IM-UFF: extending the Universal Force Field for interactive molecular modeling

Léonard Jaillet, Svetlana Artemova, Stephane Redon

IM-UFF: extending the Universal Force Field for interactive molecular modeling

Léonard Jaillet[a,*], Svetlana Artemova[a], Stephane Redon[a]

[a] *NANO-D, INRIA,*
Univ. Grenoble Alpes, LJK, F-38000 Grenoble, France,
CNRS, LJK, F-38000 Grenoble, France

Abstract

The Universal Force Field (UFF) is a broadly applicable classical force field that contains parameters for almost every atom type of the periodic table. This force field is non-reactive, *i.e.* the topology of the system under study is considered as fixed and no creation or breaking of covalent bonds is possible. This paper introduces IM-UFF (Interactive Modeling - UFF), an extension of UFF that combines the possibility to significantly modify molecular structures (as with reactive force fields) with a broad diversity of supported systems thanks to the universality of UFF. Such an extension lets the user easily build and edit molecular systems interactively while being guided by physics based interatomic forces. This approach introduces weighted atom types and weighted bonds, used to update topologies and atom parameterizations at every time step of a simulation. IM-UFF has been evaluated on a large set of benchmarks and is proposed as a self-contained implementation integrated in a new module for the SAMSON software platform for computational nanoscience available at http://www.samson-connect.net.

Keywords: Universal Force Field, Interactive modeling, Interactive simulation, Empirical force fields, Molecular perception

*Corresponding author

1. Introduction

Force fields estimate the potential energy of the system under study and compute the interaction forces acting on the atoms involved. Hence, they are key elements for modeling inter-atomic interactions of molecular systems and are used as core components for a large spectrum of molecular simulation methods [1, 2], from molecular dynamics [3] to Monte-Carlo approaches [4].

The Universal Force Field [5] (UFF) is an all-atom force field that has parameterizations for every atom of the periodic table with atomic number lower than 103. Such a flexibility makes UFF applicable to a broad spectrum of systems, which has been demonstrated through evaluations on organic molecules [6], main group compounds [7], and metal complexes [8]. Recently, new parameterizations have been introduced to treat transition metals that appear, in particular, in Metal-Organic Frameworks (MOFs) [9]. However, this universality naturally comes at the expense of a lower accuracy: for example, it has been shown that UFF is not well-adapted for condensed-phase simulations [10].

UFF assumes the topological invariance of the simulated system, and hence can be categorized among *non-reactive* force fields. Non-reactive force fields allow for small geometric changes in the system, but prohibit any large rearrangements involving creation or breaking of covalent bonds, as well as changes in atoms' hybridization states. That is why, for UFF, the topology of the system (*i.e.* covalent bonds and their order) as well as the proper set of parameters to be used for each atom are established only once before any energy or force computation is launched.

Contrarily to non-reactive force fields, *reactive* force fields do not consider the system's topology as fixed: during the simulation, covalent bonds may be created or broken, their order may change, which might also induce a change in the atoms hybridizations. Popular reactive force fields include the Stilling-Weber [11], Tersoff [12] and Brenner [13] potentials, as well as the ReaxFF potential [14]. A detailed analysis of the advantages and limitations of these force fields can be found in Kocbach et al. [15]. Reactive force fields are typically precise,

but their parameters are tuned to work on specific systems and/or a limited set of atoms. Recently, extensions of these reactive force fields have been provided to cover more chemical elements of the periodic table. For example, ReaxFF has been parameterized for a growing set of applications [16], which may involve many different types of atoms. However, obtaining these parameters for new chemical types is a tedious and time-consuming task, and each application still corresponds to a specific context (and software distribution) of the potential. Moreover, the high number of parameters involved in these force fields (which is also the reason of their precision) makes them, in general, computationally more expensive than non-reactive methods. Various methods have been developed in order to improve the performance of reactive force fields: for example, incremental algorithms were used to perform interactive modeling of hydrocarbon systems with Brenner potential [17]. However, reactive force fields cannot be easily extended to a broader variety of systems due to the aforementioned restrictions.

Interactive simulation, where a user can directly modify a system through simple interactions during minimization or simulation, has become a very popular subject of research in various fields, including quantum chemistry [18, 19, 20, 21, 22, 23, 24] and structural biology (see e.g. [25, 26]). A well-known example is the Foldit application [27], that proposes a *serious game* to address the protein structure prediction problem by allowing users to interactively manipulate proteins. Interactive approaches which merge calculation and visualization complement pure simulation methods, and can actually be combined with them (see e.g. [28]). Potentially enhanced by intuitive computer-human interfaces [18], these methods provide users with a simple way to act on a system and achieve a given goal, while respecting some physical laws. Moreover, interactive simulation methods may help users better apprehend the rules that govern the system's behavior, since they make it possible to directly see the system's response to the performed actions.

In this paper, we propose an extension of the Universal Force Field called IM-UFF (Interactive Modeling - UFF), that is able to smoothly handle topo-

logical changes of the modeled systems. More specifically, IM-UFF allows for creation and breaking of covalent bonds, changes in the order of covalent bonds as well as changes of atom typizations. With this extension, molecular structures can go through significant modifications while being simulated or edited, in order to reach a topology that better reflects the current organization of the involved atoms. Hence, simple systems can be built with a few mouse actions, and the construction of more complex systems, for which the exact topology is not necessary known in advance, can be guided by physics-based inter-atomic forces. The interest of such a methodology has already been demonstrated for hydrocarbon systems in [17]. In our case, IM-UFF benefits from the large diversity of supported systems resulting from the universality of UFF.

Note that it may be possible to convert UFF (and any non-reactive force field) to an interactive force field by directly computing the standard (UFF) topology and atoms types for the structure after each structural change. However, the resulting force-field would be non-smooth, with possibly strong gaps of forces between two arbitrarily close conformations, thus hindering any possibility of interactive modeling.

It is worth comparing IM-UFF with existing quantum chemistry methods. *Ab initio* methods are reliable, but they are the most time-consuming approaches, hence not suitable for real-time applications. DFT approaches (with in particular DFTB variants [29, 30]) and semi-emprirical methods which rely on more simplified formulations, have shown to be more efficient. For appropriate settings, these methods have allowed real-time calculations for systems comprising up to a few dozens of atoms [22, 21, 23]. However, quantum chemistry methods remain in general computationally demanding methods that, for now, can hardly address larger systems at interactive rates. Moreover, special care must typically be taken regarding the parameterization choice since some methods may lead to incorrect results if the system under study is not similar enough to the one in the database used to parametrize the method [31]. Unlike quantum chemistry methods, IM-UFF is not based on electronic structure calculations. As shown later in this paper, it is constructed such that its local minima mostly

coincide with those of UFF and hence, correct UFF equilibrium structures can be obtained through interactive modeling. Outside equilibrium, IM-UFF makes no attempt to provide realistic energy and forces (even though we show some numerical experiments that suggest that IM-UFF may provide fast approximate statistics). However, we believe that the universality of IM-UFF, its efficiency and its simplicity of use (it does not require any user-expert system preparation) make it an interesting alternative to existing quantum chemistry methods or preliminary system modeling.

Figure 1 shows, on a simple example of a molecular system being manipulated, the difference between UFF and IM-UFF. In this example, the user pulls away one of the oxygen atoms of a carbonate ion CO_3^{2-}. When this oxygen atom is sufficiently far from the carbon atom, UFF either shifts the whole carbonate ion (if covalent bonds are represented with harmonic constraints), or allows for an unrealistically long covalent bond with no structural rearrangement of the atoms (if covalent bonds are represented with Morse-like potentials). IM-UFF, on the contrary, breaks the covalent bond between the atoms, which results in a new structure.

The rest of the paper is organized as follows. Section 2 provides an overview of the methodology underlying IM-UFF, in particular the use of *partitions of unity*, weighting functions used to transition between different parameterizations. Section 3 describes the algorithms used to compute the weights for IM-UFF, and Section 4 presents how these weights are applied. In Section 5, IM-UFF is evaluated on a set of scenarios and benchmarks, which also opens discussions. Finally, Section 6 concludes this work.

2. Overview of the IM-UFF approach

In this section, we present in broad lines the concepts used by IM-UFF for the perception phase and the computation of energies and forces.

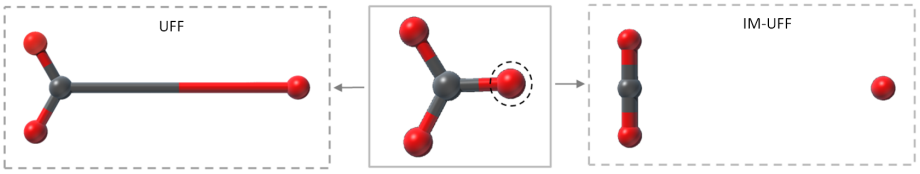

Figure 1: An oxygen atom (dashed circle) of the carbonate ion CO_3^{2-} is displaced using the interactive simulation framework in SAMSON software (center). With UFF where the bond stretch is implemented with a Morse potential, the topology remains unchanged, which leads to unrealistic geometries (left). With IM-UFF, the covalent bond is broken forming a carbon dioxide CO_2 and an isolated oxygen (right).

2.1. Continuous molecular system perception

With the Universal Force Field (UFF), interaction energies and forces cannot be directly deduced from the atoms' elements and positions: an initialization step is required to *perceive* the molecular system. Precisely, this perception step determines the topology of the system based on the atoms' elements and positions, as well as the parameters used to compute the UFF energies and forces. In particular, this step determines the set of covalent bonds, along with their bond orders, and assigns to each atom a *UFF type*. Once this one-time initialization step is performed, simulation can proceed and energies and forces may be updated whenever atoms positions change. However, the topology of the molecular system cannot change.

With the Interactive Modeling UFF (IM-UFF), perception is performed at *every time step* in order to allow for changes in the system's topology. Moreover, to make these changes *continuous*, perception produces bonds that a) are more numerous and b) may appear at larger distances than bonds in equilibrium states. Furthermore, perception assigns to each atom a *mixture of UFF types*.

In the literature, a large amount of work has been done on automatic perception of molecular systems [32, 33, 34, 35, 36, 37]. Despite such a diversity, perception methods may be limited by their computational complexity if no assumption on the system is made. Moreover, as for available perception soft-

ware, it appears that only a few free tools exist [38, 39, 40], and they mostly focus on organic molecules. Recently, we have proposed an automatic structure perception that is well adapted for UFF [41], and on which we rely to extend UFF to IM-UFF.

2.2. Weight-based approach

Even though more bonds and UFF atoms types are considered at each time step in IM-UFF, these are given *weights*, which indicate their relative importance and are involved in the computation of energies and forces. Precisely, two lists are continuously maintained: a list of *weighted bonds* and a list of *weighted atom types*, defined as follows:

- A **weighted bond** b_{ij} is a pair of atoms i and j with an associated weight $\omega_{ij} \in [0, 1]$. When ω_{ij} is equal to 0, there is no bonded interaction between atoms i and j. When ω_{ij} is equal to 1, the weighted bond represents a covalent bond between atoms i and j. Intermediate weights $\omega_{ij} \in (0, 1)$ correspond to *partial bonds* being either created or destroyed during simulation with the IM-UFF potential[1].

- A **weighted atom type** is an UFF atom type associated to a weight $\lambda \in [0, 1]$. If a given atom in the molecular system has $n+1$ possible UFF types, we note as λ_i, $0 \leqslant i \leqslant n$, the weight of the i^{th} atom type. Moreover, for each atom, we impose $\sum_i \lambda_i = 1$, so that the set of λ_i constitutes a partition of unity. As a result, each atom in the simulation is associated to a set of weighted UFF atom types, with weights varying during the simulation as bonds are created or destroyed.

As we will show, weighted atom types are computed based on weighted bonds. The weights of bonds and atoms are then used to compute IM-UFF energies and forces. We ensure that, by construction, local minima in UFF

[1] For a given pair of atoms i and j, there is at most one corresponding weighted bond b_{ij}, which may also be denoted by b_{ji}, with corresponding weight ω_{ji}.

mostly coincide with local minima in IM-UFF, so that stable structures that can be obtained with UFF can also be built through interactive modeling using IM-UFF. This makes IM-UFF usable in all situations where UFF is appropriate.

The next section describes the workflow for computing weighted bonds and weighted atom types. Then, Section 4 explains how to use these weights in order to compute the interaction energies and forces that rule the behavior of IM-UFF.

3. Computing weights

In this section, after defining a few quantities used throughout the paper, we describe our approach to computing weights associated to bonds and atoms types.

3.1. Definitions

We use the following definitions:

- r_{ij}^C is the *covalent length*, *i.e.* the distance below which a bond between atoms i and j may be covalent, provided that considering the bond as covalent is compatible with the maximum coordination and maximum valence of atoms i and j. As in Artemova et al. [41], we use:

$$r_{ij}^C = r_{ij}^{EQ} + \epsilon_{th}, \tag{1}$$

  where r_{ij}^{EQ} represents an equilibrium bond length taken as the sum of covalent radii of the atoms: $r_{ij}^{EQ} = r_i + r_j$. The covalent radii used are the UFF radii associated to a given typization (see Supplementary Material). As in Artemova et al. [41], we use $\epsilon_{th} = 0.4\mathring{A}$.

- $r_{ij}^I \geqslant r_{ij}^C$ is the *interaction length* of a weighted bond, *i.e.* the length below which a (weighted) bonded interaction is considered between atoms i and j. In this work, this interaction length is the van der Waals equilibrium distance, *i.e.* $r_{ij}^I = x_{IJ}$ in equation (20) of Rappe et al. [42]. As a result,

beyond the van der Waals equilibrium distance, the weighted bond entirely disappears and the interaction energy is purely a van der Waals energy.

- $co(i)$ is the *weighted coordination* of atom i:

$$co(i) = \sum_j \omega_{ij} \tag{2}$$

- $va(i)$ is the *weighted valence* of atom i:

$$va(i) = \sum_j bo_{ij}\omega_{ij}. \tag{3}$$

where bo_{ij} is the bond order of bond b_{ij} when the bond is covalent, and bo_{ij} is set to 1 for partial bonds.

- $co_{max}(i)$ and $va_{max}(i)$ are the *maximum coordination* and *maximum valence*, respectively, of atom i. Their values are the same as the ones used in Artemova et al. [41].

3.2. Computing bond weights

In this section, we detail how bond weights ω_{ij} are computed. These weights vary in the range $[0, 1]$ and have the following properties:

$$\omega_{ij}(r_{ij}) = \begin{cases} 0, & \text{if } r_{ij} \geqslant r_{ij}^I \\ 1, & \text{if } b_{ij} \text{ is a covalent bond} \end{cases} \tag{4}$$

where r_{ij} corresponds to the length of bond b_{ij}. The first property ensures that weighted bonds smoothly vanish when their length r_{ij} becomes larger than their interaction length r_{ij}^I. The second property, which states that a covalent bond has a weight of 1, allows us to obtain energies and forces that are similar to those of UFF for the systems at equilibrium.

The overall process used to build weighting functions ω_{ij} requires several intermediate steps that are summarized in Figure 2. These steps are detailed in the following subsections.

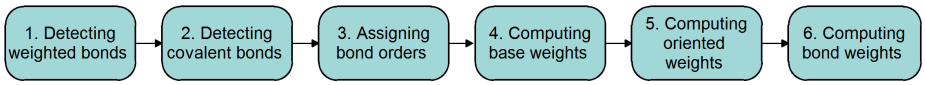

Figure 2: Workflow used for computing the weight ω_{ij} for a bond connecting atoms i and j.

3.2.1. Detecting weighted bonds

In theory, we could associate a weighted bond to any pair of atoms i and j such that $i \neq j$. We would then set $\omega_{ij} = 0$ for bonds whose length r_{ij} is larger than their interaction length r_{ij}^I. For efficiency, though, we only consider weighted bonds as created when distances r_{ij} are smaller than interaction distances $\leqslant r_{ij}^I$. To rapidly determine all pairs of atoms that may form a bond, we use a cell-list approach [2] where atoms are placed in a regular grid such that a cell size is equal to the largest possible bond distance[2]. By default, the grid is updated at each time step, and only atoms belonging to the same cell or to neighboring cells are tested for bond creation. Note that, in this step, the number of weighted bonds attached to a given atom does not have to satisfy the maximum coordination rule, since this constraint is enforced later on.

3.2.2. Detecting covalent bonds

Once weighted bonds are known, we detect covalent bonds. We use an approach very similar to the one used for UFF in Artemova et al. [41]. The only difference is that, instead of relying on a cell-list approach to find covalent bonds, we search them among the weighted bonds computed at the previous step. Precisely, for each atom i, we sort its weighted bonds according to the ratios r_{ij}/r_{ij}^{EQ}. Then, starting first with the bond with the lowest ratio, we tag these bonds as covalent if their distance is lower than the maximum covalent radius r_{ij}^C, as long as the number of covalent bonds for each atom does not

[2]In our case, this is 4.765 Å, corresponding to the maximum x_I parameters appearing in the UFF parameter table of Rappe et al. [5].

exceed its maximum coordination[3]. Once a bond has been detected as covalent, we directly set its weight ω_{ij} to 1 to satisfy the second property in equation 4.

3.2.3. Assigning bond orders

After covalent bonds are known, we determine which of them may be considered as double or triple bonds. We restrict the search to bonds where both atoms are among the six following elements: carbon (C), nitrogen (N), oxygen (O), silicon (Si), phosphorus (P) and sulfur (S). Bond-order assignment uses the method described in Artemova et al. [41], which extends the method based on *molecular penalty scores* introduced by Wang et al. and implemented in the popular Antechamber package [43]. One advantage of this extension is that a local bond-order estimation strategy propagated along the molecular system is used to limit the computational effort. Once this step is complete, we may compute the *covalent coordination* $\widetilde{co}(i)$ and the *covalent valence* $\widetilde{va}(i)$ of atom i, *i.e.* the coordination and valence of atom i when only its covalent bonds (and not its partial bonds) are considered.

3.2.4. Computing base weights

For each partial bond detected in step 1 (*i.e.* for each weighted bond that is not a covalent bond), we compute a *base weight*, *i.e.* the weight that the bond would have if the maximum coordination and maximum valence rules were not enforced.

Precisely, a twice-differentiable base weight ω_{ij}^0 is computed from the bond length r_{ij} as follows:

$$\omega_{ij}^0(r_{ij}) = \begin{cases} 1, & \text{if } r_{ij} \leqslant r_{ij}^C \\ 0, & \text{if } r_{ij} \geqslant r_{ij}^I \\ s(r_{ij}), & \text{otherwise} \end{cases} \tag{5}$$

[3]The maximum coordination co_{max} only depends on the element type, except for hydrogen, where $co_{max} = 1$ except when it is bonded to two boron elements. In this case, $co_{max} = 2$, allowing for the creation of a bridging hydrogen.

where $s(r_{ij})$ is an order-five, twice-differentiable interpolation spline:

$$s(r_{ij}) = -6\eta^5 + 15\eta^4 - 10\eta^3 + 1, \quad \eta = \frac{r_{ij} - r_{ij}^C}{r_{ij}^I - r_{ij}^C}. \tag{6}$$

The function ω_{ij}^0 is plotted in Figure 3 for a bond linking two hydrogen atoms.

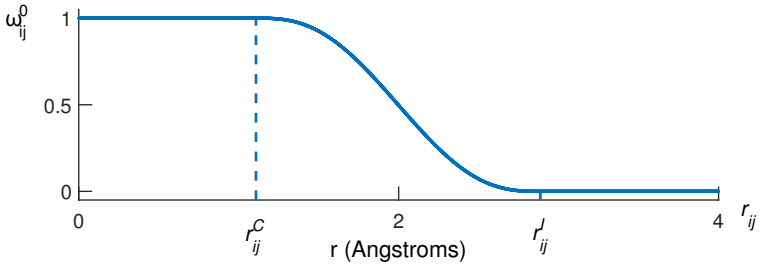

Figure 3: Weight ω_{ij}^0 as a function of r_{ij} (in Angstroms) for two hydrogen atoms ($r_{ij}^C = 0.86$ Å, $r_{ij}^I = 2.886$ Å).

3.2.5. Computing oriented weights

Once base weights are known, we compute for each partial bond b_{ij} two *oriented weights* $\omega_{i\to j}$ and $\omega_{j\to i}$ in order to satisfy the maximum coordination and maximum valence rules for atoms i and j. Precisely, we set $\omega_{i\to j} = \omega_{j\to i} = 1$ for all covalent bonds, and we want:

$$\begin{cases} \sum\limits_j \omega_{i\to j} \leqslant co_{max}(i) & \text{and} \quad \sum\limits_j bo_{ij}\omega_{i\to j} \leqslant va_{max}(i) \\ \sum\limits_i \omega_{j\to i} \leqslant co_{max}(j) & \text{and} \quad \sum\limits_i bo_{ji}\omega_{j\to i} \leqslant va_{max}(j) \end{cases}. \tag{7}$$

The approach used to compute these oriented weights is described in Algorithm 1. For each atom i, the first step is to compute its *available coordination* $co_{avail}(i)$ after considering the covalent bonds. For this, we compute a) the difference between the maximum coordination $co_{max}(i)$ and the covalent coordination $\widetilde{co}(i)$, and b) the difference between the maximum valence $va_{max}(i)$ and the covalent valence $\widetilde{va}(i)$, and we set the available coordination $co_{avail}(i)$ as the minimum of these two values (line 2). Then, for each partial bond b_{ij}

Algorithm 1: Computing oriented weights

input : A list of all the atoms. For each atom i, a list of partial bonds b_{ij}, their weights ω_{ij}^0, the coordination of covalent bonds $\widetilde{co}(i)$ and the valence of covalence bonds $\widetilde{va}(i)$.

output: Two oriented weights $\omega_{i\to j}$ and $\omega_{j\to i}$ per bond b_{ij}.

1 **foreach** *atom* i **do**
2 $\quad co_{avail}(i) \leftarrow min\left(co_{max}(i) - \widetilde{co}(i), va_{max}(i) - \widetilde{va}(i)\right)$;
3 $\quad$ **foreach** *partial bond* b_{ij} **do**
4 $\quad\quad \omega_{i\to j} \leftarrow \max\left(0, \min\left(\omega_{ij}^0, co_{avail}(i)\right)\right)$;
5 $\quad\quad co_{avail}(i) \leftarrow co_{avail}(i) - \omega_{i\to j}$;

connected to atom i (in the order used when determining covalent bonds), the oriented weight $\omega_{i\to j}$ is set (line 4). At this stage, the min function ensures that the total coordination and valence of atom i is always below its maximum values, whereas the max function ensures that $\omega_{i\to j}$ is always positive. We then subtract $\omega_{i\to j}$ from the atom's available coordination (line 5).

3.2.6. Computing bond weights

For a partial bond b_{ij}, oriented weights $\omega_{i\to j}$ and $\omega_{j\to i}$ vary continuously from 0 to 1 and are in agreement with maximum coordinations and valences of atoms i and j, respectively. To get a weighting function that satisfy coordinations and valences of both connected atoms, we take the minimum of these oriented weights:

$$\omega_{ij} = \min(\omega_{i\to j}, \omega_{j\to i}). \tag{8}$$

3.2.7. Special cases

The steps above ensure that the properties in equation (4) are enforced. In cases where covalent bonds describe a system at equilibrium, however, we would like to have the weights of non-covalent bonds to be zero. This happens when either the weighted coordination or the weighted valence is equal to its maximum

value, but one may find other cases where this property is not true. Consider for example the case of Figure 4 involving two carbonyl groups (terminal oxygen double bonded with a carbon atom). Since, in our framework, the maximum valence of oxygen is set to 3 (to take into account the possible presence of a dative bond, e.g. in carbon monoxide), there is the risk of creating a meaningless attraction between the two oxygen atoms through a bond of weight $\omega_{OO} \neq 0$. Hence, we use a post-treatment phase to set the weight of some non-covalent weighted bonds to 0, for several known equilibrium situations:

- an oxygen doubly bonded to a carbon, a phosphorus or a sulfur, the bonded atom having 3 covalent bonds.
- a sulfur doubly bonded to a carbon that has 3 covalent bonds.
- a sulfur with 2 bonds, one of these bonds connecting a thallium, a boron or an arsenic.
- a boron with 3 bonds connecting at least 2 sulfur atoms.

Note that these cases cover the situations encountered in the UFF benchmarks that were used to validate our approach, but they do not necessarily cover all the possible situations that may appear. However, additional restrictions can easily be added to the method if necessary.

Another situation where the weight of non-covalent bonds is set to 0 is the case of resonant structures. This stage that is performed after the computation of the weighted types will be described more in details further on.

3.3. Computing weighted atom types

In this section, we explain how to obtain the weights $\lambda_0, \ldots, \lambda_{n-1}$ for the n available typizations of a given atom. First, let us notice that among the 103 type of chemical elements considered in UFF, only 12 may have multiple UFF atom types: hydrogen (H), boron (B), carbon (C), nitrogen (N), oxygen (O), phosphorus (P), sulfur (S), titanium (Ti), iron (Fe), molybdenum (Mo), tungsten (W) and rhenium (Re). Hence, a weight $\lambda_0 = 1$ is associated to the unique

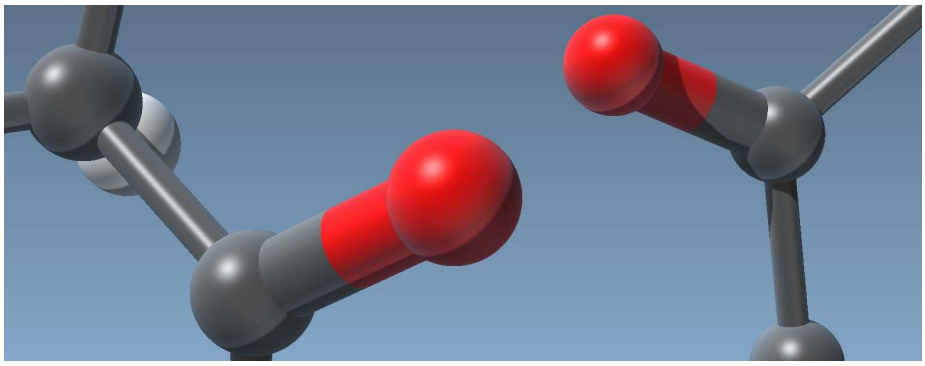

Figure 4: Two carbonyl groups, each one composed of a terminal oxygen doubly bonded with a carbon atom. In that case the unfilled oxygen valences lead to a weighted bond between oxygen such that $\omega_{OO} \neq 0$. Hence, a postprocessing is required to remove this weight which does not have physical meaning.

typization available for all other atom types. It is also important to remark that an atom typization may depend on its geometry/hybridization (*e.g.* carbon, nitrogen), its oxidation number (phosphorus) or on both characteristics (sulfur and tungsten). Moreover, UFF has two special cases for bridging hydrogen atoms (labeled H_b) and oxygen atoms in zeolite lattices (labeled O_3_z).[4]

In the following, we rely on the automatic typization process for UFF proposed in Artemova et al. [41], that we extend to obtain associated weights IM-UFF in three main steps: first, we perceive the hybridizations/geometries for atoms whose types depend on such characteristics; then, we estimate the oxidation numbers when these are used to discriminate between possible types; finally, we assign atom types and weights. We detail these steps in the following.

[4]Note that for some element types, some geometries (*e.g.* trigonal bipyramidal, square pyramidal, tricapped trigonal) are not available from the possible typizations of UFF and the extension to such kind of geometry is out of the scope of the current paper.

3.3.1. Perceiving hybridization/geometry

The atom hybridization/geometry is determined at the beginning of the type perception for 10 atomic elements: *main-group elements* carbon (C), oxygen (O), nitrogen (N), sulfur (S) and boron (B), for which the typization is related to the atom's hybridization state (sp, sp^2, sp^3), and *metallic elements* titanium (Ti), iron (Fe), molybdenum (Mo), tungsten (W) and rhenium (Re), for which the typization depends on the geometric arrangement (either trigonal or octahedral in UFF).

Main-group elements: based on the octet rule and following the same scheme as in [41], we compute an $sp^*(i)$ index that represents the hybridization tendency of the atom and that can be obtained from the previously computed weighted coordinations and valences.

The first step is to estimate the number of lone pairs $lp(i)$ associated to the atom. For that, we compute the expression $[ve(i)-va(i)]$ where $[\cdot]$ stands for the the nearest integer function and $ve(i)$ is the number of the valence electrons. If the resulting value is odd, then we add or remove 1 to the expression to estimate $lp(i)$. As in [41], the decision to remove or add an electron is taken in order to satisfy as much as possible the octet rule (*i.e.* each atom should be surrounded by 8 electrons). Hence, the number of lone pairs is estimated as:

$$lp(i) = \begin{cases} \frac{[ve(i)-va(i)]}{2} & \text{if } [ve(i)-va(i)]\%2 = 0; \\ \frac{[ve(i)-va(i)](\pm)1}{2} & \text{otherwise;} \end{cases} \tag{9}$$

This allows us to estimate the weighted hybridization as follow:

$$sp^*(i) = co(i) + lp(i) - 1. \tag{10}$$

Note that, since co can have non integer values, sp^* can also have non-integer values. For carbon, nitrogen, oxygen and sulfur atoms, these non-integer sp^* values are used to build three weights, denoted by λ_l, λ_{tr}, and λ_{te}, respectively associated to linear, trigonal and tetrahedral hybridizations/geometries. By forming a partition of the unity, these weights represent the tendency of the atom to adopt one of these three hybridizations/geometries. To determine these weig-

	λ_l	λ_{tr}	λ_{te}
$sp^* \leqslant 1$	1	0	0
$1 < sp^* \leqslant 2$	$\frac{1}{2}\left(1+\cos\left(\pi\left(sp^*-1\right)\right)\right)$	$1-\lambda_l$	0
$2 < sp^* \leqslant 3$	0	$\frac{1}{2}(1+\cos(\pi(sp^*-2)))$	$1-\lambda_{tr}$
$3 < sp^*$	0	0	1

Table 1: Partition functions λ for carbon, nitrogen, oxygen and sulfur atoms, describing the weighted typization hybridizations/geometries, in function of the sp^* parameter.

hts, we define the λ functions as described in Table 1 and plotted (as functions of sp values) in Figure 5-left. This construction allows us to smoothly switch from linear to trigonal and then to tetrahedral hybridizations/geometries for main-group elements.

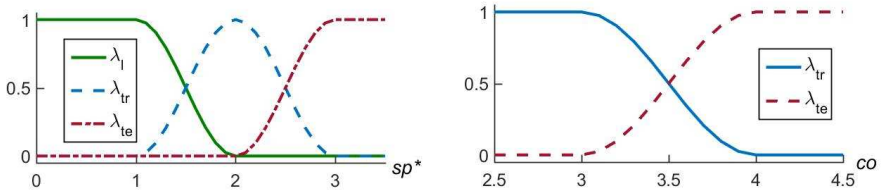

Figure 5: Partition functions for main group elements. Left: functions which allow to smoothly switch from linear to trigonal and then tetrahedral hybridizations/geometries for carbon, oxygen, nitrogen, sulfur elements. Right: functions that allow to switch from trigonal to tetrahedral for boron.

Similarly, for boron atoms, non-integer sp^* values are used to associate weights to the two possible typizations corresponding to either trigonal or tetrahedral geometries/hybridizations. Since bond orders greater than one are not considered for boron, the valence is equal to the coordination. Thus, we use the boron coordination to design the partition functions λ_{tr} and λ_{te} as described in Table 2. Figure 5-right plots the λ functions according to the boron co values.

Metallic elements: for transition metals (titanium (Ti), iron (Fe), molybdenum (Mo), tungsten (W) and rhenium (Re)), tetrahedral (λ_{te}) and octahedral (λ_{oc}) typizations are available. An additional difficulty for the choice of the typi-

	λ_{tr}	λ_{te}
$co \leqslant 3$	1	0
$3 < co \leqslant 4$	$\frac{1}{2}\left(1 + \cos\left(\pi\left(co - 3\right)\right)\right)$	$1 - \lambda_{tr}$
$4 < co$	0	1

Table 2: Partition functions for boron according to its coordination.

zation for these atoms is that the octahedral typization can serve for both square planar and octahedral structures. Thus, the strategy retained is the following. First, based on the coordination, we build weight functions which discriminate the tetrahedral and square planar contribution λ_{te+sp}, from the octahedral contribution λ^*_{oc}. The partition functions used are described in Table 3 and are represented in Figure 6. Additionally, we use a shape-checking mechanism to discriminate between the tetrahedral and the square planar contributions. For this, we define two error functions E_{te} and E_{sp} that sum the deviations from the ideal reference angles of the respective geometries, while weighting these errors by the bond weights:

$$\begin{aligned} E_{te} &= \textstyle\sum_i \omega_i (\theta_i - 109.47°)^2, \\ E_{sp} &= \textstyle\sum_i \omega_i \min\left((\theta_i - 90°)^2, (\theta_i - 180°)^2\right), \end{aligned} \tag{11}$$

where i sums over all the possible angle bends that can be formed from neighbors of the central metallic atom, and ω_i is the product of the two weights of the bonds involved in the angle θ_i. From these values, we deduce the relative tetrahedral and square planar contributions from the overall tetrahedral plus square planar contribution:

$$\begin{aligned} \lambda_{te} &= \lambda_{te+sp} \tfrac{E_{te}}{E_{te}+E_{sp}}, \\ \lambda_{sp} &= \lambda_{te+sp} \tfrac{E_{sp}}{E_{te}+E_{sp}}. \end{aligned} \tag{12}$$

Finally, since the octahedral typization is used for both the square planar and the octahedral shapes, the weight of such typization is equal to the sum of each elementary contribution. Hence, the final octahedral typization is computed as $\lambda_{oc} = \lambda_{sp} + \lambda^*_{oc}$.

	λ_{te+sp}	λ^*_{oc}
$co \leqslant 4$	1	0
$4 < co \leqslant 6$	$\frac{1}{2}\left(1+\cos\left(\pi\frac{(co-4)}{2}\right)\right)$	$1-\lambda_{te+sp}$
$6 < co$	0	1

Table 3: Functions λ for metallic elements describing the weight of each typization according to the atom coordination.

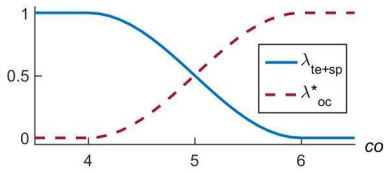

Figure 6: Partition functions allowing to smoothly switch from tetrahedral/square planar to octahedral geometries for metallic elements.

3.3.2. Oxidation numbers

In the case of phosphorus, sulfur and tungsten elements, several typizations are possible for a given hybridization/geometry, depending on the atom's oxidation number. The oxidation number is estimated using the same approach as described in Ref. [41] and is based on the atoms electronegativities.

When the typization for the exact computed oxidation state is not available (for example, there is no UFF type S_3+1), a default oxidation number is set (see Table 4).

Note, finally, that the P_3+5 and P_3+q typizations have exactly the same set of UFF parameters, so we just consider one of them to be the default typization representing the tetrahedral hybridization. The typization scheme based on the oxidation number is summarized in Table 4.

3.3.3. Specific typizations

Typizations for oxygen in zeolite lattices and for bridging hydrogens are set when specific patterns are detected in the molecular topology.

	Oxidation number	λ		
Phosphorus	3	$\lambda_{\text{P\_3+3}}$	=	1
	other cases	$\lambda_{\text{P\_3+5/P\_3+q}}$	=	1
Sulfur	4	$\lambda_{\text{S\_3+4}}$	=	λ_{te}
	6	$\lambda_{\text{S\_3+6}}$	=	λ_{te}
	other cases	$\lambda_{\text{S\_3+2}}$	=	λ_{te}
Tungsten	6	$\lambda_{\text{W\_3+6}}$	=	λ_{te}
	other cases	$\lambda_{\text{W\_3+4}}$	=	λ_{te}

Table 4: Typization scheme of atoms based on the oxidation number.

The λ_{te} weight of oxygen is attributed to the $\lambda_{O\_3\_z}$ corresponding to oxygen in zeolite lattices when the oxygen has two covalent bonds with either two silicon or two boron atoms.

Apart from its default typization H_ , hydrogen has a specific typization H_b when it is a bridging hydrogen. For simplicity, this case is considered only for the most common situation where the hydrogen atom is bonded to two boron atoms. In this case, we set $\lambda_{\text{H\_}} = 0$, $\lambda_{\text{H\_b}} = 1$ whereas in other cases we set $\lambda_{H\_} = 1$, $\lambda_{\text{H\_b}} = 0$.

3.3.4. Resonances

In IM-UFF, as in UFF, a specific typization is associated to atoms involved in resonance phenomena due to the presence of delocalized electrons. Similarly to the perception scheme proposed in Ref. [41], atoms resonances are considered for some functional groups (amide, nitro, carboxylate and enol-ether groups) and in aromatic rings. The global scheme used is the same as the one used in Ref. [41], applied only on covalent bonds ($\omega_{ij} = 1$). Once a resonant pattern is detected, the typizations of all its involved atoms are assigned to resonant, overwriting the previously-assigned typizations (for the current round of perception only), and the weights of non-covalent bonds of these atoms are set to 0. After this phase, weights have been assigned to the bonds and the types and they can be

used in IM-UFF to compute energies and forces. This is the subject of the next section.

4. Applying weights

We now describe how the weights are used in IM-UFF for energies and forces computations. In particular, we show how we weigh the various energy terms in the UFF potential energy function, based on the bond weights and atom types weights computed in the previous steps.

4.1. Parameters computation

Among UFF parameters, only three depend on the choice of the typization for a given atom element: the covalent radius r_I, the equilibrium angle θ_0 and the parameter V_J involved in the computation of torsional barriers (see Ref. [42] II.D.2). In IM-UFF, these parameters are directly computed as weighted sums of parameters corresponding to possible typizations:

$$r_I = \sum_{k=0}^{n} \lambda_k r_I^k, \quad \theta_0 = \sum_{k=0}^{n} \lambda_k \theta_0^k, \quad V_J = \sum_{k=0}^{n} \lambda_k V_J^k. \tag{13}$$

Recall that the *lambda* weights form a partition of unity, so that these weighted sums do not have to be normalized.

4.2. Weighted energies and forces

The total interaction energy in UFF is written as a sum of two-body, three-body and four-body interactions:

$$E = E_R + E_\theta + E_\phi + E_\Omega + E_{vdw} + E_{el}, \tag{14}$$

where E_R represents bond stretching interactions, E_θ describes angle bending energy terms, E_ϕ stands for dihedral angle torsions, and E_Ω corresponds to the inversion contribution. The non-bonded interactions comprise van der Waals terms E_{vdw} and electrostatic terms E_{el}. Note that this last electrostatic term is not considered in IM-UFF following the arguments in Ref. [44] stating that

UFF parameters are derived without the use of point charges on the atoms and that the consensus of the original authors is to use this force field without any additional charges.

The weights ω_{ij} computed in Section 3 represent the influence of bonds attached to atoms, while ensuring that coordinations and valences make physical sense. We thus use these weights to compute IM-UFF energies and forces. Precisely, energy terms E in UFF are transformed into energy terms $\tilde{E}$ in IM-UFF as follows:

$$\tilde{E} = g\left(\omega\right) E, \tag{15}$$

where $g(\omega) = e^{\frac{\omega-1}{\omega}}$. The g function rapidly converges to 0 when a weight ω tends towards 0. Hence, compared to ω, it allows steeper transition from 1 to 0, happening for lower radii (see figure 7), which in practice appears to be convenient when interactively manipulating molecular systems.

Bond stretching. In UFF, bond stretching energy interactions E_R are modeled either with a harmonic oscillator or with the Morse potential. In IM-UFF, we only rely on the Morse formulation which is more accurate and leads to finite energies when breaking bonds. Thus, we compute the weighted energy $\tilde{E}_R^{ij}$ as follows:

$$\tilde{E}_R^{ij} = g\left(\omega_{ij}\right)\left(E_R^{ij} - D_{ij}\right), \tag{16}$$

where D_{ij} is the bond dissociation energy as it appears in Rappe et al. [5], equation (1b). With such a formulation that directly links the bond weight to the energy, the subtraction of the bond dissociation energy allows the non-weighted contribution to vanish to 0 when the bond length tends to infinity. Figure 7 represents the weighting factor $g\left(\omega_{ij}\right)$ (left - solid line) compared to the ω_{ij} value (left - dashed line) and the resulting bond stretch function (right) for two interacting hydrogen atoms.

Angle bend. The angle bend energy E_θ involves 3 atoms and 2 bonds. Denoting by i, j and k these three atoms, and assuming j is the central atom, the modified

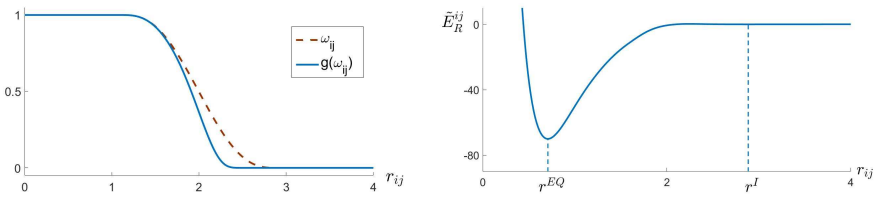

Figure 7: The energy weighting factor (left) and the resulting bond stretch energy function (right) representing the interaction of two hydrogen atoms.

contribution $\tilde{E}_\theta^{ijk}$ is:

$$\tilde{E}_\theta^{ijk} = g\left(\min\left(\omega_{ij}, \omega_{jk}\right)\right) E_\theta^{ijk}. \tag{17}$$

Dihedral torsion. The dihedral torsion energy E_ϕ involves 4 atoms connected through 3 successive bonds. Denoting by i, j, k and l these four atoms, and assuming the three bonds are b_{ij}, b_{jk} and b_{kl}, the modified contribution $\tilde{E}_\phi^{ijkl}$ is:

$$\tilde{E}_\phi^{ijkl} = g\left(\min\left(\omega_{ij}, \omega_{jk}, \omega_{kl}\right)\right) E_\phi^{ijkl}. \tag{18}$$

Inversion. For the inversion energy E_Ω^{ijkl}, 3 atoms are connected to a 4^{th} one through 3 bonds. Then, if j is the central atom, the contribution $\tilde{E}_\Omega^{ijkl}$ is computed such as:

$$\tilde{E}_\Omega^{ijkl} = g\left(\min\left(\omega_{ij}, \omega_{jk}, \omega_{jl}\right)\right) E_\Omega^{ijkl}. \tag{19}$$

Van der Waals. In UFF, a van der Waals interaction between two atoms is considered if two *van der Waals conditions* are satisfied: a) the atoms are not connected and b) they do not have a common neighbor. To transcribe these conditions in IM-UFF, we consider both the weight ω_{ij} of a bond between two atoms i and j, and a weight $\omega_n = \max_k(\omega_{ik}\omega_{kj})$ representing the strongest connection through a neighboring atom.

The van der Waals contribution is then:

$$\tilde{E}_{vdw}^{ij} = g\left(1 - \max\left(\omega_{ij}, \omega_n\right)\right) E_{vdw}^{ij}, \tag{20}$$

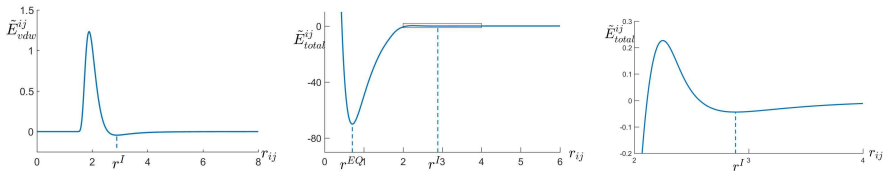

Figure 8: The van der Waals contribution (left), the bond stretch plus van der Waals contribution (middle) and a close-up of the van der Waals barrier (right) that appears as a red box in the middle picture, for the simple case of two Hydrogen atoms.

indicating that the van der Waals interaction becomes stronger as the van der Waals conditions become more and more satisfied.

Figure 8 illustrates the resulting van der Waals contribution (left), the total sum of bond stretch and van der Waals contributions (middle) and a close-up on this contribution showing the van der Waals barrier (right) for the simple case of two hydrogen atoms. Note how, with our formulation, the bond stretch and van der Waals minima are preserved, as well as the energy profiles (and thus forces) around these minima.

Forces. Forces are expressed as the negative of the derivative of the potential energy: $F = -\nabla E$. In IM-UFF, forces derived from each contribution regarding some displacement r are given by:

$$\tilde{F}_r = -\frac{\partial \tilde{E}}{\partial r} = -\frac{\partial \left(g(\omega)E\right)}{\partial r} = -\frac{g(\omega)}{\omega^2}\frac{\partial \omega}{\partial r}E - g(\omega)\frac{\partial E}{\partial r}. \tag{21}$$

The derivatives $\partial\omega/\partial r$ are obtained from the derivatives of the weights ω_{ij} (see Supplementary Material).

Note that the parameters r_I, θ_0 and V_J used to compute IM-UFF energies (cf. equation (14)) are also functions of r, since the weighted typizations λ are functions of the coordination which themselves depend on ω that are functions of r. For simplicity, however, we *do not consider these derivatives in our computations*, and the terms $\partial E/\partial r$ are computed as the normal derivatives in UFF.

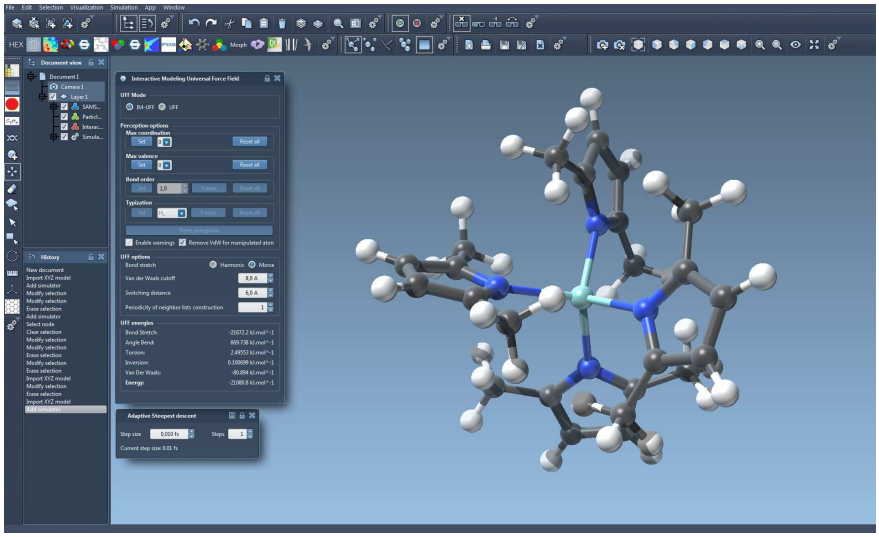

Figure 9: One of the benchmarks interactively modeled in SAMSON thanks to the IM-UFF module. The interface which appears on the left makes it possible to switch between IM-UFF and UFF, with UFF options described in Ref. [41]. Note that, when IM-UFF is set, manually setting typizations and bond orders is disabled.

5. Results and discussions

This section analyzes the performance of the IM-UFF approach. We first consider possible sources of discrepancies between UFF and IM-UFF from a theoretical point of view. Then, we compare minimum-energy structures produced by UFF and IM-UFF on a set of benchmarks. We demonstrate how IM-UFF may be used to interactively build molecular systems. We also show that IM-UFF allows us to obtain statistical properties that approximate well those obtained with a reactive force field. Finally, we discuss the obtained results. As for UFF described in Ref. [41], IM-UFF has been implemented in C++ and integrated in the SAMSON software platform [45] (see Figure 9).

5.1. Theoretical discrepancies analysis

Ideally, we would like equilibrium structures produced with IM-UFF to only include covalent bonds, with weights $\omega = 1$. In this case, the atoms' coordinations and valences would be exactly as in UFF, resulting in a weighted typization

$\lambda = 1$ for atom types. Such a perception would produce equilibrium energies and forces in IM-UFF identical to those in UFF, hence ensuring that UFF structures are preserved. In practice, however, discrepancies between IM-UFF and UFF may appear at equilibrium in the two following cases:

- As stated in subsection 3.2.7, one may encounter situations where some non-covalent bonds retain a non-zero weight, although considering only the covalent bonds would produce an equilibrium structure. We have seen that a post-treatment is proposed to treat such situations, but it may not cover all the cases that can be encountered.

- As with UFF, the interplay between energy components (bond terms, angle terms, etc.) is such that, even when the total potential energy is minimized, some individual components may not be. For example, van der Waals forces may lead to stretch a covalent bond, leading to an equilibrium distance between the atoms involved that is larger than the bond stretch equilibrium. In IM-UFF, this may lead to bond weights that are different from 1, leading to equilibrium structures that may be slightly different from the ones obtained with UFF.

In the following, we show that these situations are not common and that possible discrepancies are very limited.

5.2. Comparison with UFF

IM-UFF has been tested on a set of 156 molecules made up of four groups of benchmarks provided by the UFF authors to test their force field: 20 molecules from the original UFF paper[5] (Ref. [5]), 47 organic molecules (Ref. [6]), 57 main group compounds (Ref. [7]) and 32 metal complexes [6] (Ref. [8]).

[5] We count here only the molecules that do not already appear in other groups of benchmarks.

[6] Actually, 34 metal complexes are proposed in Ref. [8], but we excluded two metallocenes systems since they involve bonds to aromatic rings that are not currently supported in the SAMSON software platform.

Reference	#	#	Distance errors (Å)			Bond angle errors (deg)			Torsion errors (deg)		
	bench.	meas.	# meas.	average	max	# meas.	average	max	# meas.	average	max
Main article [5]	20	103	84	$1.2e^{-7}$	$1e^{-5}$	19	0	0	0	-	-
Organic [6]	47	263	142	$1.5e^{-6}$	$1.1e^{-4}$	97	$6.5e^{-4}$	$3e^{-2}$	24	$2.5e^{-5}$	$3.9e^{-4}$
Main group [7]	57	216	120	$1.0e^{-4}$	$2.6e^{-3}$	95	$1.9e^{-2}$	0.19	1	$1.0e^{-4}$	$1.0e^{-4}$
Metallic [8]	32	177	104	$2.0e^{-4}$	$6.1e^{-3}$	72	$2.8e^{-2}$	0.37	1	$2.8e^{-2}$	$2.8e^{-2}$

Table 5: Summary of the tests evaluating the difference of distances, bond angles and torsions obtained between UFF and IM-UFF for structures at equilibrium.

Distances, bond angles, and torsion angles are measured for systems at equilibrium with the IM-UFF potential energy, and compared with the values obtained with UFF, using the automatic typization method proposed in Ref. [41]. The differences obtained between UFF and IM-UFF are summarized in Table 5. The detail of the measures obtained with IM-UFF and compared with classical UFF is also shown in the Supplementary Material.

As one can see, average and maximum errors in distances, bond angles and torsions are extremely limited in all benchmarks.

Note, however, that since IM-UFF is very close to our UFF implementation at equilibrium, it retains the (limited) discrepancies identified in Ref. [41] with the reference UFF papers [5, 6, 7]. In particular, when our UFF implementation leads to an incorrect typization, IM-UFF also gives a different value than when using the correct typization. For a detailed description of the discrepancies between UFF and the reference papers, we refer the reader to Ref. [41].

Additionally, we designed a scenario to evaluate the potential discrepancies in case of non-covalent interaction involving several molecules interacting between each other. For this, we tested a system made of four molecules, each molecule belonging to one of the benchmarks groups used in the comparison with UFF above. The system was initialized with molecules internally at equilibrium, but arbitrarily arranged within the space. Then, the interactions where modeled either with UFF or with IM-UFF, and the global system was minimized thanks to the FIRE approach [46], a fast local minimization method. Finally, the RMSD between the equilibrium structures obtained with UFF and with IM-UFF were evaluated (see Figure 10). This process was repeated on 5 arbi-

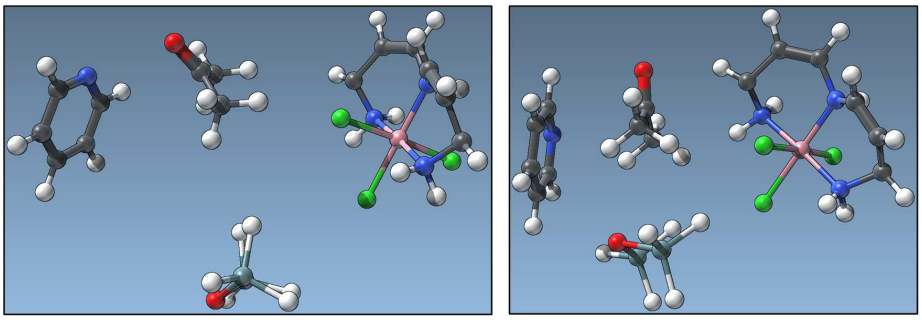

Figure 10: A benchmark made of four molecules to test IM-UFF in case of non-covalent interactions. The four molecules come from the four groups of benchmarks tested in subsection 5.2 and are the acetone, the H3Ge-O-GeH3 system, the mer-Trichloro[N-(3-aminopropyl)-1,3-diaminopropane]cobalt(III) and the pyridine. From an arbitrary initial position (left), the equilibrium states reached with UFF and with IM-UFF are essentially the same, with an average RMSD of $4.5 \times 10^{-7} \mathring{A}$ (right).

trary initial states such that the 4 molecules were always interacting between each other. On average, the RMSD between equilibrium structures obtained with UFF and IM-UFF was $4.5 \times 10^{-7} \mathring{A}$. By comparison, the average RMSD between the initial state and the equilibrium state was of $2.02 \mathring{A}$ after global system alignment. This illustrates how IM-UFF nicely preserves the non-covalent interactions of UFF in case of multiple systems.

5.3. Interactive system modeling

We have used IM-UFF to construct many benchmarks from the previous section from a set of non-bonded atoms[7]. Precisely, we were displacing individual atoms via mouse interactions, while IM-UFF forces were used to continuously minimize the system being constructed. Such a protocol had already been used in SAMSON to do interactive quantum chemistry [19, 21, 23] and interactive modeling of hydrocarbon systems [17]. One difficulty that may arise in some cases is that van der Waals forces may strongly repel atoms from each other

[7]The systems were containing up to almost 100 atoms.

while they are being manipulated by the user, even though the user's goal may actually be to overcome energy barriers and form covalent bonds. To deal with this issue, the interface of the IM-UFF module makes it possible to temporarily deactivate van der Waals forces involving the atoms currently manipulated (i.e. while they are displaced by the user thanks to an edition method). Moreover, the IM-UFF module allows the user to switch back and forth between IM-UFF and UFF, to keep the typizations and topologies constant when necessary.

Figure 11 presents examples of molecules being interactively modeled thanks to IM-UFF (see also Online Resources 1 and 2). In particular, we see how the bond order of the connected systems evolves during edition, and how interactions appear at large distances even when weighted bonds are not covalent yet. Thanks to IM-UFF, such systems can be constructed from atoms initially scattered in the 3D scene. These modeling may take from few seconds for small systems **(a, b)** to a few minutes for intermediate systems **(c, d)** and few dozens of minutes for larger systems **(e, f)**[8]. Naturally, IM-UFF can also be used to edit an existing molecule, in order, for example, to set the initial and final states of a given reaction, or just to probe the effects of some topological change. Figure 12 shows a simple case where a molecule of ethanol is edited. Depending on the hydrogen removed from this system, the user can either obtain the ethanolate ot the oxonium methyl methylene, each case corresponding to distinct topologies and shapes.

5.4. Comparison with a Reactive Force Field

IM-UFF enables smooth transitions between different topologies. Here, we show that, IM-UFF may also be used to get some statistical properties that correspond well to those obtained with a reactive force field (even though IM-UFF was designed to accurately model energy barriers). For this, we consider

[8]Note that this time could be reduced by directly adding functional groups instead of individual atoms, but our goal was to demonstrate the feasibility of constructing complex molecules from scratch with interactive physics based modeling.

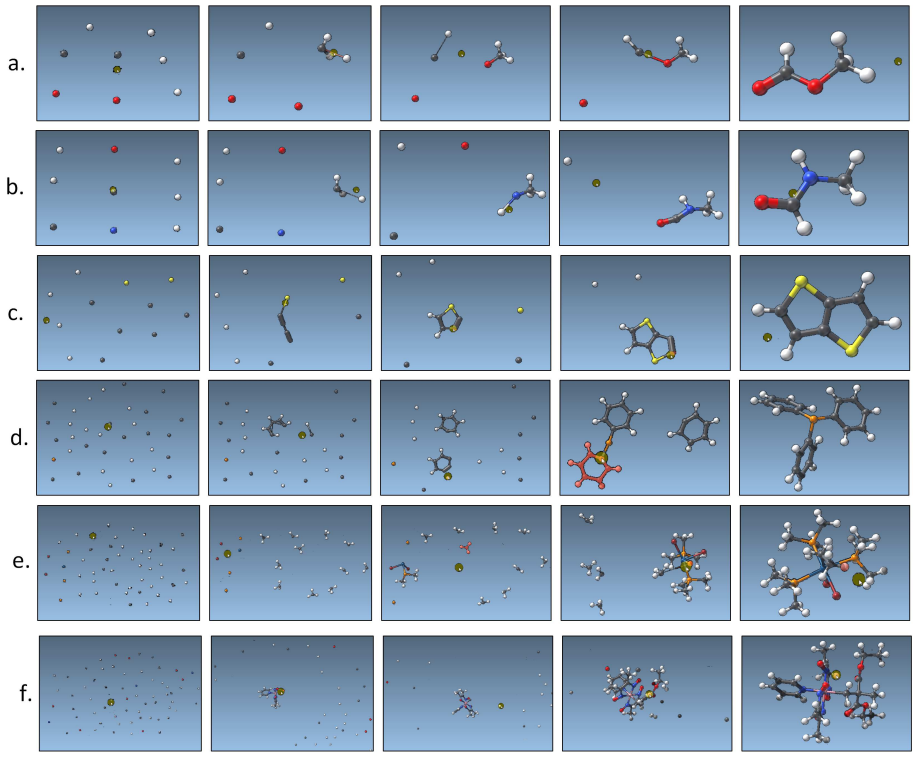

Figure 11: Examples of systems interactively modeled thanks to IM-UFF. **a** and **b** from Ref. [5] correspond respectively to the Methyl formate and the N-Methylformamide. **c** and **d** from Ref. [6] correspond respectively to the thiophene and the trimethylphosphine. Finally, **e** and **f** from Ref. [8] correspond respectively to the mer-Dibromoethyltris(trimethylphosphine)iridium and the Chloro(methyl)[(+)-(2S,3S)-O-isopropylidene-2,3-dihidroxy-1,4-bis(diphenylphosphino)butane]platinum(II).

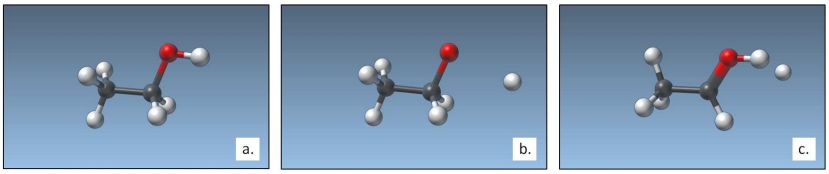

Figure 12: Interactively editing an ethanol molecule (a.). By pulling on the hydrogen linked to the oxygen, the user forms the ethanolate that mostly preserved the initial shape (b.). If instead, the user pulls on one hydrogen linked to the central carbon, the equilibrium position of the other hydrogen is displaced and the user forms the oxonium methyl methylene that contains a double bound between the carbon and the oxygen (c.). The smooth transitions enabled by IM-UFF allow users to have immediate feedback on the consequences of modeling decisions.

a system made of 10 methane molecules constrained to remain within a given fixed volume, and that we simulate using a Monte-Carlo method (see Figure 13). Four simulations are performed, using either the Brenner or the IM-UFF force field, and for a temperature of either $500K$ or $7000K$. Temperatures are chosen such that, at low temperature, covalent bonds are typically preserved whereas, at high temperature, methane molecules may dislocate and form other compounds. Figure 13 shows that the Radial Distribution Functions obtained with IM-UFF are very similar to the ones obtained with Brenner, for both temperature values. Hence, we can assume that the compounds obtained with Brenner are in proportion the same as those obtained with IM-UFF. This is an interesting feature, since such statistical measures would have been impossible with UFF alone, since it does not allow changes in molecular topologies. We think this is an encouraging result, and we are considering extending IM-UFF in the future to obtain a fully reactive force field that accurately models energy barriers.

5.5. Discussion

The experiments conducted have shown that IM-UFF allows us to recover the same structures at equilibrium as when using UFF, up to minimal geometrical differences. Moreover, the mechanism of weighted bonds and weighted typizations that we have introduced allows us to easily and interactively modify

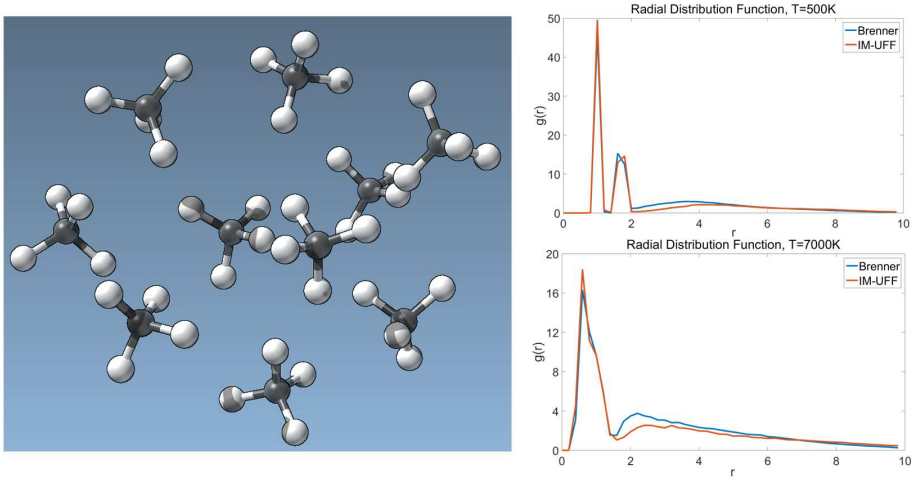

Figure 13: The interaction between 10 methane molecules restrained in a fixed volume is simulated through Monte Carlo simulation (left). The Radial Distribution Functions (RDF) obtained with Brenner (blue curves) are qualitatively the same as those obtained with IM-UFF (red curves), in case of a $500K$ temperature setup (top) as well as a $7000K$ temperature setup (bottom). It means that, unlike UFF, which does not allow changes in the molecular topologies, it might be possible to use IM-UFF to obtain statistical measures that are normally only accessible to reactive force fields.

the topology and, as a result, the geometry of the created structures. Hence, IM-UFF allows to perform interactive modeling of moderately sized molecular systems (up to about one hundred atoms), while taking advantage of the large variety of systems that can be considered with UFF. Obviously, IM-UFF suffers from a few limitations:

- The restrictions present in the automatic perception scheme proposed for UFF naturally appear in IM-UFF. Hence, as already stated, if the automatic perception that initiate UFF fails, the structure in IM-UFF will be incorrect since its perception is derived from the one of our UFF implementation. For example, the bond order assignment may be sub-optimal, the aromatic ring detection method witch relies on a specific set of patterns may appear to be insufficient, the oxidation number may be evaluated incorrectly, etc. We refer the reader to Ref. [41] for a detailed evaluation of the perception performed in UFF.

- We have experimentally shown that the minima of UFF are well reproduced with IM-UFF. However, IM-UFF also introduces minima that are not present in UFF. Fortunately, an extensive use of IM-UFF allowed us to observe that these minima only happen in very specific circumstances and always when the system under manipulation is far away from a stable structure, i.e. when the current coordination/valence of the atoms is lower than their equilibrium coordination/valence. Hence, these *false positive* stable structures are easy to detect and, in practice, they do not prevent users from modeling the correct minimal structures. An example of such a structure is shown in Supplementary Material.

- IM-UFF needs to perform the perception of the system topology at each time step of the interactive simulation, while in UFF this perception step is performed only once. Moreover, IM-UFF forces and energies correspond to more complex expressions than in UFF. These operations clearly introduce an additional computational cost. Experimentally, we have observed that IM-UFF can be used for interactive modeling of systems containing up

to around one hundred atoms. For more complex systems, interactive modeling may present lags.

- Even though IM-UFF allows for continuous topological changes, it has not been parameterized to realistically model energy barriers. Hence, even though barriers *are* present between stable minimum-energy structures, neither their position nor their height should be expected to conform to experience or calculations performed with more sophisticated models. In a preliminary study described above, however, we have shown that it may still be possible to predict statistical properties such as radial distributions.

6. Conclusions and future work

This paper extends the Universal Force Field to allow for continuous topological changes in molecular systems during interactive modeling. This approach, that we called IM-UFF, combines the possibility to significantly modify molecular structures (as with reactive force fields) with a broad diversity of supported systems thanks to the universality of UFF. Such an extension incorporated in an interactive modeling process allows us to easily and interactively build and edit molecular systems, while being guided by physics based inter-atomic forces.

As future work, we would like to accelerate the method computationally by making forces and energy updates incremental (*i.e.* only compute energies and forces that should be updated when only some atoms have moved since the previous time step). Another direction of work is to extend IM-UFF to the new atoms typizations proposed in Ref. [9]. In addition, as stated in Ref. [7], fractional bond orders may give results closer to the experiments (for example for dative bonds). Hence, we would like to correctly detect such cases and introduce fractional bond orders in IM-UFF.

In complement, we would like to propose additional tools to interactively model molecular systems. For example, we would like to allow the user to freeze some of the atoms, align them on a given plane or attract a selected set toward a given regions. We will also let users directly add functional groups.

Beyond these simple examples, we think that the integration of IM-UFF within the modular architecture of SAMSON will allow, in the future, to combine this force field with all the modules and editors that will be integrated in this platform, extending even further its molecular modeling capabilities.

We believe that the approach presented in this paper, which consists in associating weights to force fields parameters in order to obtain a continuous parameterization and allow for topological changes, might be generalizable, and we would like to investigate its applicability to other force fields. In particular, it would be interesting to apply this methodology on force fields with explicit charge distributions thanks to parameterizations allowing continuous transitions of charges when switching from one stable state to another.

Finally, we have shown that it might be possible to use IM-UFF to obtain statistical properties that are normally only accessible to reactive force fields. As a result, even though it was not its initial purpose, we would like to investigate the possibility of extending IM-UFF to accurately model energy barriers and obtain a fully reactive force field. This would potentially allow many new applications, including the estimation of free energies along reaction paths, or the computation of reaction rate constants.

Acknowledgments

We would like to gratefully acknowledge funding from the European Research Council through the ERC Starting Grant n. 307629.

References

[1] A. R. Leach, Molecular Modelling: Principles and Applications (Second Edition), Pearson Education, Harlow, 2001.

[2] D. Frenkel, B. Smit, Understanding Molecular Simulations: From Algorithms to Applications (second edition), Academic Press, San Diego, 2002.

[3] K. Binder, J. Horbach, W. Kob, W. Paul, V. Fathollah, Molecular dynamics simulations, Journal of Physics: Condensed Matter 16(5) (2004) S429–S453.

[4] K. Binder, D. Heermann, Monte Carlo Simulation in Statistical Physics: An Introduction (fifth edition), Springer, Berlin Heidelberg, 2010.

[5] A. K. Rappé, C. J. Casewit, K. Colwell, W. A. Goddard, W. Skiff, Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations, Journal of the American Chemical Society 114 (25) (1992) 10024–10035.

[6] C. Casewit, K. Colwell, A. Rappé, Application of a universal force field to organic molecules, Journal of the American Chemical Society 114 (1992) 10035–10046.

[7] C. Casewit, K. Colwell, A. Rappe, Application of a universal force field to main group compounds, Journal of the American Chemical Society 114 (25) (1992) 10046–10053.

[8] A. K. Rappé, K. Colwell, Application of a universal force field to metal complexes, Inorg. Chem 32(16) (1993) 3438–3450.

[9] M. A. Addicoat, N. Vankova, F. Ismot Akter, T. Heine, Extension of the universal force field to metal-organic frameworks, Journal of Chemical Theory and Computation 10:2 (2014) 880–891. doi:10.1021/ct400952t.

[10] M. G. Martin, Comparison of the amber, charmm, compass, gromos, opls, trappe and uff force fields for prediction of vaporliquid coexistence curves and liquid densities, Fluid Phase Equilibria 248(1) (2006) 50–55.

[11] F. H. Stillinger, T. A. Weber, Computer simulation of local order in condensed phases of silicon, Physical Review B 31 (1985) 5262–5271.

[12] J. Tersoff, New empirical approach for the structure and energy of covalent systems, Physical Review B 37 (1988) 6991–7000.

[13] D. W. Brenner, Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamond films, Physical Review B 42 (1990) 9458.

[14] A. C. T. van Duin, S. Dasgupta, L. Francois, W. A. Goddard, Reaxff: a reactive force field for hydrocarbons, Journal of Physical Chemistry A 105(41) (2001) 9396–9409.

[15] L. Kocbach, S. Lubbad, Reactive interatomic potentials and their geometrical features, eprint arXiv:0908.1540.

[16] [link].
URL `http://www.engr.psu.edu/adri/ReaxffEntry.aspx`

[17] M. Bosson, S. Grudinin, X. Bouju, S. Redon, Interactive physically-based structural modeling of hydrocarbon systems, Journal of Computational Physics 231 (2012) 2581–2598.

[18] K. H. Marti, M. Reiher, Haptic quantum chemistry, Journal of computational chemistry 30 (13) (2009) 2010–2020.

[19] M. Bosson, C. Richard, A. Plet, S. Grudinin, S. Redon, Interactive quantum chemistry: A divide-and-conquer ased-mo method, Journal of computational chemistry 33 (7) (2012) 779–790.

[20] M. P. Haag, M. Reiher, Real-time quantum chemistry, International Journal of Quantum Chemistry 113 (1) (2013) 8–20.

[21] M. P. Haag, A. C. Vaucher, M. Bosson, S. Redon, M. Reiher, Interactive chemical reactivity exploration, ChemPhysChem 15 (15) (2014) 3301–3319.

[22] M. P. Haag, M. Reiher, Studying chemical reactivity in a virtual environment, Faraday discussions 169 (2014) 89–118.

[23] A. C. Vaucher, M. P. Haag, M. Reiher, Real-time feedback from iterative electronic structure calculations, Journal of computational chemistry 37 (9) (2016) 805–812.

[24] A. H. Mühlbach, A. C. Vaucher, M. Reiher, Accelerating wave function convergence in interactive quantum chemical reactivity studies, Journal of chemical theory and computation 12 (3) (2016) 1228–1235.

[25] J. Heyd, S. Birmanns, Immersive structural biology: a new approach to hybrid modeling of macromolecular assemblies, Virtual Reality 13 (4) (2009) 245–255.

[26] S. Doutreligne, T. Cragnolini, S. Pasquali, P. Derreumaux, M. Baaden, Unitymol: Interactive scientific visualization for integrative biology, in: Large Data Analysis and Visualization (LDAV), 2014 IEEE 4th Symposium on, IEEE, 2014, pp. 109–110.

[27] S. Cooper, F. Khatib, A. Treuille, J. Barbero, J. Lee, M. Beenen, A. Leaver-Fay, D. Baker, Z. Popović, et al., Predicting protein structures with a multiplayer online game, Nature 466 (7307) (2010) 756–760.

[28] J. E. Stone, J. Gullingsrud, K. Schulten, A system for interactive molecular dynamics simulation, in: Proceedings of the 2001 symposium on Interactive 3D graphics, ACM, 2001, pp. 191–194.

[29] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, G. Seifert, Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties, Physical Review B 58 (11) (1998) 7260.

[30] M. Gaus, Q. Cui, M. Elstner, Dftb3: extension of the self-consistent-charge density-functional tight-binding method (scc-dftb), J. Chem. Theory Comput 7 (4) (2011) 931–948.

[31] K. Ramachandran, G. Deepa, K. Namboori, Computational chemistry and molecular modeling: principles and applications, Springer Science & Business Media, 2008.

[32] P. Labute, On the perception of molecules from 3d atomic coordinates, Journal of Chemical Information and Modeling 45 (2) (2005) 215–221.

[33] J. Wang, W. Wang, P. Kollman, D. Case, Automatic atom type and bond type perception in molecular mechanical calculations, Journal of Molecular Graphics and Modelling 25(2) (2006) 247–260.

[34] Y. Zhao, T. Cheng, R. Wang, Automatic perception of organic molecules based on essential structural information, Journal of Chemical Information and Modeling 47(4) (2007) 1379–1385.

[35] Q. Zhang, W. Zhang, Y. Li, J. Wang, L. Zhang, T. Hou, A rule-based algorithm for automatic bond type perception, Journal of Cheminformatics 4:26.

[36] S. Urbaczek, A. Kolodzik, I. Groth, S. Heuser, M. Rarey, Reading pdb: perception of molecules from 3d atomic coordinates, Journal of Chemical Information and Modeling 53(1) (2013) 76–87.

[37] R. Sayle, Pdb: Cruft to content (perception of molecular connectivity from 3d coordinates).
URL `http://www.daylight.com/meetings/mug01/Sayle/m4xbondage.html`

[38] A. W. Schüttelkopf, D. M. F. van Aalten, PRODRG: a tool for high-throughput crystallography of protein-ligand complexes, Acta Crystallogr. Sect. D-Biol. Crystallogr. 60 (2004) 1355–1363.

[39] A. K. Malde, L. Zuo, M. Breeze, M. Stroet, D. Poger, P. C. Nair, C. Oostenbrink, A. E. Mark, An automated force field topology builder (atb) and repository: Version 1.0, J. Chem. Theory Comput. 7(12) (2011) 4026–4037.

[40] A. Ribeiro, B. Horta., R. de Alencastro, MKTOP: a program for automatic construction of molecular topologies, J. Braz. Chem. Soc. 19(7) (2008) 1433–1435.

[41] S. Artemova, L. Jaillet, S. Redon, Automatic molecular structure perception for the Universal Force Field, Journal of computational chemistry 13 (2016) 1191–1205.

[42] A. K. Rappé, W. A. Goddard, Charge equilibration for molecular dynamics simulations, The Journal of Physical Chemistry 95 (8) (1991) 3358–3363.

[43] J. Wang, W. Wang, P. A. Kollman, D. A. Case, Antechamberm, an accessory software package for molecular mechanical calculations, Journal of Computational Chemistry 25 (2005) 1157–1174.

[44] Towhee report http://towhee.sourceforge.net/forcefields/uff.html.
URL `http://towhee.sourceforge.net/forcefields/uff.html`

[45] INRIA., SAMSON: Software for Adaptive Modeling and Simulation Of Nanosystems. Version 0.6.0. (2017).
URL `https://www.samson-connect.net`

[46] E. Bitzek, P. Koskinen, F. Gähler, M. Moseler, P. Gumbsch, Structural relaxation made simple, Physical review letters 97 (17) (2006) 170201.