure FG or CG scenario. If the molecules are near to each other, they are both described by the FG model, and if the molecules are far from each other, they are both described by the CG model (of course, there are subtleties for moderate relative separations). Nevertheless, in the case of a molecular simulation of a realistic liquid, the situation is completely different, with the system being neither purely FG nor purely CG, but instead having a hybrid nature. Each molecule simultaneously embodies both models, and a given molecule interacts with its near neighbors via the FG potential and with its far neighbors via the CG potential; the resolution of a given molecule is always relative to its observer. This is in fact the aspect which makes RelRes computationally advantageous: the few near neighbors must be evaluated via the detailed FG potential, but the many far neighbors can be evaluated via the simplified CG potential.

Let us also discuss RelRes in terms of its r_s parameter, labeling such a dependence as \tilde{U}_{r_s} . The two limits of r_s are of particular importance: \tilde{U}_{∞} is identical with the corresponding energy function of the pure FG system, and \tilde{U}_0 is identical with the corresponding energy function of the pure CG system. We specifically name the system with $r_s \rightarrow \infty$ (i.e., the one with just FG and no CG interactions) the reference system. Adequately describing its (configurational) Hamiltonian, as a function of the relevant degrees of freedom, is the fundamental aim of RelRes

$$\tilde{U}_{r_s} \approx \tilde{U}_{\infty} \quad (70)$$

and this can be achieved with the appropriate choice of the switching distance, together with the MPIL. This can in turn enable RelRes to capture the static and dynamic behavior of the reference system throughout state space. We depict this multiscale goal in the bottom panel of Fig. 2.

For completeness, we also now define the entire system Hamiltonian:

$$\tilde{E} = K + U_{\psi} + \tilde{U}. \quad (71)$$

K is the kinetic energy, while U_{ψ} accounts for all intramolecular energetics. Note that these do not have a tilde above them. This is because RelRes maintains all degrees of freedom and, in turn, its K , as well as its U_{ψ} , is strictly unaltered with the corresponding functions of the reference system (e.g., $K_{r_s} = K_{\infty}$).

III. COMPUTATIONAL VALIDATION

We now turn our attention to testing the efficacy of this multiscale framework in describing reference systems. We do this via molecular simulations of idealized systems (i.e., liquids composed of dumbbell-like or butterflylike molecules), complementing in turn the initial examination of our previous publication [37]. Our original work was a proof of concept that RelRes is very successful in describing structural correlations and thermal properties of (nonpolar) multicomponent and multiphase systems across state space (e.g., temperature and density); by invoking a tuning parameter in the Hamiltonian, we importantly showed that MPIL is the ideal choice for the RelRes potential. Consequently, for simplicity in this current work, we only examine uniform liquids, taking MPIL

for granted in all of our systems. Indeed, the main focus here is in systematically varying the distance at which the FG and CG potentials are switched between each other. Besides, we also complement our initial work by showing that RelRes with MPII captures well not only static behavior, but also dynamic behavior.

A. Numerical protocol

Our numerical results here are based on molecular simulations conducted with the GROMACS package, specifically its 4.6 version [46]; unless otherwise specified, we use the default values of the package. The general implementation here is almost identical with the one we had in our original publication [37]. Importantly, we examine two separate sets of molecular simulations: one is based on dumbbell-like molecules, and one is based on butterflylike molecules.

The LJ potential, as given by Eqs. (51) or (54), is the sole basis for the interaction between FG or CG sites, respectively. In most of the case studies here, the molecules embody identical sites, with σ and ϵ as their respective length and energy parameters; consequently, we most often invoke Eqs. (65) and (66) for the FG parameters and Eqs. (67) and (68) for the CG parameters. We employ this LJ functionality together with the continuous versions of the modified FG and CG potentials, as given by Eqs. (44) and (45). For the corresponding Eq. (50), we set $\{m_1, m_2, m_3, m_4\} = \{6, 12, 18, 24\}$ (for convenience with the LJ potential); we solve for the FG γ_m via Eqs. (46) and (48), and we solve for the CG γ_m via Eqs. (47) and (49). In most cases, we vary r_s between 1.1σ and 1.9σ at intervals of 0.2σ , and we set $\delta r_s = 0.0625\sigma$. On a relevant note, rather than abruptly truncating all potentials at 2.5σ , we analogously apply the sigmoid function of Eq. (50) between 2.25σ and 2.75σ . Note that all interactions in our reference system, denoted by $r_s \rightarrow \infty$, practically vanish beyond this distance. All of these potentials are implemented via tabulated functions. The intramolecular energetics are mostly governed by elastic bonds, which is unlike our previous publication that made all bonds rigid. The spring constant of each bond is always $2000\epsilon/\sigma^2$, and except for a few special cases, the inherent size ℓ of each bond is set at 0.5σ .

All of our systems in this work contain, in a periodic box, a total of 2000 (identical) molecules. Their actual FG coordinates have mass m , and their virtual CG centers have mass zero; the latter is generally constructed in terms of equal weights of the former. The molecular simulations propagate in time via Newtonian equations of motion. We employ neighbor lists for calculating the respective forces, updating them every five steps. Each molecular simulation starts with an equilibration phase of 5000 steps of size 0.002τ and ends with a production phase of 1 000 000 steps of size 0.001τ ; the decrease in the time step is implemented because of the importance of energy conservation in the later stage. Note that $\tau = \sqrt{m\sigma^2/\epsilon}$ is our unit of time.

In all cases, the protocol that we use involves a sequence of molecular simulations that employ the thermostat of Bussi *et al.* [47]; we set the temperature at $1.0\epsilon/k$, with k being Boltzmann's constant. The initial molecular simulation is of the reference liquid, and it is distinct in that it is also coupled to a barostat [48], whose pressure is $1.0\epsilon/\sigma^3$ together with

a respective compressibility of $0.2\sigma^3/\epsilon$. While the system size naturally fluctuates in this case, the purpose of this initial molecular simulation is to fix the box dimension for the rest of the sequence; we otherwise disregarded all of its respective results. The rest of the molecular simulations are subsequently in the canonical ensemble: one of them is of the reference liquid, while the rest are RelRes systems with different switching distances.

Above all, we thoroughly examine in our work several structural correlations, often in terms of the relative separation r_{ij} (we frequently omit the ij indices). For this purpose, we record the positions of all molecules every 100 steps of the molecular simulation. For all of the functions we examine here, r goes from the origin to the edge of the system box; we discretize it by 1000 bins.

Foremost, we look at radial distributions g between the midpoints of various bonds. For comparison between radial distributions, it is convenient to invoke a functional for the disparity between a pair of such functions. We particularly choose the Jensen-Shannon entropy [49], which has been suggested as one of the most fundamental metrics in information theory [50]:

$$S_{JS} = S\left[\frac{1}{2}(g_{r_s} + g_{\infty})\right] - \frac{1}{2}(S[g_{r_s}] + S[g_{\infty}]) \quad (72)$$

with the conventional entropy taking on this definition:

$$S[g] = - \int \kappa g r^2 \ln(\kappa g r^2) dr. \quad (73)$$

κ is just the normalization constant (i.e., $\int \kappa g r^2 = 1$), which is inversely proportional with the system volume. In Eq. (72), the term on the left [i.e., $S[\frac{1}{2}(g_{r_s} + g_{\infty})]$] is one entropy, altogether for an average of two distributions, and the term on the right [i.e., $\frac{1}{2}(S[g_{r_s}] + S[g_{\infty}])$] is an average of two entropies, each for one distribution. Basically, the Jensen-Shannon entropy measures the discrepancy between a pair of radial distributions: Obviously, if $g_{r_s} \approx g_{\infty}$, $S_{JS} \approx 0$, and S_{JS} increases as the discrepancy between g_{r_s} and g_{∞} increases.

We also examine the orientational correlations of various bonds. By letting a bond of molecule i or j be a vector, we respectively define \bar{s}_i or \bar{s}_j as its direction. We particularly focus on computing the moments of the dot product $\bar{s}_i \cdot \bar{s}_j$ as a function of the relative separation between the midpoints of the bonds. Most of our case studies have symmetry in their molecules, and thus for $\bar{s}_i \cdot \bar{s}_j$, the average vanishes but the variance persists. We consequently present only the latter, $\langle(\bar{s}_i \cdot \bar{s}_j)^2\rangle$; note that $\langle(\bar{s}_i \cdot \bar{s}_j)\rangle = \frac{1}{3}$ for decorrelated cases in which each molecular orientation is equally likely. Moreover, the bond vectors can be cast as a linear decomposition in terms of two components $\bar{s}_i = \bar{s}_i^{\parallel} + \bar{s}_i^{\perp}$ and $\bar{s}_j = \bar{s}_j^{\parallel} + \bar{s}_j^{\perp}$: In terms of the relative separation between molecules i and j , $\{\bar{s}_i^{\parallel}, \bar{s}_j^{\parallel}\}$ are the components parallel to \bar{r}_{ij} , and $\{\bar{s}_i^{\perp}, \bar{s}_j^{\perp}\}$ are the components perpendicular to \bar{r}_{ij} . In turn, we have the following:

$$\langle(\bar{s}_i \cdot \bar{s}_j)^2\rangle = \langle(s_i^{\parallel} s_j^{\parallel})^2\rangle + 2\langle s_i^{\parallel} s_j^{\parallel} s_i^{\perp} s_j^{\perp} \rangle + \langle(s_i^{\perp} s_j^{\perp})^2\rangle. \quad (74)$$

In most of our case studies, we observe that the cross term is rather negligible, so

$$\langle(\bar{s}_i \cdot \bar{s}_j)^2\rangle \approx \langle(s_i^{\parallel} s_j^{\parallel})^2\rangle + \langle(s_i^{\perp} s_j^{\perp})^2\rangle \quad (75)$$

and thus, we only present the two main components, $\langle (s_i^{\parallel} s_j^{\parallel})^2 \rangle$ and $\langle (s_i^{\perp} s_j^{\perp})^2 \rangle$, of this orientational function; realize that $\langle (s_i^{\parallel} s_j^{\parallel})^2 \rangle = \frac{1}{9}$ and $\langle (s_i^{\perp} s_j^{\perp})^2 \rangle = \frac{2}{9}$ if molecules are decorrelated.

Aside from these static functions, we also examine dynamic functions, which notably complement the examination of our original communication. Specifically, we look at the squared displacement $\langle (\bar{r}_i - \bar{r}_0)^2 \rangle$, together with its time derivative $\partial_t \langle (\bar{r}_i - \bar{r}_0)^2 \rangle$. \bar{r}_i is the position of an arbitrary molecule, i or j , at a given time t , and the index 0 is of course for zero time. Besides, we also examine an orientational correlation $\langle \bar{s}_i \cdot \bar{s}_0 \rangle$. \bar{s}_i is analogous with the bond directions discussed earlier; just as with the squared displacement, we have here the time index. Importantly in calculating these transport functions, we employ an algorithm of multiple time origins, using each recorded step of a molecular simulation.

We furthermore examine thermal properties in our work. We foremost look at the average $\langle \tilde{U} \rangle$ and variance $\langle \delta \tilde{U}^2 \rangle$ of the defining equation for RelRes. For convenience, we normalize both by the corresponding values of the reference liquid: $\langle \tilde{U} \rangle_{r_s}^* = \langle \tilde{U} \rangle_{r_s} / \langle \tilde{U} \rangle_{\infty}$ and $\langle \delta \tilde{U}^2 \rangle_{r_s}^* = \langle \delta \tilde{U}^2 \rangle_{r_s} / \langle \delta \tilde{U}^2 \rangle_{\infty}$; as we frequently do, we omit the index of r_s for these. Besides, we also look at a transport function of the total energy $\langle \delta \tilde{E}_t \delta \tilde{E}_0 \rangle / \langle \delta \tilde{E}^2 \rangle$, computing it again via multiple time origins. All thermal properties are based on probing the molecular simulations every 10 steps.

B. Dumbbell systems

Our preliminary systems are based on dumbbell-like molecules. We analyzed a liquid-liquid mixture of two distinct dumbbells in our initial communication; here, we continue with an exhaustive investigation of several variants of such dumbbells, with all of the current systems being single-component and single-phase liquids. In general, all dumbbells have two FG sites which map on one CG site, meaning that $n = 2$; by equal weighting, the latter one is exactly in the middle of the former two. In all systems here, we focus on the systematic variation of r_s in RelRes.

1. Base scenario for the dumbbells

We begin by examining a system of elementary dumbbells. One such molecule is depicted in Fig. 4. Such a dumbbell has two identical FG sites, with σ and ϵ for its LJ parameters [note Eqs. (65) and (66)]. By the MPIL, this means that the respective CG site has σ and 4ϵ for its LJ parameters since $n = 2$ [note Eqs. (67) and (68)]. The inherent bond of these elementary dumbbells is set at a distance of 0.5σ . The density of this system is $1.00m/\sigma^3$.

The radial distribution between the dumbbells is given in the top panel of Fig. 4. This structural correlation for the reference liquid is given here as the solid dark-gray curve; the remaining dashed curves are for the RelRes systems, with each color representing a different switching distance. All of these switching distances give a sufficient description of the radial distribution of the reference system, and it is clear that as r_s increases, the capability of RelRes improves, in an apparently asymptotic manner. While for $r_s = 1.3\sigma$ (i.e., the blue

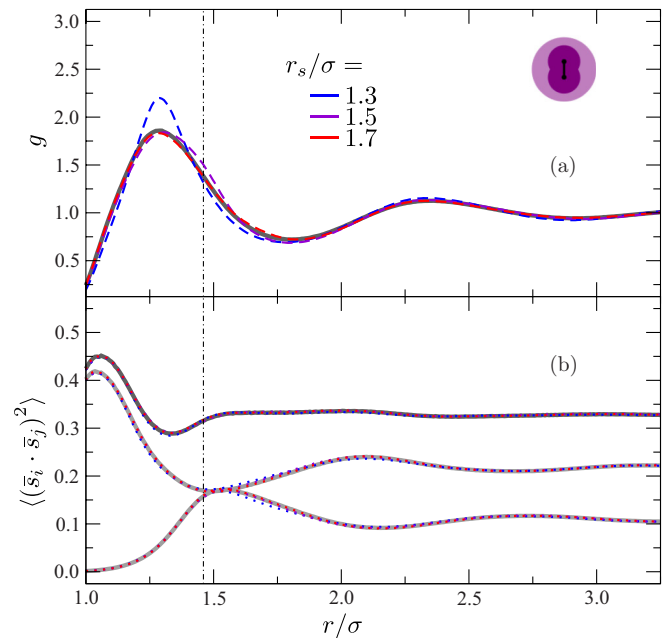


FIG. 4. Static correlations for the elementary dumbbell scenario. These molecules have uniform LJ parameters for their two sites σ and ϵ , and the bond length between them is 0.5σ . Here, all the gray curves correspond with the reference system. The colored curves correspond with the RelRes strategy, with each color representing a different switching distance r_s . Note particularly the violet curve, which is for the ideal switching distance in capturing the behavior of the reference liquid; the blue or red curve has a shorter or longer switching distance, respectively. Everything is plotted as a function of the distance between the midpoints of the dumbbells r . The top panel (a) plots the corresponding radial distribution g , with the solid curve being for the reference system, while the dashed curves being for the RelRes scenarios. The bottom panel (b), for the reference liquid, plots the orientational correlation $\langle (\bar{s}_i \cdot \bar{s}_j)^2 \rangle$ as the dark-gray curve, together with its two main components $\langle (s_i^{\parallel} s_j^{\parallel})^2 \rangle$ and $\langle (s_i^{\perp} s_j^{\perp})^2 \rangle$ as the lower and higher light-gray curves, respectively; the functionality of the dotted curves corresponds with the functionality of their neighboring solid curves; note that all of these curves are almost identical. In both panels, the vertical line represents the inflection of g , which is located here at 1.46.

curve), the replication is fairly decent, for $r_s = 1.7\sigma$ (i.e., the red curve), the replication is essentially perfect; in fact, $r_s = 1.5\sigma$ (i.e., the violet curve) already captures the translational correlation of the reference system as well as one may desire. The systematic observation here is reminiscent with the one we had in the original publication [37]: RelRes captured structural correlations with $r_s \approx 1.1\sigma$ equitably yet with $r_s \approx 1.6\sigma$ superbly. Another aspect of Fig. 4 is that RelRes can create a suction effect in the middle of the dumbbell. With decreasing r_s , nearest neighbors tend to stick more to each other, as evidenced by the increasing peak; with a sufficient choice for the switching distance, the suction effect disappears.

For the purpose of better clarifying this asymptotic behavior, we invoke the Jensen-Shannon entropy, as defined by Eq. (72). For each of the relevant curves in Fig. 4, we calculate S_{JS} , and we plot it (i.e., an indigo circle) in terms of the switching distance in Fig. 5. Clearly, this entropy is monotonic

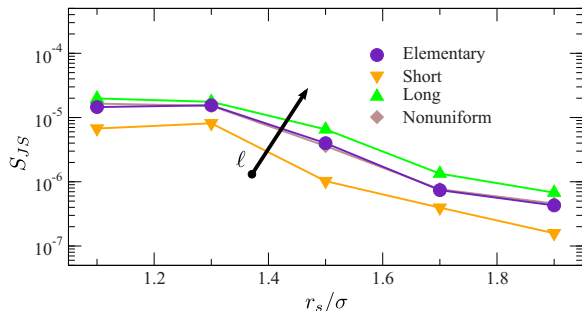


FIG. 5. The Jensen-Shannon entropy S_{JS} for the radial distribution, plotted as a function of the switching distance r_s in RelRes. Each set of points, denoted by a specific color, is for a different system. The indigo is for the molecules which have identical LJ parameters between their two sites ϵ and σ , and the corresponding bond length is 0.5σ . The orange and green refer to the systems with bond lengths of 0.3σ and 0.7σ , respectively, while their LJ parameters are also uniform. The almost hidden brown corresponds with $\{1.61\epsilon, 0.62\epsilon\}$ and $\{0.85\sigma, 1.18\sigma\}$ for its LJ parameters, while its bond is again 0.5σ . Importantly, note that the entropy is on logarithmic scale. Besides, the lines here just serve as guides.

throughout most of r_s , and considering the logarithmic scale here, S_{JS} is roughly an exponential decay of the distance; this is especially true around $r_s = 1.5\sigma$, which we mentioned in the context of Fig. 4 as the ideal switching distance for adequately capturing the radial distribution of the reference system. Thus, we confirm here our earlier observation: RelRes improves asymptotically as r_s approaches infinity. Importantly, it appears that the asymptote is practically reached at $r_s \approx 1.5\sigma$: at this distance, the discrepancy between the radial distributions almost vanishes, and in turn, S_{JS} becomes almost negligible.

So, why is 1.5σ the critical distance for switching between the FG and CG potentials in RelRes? For this purpose, we now look at orientational correlations in the bottom panel of Fig. 4. Let us begin by discussing the dark-gray solid curve, which is $\langle(\bar{s}_i \cdot \bar{s}_j)^2\rangle$ of the reference liquid. Notice that while it clearly has a maximum and a minimum early on, it quickly flattens out after the ensuing inflection at 1.42σ . This consequently explains why 1.5σ is an excellent choice for r_s in RelRes: at this distance, the dumbbell directions apparently become decorrelated, and the influence of these rotational degrees of freedom becomes negligible on the various structural correlations of the system (e.g., radial distributions). Interestingly, the flattening of $\langle(\bar{s}_i \cdot \bar{s}_j)^2\rangle$ in the bottom panel happens just around the middle between the respective maximum and minimum in g of the corresponding top panel (i.e., the relative separation of 1.54σ); this is also around the respective inflection of g (i.e., the relative separation of 1.46σ), and we mark this value by a vertical line in both panels. This means that as soon as nearest neighbors depart from each other, their directions quickly become decorrelated.

Let us continue examining the two main components of $\langle(\bar{s}_i \cdot \bar{s}_j)^2\rangle$, as given in Eq. (75). For the reference system, these are plotted as light-gray curves in the bottom panel of Fig. 4, the lower one is for $\langle(s_i^{\parallel} s_j^{\parallel})^2\rangle$ and the higher one is for $\langle(s_i^{\perp} s_j^{\perp})^2\rangle$; realize that their summation essentially retrieves

the dark-gray curve. Surprisingly, the behavior of $\langle(\bar{s}_i \cdot \bar{s}_j)^2\rangle$ is very distinct from its two components: While $\langle(\bar{s}_i \cdot \bar{s}_j)^2\rangle$ becomes mostly decorrelated as the molecules leave the primary coordination shell, its components are still correlated even as the molecules enter the secondary coordination shell. In fact, $\langle(s_i^{\parallel} s_j^{\parallel})^2\rangle$ and $\langle(s_i^{\perp} s_j^{\perp})^2\rangle$ appear as mirror images of each other, especially beyond $r \approx 1.5\sigma$, with their extrema occurring at about the same locations. In turn, once we sum these two functions, these fluctuations cancel each other, yielding the overall function which is basically flat beyond $r \approx 1.5\sigma$. The locations of the initial maximum of $\langle(s_i^{\parallel} s_j^{\parallel})^2\rangle$ and the initial minimum of $\langle(s_i^{\perp} s_j^{\perp})^2\rangle$ occur at 1.49σ and 1.54σ , respectively, which is again almost identical with the ideal r_s of 1.5σ .

Finally, there are also dotted curves in the bottom panel of Fig. 4. Their coloring is equivalent with that of the corresponding top panel, with each color representing a different r_s in RelRes; besides, the function to which each dotted curve pertains corresponds with the function of its neighboring solid curve. Importantly, RelRes captures the orientational correlation with its components perfectly well, regardless of which switching distance is used. This is of course in contrast with our observations for the respective translational correlation. As discussed earlier, the suction effect, in the middle of the dumbbell, is substantially influenced by the switching distance; if an inadequate r_s is chosen, the suction can significantly alter the radial distribution. In any case, rotational functions are expected to be mostly governed by the shape of molecules, and r_s has negligible effect on this; so, even if the suction effect makes the dumbbells erroneously stick to each other, the orientational correlations are roughly the same as those of the reference system.

We now move on to examine thermal properties of these elementary dumbbells. We specifically look at functionals of the configurational Hamiltonian of RelRes, presenting, as a function of r_s , its normalized average $\langle\bar{U}\rangle$ and variance $\langle\delta\bar{U}^2\rangle$ in the top and bottom panels of Fig. 6, respectively. The coloring here is analogous with that of Fig. 5. Focus again on the indigo circles: because of normalization, they fluctuate about unity (i.e., about the horizontal line). In fact, for all switching distances, the average and variance are essentially both within a fraction of 0.05 from their reference values. Importantly, with increasing r_s , we notice a dampening effect of the fluctuation, with both the average and variance eventually reaching unity. Besides, the fluctuating characteristic of the symbols in this figure is in striking contrast to the observation we made for the entropic functional of Fig. 5, which showed a firm monotonic behavior with r_s . Bearing also in mind Fig. 4, this means that if one is fine with a decent yet rough description of structural correlations in a liquid, one can still adequately capture thermal properties, using just a modest switching distance in RelRes that does not entirely account for all nearest neighbors. Finally, these systematic findings are analogous with the observations we made in our previous work for the pressure, together with its corresponding response function: for both of our switching distances there, $r_s \approx 1.1\sigma$ and $r_s \approx 1.6\sigma$, RelRes was very successful once we used it together with MPIL [37]. In general, the excellent replication of thermal properties reiterates the validity of the

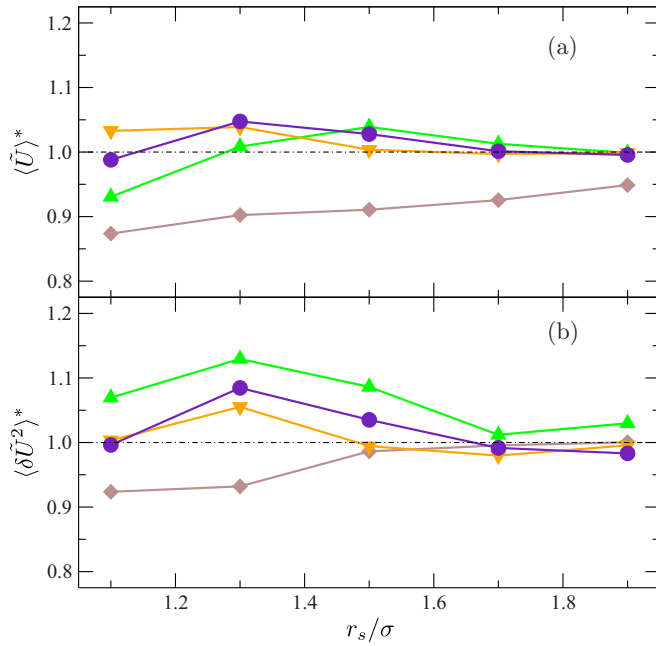


FIG. 6. Energetic functionals plotted as a function of the switching distance r_s in RelRes. Particularly, the average $\langle \tilde{U} \rangle$ and variance $\langle \delta \tilde{U}^2 \rangle$ of the configurational energy is given in the top (a) and bottom (b) panels, respectively. The asterisk denotes that these functionals are normalized by the respective values of the reference systems. Besides, the color coding here is identical with that of Fig. 5.

MPIL approximation, which is essentially based on energy conservation.

So far, considering our original publication as well, we have only examined static behavior; we now turn our attention to dynamic behavior. We consequently plot various transport functions in terms of time t in Fig. 7. The coloring here is again analogous with Fig. 4: while all solid curves are for the reference liquid, the dashed and dotted lines are for the RelRes systems with varying r_s . In the top panel for the reference liquid, we give $\langle (\bar{r}_t - \bar{r}_0)^2 \rangle$ as the dark-gray line, with its negative derivative $-\partial_t \langle (\bar{r}_t - \bar{r}_0)^2 \rangle$ as the light-gray one. In the bottom panel for the reference liquid, the orientational correlation $\langle \bar{s}_t \cdot \bar{s}_0 \rangle$ is given as the dark-gray line. Aside from these structural functions, we also present here a thermal function, specifically for the total energy $\langle \delta \tilde{E}_t \delta \tilde{E}_0 \rangle / \langle \delta \tilde{E}^2 \rangle$ as the light-gray line. The functionality of a colored line is the same as that of its neighboring gray curves.

For all transport functions of Fig. 7, RelRes satisfactorily describes the reference system irrespective of the value of r_s . Of course, there are some nuances between the various transport functions in terms of r_s , and these dynamic characteristics are reminiscent of our observations for the static features. For example, $r_s = 1.7\sigma$ is required for a flawless replication of the translational $\langle (\bar{r}_t - \bar{r}_0)^2 \rangle$, yet for the orientational $\langle \bar{s}_t \cdot \bar{s}_0 \rangle$, only $r_s = 1.3\sigma$ cannot excellently capture the behavior of the reference liquid. Besides, for $\langle \delta \tilde{E}_t \delta \tilde{E}_0 \rangle / \langle \delta \tilde{E}^2 \rangle$, all switching distances yield a perfect description of this transport function. In summary, in terms of r_s , we notice the same trends for dynamic behavior as we do for static behavior: Thermal properties can be superbly captured even with a surprisingly

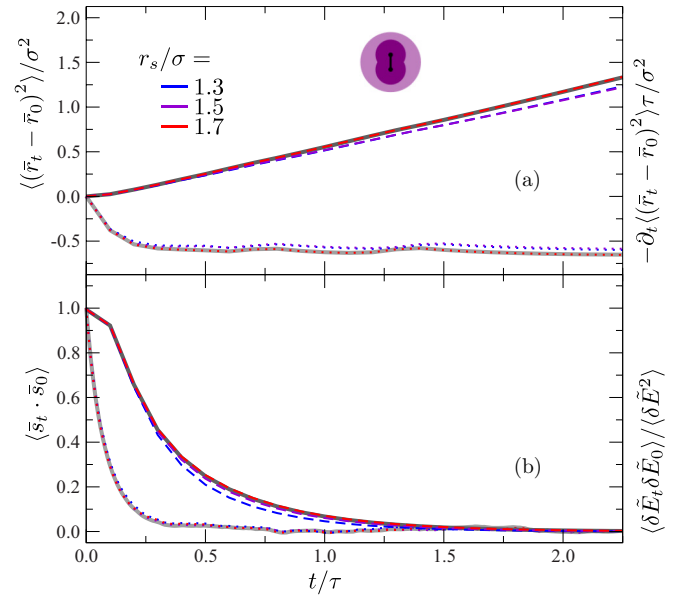


FIG. 7. Dynamic correlations for the elementary dumbbell scenario. Much of the coding here is equivalent with Fig. 4, but all of the current functions are strictly different. Everything is plotted in terms of time t . The top panel (a) plots the squared displacement $\langle (\bar{r}_t - \bar{r}_0)^2 \rangle$ of a given molecule as the dark-gray curve for the reference system, together with its neighboring dashed lines for the RelRes systems (i.e., the higher set of curves, which is labeled by the left ordinate axis); the corresponding negative derivative of this function $-\partial_t \langle (\bar{r}_t - \bar{r}_0)^2 \rangle$ is given as the light-gray curve, together with its dotted lines (i.e., the lower set of curves, which is labeled by the right ordinate axis). In an analogous manner, the bottom panel (b) plots the orientational function $\langle \bar{s}_t \cdot \bar{s}_0 \rangle$ as the dark-gray curve, together with its neighboring dashed lines (i.e., the higher set of curves, which is labeled by the left ordinate axis); the light-gray curve, together with its dotted lines, represents the energetic function $\langle \delta \tilde{E}_t \delta \tilde{E}_0 \rangle / \langle \delta \tilde{E}^2 \rangle$ (i.e., the lower set of curves, which is labeled by the right ordinate axis).

small r_s , but structural correlations require a relatively large r_s ; regarding the latter, orientational functions are more feasibly captured than translational functions. Specifically for these elementary dumbbells, it appears that $r_s = 1.5\sigma$ can capture the entire behavior of the reference liquid very well and, thus, this is our recommended switching distance for these molecules. This in turn reiterates one of the main findings of our original publication: RelRes works best if molecules interact with each via a FG potential between nearest neighbors and a CG potential between other neighbors.

2. A variation in the bond length of the dumbbells

We continue by examining dumbbells in which we vary the length of their bonds. In particular, we construct two systems, one with short dumbbells (i.e., $\ell = 0.3\sigma$) and one with long dumbbells (i.e., $\ell = 0.7\sigma$); representative sketches are given in Figs. 8 and 9, respectively. The modification of the bond length does not alter the LJ parameters of the FG and CG sites: they are again σ and ϵ for the former, and σ and 4ϵ for the latter. The system density is 1.34 for the short molecules and 0.78 for the long molecules, in units of m/σ^3 .

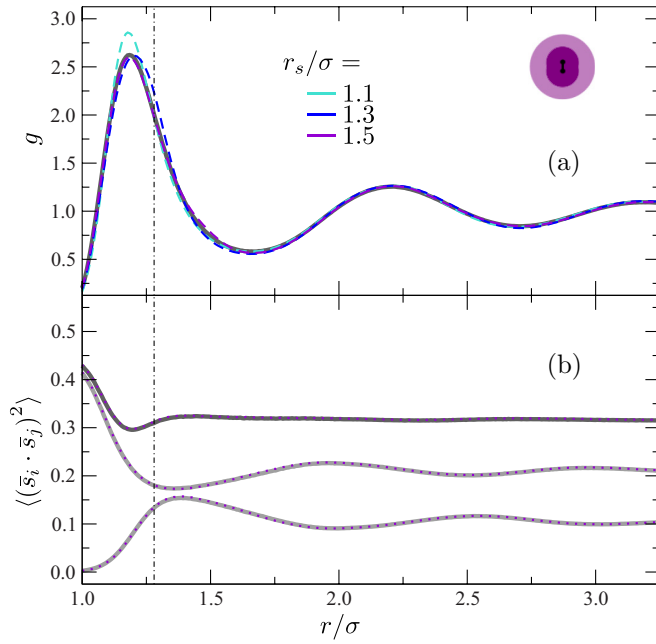


FIG. 8. Static correlations for the short dumbbells with a bond length of 0.3σ . Here, the coding, as well as the functions, is essentially equivalent to Fig. 4. However, note the turquoise curve which is for RelRes with a switching distance of 1.1σ . The vertical line here goes through 1.28 . We plot the orientational function only for $r_s = 1.5\sigma$ since all other curves are basically identical with it.

The structural correlations for the short and long dumbbells are given in Figs. 8 and 9, respectively, which are basically analogous with those of Fig. 4. All rotational functions of the reference liquids are perfectly captured by RelRes, regardless of the switching distance that we use; we consequently give in the respective bottom panels only the $r_s = 1.5\sigma$ curves. Conversely, the radial distributions are given in the respective top panels; notice importantly the turquoise curve (i.e., $r_s = 1.1\sigma$) for the $\ell = 0.3\sigma$ case and the magenta curve (i.e., $r_s = 1.9\sigma$) for the $\ell = 0.7\sigma$ case (i.e., the colors which do not appear in Fig. 4). In essence for each case, we have shifted our systematic r_s examination by 0.2σ . This is because we observe that RelRes captures the radial distributions at a switching distance that differs by roughly 0.2σ as compared with the $\ell = 0.5\sigma$ of Fig. 4: the apparently asymptotic r_s is 1.3σ in Fig. 8 and 1.7σ in Fig. 9. It is of course natural that the ideal r_s for the short dumbbells is less than the ideal r_s for the long dumbbells since the former are more “spherelike” than the latter. This further means that RelRes works better as the molecular size decreases. As a further analysis, we compute S_{JS} for these radial distributions, plotting them in Fig. 5 as orange (downward) triangles for $\ell = 0.3\sigma$ and as green (upward) triangles for $\ell = 0.7\sigma$. For these two cases, we again observe an essentially exponential decay. We also reaffirm that the efficacy of RelRes is more deficient for the long molecules than for the short molecules, considering that S_{JS} for $\ell = 0.3\sigma$ is less than that for $\ell = 0.7\sigma$ by about an order of magnitude; remember that $S_{JS} \rightarrow 0$ means that we are approaching perfect replication of the radial distribution.

In the context of Fig. 4 for $\ell = 0.5\sigma$, we have thoroughly discussed that the recommended value for r_s is signaled by structural correlations. We find that this is still the case

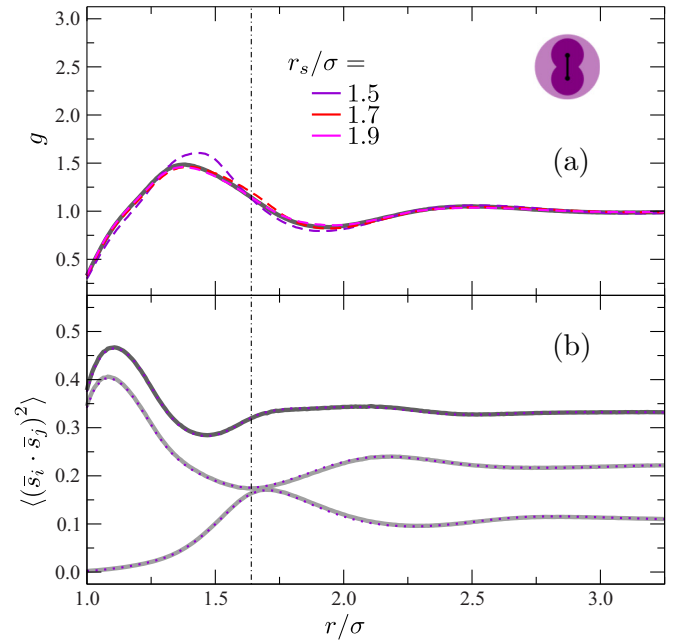


FIG. 9. Static correlations for the long dumbbells with a bond length of 0.7σ . Here, the coding, as well as the functions, is essentially equivalent to Fig. 4. However, note the magenta curve which is for RelRes with a switching distance of 1.9σ . The vertical line here goes through 1.64 . We plot the orientational function only for $r_s = 1.5\sigma$ since all other curves are basically identical with it.

here once we vary the bond length of the dumbbells. Most importantly, in the bottom panels of Figs. 8 and 9, we again notice the flattening of $\langle(\bar{s}_i \cdot \bar{s}_j)^2\rangle$ beyond a certain distance, which stems in the mirroring behavior of the maxima and minima of its components $\langle(s_i^\perp s_j^\parallel)^2\rangle$ and $\langle(s_i^\parallel s_j^\perp)^2\rangle$ that cancel each other. We summarize all of the critical distances associated with these orientational correlations, as well as with the translational ones in the respective top panels, in Table I. Among these different bond lengths, we generally notice the same trends: the inflection in $\langle(\bar{s}_i \cdot \bar{s}_j)^2\rangle$ comes just before the inflection in g , and this is followed consecutively by the minimum in $\langle(s_i^\perp s_j^\perp)^2\rangle$ and the maximum in $\langle(s_i^\parallel s_j^\parallel)^2\rangle$, with the midpoint between the extrema in g being around there as well. For a given bond length, the most important aspect of these distances is that they almost all occur within 0.1σ of each other, and they are roughly the same as our estimate for the asymptotic r_s via the systematic RelRes examination.

As such, it appears that if one is interested in determining the ideal switching distance, there may be no need for constructing many RelRes systems with different r_s : one can just look at the structural correlations of the reference liquid, with the radial distribution being arguably the most feasible choice. As such, we consequently again draw vertical lines in Figs. 8 and 9 that represent their inflections in g . Nevertheless, we can even go further: For such dumbbells of an arbitrary bond length, we may not even need to do any molecular simulations of the reference liquid since it is quite obvious here that the following linear approximation holds for the ideal value of the switching distance:

$$r_s \approx \sigma + \ell. \quad (76)$$

TABLE I. Signature distances in the structural correlations of the various dumbbell scenarios. Rows 1 and 2 give us the molecule type, as well as its respective bond length, with which each column corresponds. Column 2 goes together with Fig. 4, columns 3 and 4 refer to Figs. 8 and 9, respectively, and column 5 goes with Fig. 10. Column 1 then tells us which distance value we are dealing with in each case. Row 3 is the switching distance that we recommend for use in RelRes, based on our systematic examination of varying r_s . Rows 4 and 5 deal with signatures of the radial distributions, while rows 6–8 deal with signatures of the orientational functions. Realize that all numbers here are given in units of σ .

Molecule type	Elementary	Short	Long	Nonuniform
Bond Length	0.5	0.3	0.7	0.5
Suggested r_s	1.5	1.3	1.7	1.5
Inflection in g	1.46	1.28	1.64	1.49
Midpoint between extrema in g	1.54	1.43	1.66	1.59
Inflection in $\langle (\bar{s}_i \cdot \bar{s}_j)^2 \rangle$	1.42	1.25	1.59	1.51
Maximum of $\langle (s_i^{\parallel} s_j^{\parallel})^2 \rangle$	1.49	1.36	1.64	1.57
Minimum of $\langle (s_i^{\perp} s_j^{\perp})^2 \rangle$	1.54	1.39	1.69	1.65

This further suggests that for an arbitrary system, one may be able to do a rough estimate for an adequate r_s in RelRes just by considering the size of the respective molecules. Note though that the nature of the molecular interactions (e.g., Coulombic energetics) may substantially influence the value of the ideal switching distance; thus, for a certain molecular type, while one may find an approximation reminiscent of Eq. (76), its parameters may be quite different.

Finally, we also calculate thermal properties for these two systems. Analogous with the elementary dumbbells, we plot their average $\langle \bar{U} \rangle$ and variance $\langle \delta \bar{U}^2 \rangle$ in the top and bottom panels, respectively, of Fig. 6, with the orange (downward) triangles for $\ell = 0.3\sigma$ and the green (upward) triangles for $\ell = 0.7\sigma$. We again notice the fluctuating trend in all sets here. Importantly, notice that the magnitude of these fluctuations decreases with bond length. This reiterates the fact the RelRes works better for “spherelike” molecules.

3. A nonuniformity across the sites of the dumbbells

We now construct another system, whose molecules again have a bond length of 0.5σ . In this case, however, we are dealing with nonuniform dumbbells in terms of their LJ parameters, although they do still have equal mass across their sites; a sketch of such a molecule is given in Fig. 10. Elaborating on this nonuniformity, one FG site is small yet strong, and one FG site is large yet weak. Their respective length parameters are $6/(1 + \sqrt{37}) \approx 0.85$ and $(1 + \sqrt{37})/6 \approx 1.18$, in units of σ , and their respective energy parameters are $(1 + \sqrt{5})/2 \approx 1.61$ and $2/(1 + \sqrt{5}) \approx 0.62$, in units of ϵ ; notice that the mixed interaction between these two still has the standard σ and ϵ parameters. The relevant MPIL parametrization in this case study is set by Eqs. (63) and (64). In turn, we obtain that the CG site has LJ parameters that are roughly

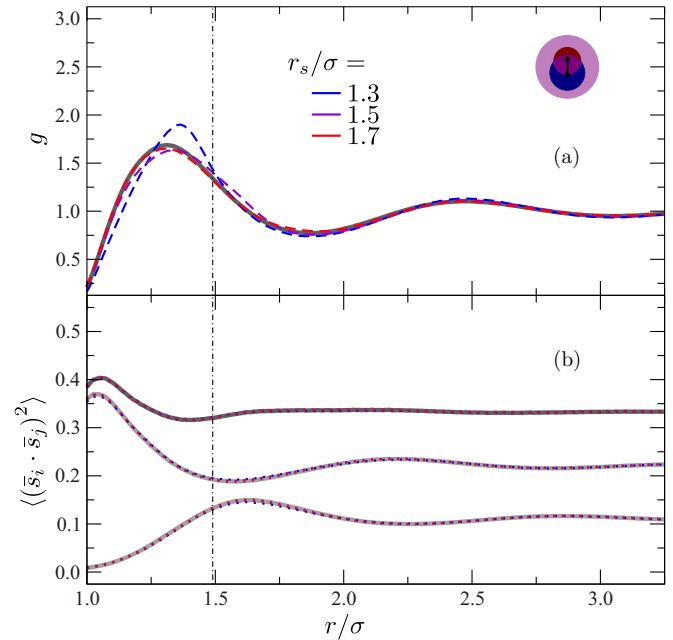


FIG. 10. Static correlations for the nonuniform dumbbells, the ones with different LJ parameters across their two sites $\{0.85\sigma, 1.18\sigma\}$ and $\{1.61\epsilon, 0.62\epsilon\}$. Everything here is essentially equivalent with Fig. 4. The vertical line is at 1.49.

1.08σ and 2.76ϵ . Importantly, realize that even though these dumbbells are nonuniform, the system itself is still uniform, being a single-component and single-phase liquid. Besides, the density of this system is $0.88m/\sigma^3$.

For this dumbbell system, we again perform a systematic examination for the switching distance. We present the corresponding structural correlations in Fig. 10, whose format is identical with that of Fig. 4. We again notice the same trends that we have observed for our elementary dumbbells. For the orientational functions of the bottom panel, the replication is perfect irrespective of the value of r_s , just as we found in all of our other scenarios. Most importantly, for the radial distributions in the top panel, we find that the asymptotic r_s is at 1.5σ , just as we have for our elementary dumbbells (i.e., $\ell = 0.5\sigma$). Once we determine all the critical distances in g , as well as in $\langle (\bar{s}_i \cdot \bar{s}_j)^2 \rangle$ with its components, we again find that they are all roughly the same as the suggested r_s for RelRes; these are summarized in Table I. Realize now that Eq. (76) still applies for this case study, even though we established it while only considering uniform dumbbells. Finally, once we invoke our functionals for the radial distributions, we show in Fig. 5 that for all r_s , S_{JS} for this nonuniform scenario (i.e., the almost hidden brown diamonds) is basically identical with its uniform counterpart (i.e., the indigo circles).

All of these observations for the structural correlations of the nonuniform dumbbells, particularly the fact that their relationship with the switching distance of RelRes is identical with that of our elementary dumbbells (i.e., those with the same bond length of 0.5σ , yet which are uniform across their sites), has significant ramifications. For an arbitrary molecule which has much nonuniformity within it, we do not need to perform any molecular simulations for it for determining its

ideal r_s ; all one must do is to construct a system of similar molecules that are strictly uniform across their sites (perhaps using the overall mean values for σ and ϵ of the nonuniform molecule). According to our observations, these uniform molecules, whose structural correlations are fairly feasible for computation, would roughly have the same r_s as the nonuniform molecules. This fact becomes especially useful once we are dealing with a set of molecules that roughly have the same topology, yet whose sites notably vary in their LJ parameters. If the molecules do not have the linear estimate for r_s as given in Eq. (76), we suggest that all one does is measure, via a single molecular simulation, the radial distribution of the reference liquid of the uniform version of these molecules, and we recommend using its inflection in g as the switching distance in RelRes for the entire family of molecules.

We now proceed with the thermal properties of the nonuniform scenario in Fig. 6, which is presented as brown diamonds. In this case, we actually do not observe the fluctuating trend which we have for all other dumbbell scenarios, and instead we have a slight asymptotic behavior; of course, it is very modest as compared with the entropic functionals of Fig. 5. Still, this is not necessarily a drawback for RelRes: If we look at our recommended r_s of 1.5σ , the average of the top panel is off by a ~ 0.1 fraction, yet the variance in the bottom panel is off by a negligible fraction. Of course, the former is still satisfactory, especially if one is not too interested in exact values of thermal properties. The discrepancy between the ability of RelRes in retrieving the average and variance may stem in the fact that the latter are related with response functions, and it is well known that these can be perfectly captured if structural correlations are also perfectly captured. In fact, the main message here is that while structural correlations of nonuniform molecules can be as feasible of retrieving as those of uniform molecules, nonuniform energetics involve much intricacies, and thus a perfect replication for them is rather difficult to achieve.

Finally, we examine the transport functions of this nonuniform scenario; we present them in Fig. 11, which is identical in format with Fig. 7. We again reaffirm our earlier observations. Foremost, RelRes perfectly captures the time behavior of the thermal property of the reference system, regardless of the switching distance. While the transport functions of the structural correlations are also adequately retrieved, there are some subtleties. For perfect replication, $\langle(\bar{r}_t - \bar{r}_0)^2\rangle$ of the top panel requires a relatively large r_s , yet $\langle\bar{s}_t \cdot \bar{s}_0\rangle$ of the bottom panel requires a relatively small r_s . In summary, just as RelRes is successful in capturing static behavior, it is also successful in capturing dynamic behavior, and we basically observe the same capability in describing the nonuniform dumbbells, as their uniform counterparts.

C. Butterfly systems

The next system which we explore is relatively complex, being representative of a plausible scenario that one may encounter in soft matter. Most of the parameters in these molecular simulations are identical with their dumbbell-like counterparts. However, the fundamental distinction here is that they are based on a butterflylike molecule, which is somewhat reminiscent of tetramethylethylene or dimethylbutadiene,

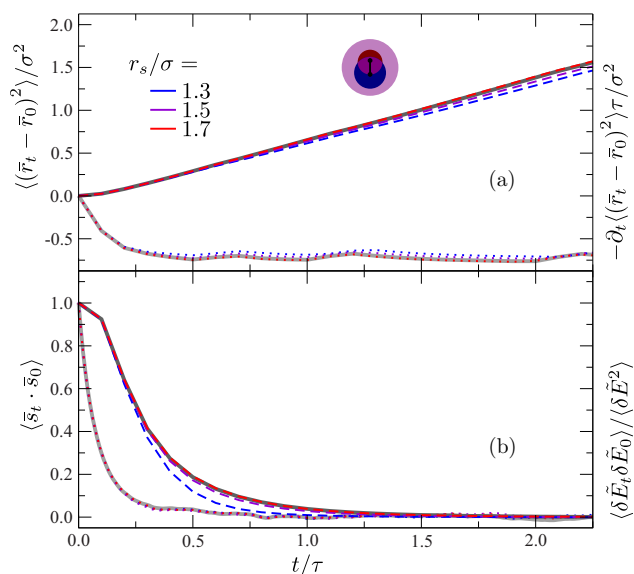


FIG. 11. Dynamic correlations for the nonuniform dumbbells, the ones with different LJ parameters across their two sites $\{0.85\sigma, 1.18\sigma\}$ and $\{1.61\epsilon, 0.62\epsilon\}$. Everything here is essentially equivalent with Fig. 7.

yet which has rotational movement about its principal axis; we present two variations of such a molecule in Fig. 12. All of the butterfly systems are single-component and single-phase liquids with a density of $1.49m/\sigma^3$. Note that, in our original communication, we did not examine this type of molecules, as well as the rather elaborate mapping which they involve.

Our main focus with the butterflies is in varying the mapping from their FG coordinates to their CG centers. Above all, we have two main classes in this case study. The naive mapping transforms between six FG sites and one CG site ($n = 6$ in RelRes), and the intricate mapping transforms

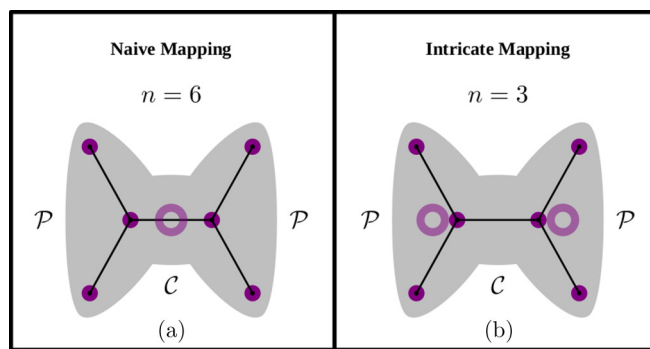


FIG. 12. Two butterflylike molecules restricted to a single plane. The gravitational centers are represented by hollow rings, and the atomistic coordinates are represented by replete disks. The left panel (a) depicts the naive mapping (i.e., $\mathcal{M} = \{1.0\}$), transforming from six sites to one site, and the right panel (b) depicts the intricate mapping transforming from six sites to two sites (i.e., $\mathcal{M} = \{1.0, 0.0\}$ and $\mathcal{M} = \{0.5, 0.5\}$). Note that the notation in both panels is the same. The gray shading does not have a physical meaning; it just delineates the shape of the molecules. The symbols \mathcal{P} and \mathcal{C} generically refer to the peripheral and central sections, respectively, of the molecules (i.e., sites, bonds, etc.).

between six FG sites and two CG sites ($n = 3$ in RelRes); we, respectively, present these two classes in the left and right panels of Fig. 12. The latter approach in essence partitions a butterfly into two groups. Importantly, we have here the set \mathcal{M} , which is the main parameter that distinguishes between the various mappings performed in this work; in Appendix B, we rigorously define this set in terms of the familiar mapping matrix, which converts from the positions of all FG sites to the positions of all CG sites [25,26]. The naive scenario is denoted by $\mathcal{M} = \{1.0\}$, and the intricate scenario is denoted by $\mathcal{M} = \{1.0, 0.0\}$ or $\mathcal{M} = \{0.5, 0.5\}$. We emphasize that the latter has two variations for \mathcal{M} ; we discuss later below the distinction between them. Besides, notice that the number of elements in this set corresponds with the number of groups the molecule contains. We also mention at this point that we present $r_s = 1.5\sigma$ and 2.0σ for the naive mapping and $r_s = 1.5\sigma$ for the intricate mapping. \mathcal{M} , together with r_s , uniquely specifies the RelRes system of our butterfly study.

Let us now thoroughly discuss the molecules given in Fig. 12. Represented by gray shading, we can think of a butterfly as composed of two congruent (isosceles) triangles, with each one of them being fairly rigid by itself. While the schematic depicts a given butterfly on a single plane, its two triangles can rotate about the axis which connects them. Each butterfly has a central section, denoted by \mathcal{C} , and a peripheral section, denoted by \mathcal{P} . We can think of the molecules, in both panels, as having two \mathcal{C} FG sites and four \mathcal{P} FG sites; we can think of the molecule, in the left panel, as having one \mathcal{C} CG site and of the molecule, in the right panel, as having two \mathcal{P} CG sites.

At the same time, in both panels of Fig. 12, we have four \mathcal{P} bonds (i.e., diagonal lines) and one \mathcal{C} bond (i.e., a straight line); the energetics of these bonds are identical with their dumbbell counterparts, having $\ell = 0.5\sigma$ together with a spring constant of $2000\epsilon/\sigma^2$. The butterflies also have fictitious restraints, which hold each pair of neighboring bonds in a triangular conformation; in molecular simulations, these are effectively treated as bonds, with the same spring constant, yet their inherent distance is set at $\sqrt{3}\ell$, which ensures that the corresponding angle is roughly $\frac{2\pi}{3}$ (i.e., a cosine of $-\frac{1}{2}$, just like in tetramethylethylene and dimethylbutadiene). Finally, the intramolecular potentials between all other FG sites are identical with their intermolecular counterparts, which are based on the LJ function. Realize that CG sites, even once there are several of them in a molecule, do not have any intramolecular energetics. For the ensuing analysis which we perform, remember that we consider the midpoint of each bond, together with its corresponding vector \bar{s} .

1. A transformation from six sites to one site for the butterflies

Of course, the most naive mapping for any molecule is collapsing all FG sites on a lone CG site. Specifically for the butterflies, this means going from six sites to one site, and it is exemplified in the left panel of Fig. 12. We specifically look at the most rudimentary weighing, in which the six FG coordinates equally contribute to the one CG center; this naive mapping is denoted by $\mathcal{M} = \{1.0\}$. Again, the FG sites have σ and ϵ for their LJ parameters. By MPIL, the CG sites have σ and 36ϵ for their LJ parameters, since $n = 6$. Just as in the

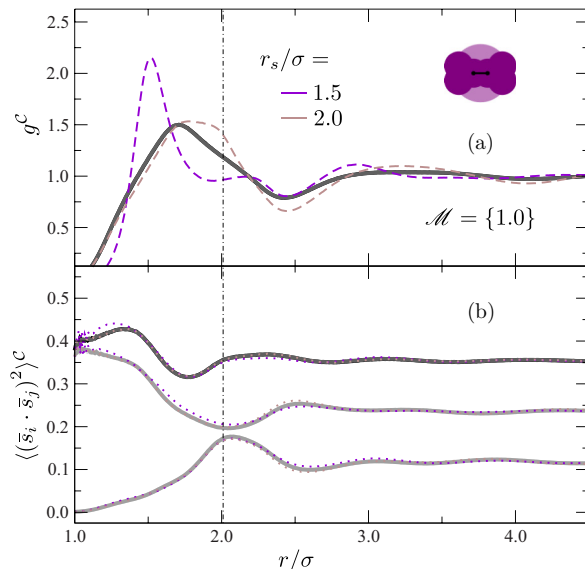


FIG. 13. Static correlations for the butterfly scenario in which we map from six FG sites to one CG site, which is denoted by $\mathcal{M} = \{1.0\}$. Realize that here g , as well as \bar{s} , corresponds with the central bond \mathcal{C} . Except for the coloring, most of the format here is identical with that of Fig. 4. Importantly, note here the brown curve for $r_s = 2.0\sigma$; just as in all other cases, the violet one is for $r_s = 1.5\sigma$. The vertical line is at 2.01.

dumbbell systems, we focus here on the effect of the switching distance on the efficacy of RelRes.

Radial distributions between the central sections \mathcal{C} of the butterflies are given in the top panel of Fig. 13. Just as with the dumbbell study, this structural correlation for the reference liquid is given here as the solid dark-gray curve. The dashed violet curve again represents the RelRes system with $r_s = 1.5\sigma$. While this is the recommended value for the dumbbells with the equivalent bond (i.e., $\ell = 0.5\sigma$), this r_s miserably fails for the butterflies of $\mathcal{M} = \{1.0\}$. This is especially clear once we compute the Jensen-Shannon entropy for this scenario, which is given in Table II; note that it

TABLE II. Metrics of structural correlations and thermal properties of the various RelRes systems with the butterfly molecules. Rows 1 and 2 define the RelRes scheme, in terms of \mathcal{M} together with r_s , which we are examining. Columns 2 and 3 are for the mapping from six FG sites to one CG site, with the only difference between them being the switching distance that we use. Columns 4 and 5 are for the mapping from six FG sites to two CG sites, with the only difference between them being the exact location of the latter sites. Column 1 then gives us the particular metric which we are focusing on. Rows 3 and 4 are for the Jensen-Shannon entropy of the radial distributions of the central and peripheral midpoints, while rows 5 and 6 are for the normalized average and variance of the configurational energy.

\mathcal{M}	{1.0}	{1.0}	{1.0, 0.0}	{0.5, 0.5}
r_s/σ	1.5	2.0	1.5	1.5
$10^3 \times S_{\text{JS}}^{\mathcal{C}}$	152.9	44.0	5.3	18.9
$10^3 \times S_{\text{JS}}^{\mathcal{P}}$	3.3	2.1	2.6	3.2
$\langle \bar{U} \rangle^*$	1.03	1.03	1.03	1.02
$\langle \delta \bar{U}^2 \rangle^*$	1.19	1.02	1.03	1.04

is about four orders of magnitude above the corresponding dumbbell value of Fig. 5. The deficiency here is actually of no surprise since the butterflies are fairly large while the dumbbells are fairly small. In any case, once we look at the corresponding orientational functions in the bottom panel of Fig. 13, we again observe the excellence of RelRes: the dotted violet curves of $r_s = 1.5\sigma$ are almost identical with their solid gray counterparts in the reference liquid. As such, we reaffirm with the butterfly systems that RelRes can better capture orientational correlations than translational correlations.

Remembering our discussion of the dumbbell systems in the context of Table I (i.e., how critical distances in structural correlations signal the ideal r_s), we try another switching distance for the butterflies. The inflection in the radial distribution of the reference system occurs at $r \approx 2.01\sigma$ (i.e., denoted by the vertical line in Fig. 13); note that the other critical distances mentioned in Table I are also located around this value. As such, we employ $r_s = 2.0\sigma$ in RelRes, presenting in the top panel of Fig. 13 its radial distribution (i.e., the dashed brown curve). This RelRes system obtains an adequate depiction of this structural correlation, approximately capturing the locations of its maxima and minima. We also note this in terms of the Jensen-Shannon entropies of Table II: increasing r_s by 0.5σ makes S_{JS} decrease by almost an order of magnitude. Once we go beyond this r_s , we do notice a further improvement in the description of the radial distribution, but in consideration of the number of neighbors beyond 2.0σ , we do not recommend going over this value for the switching distance in RelRes. In the bottom panel of Fig. 13, we also present the corresponding orientational functions of $r_s = 2.0\sigma$ (i.e., the dotted brown curves): We have almost flawless replication of this structural correlation, that is even slightly better than the scenario of 1.5σ . The ability of RelRes, with $r_s = 2.0\sigma$, of sufficiently describing all structural correlations basically reaffirms the argument we made earlier for the dumbbells: A switching distance, that equals the value of the relative separation between the maximum and minimum of the radial distribution of the reference liquid, enables RelRes in attaining adequate efficacy as a multiscale algorithm.

We also look at the structural correlations associated with the peripheral sections \mathcal{P} of the butterflies. In the top panel of Fig. 14, we present their radial distributions. These are actually captured almost flawlessly, with $r_s = 2.0\sigma$ doing just slightly better than $r_s = 1.5\sigma$. The excellent capability of RelRes for these structural correlations is in striking contrast to our observation in the top panel of Fig. 13. We also give the respective S_{JS} in Table II; the \mathcal{C} values are over one order of magnitude above their \mathcal{P} counterparts. We suspect that this discrepancy stems from the positioning between the relevant FG and CG sites. In the case of the \mathcal{C} functions, there is a significant perturbation here caused by the strong interaction of their respective site; in the case of the \mathcal{P} functions, there is no such perturbation, and only the weak interactions of their respective sites are marginally modified. Besides, note the angular distributions in the bottom panel of Fig. 14, which are again excellently retrieved.

Finally, we examine thermal properties of $\mathcal{M} = \{1.0\}$, presenting them in Table II. Rather surprisingly, despite the deficiency in describing all structural correlations, the system with $r_s = 1.5\sigma$ adequately retrieves the energetic metrics for

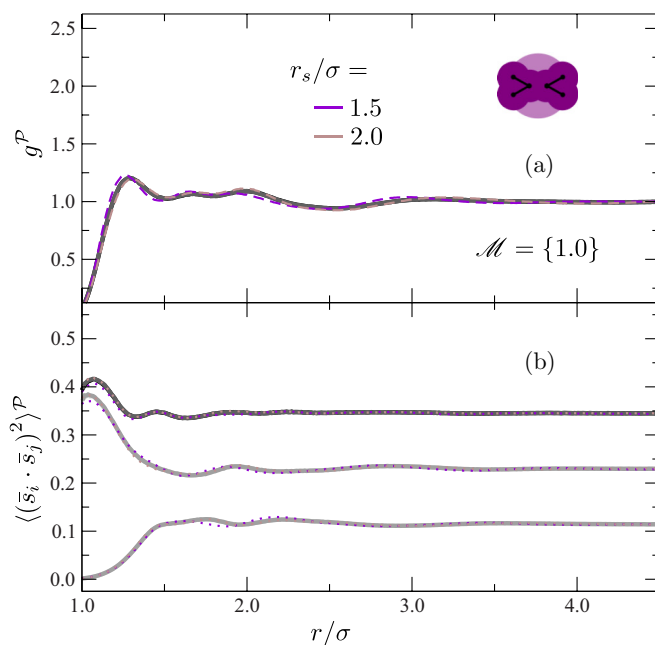


FIG. 14. Static correlations for the butterfly scenario in which we map from six FG sites to one CG site, which is denoted by $\mathcal{M} = \{1.0\}$. Everything here is essentially equivalent with Fig. 13, except that here g , as well as \bar{s} , corresponds with the peripheral bond \mathcal{P} .

the reference liquid: The replication is essentially perfect for the average $\langle \bar{U} \rangle^*$ and fairly decent for the variance $\langle \delta \bar{U}^2 \rangle^*$; it is natural that the latter is off by a rather noticeable fraction since such response functions are inherently tied to structural correlations. Regardless, once we employ $r_s = 2.0\sigma$ in RelRes, we capture almost flawlessly both of these functionals of the configurational energy. Just as we already mentioned in the dumbbell scenarios, the butterflies also exhibit a notable feasibility in capturing thermal properties, especially in comparison with radial distributions.

2. A transformation from six sites to two sites for the butterflies

In the above naive mapping of $\mathcal{M} = \{1.0\}$, RelRes is quite successful in capturing all structural correlations and thermal properties of the reference butterflies, once 2.0σ is employed for r_s . Nevertheless, this value for the switching distance is not too desirable since, during a molecular simulation, the interactions of not just the primary coordination must be fully included, but also the interactions of the secondary coordination must be partially included. Can we execute RelRes for butterflies in a way other than $\mathcal{M} = \{1.0\}$, while having the desired r_s of 1.5σ ?

We consequently perform an intricate mapping that is based on transforming from six to two sites, and this is depicted in the right panel of Fig. 12. In fact, we have two variants of this mapping: the two solely differ in the positioning between the two gravitational centers in terms of the six atomistic coordinates. The first mapping, denoted by $\mathcal{M} = \{1.0, 0.0\}$, considers just each triangle of Fig. 12 in isolation. A given triangle contains three FG sites, so we equally weigh them on one CG site. The second mapping,

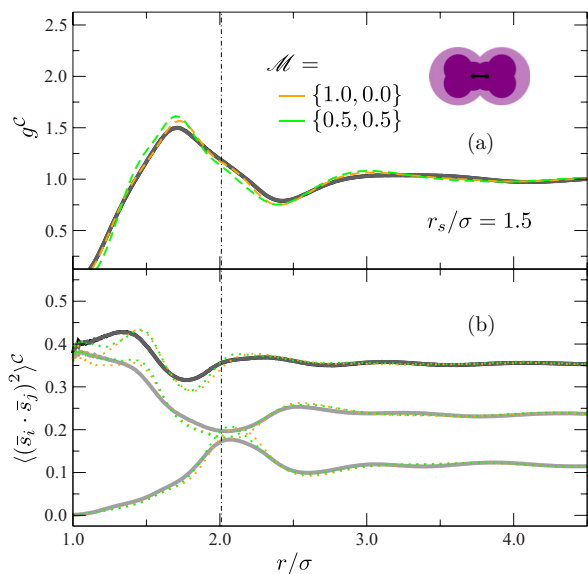


FIG. 15. Static correlations for the butterfly scenario in which we map from six FG sites to two CG sites. Realize that here g , as well as \bar{s} , corresponds with the central bond C . Again, much of the format here is analogous with that of Fig. 4. However, the coloring here is distinct, with each color relating to a subtle difference in the mapping. The orange case, denoted by $\{1.0, 0.0\}$, maps from three sites to one site, with all sites equally contributing their masses, while the green case, denoted by $\{0.5, 0.5\}$, maps from four sites to one site, with two of the former contributing just half of their mass for the latter. Note that in all of these RelRes systems, the switching distance is 1.5σ . Besides, since the reference system here is identical with that of Fig. 13, the vertical line is again at 2.01.

denoted by $\mathcal{M} = \{0.5, 0.5\}$, accounts also for the region between the two triangles of Fig. 12. In creating one CG site, we equally weigh the two \mathcal{P} FG sites of a given triangle, together with an equal partitioning of the two \mathcal{C} FG sites (these two effectively count as one site). Again, the FG sites have σ and ϵ for their LJ parameters. By MPIL, the CG sites have σ and 9ϵ for their LJ parameters since $n = 3$. Besides for both, we fix the switching distance at 1.5σ .

In Fig. 15, we present static features which are associated with the structural correlations of this intricate mapping. As usual, the gray curves are for the reference liquid; these functions are identical with their counterparts of Fig. 13. The orange curves are for $\mathcal{M} = \{1.0, 0.0\}$, while the green curves are for $\mathcal{M} = \{0.5, 0.5\}$. In the top panel, both RelRes systems excellently capture radial distributions, especially in comparison with the earlier naive mapping of $\mathcal{M} = \{1.0\}$. In the bottom panel, both RelRes systems adequately describe angular distributions, yet not as flawlessly as with the earlier naive mapping of $\mathcal{M} = \{1.0\}$. Furthermore, note that $\mathcal{M} = \{1.0, 0.0\}$ does slightly better with radial distributions, but $\mathcal{M} = \{0.5, 0.5\}$ does slightly better with angular distributions. The various entropic values of Table II reaffirm the observations we make here for Fig. 15. Table II also shows that both $\mathcal{M} = \{1.0, 0.0\}$ and $\{0.5, 0.5\}$ capture thermal properties with negligible error.

The various discrepancies between the mappings obviously stem in the locations of the CG sites in terms of the

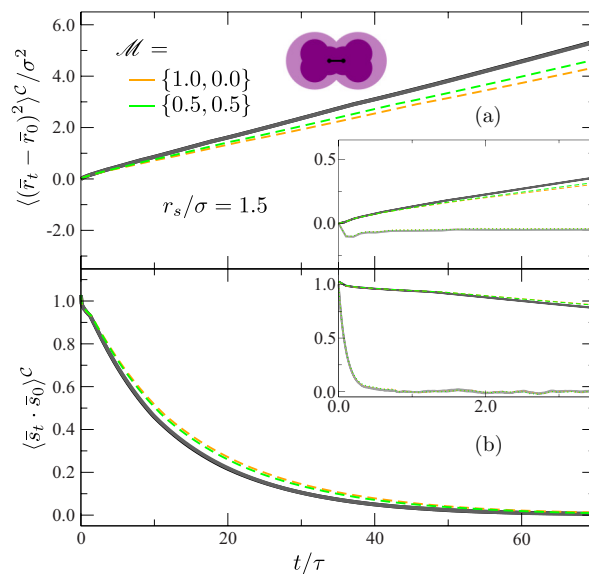


FIG. 16. Dynamic correlations for the butterfly scenario in which we map from six FG sites to two CG sites. Realize that here, everything corresponds with the central bond C . While the coloring of the curves here is the same as in Fig. 15, the format, together with all of the functions, is basically equivalent with Fig. 7. The key distinction here is that there are inset panels: these just focus on the functions at short times, while the main panels probe the functions at long times. As such, a dark-gray curve is identical between an inset panel and its main panel; for clarity, the light-gray curves are omitted in the latter.

FG sites. For the ensuing discussion, think of $\mathcal{M} = \{1.0\}$ also as a mapping from six FG to two CG sites. In the context of such a generalized mapping, while the FG sites are fixed, the CG sites slide back and forth on the middle axis of the butterflies. As the distance between the two \mathcal{P} CG sites decreases, the suction effect in the middle between the two \mathcal{C} FG sites increases. An increase in this suction squeezes neighbors, erroneously strengthening translational correlations; conversely, a decrease in this suction loosens neighbors, erroneously weakening orientational correlations. Such logic can particularly explain the slight differences of the $\mathcal{M} = \{1.0, 0.0\}$ and $\{0.5, 0.5\}$ structural correlations, as well as the substantial differences of these intricate mappings with the naive mapping of $\mathcal{M} = \{1.0\}$.

Finally, in Fig. 16, we present dynamic features that are associated with this intricate mapping. The color coding here is identical with that of Fig. 15, while the format is actually analogous with Fig. 7. However, Fig. 16 has an inset in each panel, which zooms on the same functions over a slight portion of the domain. Just as in the dumbbell systems, we again show that RelRes with $\mathcal{M} = \{1.0, 0.0\}$ and $\{0.5, 0.5\}$ can excellently describe the dynamic functions of the butterflies.

Of course, as we showed above, the mapping from six sites to one site must be employed together with $r_s = 2.0\sigma$ for a sufficient performance of RelRes. A much better description of the reference liquid is obtained with either variation of the intricate mapping, from six to two sites, with $r_s = 1.5\sigma$. It may thus offer the most computationally feasible route for RelRes. In a molecular simulation of the latter case, we

obviously have less FG interactions, in consideration of the value of the switching distance, but more CG interactions, in consideration of the dimensions of the mapping matrix.

IV. CONCLUSION

In this current work, we continue our initial effort [37], presenting in turn a comprehensive picture of RelRes. As shown in our original communication, our hybrid formalism can be thought of as an extension of the common approach in statistical mechanics, which truncates all forces beyond a certain distance, replacing them with a mean field [40–43]. The main aspect of RelRes is that molecular resolution switches in terms of relative separation: each molecule embodies both FG and CG models, and a given molecule then interacts with near neighbors via its FG potential and with far neighbors via its CG potential. Via energy conservation, the FG and CG potentials can be connected by an arithmetic parametrization, and we call this analytical expression MPIL. In turn, we already showed that RelRes with MPIL describes very well the structural and thermal behavior of nonpolar multicomponent and multiphase mixtures, as complex as a generic tetrachloromethane-thiophene system which contains a vacuum cavity [37].

In this paper, we specifically define MPIL in terms of a multipole approximation. Reminiscent of the familiar scenario for the Coulombic interaction, we do this derivation by assuming that the potential can be cast as a linear combination of inverse powers of the relative separation, and we then perform the corresponding Taylor expansion. We define the leading nonzero term as the MPIL (\mathcal{N}^* in our notation), and we show that it naturally fits for use with RelRes. Note that in our original publication, we naively assumed an infinite limit for deriving the MPIL [37]; in the current work, we show that it is identical with the zero-order term of the multipole expansion. Regardless, the derivation here is in fact the comprehensive one, and we notably present the first and second terms of the Taylor series. In general, these latter terms are necessary for polar molecules, and they may be also used for correction purposes in nonpolar molecules.

Furthermore, we extend here our earlier results for nonpolar liquids, exhaustively showing via molecular simulations that RelRes with MPIL is successful in capturing not just statics but also dynamics. On a fundamental level, we notably focus on systematically examining the role of the switching distance in uniform liquids of dumbbell-like molecules. Mapping from two FG sites to one CG site, we importantly find that the appropriate distance for switching between the FG and CG potentials corresponds with the relative separation between the dumbbells at which their molecular orientation decorrelates, and this is about equivalent with the boundary between their respective primary and secondary coordination shells. We also examine butterflylike molecules, in which we vary the type of mapping that we employ: We successfully map not just from six FG sites to one CG site, but also from six FG sites to two CG sites. The success of these case studies suggests that RelRes, with just the zero-order term in the multipole approximation, can adequately describe the structural and thermal behavior of other nonpolar systems as well (carbon dioxide, benzene, etc.).

We now thoroughly describe a practical implementation of RelRes in molecular simulations, focusing on the LJ potential. In the general case for RelRes with Newtonian evolution in time, the FG interaction between its actual coordinates is set by Eq. (44), together with Eq. (51), and the CG interaction between its virtual centers is set by Eq. (45), together with Eq. (54). While the parameters of the FG models can be readily found in various packages, the parameters of the CG models can be readily retrieved through the MPIL expressions. In the LJ case, it is given by Eqs. (55) and (56). If the product rule for the FG coefficients applies [i.e., Eqs. (57) and (58)], then the product rule for the CG coefficients applies [i.e., Eqs. (61) and (62)]; in turn, one can use instead Eqs. (63) and (64) for parametrizing between the various LJ parameters. Importantly, if all LJ interactions among molecules are identical, we have an immense simplification: the relevant parameters are Eqs. (65) and (66) for the FG interaction and Eqs. (67) and (68) for the CG interaction. Otherwise, if the RelRes system evolves via the Monte Carlo approach, the step functions of Eqs. (42) and (43) can be used instead of Eqs. (44) and (45), respectively.

The main goal of our work was in showing that RelRes with MPIL can correctly capture the behavior of various liquids. As such, we did not explore the computational efficiency associated with the hybrid approach. Indeed, our implementation here in GROMACS, which uses a single-neighbor list for both FG and CG sites, did not formally reduce the computational time of the molecular simulations. Nevertheless, if two independent neighbor lists, with two different cutoff values, are separately implemented for the FG and CG models, computational efficiency can be actually realized, and this is despite the fact that RelRes maintains all degrees of freedom of the reference system. The main reason behind this stems in the fact that we generally find that the ideal r_s is roughly between the primary and secondary coordination shells. Such a switching distance means that an arbitrary FG site must consider just the interactions with its $O(10)$ nearest neighbors, yet an arbitrary CG site must consider all the interactions with its $O(100)$ other neighbors. RelRes is thus significantly dominated by the calculation associated with gravitational centers rather than atomistic coordinates. Continuing this logic specifically for molecules of mapping n , if neighbor lists are implemented separately for the FG and CG sites, the efficiency of a molecular simulation can roughly improve by a factor of n (remember that, with neighbor lists, the computational cost scales linearly with the total number of sites). The entire argument here does rely on the fact that r_s is roughly equal to $1\frac{1}{2}$ (in dimensionless units). This readily gives satisfactory results for small molecules (e.g., our dumbbells) yet not for large molecules (e.g., our butterflies). Thus, once one encounters oligomeric or polymeric systems, one must partition a molecule into several groups, as we have done in the later case study (i.e., a mapping of 6:2 instead of 6:1). On a heuristic level for partitioning, we have noticed that RelRes with MPIL yields sufficient results if the distance between the FG and CG sites is about $\frac{1}{2}$ (in dimensionless units). In summary, regarding the computational cost, although RelRes with MPIL does not reduce the number of molecular sites, it does reduce the number of molecular interactions,

implying that our multiscale framework can make molecular simulations quite efficient.

The crucial course of action for RelRes is implementing the multiscale protocol for polar systems, specifically for aqueous solutions. Importantly, all of the necessary mathematical ingredients have been already derived here. Regarding neutral water, whose Coulombic monopole is strictly zero, Eq. (34) yields $c_{ij,1} = 0$, and Eq. (37) yields $c_{ij,2} = \frac{1}{r_{ij}^2}[(\bar{\mathbf{p}}_{i,1} \cdot \bar{\mathbf{p}}_{j,1}) - 3(\hat{r}_{ij} \cdot \bar{\mathbf{p}}_{i,1})(\hat{r}_{ij} \cdot \bar{\mathbf{p}}_{j,1})]$. This MPIL is in line with various multipolar models of water [51,52], except that in the context of RelRes, the interaction is actually hybrid, being a function of pairwise distance. At the same time, RelRes for hydrated ions is rather straightforward since the corresponding Coulombic monopoles are finite, and Eq. (28) becomes the applicable one for parametrization. Keep in mind that with polar molecules, the ideal r_s for the Coulombic potential may not necessarily equal the ideal r_s for the LJ potential; however, for facilitating the implementation of the neighbor lists, we recommend using a single switching distance that encompasses both of these values. Realize also that with RelRes, one must still deal with Coulombic interactions beyond the periodic box, for example, via Ewald summation. Besides, as we move to aqueous systems that are representative of biology, complex mappings may be required (e.g., from 200 FG sites to 50 CG sites).

Of course, RelRes is only one route for enhancing the efficiency of molecular simulations, and we recommend that one employs it concurrently with other computational approaches for optimal results. For biological systems (e.g., the assembly of proteins in a water medium), we expect that combining RelRes with Adaptive Resolution may be particularly useful. While RelRes is ideal for modeling the binding sites of proteins, Adaptive Resolution is useful for eliminating many uninteresting degrees of freedom which bulk water possesses. In fact, various explicit-implicit approaches, reminiscent of Adaptive Resolution [53], can further facilitate such a task. These algorithms switch from an explicit solute near to the origin to an implicit solvent far from the origin [54] and, thus, they can also substantially reduce the computational time required for molecular simulations.

Interestingly, RelRes with MPIL bears much resemblance with the famous “cell-multipole” formalism, the algorithm which is frequently mentioned as one of the most computationally powerful methods [55]. For an arbitrary potential of an inverse power, the “cell-multipole” approach also invokes a multipole approximation at appropriate distances: Inside a “cell” at small separations, one evaluates interactions between discrete points, and outside a “cell” at large separations, one evaluates interactions between continuum patches. A crucial distinction with our multiscale scheme is that the “cell-multipole” strategy considers interaction sites that move freely of each other, in turn lumping them together in arbitrary “cells” in space. In molecular simulations, interaction sites do not move freely of each other as they are constrained by bonds; RelRes, just as other multiscale approaches [9–24,38,39], consequently invokes the natural mapping embodied by molecules, connecting FG atomistic coordinates with CG gravitational centers. As such, one can think of RelRes with MPIL as a natural modification of the

“cell-multipole” approach for molecular systems. This can be extremely useful since, despite the fact that the “cell-multipole” method is frequently used in a myriad of systems, it is rarely employed in molecular fluids. In summary, RelRes can become a beneficial tool for molecular simulations, and it has already been acknowledged as such by several review articles [56,57].

ACKNOWLEDGMENTS

Foremost, we are thankful for partial funding by the Alexander von Humboldt Foundation, as well as the financial support of the Melvin J. Berman Hebrew Academy. We also appreciate the insightful conversations with D. Andrienko regarding polar systems. We further thank A. Markina and O. Valsson for constructive comments regarding the manuscript. Finally, we are grateful for the biological consultations of M. Wilson.

APPENDIX A: THEORETICAL SECTION

For proceeding beyond the monopole-monopole term of Eq. (30), we must define dimensionless variables, reminiscent of Eqs. (3) and (4). Here, they are

$$\bar{\xi}_{v_l} = \frac{\bar{\Delta}_{v_l}}{r}, \quad (\text{A1})$$

$$\cos \theta_{v_l} = \frac{\bar{r} \cdot \bar{\Delta}_{v_l}}{r \Delta_{v_l}}. \quad (\text{A2})$$

For compactness, we mostly use here the index v , together with the index l : the Latin index corresponds with either i or j , and the Greek index corresponds with either μ or ν . Besides, we also frequently omit yet imply the combination of the ij indices (e.g., $\bar{r} = \bar{r}_{ij}$). Using these, together with Eqs. (3) and (4), we derive useful expressions for the dimensionless variables involved in the energy function of Eq. (10):

$$\xi_{\mu_i \nu_j} \cos \theta_{\mu_i \nu_j} = \xi_{\nu_j} \cos \theta_{\nu_j} - \xi_{\mu_i} \cos \theta_{\mu_i}, \quad (\text{A3})$$

$$\xi_{\mu_i \nu_j}^2 = \xi_{\nu_j}^2 - 2(\bar{\xi}_{\nu_j} \cdot \bar{\xi}_{\mu_i}) + \xi_{\mu_i}^2. \quad (\text{A4})$$

Note that we also used here the definition of $\bar{\Delta}_{\mu_i \nu_j}$ of Eq. (2).

Let us now proceed by invoking these variables in $c_{ij,\aleph}$. For $\aleph = 1$, we specifically substitute Eq. (A3) in Eq. (21), attaining

$$\begin{aligned} c_{ij,1}(\{\xi_{\mu_i \nu_j}\}, \{\cos \theta_{\mu_i \nu_j}\}) \\ = m \sum_{\mu_i \nu_j} [c_{\mu_i \nu_j} \xi_{\nu_j} \cos \theta_{\nu_j}] - m \sum_{\mu_i \nu_j} [c_{\mu_i \nu_j} \xi_{\mu_i} \cos \theta_{\mu_i}] \end{aligned} \quad (\text{A5})$$

and by invoking the product assumption of Eq. (26), we derive this:

$$\begin{aligned} c_{ij,1} = m \left(\sum_{\mu_i} c_{\mu_i} \right) \left(\sum_{\nu_j} [c_{\nu_j} \xi_{\nu_j} \cos \theta_{\nu_j}] \right) \\ - m \left(\sum_{\nu_j} c_{\nu_j} \right) \left(\sum_{\mu_i} [c_{\mu_i} \xi_{\mu_i} \cos \theta_{\mu_i}] \right). \end{aligned} \quad (\text{A6})$$

For facilitating the ensuing mathematics, we introduce now $c_{ij,2p^\pm}$ and $c_{ij,2q^\pm}$, whose summation yields $c_{ij,2^\pm}$:

$$c_{ij,2^\pm} = c_{ij,2p^\pm} + c_{ij,2q^\pm}. \quad (\text{A7})$$

For $\aleph = 2^+$, we specifically substitute Eq. (A3) in Eq. (23), attaining

$$c_{ij,2p^+}(\{\xi_{\mu_i v_j}\}, \{\cos \theta_{\mu_i v_j}\}) = -m \sum_{\mu_i v_j} [c_{\mu_i v_j} \xi_{v_j} \xi_{\mu_i} \cos \theta_{v_j} \cos \theta_{\mu_i}], \quad (\text{A8})$$

$$c_{ij,2q^+}(\{\xi_{\mu_i v_j}\}, \{\cos \theta_{\mu_i v_j}\}) = \frac{m}{2} \sum_{\mu_i v_j} [c_{\mu_i v_j} \xi_{v_j}^2 \cos^2 \theta_{v_j}] + \frac{m}{2} \sum_{\mu_i v_j} [c_{\mu_i v_j} \xi_{\mu_i}^2 \cos^2 \theta_{\mu_i}] \quad (\text{A9})$$

and by invoking the product assumption of Eq. (26), we derive this:

$$c_{ij,2p^+} = -m \left(\sum_{v_j} [c_{v_j} \xi_{v_j} \cos \theta_{v_j}] \right) \left(\sum_{\mu_i} [c_{\mu_i} \xi_{\mu_i} \cos \theta_{\mu_i}] \right), \quad (\text{A10})$$

$$c_{ij,2q^+} = \frac{m}{2} \left(\sum_{\mu_i} c_{\mu_i} \right) \left(\sum_{v_j} [c_{v_j} \xi_{v_j}^2 \cos^2 \theta_{v_j}] \right) + \frac{m}{2} \left(\sum_{v_j} c_{v_j} \right) \left(\sum_{\mu_i} [c_{\mu_i} \xi_{\mu_i}^2 \cos^2 \theta_{\mu_i}] \right). \quad (\text{A11})$$

For $\aleph = 2^-$, we specifically substitute Eq. (A4) in Eq. (24), attaining

$$c_{ij,2p^-}(\{\xi_{\mu_i v_j}\}, \{\cos \theta_{\mu_i v_j}\}) = -m \sum_{\mu_i v_j} [c_{\mu_i v_j} (\bar{\xi}_{v_j} \cdot \bar{\xi}_{\mu_i})], \quad (\text{A12})$$

$$c_{ij,2q^-}(\{\xi_{\mu_i v_j}\}, \{\cos \theta_{\mu_i v_j}\}) = \frac{m}{2} \sum_{\mu_i v_j} [c_{\mu_i v_j} \xi_{v_j}^2] + \frac{m}{2} \sum_{\mu_i v_j} [c_{\mu_i v_j} \xi_{\mu_i}^2] \quad (\text{A13})$$

and by invoking the product assumption of Eq. (26), we derive this:

$$c_{ij,2p^-} = -m \left[\left(\sum_{v_j} [c_{v_j} \bar{\xi}_{v_j}] \right) \cdot \left(\sum_{\mu_i} [c_{\mu_i} \bar{\xi}_{\mu_i}] \right) \right], \quad (\text{A14})$$

$$c_{ij,2q^-} = \frac{m}{2} \left(\sum_{\mu_i} c_{\mu_i} \right) \left(\sum_{v_j} [c_{v_j} \xi_{v_j}^2] \right) + \frac{m}{2} \left(\sum_{v_j} c_{v_j} \right) \left(\sum_{\mu_i} [c_{\mu_i} \xi_{\mu_i}^2] \right). \quad (\text{A15})$$

While the monopoles of Eq. (29) can be readily substituted in most of the above expressions, further manipulation must be performed for employing here the dipoles, as well as the quadrupoles.

As such, let us evaluate some of the summations which appear above, by invoking the dimensionless variables of Eqs. (A1) and (A2). Here are summations which can be conveniently cast in terms of the dipoles of Eq. (31):

$$\sum_{v_l} [c_{v_l} \xi_{v_l} \cos \theta_{v_l}] = \frac{1}{r^2} \sum_{v_l} [c_{v_l} (\bar{r} \cdot \bar{\Delta}_{v_l})] = \frac{1}{r^2} \left[\bar{r} \cdot \left(\sum_{v_l} [c_{v_l} \bar{\Delta}_{v_l}] \right) \right] = \frac{1}{r} (\hat{r} \cdot \bar{p}_{l,1}), \quad (\text{A16})$$

$$\sum_{v_l} [c_{v_l} \bar{\xi}_{v_l}] = \frac{1}{r} \sum_{v_l} [c_{v_l} \bar{\Delta}_{v_l}] = \frac{1}{r} \bar{p}_{l,1}, \quad (\text{A17})$$

while here are summations which can be conveniently cast in terms of the quadrupoles of Eqs. (32) and (33):

$$\sum_{v_l} [c_{v_l} \xi_{v_l}^2 \cos^2 \theta_{v_l}] = \frac{1}{r^4} \sum_{v_l} [c_{v_l} (\bar{r} \cdot \bar{\Delta}_{v_l})^2] = \frac{1}{r^4} \left[\bar{r} \bar{r} : \left(\sum_{v_l} [c_{v_l} \bar{\Delta}_{v_l} \bar{\Delta}_{v_l}] \right) \right] = \frac{1}{r^2} (\hat{r} : \bar{q}_{l,2}), \quad (\text{A18})$$

$$\sum_{v_l} [c_{v_l} \bar{\xi}_{v_l}^2] = \frac{1}{r^2} \sum_{v_l} [c_{v_l} \Delta_{v_l}^2] = \frac{1}{r^2} q_{l,0}; \quad (\text{A19})$$

note that we used here a common tensor identity $(\bar{r} \cdot \bar{\Delta}_{v_l})^2 = (\bar{r} \bar{r} : \bar{\Delta}_{v_l} \bar{\Delta}_{v_l})$.

By employing these, together with Eq. (29), we can evaluate the relevant $c_{ij,\aleph}$. For $\aleph = 1$, substituting Eq. (A16) in Eq. (A6), we obtain the following:

$$c_{ij,1} = \frac{m}{r} c_{i,0} (\hat{r} \cdot \bar{p}_{j,1}) - \frac{m}{r} c_{j,0} (\hat{r} \cdot \bar{p}_{i,1}). \quad (\text{A20})$$

This is compactly presented in the main text as Eq. (34). For $\aleph = 2^+$, substituting Eqs. (A16) and (A18) in Eqs. (A10) and (A11), respectively, we obtain the following:

$$c_{ij,2p^+} = -\frac{m}{r^2} (\hat{r} \cdot \bar{p}_{j,1}) (\hat{r} \cdot \bar{p}_{i,1}), \quad (\text{A21})$$

$$c_{ij,2q^+} = \frac{m}{2r^2} c_{i,0} (\hat{r} : \bar{q}_{j,2}) + \frac{m}{2r^2} c_{j,0} (\hat{r} : \bar{q}_{i,2}). \quad (\text{A22})$$

For $\aleph = 2^-$, substituting Eqs. (A17) and (A19) in Eqs. (A14) and (A15), respectively, we obtain the following:

$$c_{ij,2p^-} = -\frac{m}{r^2} (\bar{p}_{j,1} \cdot \bar{p}_{i,1}), \quad (\text{A23})$$

$$c_{ij,2q^-} = \frac{m}{2r^2} c_{i,0} q_{j,0} + \frac{m}{2r^2} c_{j,0} q_{i,0}. \quad (\text{A24})$$

Invoking Eq. (A7), these are compactly presented in the main text as Eqs. (35) and (36). Besides, in an analogous manner which we performed here for the first and second terms of $c_{ij,\aleph}$, other terms of this coefficient can be evaluated as well.

APPENDIX B: COMPUTATIONAL SECTION

Before defining the mapping matrix, let us introduce some notation that better elaborates on the topology of the

butterflies in Fig. 12. In each panel, we denote the left triangle by \mathcal{L} and the right triangle by \mathcal{R} , and for each triangle, we denote its upward portion by \mathcal{U} and its downward portion by \mathcal{D} ; besides, we use \mathcal{M} for the middle axis of the entire molecule. We can now introduce indices for the various \mathcal{P} and \mathcal{C} sites. We call, in both panels, the two central FG sites $\{\mathcal{C}_{\mathcal{L},\mathcal{M}}, \mathcal{C}_{\mathcal{R},\mathcal{M}}\}$ and the four peripheral FG sites $\{\mathcal{P}_{\mathcal{L},\mathcal{U}}, \mathcal{P}_{\mathcal{L},\mathcal{D}}, \mathcal{P}_{\mathcal{R},\mathcal{U}}, \mathcal{P}_{\mathcal{R},\mathcal{D}}\}$; we call, in the left panel, the one central CG site $\{\mathcal{C}_{\mathcal{M}}\}$ and, in the right panel, the two peripheral CG sites $\{\mathcal{P}_{\mathcal{L}}, \mathcal{P}_{\mathcal{R}}\}$. Realize that in our notation, a FG site is always labeled by a binary index, and a CG site is always labeled by a unitary index. As sketched in Fig. 12, the larger index (e.g., \mathcal{L} or \mathcal{R}) is for the horizontal location in the molecule, and the smaller index (e.g., \mathcal{U} or \mathcal{D}) is for the vertical location in the molecule; if the latter index is omitted, this means that the site is on the middle axis of the butterfly.

We now define the familiar mapping matrix \mathbb{M} for translating between the positions of the FG and CG sites. The mapping operator in the naive case, which transforms from six sites to one site, can be cast as the following:

$$\mathbb{M} = [w_{\mathcal{L},\mathcal{U},\mathcal{M}} \ w_{\mathcal{L},\mathcal{D},\mathcal{M}} \ w_{\mathcal{L},\mathcal{M},\mathcal{M}} \ w_{\mathcal{R},\mathcal{M},\mathcal{M}} \ w_{\mathcal{R},\mathcal{U},\mathcal{M}} \ w_{\mathcal{R},\mathcal{D},\mathcal{M}}] \quad (\text{B1})$$

and the mapping operator in the intricate case, which transforms from six sites to two sites, can be cast as the following:

$$\mathbb{M} = \begin{bmatrix} w_{\mathcal{L},\mathcal{U},\mathcal{L}} & w_{\mathcal{L},\mathcal{D},\mathcal{L}} & w_{\mathcal{L},\mathcal{M},\mathcal{L}} & w_{\mathcal{R},\mathcal{M},\mathcal{L}} & w_{\mathcal{R},\mathcal{U},\mathcal{L}} & w_{\mathcal{R},\mathcal{D},\mathcal{L}} \\ w_{\mathcal{L},\mathcal{U},\mathcal{R}} & w_{\mathcal{L},\mathcal{D},\mathcal{R}} & w_{\mathcal{L},\mathcal{M},\mathcal{R}} & w_{\mathcal{R},\mathcal{M},\mathcal{R}} & w_{\mathcal{R},\mathcal{U},\mathcal{R}} & w_{\mathcal{R},\mathcal{D},\mathcal{R}} \end{bmatrix}. \quad (\text{B2})$$

Each w is the mass fraction of a specific FG site, denoted by the first (binary) index, that weighs in a specific CG site, denoted by the second (unitary) index. These mapping matrices are obviously presented here for a single molecule. Their columns are for the FG sites, and their rows are for the CG sites; n equals the ratio between the number of columns and the number of rows. For purposes of abbreviated notation, we introduce the useful set \mathcal{M} which we mentioned in the main text. In particular, this set is defined as follows for the

naive scenario

$$\mathcal{M} = \{3(w_{\mathcal{L},\mathcal{M},\mathcal{M}} + w_{\mathcal{R},\mathcal{M},\mathcal{M}})\}, \quad (\text{B3})$$

and as follows for the intricate case

$$\mathcal{M} = \{\frac{3}{2}(w_{\mathcal{L},\mathcal{M},\mathcal{L}} + w_{\mathcal{R},\mathcal{M},\mathcal{R}}), \frac{3}{2}(w_{\mathcal{L},\mathcal{M},\mathcal{R}} + w_{\mathcal{R},\mathcal{M},\mathcal{L}})\}. \quad (\text{B4})$$

The numerical constants 3 and $\frac{3}{2}$, in these linear combinations, just guarantee, for our current work, that the sum of all elements in a particular \mathcal{M} equals unity. In any case, the value of each element here just describes the relationship between the positions of all sites along the middle axis of the butterfly.

How about the choice for the weights in the above expressions? For the naive mapping, we just look at the case in which all sites have an equal contribution, so the mapping operator is the following,

$$\mathbb{M} = [\frac{1}{6} \ \frac{1}{6} \ \frac{1}{6} \ \frac{1}{6} \ \frac{1}{6} \ \frac{1}{6}], \quad (\text{B5})$$

which in turn, by Eq. (B3), means:

$$\mathcal{M} = \{1.0\}. \quad (\text{B6})$$

In the intricate scenarios, we have two types of weighting, with the mapping matrices having slightly different values between their elements. Here are the two variations of our mapping operator,

$$\mathbb{M} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}, \quad (\text{B7})$$

$$\mathbb{M} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} & 0 & 0 \\ 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}, \quad (\text{B8})$$

and by Eq. (B4), we respectively have:

$$\mathcal{M} = \{1.0, 0.0\}, \quad (\text{B9})$$

$$\mathcal{M} = \{0.5, 0.5\}. \quad (\text{B10})$$

The meaning of the label \mathcal{M} is now clarified: its first element signifies the fraction a given \mathcal{C} FG weighs in a \mathcal{P} CG of the same side, and its second element signifies the fraction a given \mathcal{C} FG weighs in a \mathcal{P} CG of the opposite side.

-
- [1] C. Jarzynski, Nonequilibrium Equality for Free Energy Differences, *Phys. Rev. Lett.* **78**, 2690 (1997).
- [2] F. Wang and D. P. Landau, Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States, *Phys. Rev. Lett.* **86**, 2050 (2001).
- [3] A. Laio and M. Parrinello, Escaping free-energy minima, *Proc. Natl. Acad. Sci. USA* **99**, 12562 (2002).
- [4] O. Valsson and M. Parrinello, Variational Approach to Enhanced Sampling and Free Energy Calculations, *Phys. Rev. Lett.* **113**, 090601 (2014).
- [5] C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, Transition path sampling and the calculation of rate constants, *J. Chem. Phys.* **108**, 1964 (1998).
- [6] R. J. Allen, P. B. Warren, and P. R. ten Wolde, Sampling Rare Switching Events in Biochemical Networks, *Phys. Rev. Lett.* **94**, 018104 (2005).
- [7] T. S. van Erp, Reaction Rate Calculation by Parallel Path Swapping, *Phys. Rev. Lett.* **98**, 268301 (2007).
- [8] Weinan E, W. Ren, and E. Vanden-Eijnden, String method for the study of rare events, *Phys. Rev. B* **66**, 052301 (2006).
- [9] M. Scott Shell, Coarse-graining with the relative entropy, *Adv. Chem. Phys.* **161**, 395 (2016).
- [10] W. G. Noid, J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen, The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models, *J. Chem. Phys.* **128**, 244114 (2008).
- [11] W. G. Noid, P. Liu, Y. Wang, J.-W. Chu, G. S. Ayton, S. Izvekov, H. C. Andersen, and G. A. Voth, The multi-scale coarse-graining method. II. Numerical implementation for coarse-grained molecular models, *J. Chem. Phys.* **128**, 244115 (2008).

- [12] R. L. Henderson, A uniqueness theorem for fluid pair correlation functions, *Phys. Lett. A* **49**, 197 (1974).
- [13] A. P. Lyubartsev and A. Laaksonen, Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach, *Phys. Rev. E* **52**, 3730 (1995).
- [14] D. Reith, M. Putz, and F. Muller-Plathe, Deriving effective mesoscale potentials from atomistic simulations, *J. Comput. Chem.* **24**, 1624 (2003).
- [15] J. W. Mullinax and W. G. Noid, Generalized Yvon-Born-Green Theory for Molecular Systems, *Phys. Rev. Lett.* **103**, 198104 (2009).
- [16] J. G. Kirkwood, Statistical mechanics of fluid mixtures, *J. Chem. Phys.* **3**, 300 (1935).
- [17] M. Neri, C. Anselmi, M. Cascella, A. Maritan, and P. Carloni, Coarse-Grained Model of Proteins Incorporating Atomistic Detail of the Active Site, *Phys. Rev. Lett.* **95**, 218102 (2005).
- [18] Q. Shi, S. Izvekov, and G. A. Voth, Mixed atomistic and coarse-grained molecular dynamics: Simulation of a membrane-bound ion channel, *J. Phys. Chem. B* **110**, 15045 (2006).
- [19] V. A. Harmandaris, N. P. Adhikari, N. F. A. van der Vegt, and K. Kremer, Hierarchical modeling of polystyrene: From atomistic to coarse-grained simulations, *Macromolecules* **39**, 6708 (2006).
- [20] T. Spyriouni, C. Tzoumanekas, D. Theodorou, F. Muller-Plathe, and G. Milano, Coarse-grained and reverse-mapped united-atom simulations of long-chain atactic polystyrene melts: Structure, thermodynamic properties, chain conformation, and entanglements, *Macromolecules* **40**, 3876 (2007).
- [21] E. Lyman, F. M. Ytreberg, and D. M. Zuckerman, Resolution Exchange Simulation, *Phys. Rev. Lett.* **96**, 028105 (2006).
- [22] M. Christen and W. F. van Gunsteren, Multigraining: An algorithm for simultaneous fine-grained and coarse-grained simulation of molecular systems, *J. Chem. Phys.* **124**, 154106 (2006).
- [23] M. Praprotnik, L. D. Site, and K. Kremer, Adaptive resolution molecular-dynamics simulation: Changing the degrees of freedom on the fly, *J. Chem. Phys.* **123**, 224106 (2005).
- [24] C. F. Abrams, Concurrent dual-resolution monte carlo simulation of liquid methane, *J. Chem. Phys.* **123**, 234101 (2005).
- [25] V. Ruhle, C. Junghans, A. Lukyanov, K. Kremer, and D. Andrienko, Versatile object-oriented toolkit for coarse-graining applications, *J. Chem. Theory Comput.* **5**, 3211 (2009).
- [26] W. G. Noid, Perspective: Coarse-grained models for biomolecular systems, *J. Chem. Phys.* **139**, 090901 (2013).
- [27] R. L. McGreevy and L. Pusztai, Reverse monte carlo simulation: A new technique for the determination of disordered structures, *Mol. Simul.* **1**, 359 (1988).
- [28] T. Head-Gordon and F. H. Stillinger, An orientational perturbation theory for pure liquid water, *J. Chem. Phys.* **98**, 3313 (1993).
- [29] A. A. Louis, Beware of density dependent pair potentials, *J. Phys.: Condens. Matter* **14**, 9187 (2002).
- [30] C. Peter and K. Kremer, Multiscale simulation of soft matter systems, *Faraday Discuss.* **144**, 9 (2010).
- [31] A. J. Rzepiela, M. Louhivuori, C. Peter, and S. J. Marrink, Hybrid simulations: Combining atomistic and coarse-grained force fields using virtual sites, *Phys. Chem. Chem. Phys.* **13**, 10437 (2011).
- [32] W. Han and K. Schulten, Further optimization of a hybrid united-atom and coarse-grained force field for folding simulations: Improved backbone hydration and interactions between charged side chains, *J. Chem. Theory Comput.* **8**, 4413 (2012).
- [33] B. Ensing, S. O. Nielsen, P. B. Moore, M. L. Klein, and M. Parrinello, Energy conservation in adaptive hybrid atomistic/coarse-grain molecular dynamics, *J. Chem. Theory Comput.* **3**, 1100 (2007).
- [34] R. Potestio, S. Fritsch, P. Espanol, R. Delgado-Buscalioni, K. Kremer, R. Everaers, and D. Donadio, Hamiltonian Adaptive Resolution Simulation for Molecular Liquids, *Phys. Rev. Lett.* **110**, 108301 (2013).
- [35] J. Zavadlav, M. N. Melo, S. J. Marrink, and M. Praprotnik, Adaptive resolution simulation of an atomistic protein in MARTINI water, *J. Chem. Phys.* **140**, 054114 (2014).
- [36] A. C. Fogarty, R. Potestio, and K. Kremer, Adaptive resolution simulation of a biomolecule and its hydration shell: Structural and dynamical properties, *J. Chem. Phys.* **142**, 195101 (2015).
- [37] A. Chaimovich, C. Peter, and K. Kremer, Relative resolution: A hybrid formalism for fluid mixtures, *J. Chem. Phys.* **143**, 243107 (2015).
- [38] S. Izvekov and G. A. Voth, Mixed resolution modeling of interactions in condensed-phase systems, *J. Chem. Theory Comput.* **5**, 3232 (2009).
- [39] L. Shen and H. Hu, Resolution-adapted all-atomic and coarse-grained model for biomolecular simulations, *J. Chem. Theory Comput.* **10**, 2528 (2014).
- [40] H. L. Frisch, The equation of state of the classical hard sphere fluid, *Adv. Chem. Phys.* **6**, 229 (1964).
- [41] B. Widom, Intermolecular forces and the nature of the liquid state: Liquids reflect in their bulk properties the attractions and repulsions of their constituent molecules, *Science* **157**, 375 (1967).
- [42] D. Chandler and J. D. Weeks, Equilibrium Structure of Simple Liquids, *Phys. Rev. Lett.* **25**, 149 (1970).
- [43] S. Toxvaerd and J. C. Dyre, Role of the first coordination shell in determining the equilibrium structure and dynamics of simple liquids, *J. Chem. Phys.* **135**, 134501 (2011).
- [44] J. N. Israelachvili, *Intermolecular and Surface Forces* (Academic, New York, 2011).
- [45] A. Stone, *The Theory of Intermolecular Forces* (Oxford University Press, Oxford, 2016).
- [46] B. Hess, C. Kutzner, D. V. Der Spoel, and E. Lindahl, GRO-MACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation, *J. Chem. Theory Comput.* **4**, 435 (2008).
- [47] G. Bussi, D. Donadio, and M. Parrinello, Canonical sampling through velocity rescaling, *J. Chem. Phys.* **126**, 014101 (2007).
- [48] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, Molecular dynamics with coupling to an external bath, *J. Chem. Phys.* **81**, 3684 (1984).
- [49] J. Lin, Divergence measures based on the Shannon entropy, *IEEE Trans. Inf. Theory* **37**, 145 (1991).
- [50] E. H. Feng and G. E. Crooks, Length of Time's Arrow, *Phys. Rev. Lett.* **101**, 090602 (2008).
- [51] D. Bratko, L. Blum, and A. Luzar, A simple model for the intermolecular potential of water, *J. Chem. Phys.* **83**, 6367 (1985).

- [52] M.-L. Tan, K. N. Tran, F. C. Pickard, A. C. Simmonett, B. R. Brooks, and T. Ichiye, Molecular multipole potential energy functions for water, *J. Phys. Chem. B* **120**, 1833 (2016).
- [53] J. A. Wagoner and V. S. Pande, A smoothly decoupled particle interface: New methods for coupling explicit and implicit solvent, *J. Chem. Phys.* **134**, 214103 (2011).
- [54] D. Beglov and B. Roux, Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations, *J. Chem. Phys.* **100**, 9050 (1994).
- [55] L. Greengard and V. Rokhlin, A fast algorithm for particle simulations, *J. Comput. Phys.* **73**, 325 (1987).
- [56] L. D. Site and M. Praprotnik, Molecular systems with open boundaries: Theory and simulation, *Phys. Rep.* **693**, 1 (2017).
- [57] M. R. Machado, A. Zeida, L. Darre, and S. Pantano, From quantum to subcellular scales: multi-scale simulation approaches and the SIRAH force field, *Interface Focus* **9**, 20180085 (2019).