JCTC Journal of Chemical Theory and Computation

ARTICLE

pubs.acs.org/JCTC

A Multibox Splitting Scheme: Robust Approximation For *ab Initio* Molecular Dynamics

Stas M. Avdoshenko*

Institute for Materials Science and Max Bergmann Center of Biomaterials, Dresden University of Technology, 01062 Dresden, Germany

ABSTRACT: In this work, the multibox (M-box) simulation scheme is introduced, which can be considered as a generalization of the QM/MM scheme for multifragment (molecular) systems. This scheme exploits the natural locality of multifragment molecular-based systems by mapping the system into force-coupled block subspaces. Where defined in this way, the entire system can be fully modeled under a quantum mechanical force field. This allows the description of each subspace explicitly by means of a robust electronic structure theory without the requirement for large computational resources. An adequate block-to-block coupling by means of shared subsystem fragments is applied to preserve the long-distance structural correlation in the system during a molecular dynamic (MD) simulation. Since electronic structure descriptions play a central role in the formulation of several parametric models for charge or energy transport, we expect that this space−time correlated scheme can become a reliable computational tool for charge/energy transport/transfer applications. The efficiency of the method is demonstrated by performing statistical and time-resolved analysis using both the multifragment box and full *ab initio* approaches. We illustrate the method using as examples the melting process of a one-dimensional benzene chain (weak interaction situation) and NVE dynamics for the C_nH_n polymeric chain (strong interaction situation). We also have extended the threshold of applicability of our model, demonstrating how it can cope with MD simulation with more complex systems and processes.

1. INTRODUCTION

The importance of molecular dynamics (MDs) in revealing the physics behind molecular and nanoscale phenomena has been recognized since the 1960s[1] when biophysicists performed the first X-ray experiments on enzymes. Nowadays, it is hard to imagine a modern X-ray data analysis of protein crystals without MD tools.[2,3] Now, in the nanoscience era, MD simulations play a crucial role in current state-of-the-art nanoscale energy/charge transport/transfer models.[4−6] The general idea in these studies is to combine a *fermionic* degree of freedom with time-resolved atomic motions (which are *bosonic*) to predict transport characteristics in molecular-based systems. As previously demonstrated,[4−6] not only is the statistical average in the ensemble of interest important for such dynamic analysis but also the specific behaviors occurring on a short time scale domain (usually ∼10 fs) are extremely relevant for a successful theoretical description of the electronic systems. On top of that, it is already known that for small systems (up to a few hundred atoms), Born−Oppenheimer (BO) *ab initio* MD techniques are enough to cope with the evolution of the electronic wave function along the dynamic trajectory of the ions and roughly estimate possible paths of the evolution of molecular orbital ensembles. However, the size of typical molecular-based devices such as organic light emitting diodes (OLED), organic field-effect transistors (OFET), or even transport/transfer effects taking place in certain biosystems generally includes many atoms (usually exceeding 1000 atoms). Although development of computational hardware broadens the ranges of the objects which can be studied by *ab initio* techniques with the use of supercomputer resources,[7,8] such techniques are still hardly applicable for describing structural dynamics of extended systems in a routine way. It means that *ab initio* models are already hardly applicable for describing structural dynamics of such extended systems. Therefore, the implementation of an effective electronic structure description capable of dealing with such intricate ingredients is widely required.

A good electronic structure description is strongly linked to an accurate structural prediction and vice versa. Any tiny error in the first stage can lead to absolutely misleading parameters at the end, interfering strongly with the final transport or transfer characteristics. For instance, depending on the level of theory, a deviation of ∼1% in the benzene molecule geometry can result in an overestimation of 0.5−1.0 eV of its molecular orbital gap size. At the same time, highly accurate approaches such as MP2, MP4, multiconfiguration models, or even hybrid functionals are impractical to employ in large systems with more than 1000 atoms, if the system is a bit complex (transition metals, conjugated systems). A reasonable solution for systems with a large amount of atoms would be the use of classical force field methods. Still, such classical techniques hold disadvantages since they are either not valid or fail for phase space integration on the short time domain.[9,13,14] Also, such rough approximations following canonical parametrization types can estimate the wrong geometrical parameters for certain physical aspects such as molecular contacts or conformations.[9]

An alternative solution to improving the results could be achieved by implementing *ab initio* refinements on the snapshots of the classical trajectories,[10] although some critical comments about this approach have been published.[11,12] Problems will arise strongly for not so complicated systems such as the DNA backbone group PO_4^{-3}, noncanonical atom types (fluorine, boron, and again metals), and heterocycle compounds.[9]

Received: April 20, 2011
Published: October 24, 2011

Reasonable results have been achieved within semiempirical models (AM1, PM3, PM6)[15] or the DFTB[16] approximations. For structures close to their equilibrium, these methods are sufficiently accurate in obtaining the geometric and electronic structures. However, these models have limitations when MD simulations are conducted following the parametrization schemes of the respective methods. In general, the parameters are tuned on the basis of *ab initio* results; i.e., they are taken from reference equilibrium points located in potential energy surfaces which contain a limited amount of canonical interaction types.[18,19] This compromises their transferability; hence nonequilibrium states or more complex molecular structures are not necessarily well described by means of semiempirical schemes.[20,21] As a consequence, a verification of the parametrization set obtained for a particular molecular system is essential and often time-consuming.

The solution to overcome all of these problematic issues that we list here seems to be a return to the foundations of *first principle* methods.

The development of linear scaling methods has a long history. Hartree−Fock (HF), the post-HF (MP2, MP4, etc.), and DFT-based models have been widely used in the scientific community. On the basis of Rokhlin's ideas aimed at electronic structure problems,[17] Scuseria and Kudin have implemented a direct space fast-multipole technique which was adapted for electronic structure problems, solving the Coulomb problem linearly for rather large system sizes.[17,22,23] In parallel, Hutter et al. have developed a similar but more elaborate method of solving electronic structure problems by means of the Gaussian and Plane-Wave (GPW) technique.[24,25] In their model, a dual density treatment of the Kohn−Sham model is applied to map the large-scale Coulomb problem onto Fourier-space through a plane-waves representation. At the same time, the one-electron terms can be efficiently treated using a local Gaussian basis. Another important contribution for the development of linear theories was performed by Gao in the X-pol[26−28] model. X-pol forms the foundation for the Kitaura−Morokuma−Fedorov energy decomposition ideas summarized later in the fragment molecular orbitals (FMO) scheme.[29−31] These models are highly suitable for a precise treatment of the local electronic structure in large-scale and macrmolecular systems. The model takes advantage of splitting a multimolecular system into coupled monomolecular quantum subsystems, which are coupled in a pure quantum manner. The model we are going to describe here has its origin in a system fragmentation detailed below. Importantly, in our model, each fragment acts as a separate QM/MM problem, and then a summation over crossed coupled subsystems is performed to obtain an approximation of the total system. In other words, the gradient of the whole system is derived from the multiforce treatment in terms of multi-QM/MM parts coupled due to shared fragments between different QM/MM problems, which closely approximates the gradients evaluated in the full *ab initio* description.

2. METHODOLOGY

2.1. Philosophy of Multibox Model: QM/MM. The Quantum Mechanics/Molecular Mechanics (QM/MM)[32−35] approach is an example of a hybrid method that is currently one of the most popular models for the theoretical treatment of various physical phenomena in complex environments. In this method, the quantum region where all of the main chemical properties and reactions take place is treated quantum mechanically, while classical or semiclassical molecular mechanics are employed to model the effect of the external potentials due to the surroundings. An example of a hybrid Hamiltonian for QM/MM problems is given by

$$\hat{H} = \hat{H}_{\mathrm{QM}} + \hat{H}_{\mathrm{MM}} + \hat{H}_{\mathrm{QM/MM}} \quad (1)$$

where $\hat{H}_{\mathrm{QM}}$ deals with the quantum region, $\hat{H}_{\mathrm{MM}}$ gives the energy of the molecular mechanics contribution, and $\hat{H}_{\mathrm{QM-MM}}$ represents the interaction between the QM and MM parts. In more detail, the term $\hat{H}_{\mathrm{QM}}$ can be a quantum operator acting on the space of N_{bs} local atomic basis functions, where the level of sophistication of the adopted theory is dictated by the size of this basis set. The computational cost required for modeling this term can be $O(N_{\mathrm{bs}}^n)$ with $n = 2-8$, depending on the level of theory. The $\hat{H}_{\mathrm{MM}}$ part consists of classical interactions located in the surroundings, and its computational complexity can reach $O(N_{\mathrm{a}}^2)$ where N_{a} is the number of atoms. In some cases, more efficient approximations for the potentials and cutoff functions are capable of reducing this numerical complexity to $N_{\mathrm{a}}\ln(N_{\mathrm{a}})$. The graphical representation of the typical QM/MM model is shown in Figure 1A.

One of the advantages of QM/MM is the possibility to describe the environmental effects in a reliable way since an explicit contribution due to the presence of solvents, for instance, can be included in the theory easily. Nevertheless, the description of the QM region, which carries most of the interesting physics, can be significantly compromised since its efficiency still scales with $O(N_{\mathrm{bs}}^n)$. However, for complex systems containing about 1000 atoms in the QM region, this model becomes impractical at more sophisticated levels of theory. In other words, the size of the QM region restricts the applicability of this approach to the same extent as pure assumed *ab initio* techniques.

However, we can easily split such big system in parts following Gao's fragment picture and construct the QM/MM problem for each subsystem and then find a way to merge these subsystems to reconstruct a more reliable approximation for whole original system space. In this fashion, all long distance effects will be defined locally but, at the end, are collected over the entire system.

2.2. Multibox Origin: X-pol Method and Fast Multipole Model. The multibox (M-box) scheme introduced here provides a way to overcome some of the restrictions imposed while dealing with large QM systems. The main idea is to apply fragment orbital concepts mixed with the QM/MM formalism[26−28] and make these fragments coupled in the sense of forces coupling instantly between shared elements of different fragments. For simplicity's sake, we illustrate this approach in linear homomolecular systems; nevertheless, the generalization of the ideas presented here to heteromolecular or topologically different species is straightforward.

Following the ideas of Gao,[26−28] it is possible to exploit the natural locality of molecule-based structures with strong localized electronic structures by treating each molecule as a separate QM system. This implementation already reduces the computational cost of the problem to $O(M^*(M_{\mathrm{bs}})^n)$, where M is the number of molecules in the system and M_{bs} is the number of basis functions for a single molecule. However, fragmenting the system into pieces can become a real problem when dynamic aspects need to be treated. First, the repartition of the system into weakly interacting QM regions removes the dependency of the electronic structure on the environment and nonlocal effects. Such a splitting scheme creates M independent trajectories in the phase space with only weak QM interaction considered. This decoupling procedure is invalid when charges/spins or vibrations are correlated over a long range in space (on a larger separation than

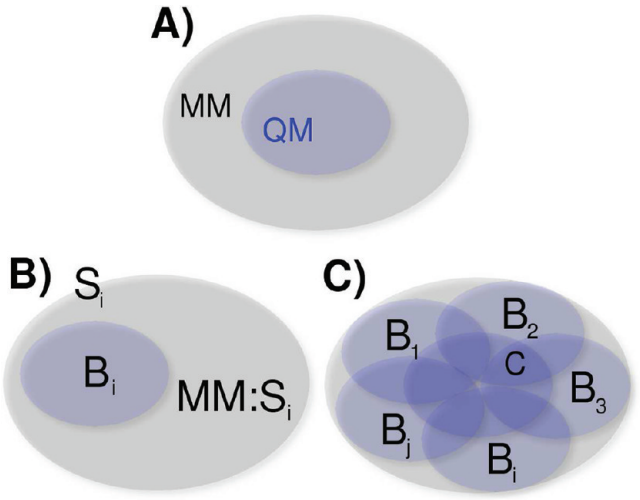

Figure 1. (A) Graphical representation of QM/MM model concepts. The graphical representation of the M-box model. (B) Definition of the subspace S_i containing the quantum block B_i using QM/MM charge effects from neighbors MM:S_i. (C) The whole task mapping, where the system is covered by active QM space. The border (*C*) is a crossing of the different QM regions.

the fragment). In this sense, one needs to preserve an effective short-range correlation description while at the same time including long-range effects without compromising the computational efficiency.

2.3. Multibox. In the M-box model, the dependence of the molecular blocks on the environment is constructed in two ways.

First, following the philosophy of the tight binding (TB) approximation, a better description of the short-range interactions can be obtained by maintaining at a least nearest neighbors contribution on each QM subsystem. Therefore, the system is not simply split into pure isolated single-molecular blocks but into effective blocks, which contain the information of the molecule "surroundings" in the full quantum sense. In this way, there is no need to treat the system within the FMO approach since neighboring effects can be naturally included as sketched in Figure 1. However, as in the FMO approach, the use of a broad spectra of theoretical methods for this effective fragment is possible.

At the same time, for each block, the influence of long-range Coulomb interaction can be implemented via the standard QM/MM method, e.g., by means of a classical electronic coupling (point charge scheme in the simplest case). It is important then to compare the alternative method presented here with the pure QM/MM description on the same QM scale. In parallel, one can show explicitly how the construction of the subsystems is done within each methodology. It is trivial that, in some cases, simple electrostatic corrections due to the influence of certain solvents are already sufficient for a proper description of the systems, and quantum contributions are not that important. Additionally, as was found for the fast multipole model (FMM)—which is rather similar to QM/MM ideas—the solution for the electronic structure part can become more accurate. It follows from the fact that, for many sophisticated and high-level theoretical chemistry models which are based on the Cauchy-Schwarz inequality, Coulomb interaction screening cuts are defined parametrically for long-range contributions. On the other hand, in the FMM/QM/MM-like picture, all nonlocal contributions are treated locally and exactly with the aid of the multipoles field.

Here, we start now describing the new splitting scheme, which is centered on placing more than one molecule in each block. The number of computational operations increases to $O(M_{bl}(N_{bl}{}^*M_{bs})^n)$, where M_{bl} is the number of blocks and N_{bl} is the number of molecules in one block. Hence, the method scales linearly with the number of blocks, and it is only limited by the level of theoretical calculations performed on each local block, for instance, HF|MP2 or more sophisticated approaches. As the maximum number of blocks is equal to the number of molecules, one can state that the method still scales linearly with *M*. Furthermore, we assume that the molecules from other blocks are coupled to the reference block by the construction of QM/MM interactions established for each subsystem. In addition, it is also possible to simply consider that each block feels a classical Coulomb response caused the MM parts of blocks' representation.

Following this setup, it is ensured that each block interacts with the whole system at least at the level of electrostatic coupling. In other words, a set of *M* complete but independent QM/MM systems, S_i, is constructed from the original system, **S**. At this point, we have to emphasize that the model described here is closely related to Gao's ideas, with the extra ingredient that, instead of decomposing the energy in the space of the molecular fragments, we have constructed topologically identical multisets which are linked through the forces in the shared parts of the fragment.

2.4. Multibox Forces and Hamiltonian. However, the time evolution of the independent blocks in the M-box approach remains uncorrelated with phase space. The pure inclusion of long-range Coulomb interaction by means of classical charge effects—all blocks mutually affect each other—is not enough to guarantee the space−time correlation. Additional ingredients must be implemented to solve this issue; in particular, one can employ broad spectra schemes for possible dispersion effects (classical VdW, partial classical fields to describe long-range interaction, classical charge effects, etc.).

For this purpose, we introduce a number of molecules with special *interconnect* properties. Instead of creating a system composed of isolated blocks, one assumes that each block shares a part of its own atomic structure (via an interconnector) with its neighbors. Formally, one can say that these molecules act on different subsystems within the same time scale, thus controlling the evolution of the whole system. This leads to the definition of a novel *coupling model* which can be redefined depending on the problem of interest, or modeling task (transport/transfer). In the simplest case, the coupling model can just be built up from a linear combination of the *ab initio* forces, $\vec{F}$, acting on the molecule *I* located in different blocks, B_J, with *J* specifying the central molecule in the block. Thus, the coupling model can be read as

$$\vec{\mathscr{F}}_I^{\mathbf{S}} = \begin{cases} \vec{F}_I & \text{if } B_I \in \mathbf{B} \\ \sum_{I \in B_J} \vec{F}_I^{B_J} & \text{otherwise} \end{cases} \tag{2}$$

with $\vec{\mathscr{F}}_I^{\mathbf{S}}$ being the final force on the molecule and **B** labeling the complete set of blocks. The fact that the molecules belong to different QM blocks indeed increases the complexity of the method. This compensates for the force linearity in *M*. In analogy to QM/MM methods, a graphical representation of the method can be seen in Figure 1.

A natural question which could be raised at this point is whether the validity of the Hamiltonian principles is maintained (e.g., conservation of energy). On the basis of the ideas described above,

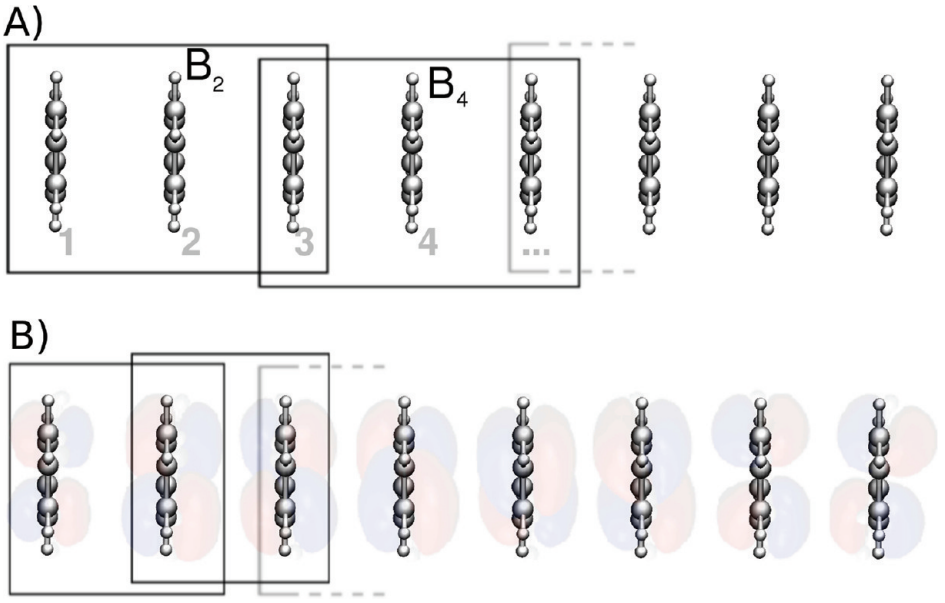

Figure 2. (A) The 1D benzene chain with highlighted QM regions according to the splitting applied in the simulations. Also shown are the coupling sites ($3, 5 \in B$) between B_2 and B_4 and between B_4 and B_6 in the frame of the M-box model. (B) The 1D benzene chain splitting with natural system symmetry preservation. Block-HOMO orbitals for one of the initial trajectory point are shown.

one can see that the M-box description operates with partial forces which are naturally close to a constrained Hamiltonian scheme.[46] In fact, this is the idea behind the coupling method: to *glue* the systems by these common parts in terms of constrained forces and try to preserve this mapping back to the original space in the integration scheme.

That is why, as we are interested in the evolution of system forces for only N particles, a reasonable choice for constrained forces can be determined by chemical intuition. An assumption could be that the potential energy of the whole system (original on $2N$-space, or N particles) can be represented as a sum of other blocks (b). Then, one can derive explicitly the partial forces ($F_i = \Sigma_{k\in b}\partial/\partial q_i U_k(q)$ for $i \in N$), and the partial moments $p_i = \int F_i \, dt$, where $U_k(q)$ is the potential energy with respect to the subsystem $k \in b$. From this, the definition of the system forces becomes a trivial problem.

3. RESULTS

3.1. Multibox vs. Full BOMD: Weak Interaction Limit. In the following, the quality of this approach will be demonstrated by a direct comparison of Mbox results with full *ab initio* BOMD results. To illustrate this point, we adopted a unidimensional (1D) periodic chain composed of benzene rings in order to avoid additional complexity in this first analysis. At the same time, a 1D system is not a totally arbitrary choice since in many organic systems (crystals), structural characteristics often have a principal axis and well-known properties such as anisotropy according to these axes. Hence, many 3D structures can be deconvoluted into a 1D stack.

Figure 2 shows the system of interest with eight molecules in total. It also depicts two possible ways in which the system can be split. In panel A, the system apparently reveals symmetry breaking (ring-to-ring distance is ± 0.005 Å) which does not occur in the full (eight molecules) system optimization. In panel B, the system is fully symmetric (ring-to-ring distance is almost the same as in the full system optimization). Since intermolecular forces are very weak and only local impacts are essential, it is not surprising that in both cases a good agreement is achieved with results for the full system QM optimization.

We now discuss the molecular dynamic simulation results based on system A, where each block contains three molecules, and there is only one molecule in the coupling region. In the coupling region, the forces for each block are read as $\vec{F}_3 = \vec{F}_3^{B_2} + \vec{F}_3^{B_4}$ in the case of the molecule labeled **3**. The initial system guess was made using a ~3.8 Å translation vector of the original equilibrium C_6H_6 structure, with preoptimization until a system gradient of 10^{-5} au/Å was obtained. In this case, the system is small enough to fully treat it using *ab initio* BOMD for comparison with the M-box result.

The benchmarks were considered using the GPW-based[36] CP2K code,[37] which allows a mapping of the system onto a different force environment. Also in this code the QM/MM interface is implemented by its developers.[38,39] All MD simulations have been performed in the microcanonical ensemble (NVE) in order to avoid random velocity renormalization and friction forces from thermostats. A time step of 0.5 fs and an initial temperature of 300 K was applied during all of the simulations. For each subsystem, the QM/MM problem was solved where the properties of the QM part were defined at the DFT|PADE|TZV2P[41] level of theory with a density cutoff of 280 Ry in the field creating by an external MM electrostatic potential.

For the classical MM part, the AMBER-99[42−44] force field was used for the aromatic benzene atoms. We emphasize beforehand the fact that in the current study the classical dispersions do not influence the evaluated forces. This can be done easily by changing the coupling properties within each block and between blocks. In accordance to the M-box scheme, each subsystem acts under QM and MM force fields, so this gives rise to the generalized scheme of coupling forces as a weighted superposition of classical and quantum forces. With these computational options, the energetics and geometrical characteristics obtained from the full

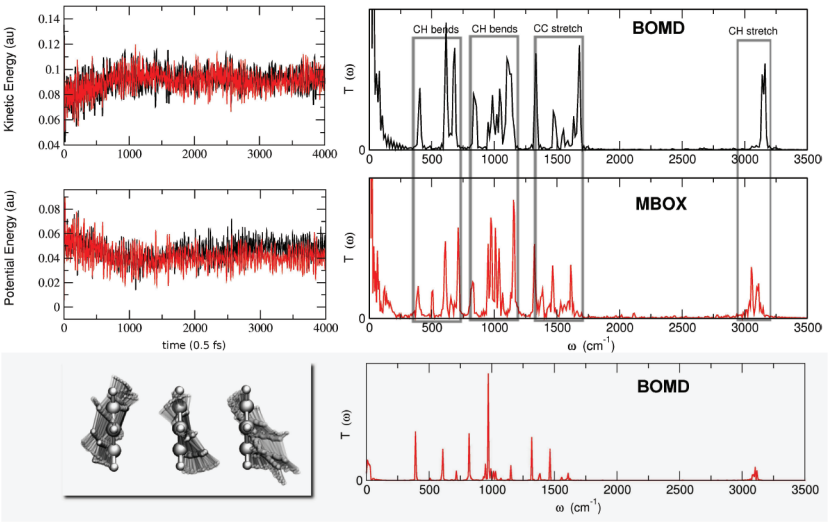

Figure 3. The evolution of the kinetic and potential energies (absolute value for both runs with respect to the start point) during BOMD and M-box run over 2.0 ps (left panel) and VDOS spectra function of the kinetic energy for 2 ps production runs for both models, where characteristic band regions are highlighted (right panel). The 2 ps region has been chosen because on this scale all other molecules of the chain are still in close vdW contact. (Lowest gray panel) VDOS based on time 2 ps NVT evolution for a three-molecule system (single block in the M-box splinting), on the left-hand real space geometry changes during the first 1 ps NVE run with 20 fs sampling.

ab initio BOMD simulation are compared to M-box results. Figure 3 presents the short time evolution of the kinetic and potential energies. The data show that the deviation for both methods covered the same energetic range with a good correlation of the on-time behavior. The initial structure guess at 0 K and the random velocities distribution at 300 K distribution results in total kinetic energy varying around an expectable average value of $T \sim 150$ K.

The collective behavior and the significance of the *subsystems coupling* cannot be easily demonstrated in the real time domain. Therefore, the Fourier spectra $[T(\omega)]$ of the kinetic energy in both simulation runs based on 2 ps trajectories (right panels in Figure 3) are analyzed. The spectral function (VDOS) can be compared with the well-known vibration spectrum for the benzene molecule, which has only four prominent IR bands in the gas phase resulting from the dipole selection rules.[45] At 1500 cm^{-1} and 3100−3000 cm^{-1}, the stretching modes associated with the aromatic carbon−carbon bonds and the aromatic CH appear, respectively. The bending motions involving carbon−hydrogen bonds appear at 1000 cm^{-1} for the *in-plane* bends and at 675 cm^{-1} for the *out-of-plane* bends. Both models are able to reproduce all of the mentioned bands. As one can see, M-box closely reproduces the spectral function obtained by the full BOMD description. It is remarkable that the low energy profiles up to 500 cm^{-1} for both methods are mostly identical to each, illustrating the close character of collective motions predicted. An additional observable band splitting can be associated with collective effects. It is closely related to the differences between the gas and liquid phase experimental IR spectra, where for the liquids, a much richer IR spectrum should be detected.

The impact of the subsystem coupling can be illustrated by analysis of the behavior of an independent subsystem behavior on the same time scale. Figure 3 (gray panel) shows the spectral function of a single subsystem which was obtained on the same time domain of 2 ps. Obviously, this spectral function differs dramatically with respect to the previous cases because of the lack of additional dispersion from the molecule−molecule interplay. Again, high and low energy bands can be seen in this VDOS spectrum, but additional band splitting (band intercombination) is small and the phonon region (up to 500 cm^{-1}) becomes extremely thin.

In order to exclude the effects of random initial guesses in the frame of the NVE ensemble, test runs were analyzed for several independent runs. In these analyses, all observables (energy mean values, local dipole moment, etc.) were nicely reproduced. The real time geometrical changes averaged over 10 independent BOMD and M-box runs are considered. Each trajectory was started from the same structural initial guess but with different initial velocities. The resulting geometrical trajectories for both techniques are presented in Figure 4 with 50 fs sampling for the graphical presentation of a 2 ps trajectory. Due to the dissociative nature of the melting process in these tasks, a detailed statistical analysis (principle component analysis, for instance) is difficult to perform because of a large contribution of irreversible types of atomic motions. Nevertheless, from Figure 4, it is obvious that those average trajectories are almost identical, and one can reasonably assume a close space−time correspondence between the M-box model and the full BOMD models. Most important is the fact that in all of the above runs, the splitting of the whole system (A−symmetry break, B−symmetrical task) leads to an artificial symmetry breaking with respect to the proposed force-coupled QM regions. While the central atoms are in a symmetric environment as QM describes all of the neighbors, the coupled atoms are not. In each of our blocks, one of their nearest neighbors is part of the QM region, while the others interact via the QM/MM coupling, which is an electrostatic interaction only in this report. Being aware of this fact, the overall agreement with the QM results indicates that this different treatment of the molecules has no significant effect on the description method of the whole system. In other words, it means that the results are almost the same, independent of the splitting scheme (see Figure 2). Nevertheless, for more complex systems or to prevent a too long extension of this problem, a different definition of the quantum blocks should be implemented according to the system symmetry, by assuming that the environment is treated in the same way for

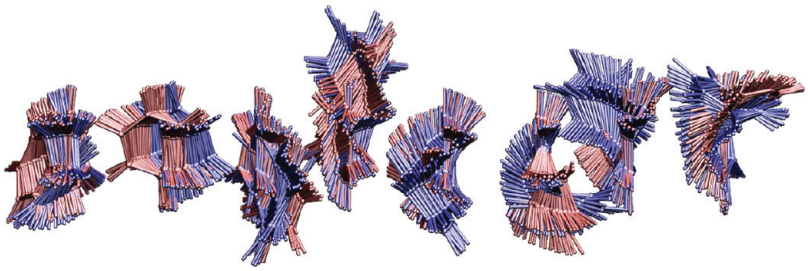

Figure 4. Superimposed real space changes for full BOMD (red) and M-box (blue) runs in a 2 ps trajectory window with 20 fs sampling (average over 10 independent runs).

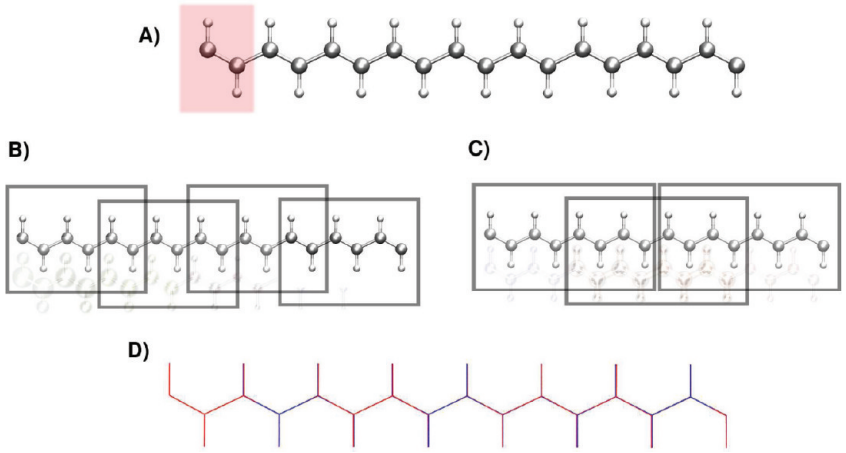

Figure 5. (A) The original systems of interest. Unit cell for the *trans*-$(CH)_x$ chain, the red rectangle highlights two different origins ($x = 18$ and $x = 16$) for whole systems where Γ-point solutions are remarkably different. (B−C) The two chosen ways of system mapping, see text for details, depending on the origin, $x = 6$ (benzene like) and $x = 8$ in the sub-block. (D) The merging optimized structures obtained by using the M-box (red) scheme and full QM method (blue) for the (C) $-x = 8$ and system (A) $x = 16$ (see text for details), where a structure comparable in the sense of *G*-point results runs. The C−C bonds length are split at 1.39 and 1.47 Å, as it is easier for (A) $-x = 16$ and (C) $-x = 6$.

each molecule (or parts) of the system. This could be achieved, e.g., by defining the blocks such that, for every molecule I, a block B_I exists.

3.2. Multibox vs BOMD: Strong-Interaction Limit. In order to address all of the possible interaction limits, we discuss an interesting set of systems where a strong or covalent interaction between the blocks is considered. According to our initial goals, it is worth mentioning that the M-box scheme is capable of treating systems where weak interactions take place, and so the method would not be expected to work well for the problem with strong interactions.

As an example, a simple polymeric $-[CH]_x-$ annulene was chosen. One can see that such a problem involves a particular complexity which is associated with π systems. It is clear that the QM/MM standard scheme may face problems while describing bonds at the edge of the fragment, even if their saturation by means of QM/MM is conveniently addressed. Extra complications can arise from Γ-point calculations. On top of that, the pathogenic density delocalization in the DFT picture itself also becomes a difficulty.

This issue has been analyzed in the same M-box spirit as our results presented before. The $-[CH]_x-$ systems are shown in Figure 5, highlighting the several ways of mappings where different variances with distinct sub-block sizes and intercoupling have been chosen. The subsystems of interest were translated in such ways as to avoid the QM/MM saturation of QM-block terminal atoms. This means that the QM cell size for each block is exactly sufficient to allow the saturation of this QM block by the translation of this block. This way, we try to reduce or completely eliminate some nonphysical contributions from the correlation spectra function which are often driven by this saturation.

We keep allowing the system to move with respect to the forces defined in each QM region and for every block and with respect to the coupling model defined for the cross-sections illustrated in Figure 5B and C.

It should be emphasized that variations of the properties of the polyacetylene as a result of the fragment taken as a translation unit is the outcome of Γ-point calculations. If the whole k space is covered, both approaches should give identical results. However, complete Brillouin zone integration at each MD step is impractical because of the high demand it would make on computational resources. Therefore, the proper choice of the translation element is crucial to make Γ-point computations reliable.

It is clear that for a hypothetical $-[CH]_x-$ annulene depending on $x = 18$ or $x = 16$, the Γ point can be related to $k = 0$ and $k = \pi/a$ states. In the chemical language, it can be explained in terms of Huckel's $4n + 2$ rule.[48] In fact, an N-site Huckel matrix is related to a nonperiodic system, such as the one addressed here. It means that the aromatic condition given by $(\alpha - \varepsilon)/\beta = -2\cos(2\pi n/N)$ is valid. For a simple tight-binding Hamiltonian and for $N = 2, 6, 10, 14, 18, ...4n + 2$, no antibonded contribution is possible until the limit is reached where the k space is reduced

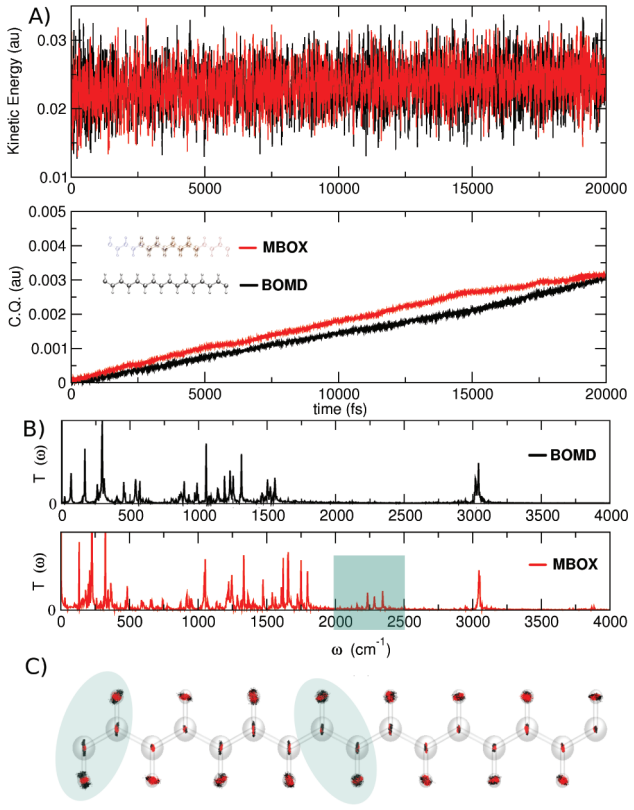

Figure 6. The deviation of the kinetic energy change during full BOMD and M-box run for 20 ps (A) and the VDOS spectra as a function of the kinetic energy for 2 ps production runs for both models, where characteristic band regions are highlighted (B). The real-time system behavior over a 10 ps trajectory with 50 fs sampling (C).

to a single point (Γ point). According to this, in the present cases, aromatic-like systems with sub-blocks containing $x = 6$ ([6]annulene) should be comparable with $x = 18$ ([18]annulene) and $x = 8$ ([8]annulene) with $x = 16$ ([16]annulene), e.g., nonaromatic systems.

The initial systems were optimized using the BFGS (Broyden–Fletcher–Goldfarb–Shanno) scheme for an equilibrium geometry search. For the search of equilibrium structures and MD study, the following splitting schemes into the blocks were applied:

- system B in Figure 5, with $x = 6$ elements in each block and $x = 2$ elements in the coupling region between blocks—the full system has $x = 18$ elements
- system C in Figure 5, with $x = 8$ elements for each block and $x = 4$ elements in the coupling region between blocks, which is logically comparable with whole structure with $x = 16$ elements structure

From our numerical observation, both choices have shown almost identically accurate final structures at 10^{-5} au/Å. It is remarkable that an intuitively rough approximation such as the one adopted in B was already enough to get the final equilibrium geometry close to the one predicted by the full QM theory. These runs have shown a good correlation for both splitting schemes, such as C–C bonds in chains being 1.37/1.49 Å (for the nonaromatic case) and 1.43 Å (for the aromatic case) after reaching the assumed tolerance factors. We will see later that such a split approximation as in B is not really applicable for the phase sampling, and case C seems to be more reliable in a statistical sense.

The MD runs for the strongly bounded regime have been integrated for a long time of ∼20 ps. This has allowed us to make statistical and local analysis in a more precise manner. We will focus on case C with $x = 8$ elements in each fragment and $x = 4$ elements in the coupling region between blocks, and the system used for comparison is $x = 16$ in Figure 5.

Figure 6 shows the long time evolution of the kinetic and potential energies of the system following the M-box and full QM-BOMD descriptions. As one can see, the time variations of the kinetic/potential (here, the potential energy is given by the module with respect to the $t = 0$ fs point for the whole and M-box tasks in such a way that the starting point is apparently different) energies are identical. However, statistics show that, in contrast to the *weakly interacting case*, additional frequencies interfere with the reference spectra function from the full BOMD problem. Moreover, the VDOS functions are different in the range of 1500–3000 cm^{-1} (Figure 6). However, all of these new features in the M-box model can be straightforwardly understood.

We now analyze the shift of the C–C stretching modes from 1500^{-1} predicted by full the BOMD to the more experimentally relevant range[49] of 1700–1800 cm^{-1} predicted by the M-box approach. The origin of this shift lies in the fundamental failure of DFT in overestimating the electronic delocalization. For instance, even in the case of the nonaromatic [16]annulene (e.g., $x = 16$), overestimation of delocalization by DFT in the full BOMD results in a noticeable weakening of the double bonds (which leads to an aromatic-like frequency at 1500 cm^{-1}), making such MD sampling problematic. In fact, this error is due to the shrinking atomic space within blocks of smaller atoms under the single particle DFT construction. Eventually at the same quality of basis (DZ/TZ), the M-box model might be more accurate in a prediction for smaller systems, where DFT methods are more suitable even if the system is strongly localized. We performed an additional analysis for [6]annulene, and results were mostly identical, and $x = 18$ ([18]annulene) around 1500 cm^{-1} is mostly identical. However from the VDOS spectra for the B approximation, one can say that the spectra shapes are similar for the M-box, and for the full BOMD simulation the characteristic bands (C–C and C–H) can be regarded from both models.

The most pronounced differences in the results can be seen around 2100–2200 cm^{-1}. Here, we need to conclude that the M-box scheme does not work perfectly well. Apparently, the periodic boundary conditions which were chosen to saturate the blocks are able to generate additional effects on the C–H dispersion (in the M-box, formally, we have eight terminal hydrogens) in the periodic reflection. These modes were not so efficient in case of the full BOMD problem (there are only two C–H bonds on the edges). This proportionality with respect to the number of the actual terminal groups can be seen as well in example B, where the intensities around 2100 cm^{-1} bands are much higher than expected.

It has been shown above that the ideas underlying this very simple scheme have the ability to retain the simplistic integration, and the system retains its Hamiltonian properties. It is well-known that numerical and energy definition problems such as the ones in DFT calculations have effects on the conservation quality (QC), which provides a drift in the sum QC = $E_{pot} + E_{kin}$ (Figure 6).

From our numerical observation for the vdW system (our benzene chain), conservation quantity has a slight deviation in comparison to the values of ∼0.00050 au/ps obtained for the M-box and for the full BOMD problem, which is only ∼0.00045 au/ps,

independent of the splitting schemes. This is due to the interplay of forces being defined only within a molecular pair. It is locally defined and affects the whole prediction, contributing additively over all errors given by BSSE, SCF accuracy, etc. at our level of theory. All of these errors in vdW complexes vary slightly with system size because of obvious reasons (one can think about this fact in terms of energy decomposition analysis, where it is clear that the contribution to BSSE given by four molecular centers, for example, is an irrelevant value). Hence, for our benzene system, the QC drift is simply defined in the benzene−benzene pair, and it is precisely the same for both BOMD and M-box.

The situation around conjugated chains develops differently, but of course it is driven by the same ideas as before. This is partially due to the impact of the bigger SCF wave function uncertainty predicted in larger systems in contrast to smaller ones, with the condition that close tolerances for the SCF procedure are used. In other words, it is simpler to obtain a more accurate approximated wave function for a small system. Because of this fact, for a comparable system size such as $x = 8$ ([8]annulene) blocks and its analog $x = 16$ ([16]annulene), they have QC deviation on the same energetic scale. However, in the short sub-block case—system $x = 6$ blocks and $x = 18$—one can even have a smaller drift in QC values. This confirms that the SCF problems are scalable for the conjugated system (due to the band gap size problem, for example).

3.3. Multibox and Boxes Crossing Processes. An interesting investigation would be to verify what will happen if an event of particle exchange between boxes takes place. In other words, how much time-dependent particle exchange between blocks and along interconnect regions (B) is needed? Can this be described using our M-box scheme? A conceptually close approach, with a real-space splitting (with respect to the level of theory as well) for on-flight QM/MM border crossing processes has been discussed already by Bulo et al.[52] and earlier by Truhlar and Heyden.[53] For a system with strong interactions, such a crossing effect has to be very rare. Besides, these processes of interest should have a relatively large time scale where *ab initio* knowledge hardly can obtain information for the whole time interval in general. In this sense, only weakly interacting systems would need this knowledge, e.g., liquids, gases, etc. Systems with "weak interaction" can undergo large spatial rearrangements on a short time domain. A combined model can be represented as the following: a protein chain coupled block-by-block. In this structure, conformational transformation is not possible, and it holds a stable conformation for all of the simulation time. An example is closely related to the C_nH_n chain we have described before. Also, we can have the surrounding solvents shell some light and mobile molecules, which due to weak of interactions between them can cross the fragment boxes with a high probability. Here, we can construct a simple example like a light gas, which will expand in view of the NVE ensemble.

For the sake of simplicity we consider the ideal system dynamics of adiabatic gas expansion. According to the example examined below, we stress that one has to upgrade our level of theory to one which is grid-independent. It could still be a DFT model, but we need to isolate the plane wave part or make a careful recount in the renormalization. For obvious reasons, if the total energy of the system is a grid function such as in the PW part of a DFT model, one cannot avoid step-like changes in the total energy due to different grids for different amounts of particles under consideration. Hence, we use the AM1[15] level of theory, which is also implemented in the CP2k code.[37] An ideal gas composed of Xe

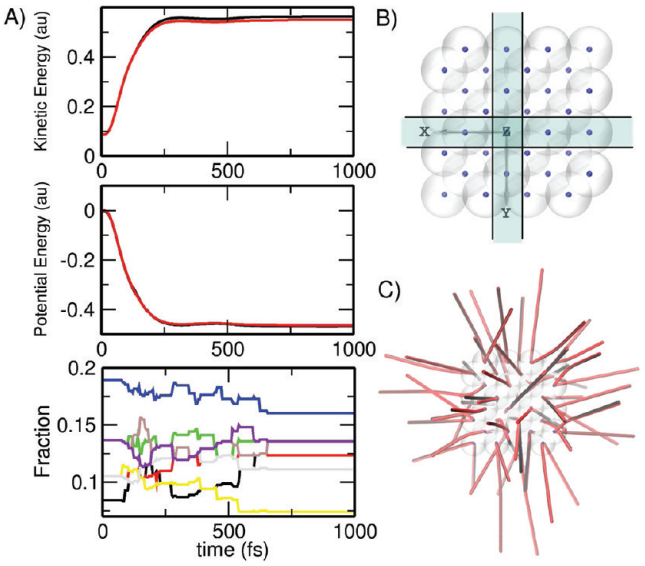

Figure 7. (A) Comparing the deviation of the kinetic and potential energies during full BOMD (AM1 level of theory) and M-box for the Xe-cluster tasks (B). Task setup with the hightlighted B region in the real space and the box organization in the *XY* projection. (C) Real space Xe-cluster evolution for full BOMD (black) and multibox (red) runs in a 2 ps trajectory window with 10 fs of sampling.

atoms was assumed as the trial system, and its original geometry as shown in Figure 7 is taken from the diamond supercell with the Xe−Xe distance equal to 1.7 Å. Eight boxes have been constructed due to geometrical reasons. The system geometry is symmetric with respect to the (0,0,0) point. Cube **1**, for example, is given by $(x > -1, y > -1, z > -1)$. Cube **2** is $(x < 1, y > -1, z > -1)$. This problem is defined as it automatically forms a 2.0 Å coupling region and B region, as can be seen in the 2D (*XY*) projection.

Our action script is next; every 5 fs, we checked which block each particle was supposed to be in with respect to the described splitting scheme. As occurred in previous cases, the blocks are coupled by the gluing (force merging), which is also dynamically mapped onto the full system size, which means that the temperature regime—or kinetic energy—is preserved by the limited number of atoms. However, within this construction (box crossing description), total potential energy given as an additive sum is not correct anymore and cannot be considered as an energetic criterion. For homogeneous systems, this inconsistency can be easily avoided by a renormalization of all of the parts which contribute to the total energy by using a weight of active parts normalized to practical numbers, e.g. $U_{\text{total}} = \Sigma_b w_b U_b$. One can estimate that if the system is equilibrated or optimized for a certain box size, the interesting event—box crossing—can take a long time. We accelerate the task by making this event highly probable on a short simulation time scale of 1 ps. In order to achieve this, we made the starting point cluster highly stressed, and then the system followed the NVE expansion. Our blank theoretical experiment involved the same initial geometric guess following the same AM1 theoretical model. The result of this construction is presented in Figure 7A (lower panel), where the evolution of kinetic and potential energies can be seen. As can be seen, the compressed gas with the high potential energy at the starting point is about to expand. The potential energy decay is accompanied by the expected growth of temperature/kinetic energy (similar time dependencies can be simply derived from

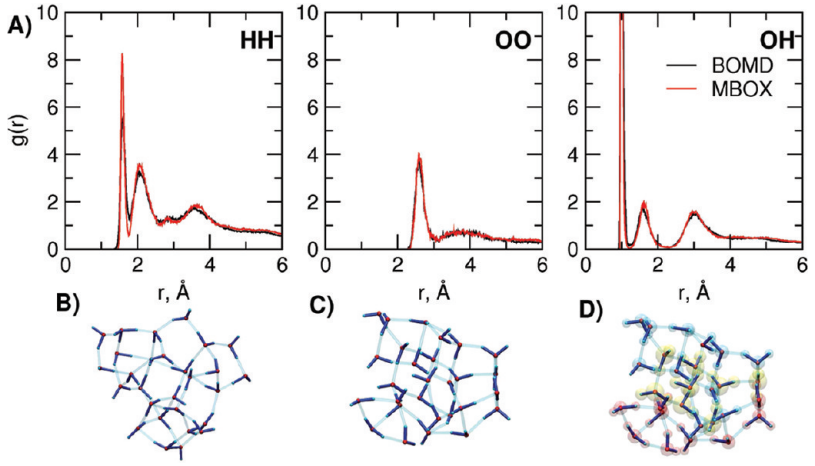

Figure 8. (A) Site−site distribution functions for water molecules (H−H, O−O, and O−H couples), reconstructed from the MD simulation with full *ab initio* and M-box models. (B) Water cluster geometrical characteristic after 4 ps of MD simulation in the case of full *ab initio*. (C) Water geometrical characteristics after 4 ps of NVT-MD simulation for the M-box model. (D) M-box splitting scheme (fragments 1, red; 2, yellow; 3, blue transparent), propagation done with the force vector defined as $\vec{F}_{tot} = \vec{F}^{1-2} + \vec{F}^{2-3} + \vec{F}^{1-3}$.

Lennard-Jones or Morses's potentials). Obviously, the pictures are mostly identical for the M-box model and full description cases. Also, the real time changes are identical, as illustrated in Figure 7C, where a snapshot representation of the trajectories is shown. The small overestimation in the saturation limit for the M-box tasks can originate in the random nature of the initial velocity guess. As shown in Figure 7 by dynamics of fractions $(N_n(\text{inblock})/\Sigma_n N_n)$, our goal was achieved, and during the simulation, the number of particles in the boxes was changed. Moreover, the potential and kinetic energy time evolution results seem to reveal that the conservation criteria are preserved, at least for the present NVE integration.

Besides the time evolution of the specific systems that were discussed previously, the statistical average and expectation values are still on the list of parameters which are mostly desired during MD runs. It is well-known that the thermodynamical characteristics can be easily derived from the radial distribution function, $g(r)$, based on the fundamental relation for the system's total energy. For example, an ideal gas would be $E_{total} = 3/2Nk_BT + 2\rho N\int r^2V(r)\,g(r)\,dr$, where $V(r)$ is the pair potential and $g(r)$ is the radial distribution function. In more complex systems such as the water cluster, it would require more effort to split $g(r)$ into parts in terms of site−site distribution functions, e.g., $g_{HH}(r)$, $g_{OO}(r)$, and $g_{HO}(r)$, as depicted in Figure 8A.

Finally, in Figure 8 we analyze the two series of site−site distribution functions obtained from the full BOMD simulation and the M-box. This last one follows the same level of theory as the one adopted for the benzene system with DFT|PADE|TZV2P within the canonical ensemble by using a Nose−Hoover global thermostat for the 4 ps time interval. The equilibration was performed by classical AMBER fields and the TIP3P water model.[41,42] The tasks ran for 28 water molecules in a gas cluster, as displayed in Figure 8B,C and in a nonperiodic 8000 Å^3 cube. Under normal conditions with 10^6 Pa and 300 K, it fills a space of 1000 Å^3. This estimation can be derived from the equation of state for the ideal gas, $n = (pV/R)N_A$. By increasing the size of the system, we artificially introduce additional degrees of freedom which can be suppressed in the periodical solution. The M-box scheme regions are shown in Figure 8D (1, red; 2, yellow; 3, blue transparent). The complete problem depends on the force vector, which is $\vec{F}_{tot} = \vec{F}^{1-2} + \vec{F}^{2-3} + \vec{F}^{1-3}$. As can be seen from Figure 8A, all site−site distribution functions are identical, and the integral values (e.g., $\int g_{ij}^{BOMD}(r)\,dr - \int g_{ij}^{MBOX}(r)\,dr$) are precisely zero for all ij from {O,H}. This means that the M-box model is adequate for modeling thermodynamical values based on the MD integration, which follows the M-model.

3.4. Multibox and Complex Cases. A key question that might arise is why would we need the M-box splitting scheme since the linear scaling methods have been established for decades in the DFT model? First of all, linearity does not take place in every DFT model. It is well-known that the applicability of DFT for certain problems needs to be certified, and the adopted functional approximation must be tested. For instance, pure LSD or LDA approximations are not suitable for treating d- or f-chemical elements where long-range electronic effects are essential. Even if such a system consists of a single molecule, those functionals have difficulty in correctly describing of d or f compounds. It is worth mentioning that when we say "long-range electronic interaction correlation", it does not mean that such influence extends over the entire system (∼10 nm). In some cases, it can be a very local interaction limited to 1−2 neighbors for one molecular site. The molecular system itself is a difficult problem for DFT. One well-known example is vdW complexes, which are hard to describe with chemically relevant accuracy within a pure DFT model. Additionally, natural SCF diagonalization limitation can be an issue for rather complex systems such as degenerate states, energy gap and convergence problems, etc. Second, a nonextrapolated SCF procedure could be very problematic when implemented for large systems, and at the same time good DIIS algorithms can have problems finding and verifying the real ground state.

Also, the issue of the diagonalization of large matrixes might lead to problems. Rapid convergence techniques such as Lanczos and analogues[54] have strong numerical instabilities. In particular, for electronic structures, this problem grows considerably, and it is also complicated to obtain excited states. These points can partially explain certain limitations that such well established codes and models might present. It seems quite contradictory: whenever "fast models" can be employed, dynamic knowledge is hardly needed; whenever dynamic knowledge plays a role, "fast models" can hardly be used. Let us exemplify this point with an artificial molecular

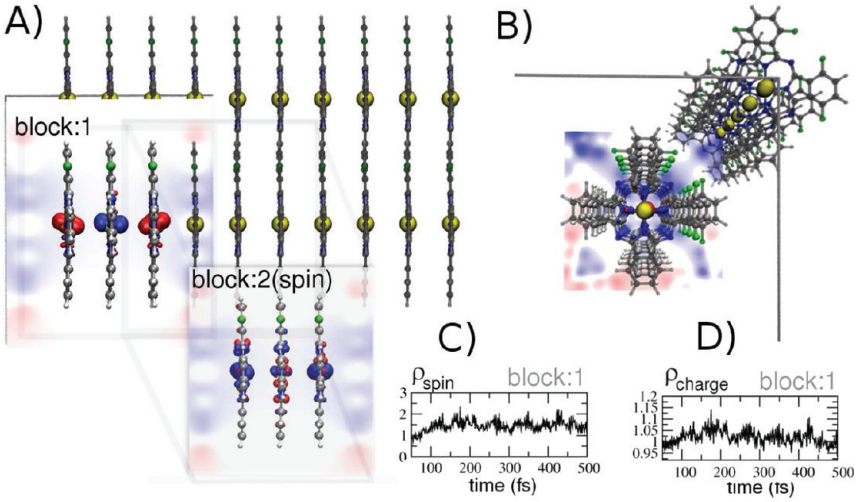

Figure 9. Equilibrium geometry of a model system of a modified Co-porphyrin stack in two projections with a profile of two blocks on top (A, B). In profile, the spin density destruction is shown for different respective blocks and the (−0.07 eV; 0.07 eV) electrostatic map given by the environment (A, B). Insets below show a short time evolution of Mulliken spin/charge density in the first block for central Co (C, D).

magnet system. We assume the system shown in Figure 9, consisting of a Co-porphyrin class 1D stack. By changing two of the hydrogen atoms in the benzene ring to flourine atoms, an internal polarization is induced. In this case, this system might be of interest for spintronic applications since rotational operations could be used to induce such polarizations.

Even if our interest is just to obtain the equilibrium geometry of the Co-based material, some default or even more sophisticated SCF procedures cannot guarantee good results. The computational cost for 10 cobalt atoms in a line is a crucial issue for SCF convergence. We might also be interested in more interesting features such as the time spin-density dynamics recently discussed by Popov et al.[55] For this BO spin-dynamic, a natural real-space length limitation exists, and it is given by an expected tunneling length of about 3−4 localized sites. Beyond this characteristic size, single particle models hardly make sense for chemically relevant electronic structure predictions so that a local M-box treatment is a good approximation for on-site behavior. Also, forces in this system along the whole MD trajectory can be defined from the M-box model as they are also defined locally.

By using the SCF option, which is suitable for the single molecule case, we were not be able to obtain a reliable wave function for whole system stack convergence (this conclusion is based on the topology of the frontal orbitals). On the basis of the Hirschfield charge prediction obtained for isolated molecules, the original electrostatic map was constructed for the QM/MM task in each block. All charges have been assumed to be equal at all MD points of the MM shells. For this task, the level of theory for the prediction of the local gradients was similar to the one used for the benzene chain problem. It means that within the each block our DFT model is also linear according to GPW ideas.[36,37,41] For this case, the M-box splitting procedure is shown in Figure 9. As one can see, blocks **1** and **2** (three molecules each) share only one coupling molecule. The electrostatic profile volume slice in the range of (−0.07 eV, 0.07 eV) is also shown at the equilibrium point.

As an illustration for possibilities in Figure 9C and D, we present the short time fluctuations (∼500 fs) for the spin and charge densities (Mulliken analysis for central Co in the first M-box block). This ability is obviously essential, for instance, in the electronics of complex molecular systems: small changes in the geometry can drastically affect spin/charge distributions, and dynamic effects are also important, as proposed by Popov et al.[55]

We highlight that a deeper analysis of these system dynamics goes far beyond the original scope of this manuscript. For a final remark, let us speculatively assume that the M-box local wave function which is defined in the local basis over all coupled fragments can represent in some sense an approximation for the whole system wave function in analogy to the force description. This idea once more goes back to the origin of these model FMO and FMM methods in such a way that its reconstruction does not seem to be unusual. For example, one can see a straightforward way of reconstructing such a function for the benzene chain splitting shown in Figure 2B, by using only odd blocks. This wave function, Φ_{system} built upon the combination $\{\Psi_{B_1}^{1-2}, \Psi_{B_3}^{3-4}, ...\Psi_{B_9}^{9-10}\}$, is an *initial guess* or *an approximation* for the whole system in the same basis set as it is for blocks. A wave function defined in this way (only odd blocks) contains vectors which are normalized and mutually orthogonal. However, since its splitting-map does have some orthogonal components, the system wave function construction requests only an additional orthogonalization step. In other words, a system wave function can be glued (as we did for forces) with respect to the each assumed splitting scheme and coupling range in a linear/nonlinear fashion depending on necessity.

By using M-box ideas, we can obtain space- and time-consistent sampling of the potential energy surface for the system with experimentally relevant sizes and chemically accurate geometrical prediction. Such approximations for the dynamical wavefuntion and system Hamiltonian are suitable for physical phenomena such as those addressed here[4−6] or even more complex structures.[56] In the future, our next step is to generalize the M-box model for the situation where system blocks can act on a potential energy surface with different multiplicity/charge sampling. This would make the local block forces couple on the electronic degrees of freedom, with the possibility of sampling along MD trajectories becoming possible.

3.5. Multibox Possible Extensions. In summary, we would like to discuss further some technical details regarding the multibox method explicitly:

(a) Importantly, there are no strict limitations on the level of theory for the QM (it can be, for example, some post-HF theory) region nor for MM potentials.

(b) It is straightforward to notice that, in the M-box model, the energies and forces of each subsystem are evaluated in the presence of the surrounding MM potentials. Therefore, classical dispersion interactions for each quantum block can be included in analogy to a DFT-D model.[47] Again, DFT-D is given as an example, if a DFT model is used for the QM region.

(c) As for all QM/MM subsystems, S_i can be computed completely independently, and it is possible to build a distribution of S_i on subgroups of processors. Because of this fact, the method allows for an almost perfect parallelization for increasing system sizes. The only peculiar aspect of this type of parallelization will be the communication of the forces, which should take a negligible amount of time and was previously implemented by CP2k developers.[37]

(d) If for any reason the wave function of the whole system matters, as was shown for forces, a quick and efficient guess can be constructed in a local basis set. Coefficients can be defined as a combination (depending on a splitting scheme) with respect to the blocks in which they are.

(e) In addition to the QM/MM interactions, a block-to-block electronic structure correction might be included according to the Fedorov−Morokuma decomposition scheme. However, this correction will not be crucial for the local BO gradients on each subsystem.

(f) The M-box includes an additional useful tool. As one can see, the M-box splitting can obtain a quick and accurate initial wave function guess, where it is possible to control its main features such as net charge or spin for a fragment of the whole system. We plan to discuss this option elsewhere. So far, it is clear from the M-box problem construction that the M-box model is partially acting on a different QM space, with real space (atomic) coupling in the local basis set, and the description can be projected to the original atomic space in the same fashion as was done for the forces in the model discussed here. A simple example: suppose that the fragment B_1 in Figure 2B has an additional *replica* with additional charge or another excitation. Final forces can include this information. We can continuously assign an amount of charge or excitation contributions to the fragment during the whole system dynamics again by controlling fragment forces with additional knowledge about an approximation to the multiconfiguration wave function.

4. CONCLUSIONS AND OUTLOOK

In summary, we conclude that the proposed M-box scheme is a convenient technique for molecular dynamic studies of multimolecular systems. This model produces linear scaling with the number of submolecular blocks of the system, $\sim O(N_{bl})$. For a simple example, such as the presented one (molecular chain or covalently bonded systems), we have demonstrated a close correspondence between full *ab initio* molecular dynamics and M-box simulations. It was shown that it is possible to describe the time-dependent box crossing process and the thermodynamical characteristic for the weakly interacting systems. The correlations of the trajectories on time and space domains were also addressed, confirming that the M-box scheme is an appropriate approximation for the space−time-resolved transport phenomena studies (like energy/charge transport/transfer). We expect that employing such an accurate methodology in structural predictions can render well-defined local electronic properties for future time-dependent transport models in large-scale systems.

AUTHOR INFORMATION

Corresponding Author

*E-mail: s.avdoshenko@gmail.com.

ACKNOWLEDGMENT

S.A. is thankful for financial support from the Erasmus Mundus programme External Co-operation (EM ECW-L04 TUD 08-11) and Dr. F. Schiffmann (UCL, London), Dr. A. Popov (IFW, Dresden), Dr. I. Ioffe (MSU, Moscow), Dr. C. Toher (TU, Dresden), and Dr. C. G. Rocha (UJ, Finland) for helpful discussions.

REFERENCES

(1) Kendrew, J. C.; Bodo, C.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. The Amino-Acid Aequence X-ray Nethods, and Its Correlation With Chemical Data. *Nature* **1958**, *181*, 662–666.

(2) Lundstrom, D. Structural Genomics For Membrane Proteins. *Cell. Mol. Life Sci.* **2006**, *63*, 2597–2607.

(3) Blundell, T. L.; Johnson, L. N. *Protein Crystallography*; Academic Press: London, 1976.

(4) Grozema, F. C.; Berlin, Y. A.; Siebbeles, L. D. A. Mechanism of Charge Migration through DNA: Molecular Wire Behavior, Single-Step Tunneling or Hopping? *J. Am. Chem. Soc.* **2000**, *122*, 10903–10909.

(5) Berlin, Y. A.; Burin, A. L.; Siebbeles, L. D. A.; Ratner, M. A. Conformationally Gated Rate Processes in Biological Macromolecules. *J. Phys. Chem. A* **2001**, *105*, 5666–5678.

(6) Woiczikowski, P. B.; Kubar, T.; Elstner, M.; Cuniberti, G.; Gutierrez, R.; Caetano, R. Charge Transport Through Biomolecular Wires in a Solvent: Bridging Molecular Dynamics and Model Hamiltonian Approaches. *Phys. Rev. Lett.* **2009**, *102*, 208102.

(7) Iannuzzia, M.; Hutter, J. Comparative study of the nature of chemical bonding of corrugated graphene on Ru(0001) and Rh(111) by electronic structure calculations. *Surf. Sci.* **2011**, *15*, 1360–1368.

(8) Ding, Y.; Iannuzzi, M.; Hutter, J. Investigation of h-BN/Rh(111) nanomesh interacting with water and atomic hydrogen. *CHIMIA* **2011**, *65*, 256–259.

(9) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(10) Sulpizi, M.; Raugei, S.; VandeVondele, J.; Carloni, P.; Sprik, M. Calculation of Redox Properties:âĂL′ Understanding Short- and Long-Range Effects in Rubredoxin. *J. Phys. Chem. B* **2007**, *111*, 3969–3976.

(11) Van Voorhis, T.; et al. The Diabatic Picture of Electron Transfer, Reaction Barriers, and Molecular Dynamics. *Annu. Rev. Phys. Chem.* **2010**, *61*, 149–170.

(12) Gamiz-Hernandez, A. P.; Kieseritzky, G.; Ishikita, H.; Knapp, E. W. Rubredoxin Function: Redox Behavior from Electrostatics. *J. Chem. Theory Comput.* **2011**, *7*, 742–752.

(13) Talbot, J.; Frenkel, D. Short-time correlations in liquids: Molecular-dynamics simulation of hard spheroids. *Phys. Rev. Lett.* **1990**, *65*, 2828.

(14) Pagonabarraga, I.; Hagen, M. H. J.; Lowe, C. P.; Frenkel, D. Short-time dynamics of colloidal suspensions in confined geometries. *Phys. Rev. E* **1999**, *59*, 4458.

(15) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. P. Development and Use of Quantum Mechanical Molecular Models. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3002.

(16) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-Consistent-Charge Density-Functional Tight-Binding Method For Simulations of Complex Materials Properties. *Phys. Rev. B* **1998**, *58*, 7260.

(17) Greengard, L.; Rokhlin, V. A fast algorithm for particle simulations. *J. Comput. Phys.* **1987**, *73*, 325.
(18) Clare, B. W.; Kepert, D. L. Structures, stabilities and isomerism in $C_{60}H_{36}$ and $C_{60}F_{36}$. A comparison of the AM1 Hamiltonian and density functional techniques. *THEOCHEM* **2002**, 195–207.
(19) Cybulski, S. M.; Bledson, T. M.; Toczylowski, R. R. Comment on "Hydrogen bonding and stacking interactions of nucleic acid base pairs: A density-functional-theory treatment. *J. Chem. Phys.* **2002**, *116*, 11039.
(20) Witek, H. A.; Irle, S.; Zheng, G.; de Jong, W. A.; Morokuma, K. Modeling carbon nanostructures with the self-consistent charge density-functional tight-binding method: Vibrational spectra and electronic structure of C_{28}, C_{60}, and C_{70}. *J. Chem. Phys.* **2006**, *125*, 214706.
(21) Zheng, G.; et al. Parameter Calibration of Transition-Metal Elements for the Spin-Polarized Self-Consistent-Charge Density-Functional Tight-Binding (DFTB) Method:âĂL′ Sc, Ti, Fe, Co, and Ni. *J. Chem. Theory Comput.* **2007**, *3*, 1349–1367.
(22) Scuseria, G. E. Linear Scaling Density Functional Calculations with Gaussian Orbitals. *J. Phys. Chem. A* **1999**, *103*, 4782–4790.
(23) Kudin, K. N.; Scuseria, G. E. Linear Scaling Density Functional Theory with Gaussian Orbitals and Periodic Boundary Conditions: Efficient Evaluation of Energy and Forces via the Fast Multipole Method. *Phys. Rev. B* **2000**, *61*, 16440–16453.
(24) Lippert, G.; Hutter, J.; Parrinello, M. The Gaussian and Augmented-Plane-Wave Density Functional Method for *Ab Initio* Molecular Dynamics Simulations. *Theor. Chem. Acc.* **1999**, *103*, 124–140.
(25) Guidon, M.; Schiffmann, F.; Hutter, J.; VandeVondele, F. *Ab Initio* Molecular Dynamics Using Hybrid Density Functionals. *J. Chem. Phys.* **2008**, 128.
(26) Gao, J. Toward a Molecular Orbital Derived Empirical Potential for Liquid Simulations. *J. Phys. Rev. B.* **1997**, *101*, 657–663.
(27) Gao, J. A molecular-orbital derived polarization potential for liquid water. *J. Phys. Rev.* **1998**, *109*, 2346.
(28) Song, L.; Han, J.; Lin, Y.-l.; Xie, W.; Gao, J. Explicit Polarization (X-Pol) Potential Using ab Initio Molecular Orbital Theory and Density Functional Theory. *J. Phys. Rev. A.* **2009**, *113*, 11656–11664.
(29) Kitaura, K.; Morokuma, K. New Energy Decomposition Scheme for Molecular-Interactions Within Hartree-Fock Approximation. *Int. J. Quantum Chem.* **1976**, *10*, 325.
(30) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment molecular orbital method: application to polypeptides. *J. Phys. Chem. A* **2007**, *111*, 6904–6914.
(31) Fedorov, D.; Kitaura, K. *The Fragment Molecular Orbital Method*; CRC Press: Boca Raton, FL, 2009.
(32) Warshel, A.; Bromberg, A. Deuterium in the Initiation Step of the Thermal Reaction With Oxyge. *J. Chem. Phys.* **1970**, *52*, 1262.
(33) Methods and Applications of Combined Quantum Mechanical and Molecular Mechanical Potentials. In *Reviews in Computational Chemistry*; Eds., Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1995; Vol. 7, pp 119−185.
(34) Altoe, P.; Stenta, M.; Bottoni, A.; Garavelli, M. A Tunable QM/MM Approach to Chemical Reactivity, Structure and Physico-Chemical Properties Prediction. *Theor. Chem. Acc.* **2007**, *118*, 219.
(35) Senn, H. M.; Thiel, W. QM/MM Methods for Biological Systems. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198–1229.
(36) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutterc, J. Quickstep: Fast and Accurate Density Functional Calculations Using a Mixed Gaussian and Plane Waves Approach. *Comput. Phys. Commun.* **2005**, *167*, 103–128.
(37) The CP2K developers group (2010). http://cp2k.berlios.de (accessed October 2011).
(38) Laino, T; Mohamed, F; Laio, A; Parrinello, M. An efficient real space multigrid OM/MM electrostatic coupling. *J. Chem. Theory Comput.* **2005**, *6*, 1176–1184.
(39) Laino, T; Mohamed, F; Laio, A; Parrinello, M. An efficient linear-scaling electrostatic coupling for treating periodic boundary conditions in QM/MM simulations. *J. Chem. Theory Comput.* **2006**, *5*, 1370–1378.
(40) Humphrey, W.; Dalke, A.; Schulte, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
(41) Goedecker, S.; Teter, M.; Hutter, J. Separable Dual-Space Gaussian Pseudopotentials. *J. Phys. Rev. B.* **1996**, *54*, 1703.
(42) Case, D. A.; Cheatham, T. A.; et al. *AMBER 10*; San Fransisco, CA, 2008.
(43) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham; et al. Amber. *Comp. Phys. Commun.* **1995**, *91*, 1112.
(44) Case, D. A.; Cheatham, T.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. Amber. *J. Comput. Chem.* **1995**, *26*, 1668.
(45) Linstrom, P. J.; Mallard, W. G. *NIST ChemistryWebBook*, NIST Standard Reference Database 69; 2010.
(46) Leimkuhler, B.; Reich, S. Symplectic Integration Of Constrained Hamiltonian Systems. *Math. Comput.* **1994**, *63*, 589–605 and references therein.
(47) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate *Ab initio* Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
(48) Huckel, E. *Z. Phys.* **1932**, *70* (3/4), 204–86. *Z. Phys.* **1932**, *72*, 310–37. *Z. Phys.* **1932**, *76*, 628–48.
(49) Piaggio, P.; Dellepiane, G.; Piseri, L.; Tubino, R.; Taliani, C. Polarized infrared spectra of highly oriented polyacetylenes. *Sol. Commun.* **1984**, 947–956.
(50) Head-Gordon, M. Quantum Chemistry and Molecular Processes. *J. Phys. Rev.* **1996**, *100*, 13213–13225.
(51) Cohen, A. J.; Mori-Sanchez, P.; Yang, W. Insights into Current Limitations of Density Functional Theory. *Science* **2008**, *321*, 792–794.
(52) Bulo, R. E.; Ensing, B.; Sikkema, J.; Lucas Visscher, L. Toward a Practical Method for Adaptive QM/MM Simulations. *J. Chem. Theory Comput.* **2009**, *5*, 2212–2221.
(53) Heyden, A.; Truhlar, D. G. Conservative Algorithm for an Adaptive Change of Resolution in Mixed AtomisticCoarse-Grained Multiscale Simulations. *J. Chem. Theory Comput.* **2008**, 217–221.
(54) Lanczos, C. *J. Res. Natl. Bur. Stand.* **1950**, *45*, 255. Gagliano, E.; et al. *Phys. Rev. B* **1986**, *34*, 1677.
(55) Popov, A. A.; Chen, C.; Yang, S.; Lipps, F.; Dunsch, L. Spin-Flow Vibrational Spectroscopy of Molecules with Flexible Spin Density: Electrochemistry, ESR, Cluster and Spin Dynamics, and Bonding in TiSc2N@C80. *ACS Nano* **2010**, *8*, 4857–4871.
(56) Lee, M. H.; Avdoshenko, S.; Gutierrez, R.; Cuniberti, G. Charge migration through DNA molecules in the presence of mismatches. *Phys. Rev. B* **2010**, *82*, 155455.