# Atomic pseudo-potentials for reproducing the valence electron behaviour of sp^2 carbon atoms

Alexander Punter,[1] Paola Nava,[1] and Yannick Carissan[1]

Aix Marseille Univ, CNRS, Centrale Marseille, iSm2, Marseille,

France

A pseudo-potential system for recreating an sp^2 carbon atom is built and tested as a building block for various pseudo-hydrocarbon chain and ring systems. This pseudo-system has a central charge of one, thus it contains only one electron. It is employed in *ab-initio* calculations on a variety of hydrocarbons, from small chains to polycyclic aromatics. The relative errors obtained with the PBE0 functional range from 1.0% on TD-DFT first excitation energy to 10% on ionisation energy.

TD-DFT Excitation
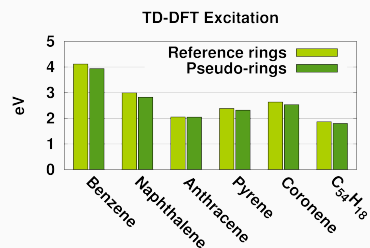
A pseudo-potential system for recreating an sp^2 carbon atom is built and tested as a building block for various pseudo-hydrocarbon chain and ring systems. This pseudo-system has a central charge of 1, thus it contains only one electron. It is employed in *ab-initio* calculations on a variety of hydrocarbons, from small chains to polycyclic aromatics. The relative errors obtained with the PBE0 functional range from 1.0% on TD-DFT first excitation energy to 10% on ionisation energy.

I. INTRODUCTION

It is a common idea that a chemical system can be thought of as comprised of two parts: an active one, where most of the chemistry takes place, and an inactive one, which must be taken into account in order to fulfill chemical requirements. Based on this general statement, many successful theoretical approaches have been developed. Among them can be cited QM/QM' or QM/MM methods (i), where the active region that usually contains several atoms is treated at a high level of calculation, while the inactive part is treated at a lower level of theory.[1] Fragmentation methods split the complete problem into smaller parts and combine individual calculations on these fragments to recover the properties of the overall system.[2,3] Frozen density embedding techniques (ii) replace part of a molecule, or its surroundings, by a frozen electronic density extracted on a reference system.[4] They are an extension of the frozen core approximation, already used to reduce the number of parameters to be optimised in the self consistent field calculation. In the framework of this work, the pioneering work of Rivail *et al.* must be mentioned: the total wave function is optimised with some molecular orbitals kept frozen.[5]

In this article we shall focus on pseudo-potential techniques, which also divide the chemical problem into two parts. Effective core potentials (iii) are commonly employed for atoms: core electrons (and effects due to the corresponding nuclear charge) are replaced by an operator which is quick to evaluate, and the active electrons are treated explicitly.[6] In the same vein, model potentials (iv) replace a frozen electronic density (computed from the atom) by a series of operators, which depend on the level of refinement required.[7]

The effective group potentials (v) bridge the frozen density and the core potential approaches: using core potential extraction techniques, effects due to the implicit electron density (and corresponding nuclear charge) are reproduced by a mono-electronic operator.[8,9] However, while effective core potentials and model potentials intend to reproduce atomic properties, effective group potentials aim at mimicking the effect of atoms involved in one or more chemical bonds. The cyclopentadienyl group has, for instance, been successfully extracted.[8] In this example, the active part consisted of six electrons (the π electrons) and six nuclear charges, with the pseudo-potential replacing all the rest, including the hydrogen atoms. Even if the σ/π separability suggested how to define the active and the inactive

parts, carbon atoms are hybridised and the core/valence distinction is not strictly defined.

The theoretical background to extract effective group pseudo-potentials can be traced back to several contributions in the 80's and 90's.[7,10–14] In 1992 Katsuki built molecular potentials based on Huzinaga's model potentials.[15,16] In his work on this subject, Huzinaga emphasises that pseudo-potentials should maintain three effects of the 'dormant' electrons on the active ones: the Coulomb, the exchange and the 'no-collapse' term.[11] Following his proposal, for an atom where active and dormant electrons have been separated, the Hamiltonian reads:

$$\hat{H} = \sum_{i=1}^{n} \hat{h}(i) + \sum_{i<j} \frac{1}{r_{ij}} \tag{1}$$

with $\frac{1}{r_{ij}}$ the bi-electronic interaction between explicitly treated active electrons and the mono-electronic operator:

$$\hat{h}(i) = -\frac{1}{2}\Delta_i - \frac{(Z - Z_c)}{r_i} + \hat{W} \tag{2}$$

where Δ_i is the Laplacian of the coordinates of electron i, and Z_C is the number of core electrons withdrawn from the reference system. The operator $\hat{W}$ contains two terms: the $\hat{\sigma}$ 'no-collapse' term that prevents active electrons collapsing into the dormant region and the operator $\hat{V}$ that reproduces the Coulomb and exchange interactions:

$$\hat{V} = \frac{1}{r}\left[\sum_I A_I \exp(-\alpha_I r^2) + \sum_J B_J r \exp(-\beta_J r^2)\right] \tag{3}$$

To take into account the fact that dormant electrons are removed from the system, the nuclear charge is modified by the Z_c value. Yet, as the effective charge felt by the active electrons is likely not to be an integer, the modification of the value of the nuclear charge is scaled in $\hat{V}$. In this operator we notice that the r^{-1} behavior is maintained for the first term (the Coulomb one).

In the present work, the aim is to propose a methodology for extracting a potential for a hybridised carbon atom (here an sp^2 carbon atom), which can then be used as a building block for constructing chemical systems. This carbon atom will contain exactly one explicit nuclear charge and one explicit electron. Moreover, the method should be usable out of the box in any standard quantum chemistry software. Thus, no modification of the source code should be done. This supplementary constraint will be fulfilled by strategically positioning in space the pseudo-potentials we intend to use. The replacement of bi-electronic interactions

with a set of mono-electronic integrals will result in drastic gains in timings, a study of which is available in the supplementary material. However, the main purpose of the current work is not to study the time gain from the use of our pseudo-potentials but to show that our approach of extracting pseudo-potentials for hybridised atoms is possible with strictly local pseudo-potentials.

Pseudo-potentials for hybridised carbon atoms have been already successfully extracted in a previous work of ours.[17] There, our attention was focused on the 'no-collapse' term, and functions in the pseudo-potential shift up unwanted orbitals and correctly place the energy of the wanted orbitals. A drawback of this method was that some pseudo-potentials had to be put precisely at the center of each bond. Thus, those pseudo-potentials could not be considered as purely local as they were not defined solely with respect to the position of the atom they applied to. This previous work should therefore be regarded as a proof-of-concept. In this new version, we include more physical meaning in our model, by placing pseudo-potentials that mimic Coulomb interactions among both the active and dormant electrons, and the shielded nuclear charge. The new pseudo-potentials are fully atomic: when replacing a hybridised atom, the atom's position and orientation can determine completely the placement of the potentials (as hybridisation destroys isotropy, the preferred orientations must be given to define the pseudo-potential).

This article is structured as follows. The first part (Methodology) gives the general definition of the pseudo-potential, which we subsequently develop. Secondly, the particular case of the $CH_3^{\bullet}$ radical is detailed as it gives access to some physically grounded values. In a third part, the optimal pseudo-potential is defined. Details of the step-by-step extraction are given in the supplementary material. The application part focuses on the performances of the optimal pseudo-potential. It is shown that the properties can be well reproduced except when dealing with systems in which spin contamination is large. For these cases, a solution is provided.

II.   METHODOLOGY

A.   General pseudo-potential definition

We make use of two kinds of gaussian pseudo-potentials[18], of s and p shapes. As we want to avoid modifying the quantum chemistry software itself, the pseudo-potentials have a semi-local form. For a pseudo-carbon atom, the multi-centered pseudo-potential that we built reads as follows:

$$\hat{W} = \underbrace{\frac{1}{r} A \exp(-\alpha r^2) \sum_{m=-1}^{m=+1} |Y_{1,m}\rangle \langle Y_{1,m}|}_{p \text{ projector}} + \tag{4}$$

$$\underbrace{\sum_J \frac{1}{r'_J} A'_J \exp(-\alpha'_J (r'_J)^2) |Y_{0,0}^J\rangle \langle Y_{0,0}^J|}_{s \text{ projectors}}$$

The p projector is centered on the pseudo-atom and it is entirely defined by the A coefficient and the α exponent, $Y_{1,m}$ are the p real solid harmonics. The s projectors are placed around the pseudo-atom. The Jth s projector is defined by its A'_J coefficient, its α'_J exponent and its position relative to the pseudo-atom, so that $r'_J = r - r_J^0$ with r_J^0 the distance to the pseudo-atom. $Y_{0,0}^J$ is the s real solid harmonic placed on the Jth center.

By analogy with Huzinaga model potentials defined in (3) we can say that our projectors have a coulombic nature. The p pseudo-potential allows us to reproduce atomic properties (see hereafter), while the s potentials mimic electrostatic electron pair repulsion, in particular they recover part of the bi-electronic interaction between the dormant and the active electrons. The pseudo-potentials are mono-electronic operators, thus they shall not be able fully to reproduce bi-electronic interactions. Knowing this, we adjust the parameters of the s potentials for the pseudo-system to have a wavefunction in which relevant properties, such as spatial extent, energy and excitation energies, are as close as possible to those of the reference system.

B.   Extraction for the CH$_3^\bullet$ radical

In a first step we use the planar CH$_3^\bullet$ ($d_{CH} = 2.0466$ $a.u.$, HF/def-SV(P)) as a case study. It is the smallest system containing one and only one sp^2 hybridised carbon atom. This will
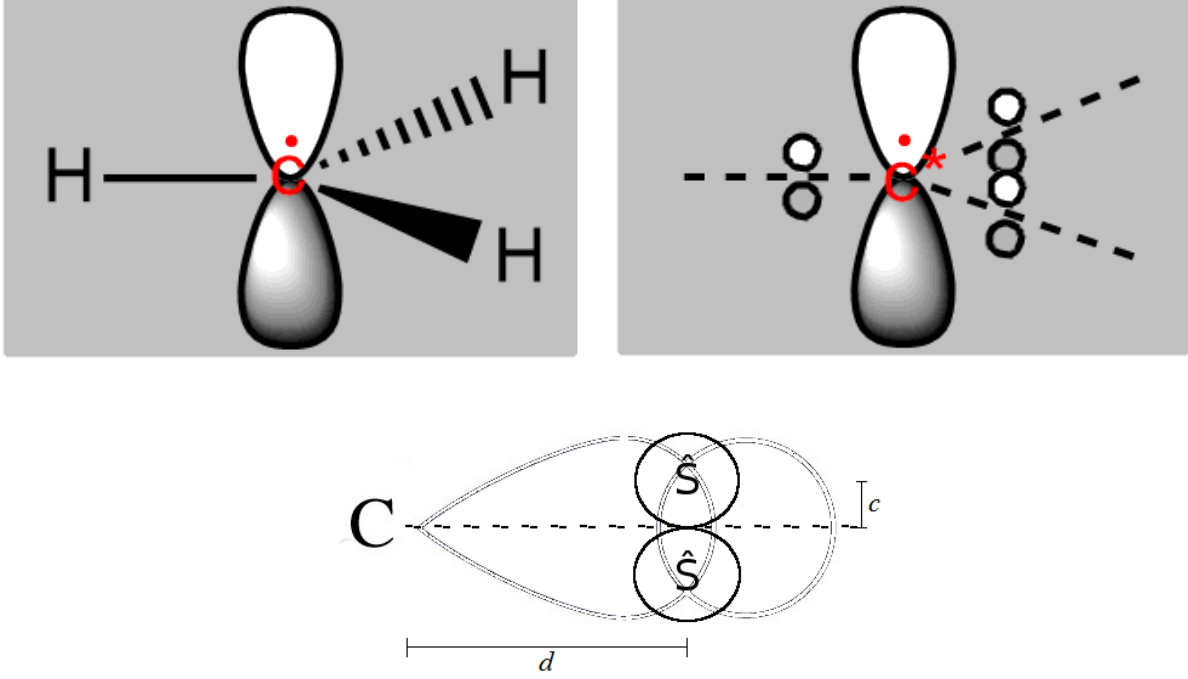
FIG. 1. Diagrams of CH$_3^{\bullet}$ (left) and pseudo-CH$_3^{\bullet}$ (right, below) molecules. The pseudo-CH$_3^{\bullet}$ diagrams display the s and p potential positions, and the distances d and c. White potentials are above the xy plane, gray potentials are below. In this article $d = 0.5$ $a.u.$ and $c = 0.25$ $a.u.$

provide us with physically meaningful parameters, which are then employed as initial guess for the extraction of a pseudo-potential for a general π carbon atom.

Figure 1 displays the final pseudo-system: the CH$_3^{\bullet}$ radical has been replaced by a hydrogen-like "pseudo-carbon", with a nuclear charge of $Z_{nucleus} = 1$, and one electron occupying the p_z orbital. There are now no H atoms, and the system is surrounded by three potential pairs at a planar distance of d, each consisting of two s-shaped potentials with a distance of c above and below the (xy) plane. A further p-shaped potential is applied directly to the pseudo-carbon. Based on Equation 4, the complete pseudo-potential reads as:

$$\hat{W} = \frac{1}{r} A \exp(-\alpha r^2) \sum_{m=-1}^{m=+1} |Y_{1,m}\rangle \langle Y_{1,m}| + \tag{5}$$

$$\sum_{J=1}^{6} \frac{1}{r'_J} A' \exp(-\alpha'(r'_J)^2) |Y_{0,0}^J\rangle \langle Y_{0,0}^J|$$

In this article, we fix $c = 0.25$ a.u.; $d = 0.5$ a.u., Table I. This gives us four variables to manipulate the properties of the system, notably the A coefficient and the α exponent for the p potential, and the A' coefficient and α' exponent, which will be common to the six s potentials.

TABLE I. Position of the s potentials around the pseudo carbon atom, $c = 0.25$ a.u. and $d = 0.5$ a.u. as in Figure 1.

x	y	z
$-d$	0	c
$-d$	0	$-c$
$\frac{d\sqrt{3}}{2}$	$\frac{d}{2}$	c
$\frac{d\sqrt{3}}{2}$	$\frac{d}{2}$	$-c$
$\frac{-d\sqrt{3}}{2}$	$\frac{d}{2}$	c
$\frac{-d\sqrt{3}}{2}$	$\frac{d}{2}$	$-c$

For this case study, we first determine the p potential parameters by using a criterion based on the concept of the effective nuclear charge felt by the explicit electron. Clearly the assumption that $Z = 1$ is unrealistic. Slater's Rules[19] suggest that, with the screening effect, the p_z electron of an isolated carbon atom should experience a charge of $Z = 2.4$. Starting from this consideration we shall show in the following paragraph how we could get a guess for the p potential parameters and, once fixed, the s potential will be extracted on some energy criteria.

1. Extraction of the p potential parameters

The p_z pseudo-potential is here to fill the gap between the pseudo and the real system for the effective charge of the nucleus. In order to make an educated guess of the parameters of this new potential, we consider the p_z component of the p potential and we write it as:

$$\widehat{p_z} = -\frac{Z_{pseudo}}{r}|\chi_{p_z}\rangle\langle\chi_{p_z}| \tag{6}$$

where χ_{p_z} is a p_z Gaussian function with exponent α, and $-Z_{pseudo}$ corresponds to the A parameter. If χ_{p_z} were exactly the π molecular orbital ψ_{p_z}, which hosts the radical, the

expected value of the $\widehat{p}_z$ operator for ψ_{p_z} would give us the effect of the screened nucleus. As χ_{p_z} is one function only, the $-Z_{pseudo}$ parameter should contain a correction for the overlap of the χ_{p_z} and the reference ψ_{p_z} molecular orbital. We define:

$$S = \langle \psi_{p_z} | \chi \rangle \tag{7}$$

Knowing that our hydrogen-like pseudo-system already contains a charge, $Z_{nucleus} = 1$, we subtract this $Z_{nucleus}$ from the desired effective nuclear charge Z_{eff}, and we correct by the overlap by writing:

$$Z_{pseudo} = -A = (Z_{eff} - Z_{nucleus})S^{-2} \tag{8}$$

We need now to find an expression for Z_{eff}. From an all-electron HF reference calculation on the $CH_3^{\bullet}$ system, the expected distance of the electron from the nucleus is computed:

$$\langle r \rangle = \langle \psi_{p_z} | r | \psi_{p_z} \rangle \approx 1.8 \ a.u. \tag{9}$$

In the pseudo-system there is only one explicit electron in a p_z orbital. The analytical form of the p_z orbital for a hydrogen-like atom is[18]

$$\phi_{210} = \frac{1}{\sqrt{\pi}} \frac{Z_{eff}}{2a_0}^{\frac{5}{2}} r e^{-\frac{Z_{eff}r}{2a_0}} \cos\theta \tag{10}$$

and from this we obtain

$$\langle r \rangle = \langle \phi_{210} | r | \phi_{210} \rangle = \frac{5a_0}{Z_{eff}} \tag{11}$$

From Equation 9 and Equation 11 it follows that $Z_{eff} \approx 3.6$.

We now have the power to choose a p_z pseudo-potential based solely on the Gaussian exponent, and the Z_{eff} then follows from the above. For this $CH_3^{\bullet}$ case study, the exponent was chosen to maximise the overlap with the molecular orbital we wished to mimic (a maximum overlap of 1 would lead to $A = Z_{nucleus} - Z_{eff}$). Hence we arrive at a p_z potential that should be physically meaningful. The optimal values that we found were a coefficient of $A = -2.81 \ a.u.$ for an exponent of 0.261 and an overlap of 0.79. It should be noted here that the choice of a different exponent would lead to another value of A via Equation 8. The overlap would differ but the effective charge would be exactly the same. We shall see later that the diffuseness of χ does matter. In the extraction of a pseudo-carbon atom which interacts with other π electron systems, more constraints on the system will apply and the exponent will be chosen to reflect this.

2. *Extraction of the six s potentials parameters*

Now that we have set up the p potential placed at the nucleus, we shall define the parameters for the s potentials surrounding the nucleus. To do so, we decide to focus on the reproduction of the energy of the singly-occupied HOMO of this system at the HF/def-SV(P) level. The energy of this orbital is $-10.537\ eV$. This value can be exactly reproduced with a range of exponents and coefficients for the s potentials. For example, by choosing an exponent of 1.0 we get a coefficient of 2.77 $a.u.$ which gives the correct HOMO value.

C. Extraction in the general case for C sp^2 atoms

The CH$_3^\bullet$ radical allowed us to test our model but cannot be used as such when dealing with π molecular interactions as it does not contain information about the π bond. As discussed earlier, the exponent extracted for CH$_3^\bullet$ is too diffuse and the p pseudo-potential interacts with neighbouring atoms. As we will see below, the tuning of the potentials to reflect other properties of the system will also require the potentials to affect other orbitals of the pseudo-carbon (the π^* orbital, for example). Thus, a reparametrisation is necessary with previous parameters used as a starting guess for optimisation. This explains the difference between the parameters obtained in the previous section and the ones of Table II. As can be seen, the optimal exponent is found to be 0.624 (it was 0.261 for CH$_3^\bullet$), which decreases the overlap with the p_z orbital. This is compensated by a larger absolute value of the coefficient ($-3.91\ a.u.\ v.s.\ -2.81\ a.u.$).

We intend to extract a pseudo-potential, which interacts properly with other π-carrying atoms, be they pseudo-systems or real systems. Furthermore, we expect our hypothesis to be grounded and to be able to extract parameters on as few data as possible. We extracted the optimal pseudo-potential by a step-by-step procedure on the smallest π system: the ethene molecule. Formally, the pseudo-potential has the same shape as in Equation 5. The position of the s potential is kept the same (Table I). We began by optimising sets of potentials from different starting guesses to obtain the exact Δ_{ST} gap (the Δ_{ST} value is the difference between the lowest π^* triplet, ground state in the triplet space, and the lowest singlet state, the ground state). We then picked the potential set that also gave the best HOMO and ionisation energies, and used this potential as the starting guess for

a final optimisation aimed at minimising the total error across all three values. Values for the optimised parameters are reported in Table II. Ultimately, the optimal potential found maintained a very accurate Δ_{ST}. See the supplementary materials for further details.

TABLE II. Optimal parameters obtained for the pseudo π carbon atom at the HF/def-SV(P) level of theory. Coordinates are relative to the position of the nucleus. In this work, $c = 0.25\ a.u.$ and $d = 0.5\ a.u..$ Coefficients are in a.u..

Type	Coeff.	Exp.
p	$A = -3.91$	$\alpha = 0.624$
s	$A' = 1.5$	$\alpha' = 0.5$

With parameters defined in Table II, the potential reproduces the properties of the ethene π cloud system with a good accuracy: relative errors on the ionisation energy and the energies of the HOMO orbital of 8% and 3% respectively (Table III).

TABLE III. Comparison of the singlet-triplet splitting (Δ_{ST}), ionisation energy (I.E.) and energy of the HOMO orbital (ε_{HOMO}) of the reference ethene molecule and its reproduction with the optimal pseudo potential obtained in this work. The Δ_{ST} values are the difference between the lowest π^* triplet (in an unrestricted formalism) and the lowest singlet state (in a restricted formalism). Calculations are performed at the HF/def-SV(P) level, values are in eV.

	Δ_{ST}	I.E.	ε_{HOMO}
Reference Values	3.533	9.091	-10.363
Optimal pseudo-potential	3.533	9.806	-10.062

III. RESULTS AND DISCUSSION

A. Computational Details

All Hartree-Fock (HF), Density Functional Theory (DFT) and Time-Dependent DFT (TD-DFT) calculations are performed with TURBOMOLE 7.1[20]. The basis set used throughout is def-SV(P)[21]. Wherever possible, planar (C_S) symmetry is used. The convergence energy is 10^{-7}H (`$scfconv = 7`) for SCF and 10^{-6}H for DFT.

Chain Alkenes, Ring Molecules. In addition to Hartree-Fock calculations, DFT is used with PBE0, PBE, TPSS and TPSSh functionals.[22–25] The multiple m4 grid is used for integration. Also performed are TD-DFT calculations, where the Tamm-Dancoff approximation (CIS)[26] is switched on to avoid triplet instability.

B.   Preliminary note

The extraction done in the previous section was at the HF/def-SV(P) level. We intend to show that there is no need for reparametrisation of the potential when used with other methods. Thus, in the following sections, we use HF and four functionals in the DFT framework: one GGA (PBE), one meta-GGA (TPSS) and their hybrid variants (PBE0 and TPSSh). As a measure of the success of the potentials extracted we examine the three properties Δ_{ST}, the ionisation energy, and the ground state HOMO energy across a series of molecules other than ethene. Furthermore, the parametrisation is done with ground state calculations only: the Δ_{ST} values are simply the difference between the lowest π^* triplet (ground state in the triplet space) and the lowest singlet (ground state). We then go on to use this potential in the TD-DFT framework, which uses the virtual orbitals. We show in the last section that results obtained within the TD-DFT framework match the reference calculations from both the quantitative (excitation energies match) and qualitative (excitations are described by the same orbitals) point of view. Finally, despite the fact that the extraction was done on a non-aromatic system, we perform calculations on fused ring aromatic systems. Again, these calculations are successful. As a consequence, one can say that the potential extracted on ethene at the HF/def-SV(P) level is free of obvious transferability problems.

As mentioned above, in addition to HF, four functionals were used. As HF was the method used to extract the parameters of the pseudo-potential, one could expect this method to outperform the others. It is not found to be true. It turns out that PBE0 performs better on average than any other method used in this work. Thus, in the following sections, we shall illustrate our discussion using the PBE0 results. The mean absolute errors for the other methods are given. The complete set of results is in the supplementary material. Additionally, the potential extracted on ethene gave an exact Δ_{ST} result, but had errors of

3% and 8% on the HOMO and ionisation energies respectively, thus we expect errors of at least a similar size in other systems in which they are used. Furthermore, ionisation energies are difficult to reproduce owing to the fact that the atomic charges of the pseudo-atoms are drastically modified.

C.   Alkene chains: $C_{2n}H_{2n+2}$, $2 \leq n \leq 6$

Taking the optimal parameters we test them against a series of chain alkenes up to length $C_{12}H_{14}$, using a variety of functionals.

In each case, the geometry of the reference system is optimised according to the method used, before taking the reference geometry and applying the pseudo-potentials from Table II. Table IV gives a breakdown of the percentage errors for each method across all molecules tested. Figure 2 shows the results. The pattern of increasing HOMO energy and decreasing cation-singlet and triplet-singlet energies seen in the reference systems is well replicated by the pseudo-alkenes, with the energies following the same gradient.

TABLE IV. Mean relative errors (in percent) across methods (HF or different functionals) for short chain alkenes (C_2-C_{12}). The basis set is def-SV(P).

%-error	HF	PBE0	PBE	TPSS	TPSSh
ε_{HOMO}	1.3	4.0	8.5	12.0	9.7
I.E.	5.3	6.0	8.1	10.1	9.0
Δ_{ST}	4.1	3.6	7.5	13.3	11.4
TD-DFT $\pi - \pi^*$	11.6	2.6	2.8	6.3	6.9

Also compared are TD-DFT results for this system and results of some of the authors' previous work, which match to within 3% (see supplementary materials).[17]

Here we discuss in more detail the results obtained with the PBE0 functional presented in Figure 2. The properties computed with the optimal pseudo-potential are in very close agreement with the reference values. Even for HOMO and ionisation energies, the agreement between pseudo and reference calculations is satisfactory (average error $\leq 6\%$). Thus, the addition or withdrawal of an electron from the pseudo-molecule is well reproduced. Furthermore, the first singlet-triplet $\pi - \pi^*$ excitations agree very well in either the Δ_{ST} or TD-DFT
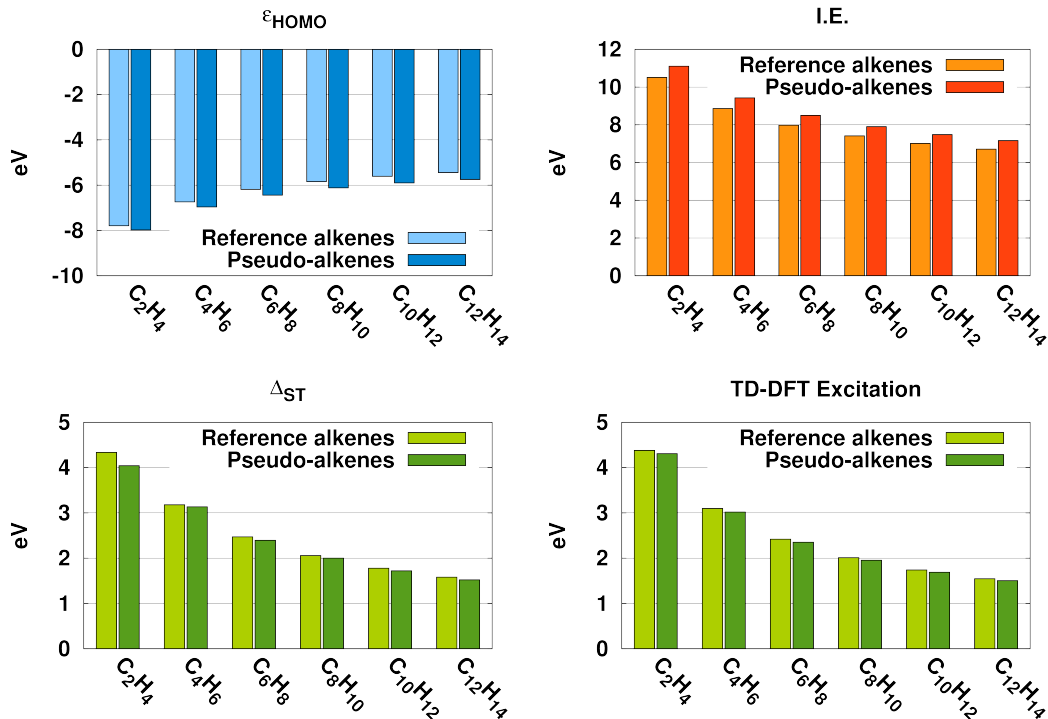
13

FIG. 2. Comparison of the HOMO (ε_{HOMO}), the ionisation (I.E.), and the first singlet-triplet $\pi - \pi^*$ excitation (Δ_{ST} and TD-DFT) energies between the all-electron reference and optimal pseudo-potential systems across a range of chain alkenes. The Δ_{ST} values are the difference between the lowest π^* triplet (unrestricted formalism) and the lowest singlet state (restricted formalism). Calculations are at the PBE0/def-SV(P) level.

framework (average error $\leq 3.6\%$).

D.   Fused rings and larger chains

The potentials derived above are also tested on larger systems. Figure 3 shows the Δ_{ST}, the IE and HOMO energy values for several ring systems, the largest of which is $C_{54}H_{18}$, shown in Figure 4. As with the short chain alkenes, the general trend of the results is well-replicated by the pseudo-systems, and the percentage errors, displayed in Table V, are similar. For PBE0, Figure 3, mean average errors are below 3.6% for excitation energies, independently of the way they are computed (Δ_{ST} or TD-DFT). For HOMO and ionisation energies, average errors are somewhat larger, up to 10%. Yet, the agreement between
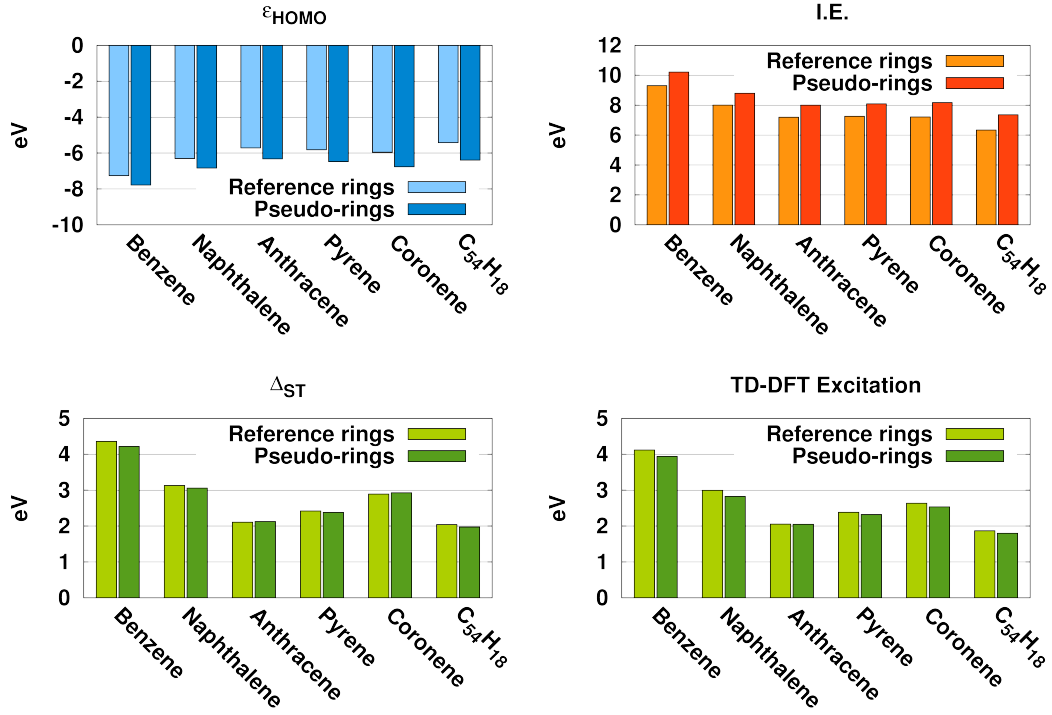
14

FIG. 3. Comparison of the HOMO (ε_{HOMO}), the ionisation (I.E.), the first singlet-triplet $\pi - \pi^*$ excitation (Δ_{ST} and TD-DFT) energies between the all electron reference system and the optimal pseudo-potential across a range of ring molecules. The Δ_{ST} values are the difference between the lowest π^* triplet (unrestricted formalism) and the lowest singlet state (restricted formalism). Calculations are at the PBE0/def-SV(P) level. The $C_{54}H_{18}$ molecule is represented in Figure 4.

FIG. 4. The $C_{54}H_{18}$ molecule.
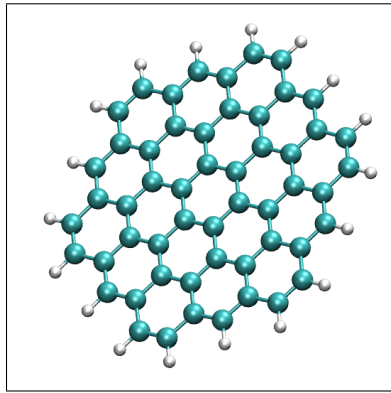
reference and pseudo-potential calculations is very good. The modeling through our pseudo-potential does not suffer from error accumulation: different pseudo-systems, whether of 2 or 54 pseudo-atoms, lead to similar errors on the computed properties.

TABLE V. Mean relative errors (in percent) across methods (HF or different functionals) for ring-systems. The basis set is def-SV(P).

%-error	HF	PBE0	PBE	TPSS	TPSSh
ε_{HOMO}	3.4	9.3	13.4	17.1	14.7
I.E.	7.4	10.0	12.5	14.4	13.0
Δ_{ST}	3.3	1.9	1.7	6.8	6.7
TD-DFT $\pi - \pi^*$	13.5	3.6	2.7	6.3	7.0

TABLE VI. Mean relative errors (in percent) across methods (HF or different functionals) for long chain alkenes (C_{20}-C_{100}). The basis set is def-SV(P).

%-error	HF	PBE0	PBE	TPSS	TPSSh
ε_{HOMO}	1.8	7.3	11.3	16.7	13.6
I.E.	25.1	6.7	9.4	10.3	11.6
Δ_{ST}	55.3	85.8	83.4	239.8	320.2
TD-DFT $\pi - \pi^*$	30.1	1.0	6.1	5.6	2.9

Figure 5 and Table VI refer to longer alkene chains (n up to 100). The pattern of decreasing ionisation energies and Δ_{ST} with increasing HOMO energy is still followed, with the absolute error remaining consistent. However, differences in the triplet-singlet energies between the reference and pseudo-systems become significant, notably for the largest case. Here again, ionisation energy and ϵ_{HOMO} agree with mean average errors lower than 10% for PBE0.

Unlike for the previous systems, the larger the system, the larger the discrepancy between Δ_{ST} and TD-DFT results. This apparent failure of the pseudo-potentials is to be found in the representation of the triplet state. The expectation values of the S^2 operator for the triplet calculations are plotted in Figure 6, which shows that the spin contamination of the triplet state computed as a single configuration (*i.e.* in a SCF framework) increases in both
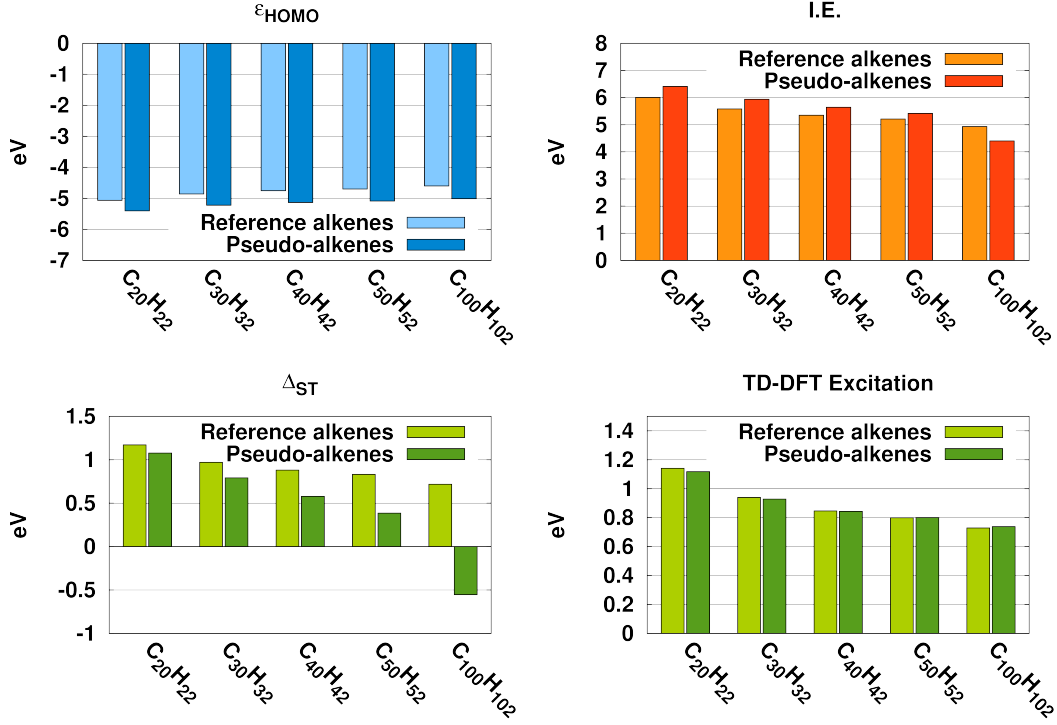
FIG. 5. Comparison of the HOMO (ε_{HOMO}), the ionisation (I.E.), the first singlet-triplet $\pi - \pi^*$ excitation (Δ_{ST} and TD-DFT) energies between the all electron reference system and the optimal pseudo-potential across a range of long chain alkenes (C_{20}-C_{100}). The Δ_{ST} values are the difference between the lowest π^* triplet (unrestricted formalism) and the lowest singlet state (restricted formalism). Calculations are at the PBE0/def-SV(P) level.

reference and pseudo-potential cases. Yet, this effect is strengthened in the pseudo-potential calculations. The triplet instability of the system represented with pseudo-potentials is exacerbated. As already shown, this is fixed by using the Tamm-Dancoff approximation, as can be seen from the excellent agreement obtained with TD-DFT.[26]

In order to show that the recovering of the agreement between the pseudo-potential and the reference calculations is not an artefact, we give in Table VII the weight and nature of each excitation (weight larger than 3%) in the description of the triplet excited state for $C_{50}H_{52}$ (other values can be found in the SI, which exhibit the same trends). As can be seen, the agreement is very good.

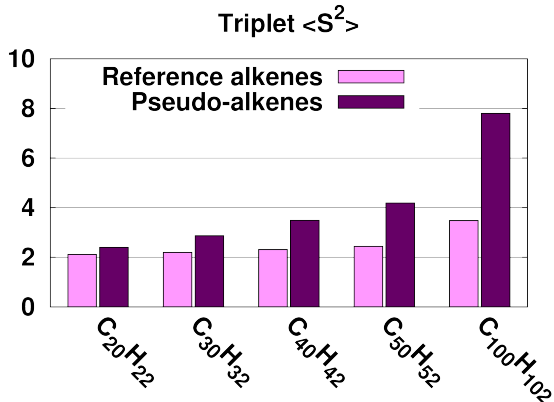These results show that the pseudo-potentials that we have extracted are able to reproduce

FIG. 6. Comparison of S^2 expectation values obtained for the calculation of the first π^* triplet configuration in a SCF formalism, for reference and pseudo-systems.

TABLE VII. Comparison of the weights (all electron *vs.* pseudo-potentials) of the excitations obtained with TD-DFT to represent the triplet excited state from the closed shell singlet state. Example case of $C_{50}H_{52}$.

Excitation	Weight(%)	
MO	Ref.	Pseudo.
25 a" $\to$ 26 a"	77.0	67.1
24 a" $\to$ 27 a"	10.5	13.1
23 a" $\to$ 28 a"	3.6	5.2

the π systems in a variety of situations which are not part of their extraction set. The molecular orbital virtual space is also well described (cf. Table VII), which demonstrates that the good agreement with reference calculations is physically grounded.

IV. CONCLUSION

In this work, we tackled the two main problems of the initial version of our molecular potentials. Firstly, the new pseudo-potentials for sp^2 hybridised atoms are completely atomic and, even if the directionality of the bonding pattern has to be fulfilled by correctly positioning the s potentials. No potentials need to be added relative to the position of more

than one atom. Secondly, we gave a physical meaning to all the pseudo-potential terms. Contrary to our previous attempt, we do not rely on the "no collapse" term. In fact, in this work we did not need such a term. We could show that not only the occupied orbitals were well-reproduced by the use of these new potentials, but also that the virtual space is of good quality for excited states calculation. The model defined here can be used in any quantum chemistry package which implements atomic pseudo-potentials as we provide all the necessary parameters. We are now working on extending our extraction method to other hybridised atoms in order to provide a library of such potentials.

V. ACKNOWLEDGMENTS

VI. SUPPLEMENTARY MATERIALS

In the supplementary materials are provided:

- a file containing discussions of the optimisation process used to generate the potentials, as well as of the computational gains found using the potential systems.

- a spreadsheet file with all the calculated energies.

- all the geometries used for the calculations.

REFERENCES

1. L. W. Chung, W. M. C. Sameera, R. Ramozzi, A. J. Page, M. Hatanaka, G. P. Petrova, T. V. Harris, X. Li, Z. Ke, F. Liu, et al., Chem. Rev. **115**, 5678 (2015).

2. M. S. Gordon, M. A. Freitag, P. Bandyopadhyay, J. H. Jensen, V. Kairys, and W. J. Stevens, J. Phys. Chem. A **105**, 293 (2001).

3. C. Steinmann, D. G. Fedorov, and J. H. Jensen, PLoS ONE **7** (2012).

4. T. A. Wesolowski, S. Shedge, and X. Zhou, Chem. Rev. **115**, 5891 (2015).

5. X. Assfeld and J.-L. Rivail, Chem. Phys. Lett. **263**, 100 (1996), ISSN 0009-2614.

6. M. Dolg and X. Cao, Chem. Rev. **112**, 403 (2012).

7. S. Huzinaga, Can. J. Chem. **73**, 619 (1995).

8. Y. Carissan, F. Bessac, F. Alary, J.-L. Heully, and R. Poteau, Int. J. Quantum Chem. **106**, 727 (2006).

9. C. Raynaud, I. del Rosal, F. Jolibois, L. Maron, and R. Poteau, Theor. Chem. Acc. **126**, 151 (2010).

10. G. Nicolas and P. Durand, J. Chem. Phys. **72**, 453 (1980).

11. S. Huzinaga, J. Mol. Struct. THEOCHEM **234**, 51 (1991).

12. N. C. Hernandez and J. F. Sanz, J. Chem. Phys. **113**, 6082 (2000).

13. J. A. M. Romero and J. F. Sanz, J. Chem. Phys. **99**, 1255 (1993).

14. J. A. Mejias and J. F. Sanz, J. Chem. Phys. **102**, 327 (1994).

15. S. Katsuki, S., Can. J. Chem. **70**, 285 (1992).

16. S. Katsuki, J. Chem. Phys. **98**, 496 (1993).

17. J. Drujon and Y. Carissan, J. Comput. Chem. **34**, 49 (2013).

18. T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic-Structure Theory* (John Wiley and Sons Ltd, Baffins Lane, Chichester, West Sussex PO19 1UD, England, 2002).

19. J. C. Slater, Phys. Rev. **1** (1930).

20. *TURBOMOLE V7.1 2016* (a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com., 2016).

21. A. Schäfer, H. Horn, and R. Ahlrichs, J. Chem. Phys. **97** (1992).

22. J. Perdew, M. Ernzerhof, and K. Burke, J. Chem. Phys. **105** (1996).

23. J. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77** (1996).

24. J. Tao, J. Perdew, V. Staroverov, and G. Scuseria, Phys. Rev. Lett. **91** (2003).

25. V. Staroverov, G. Scuseria, J. Tao, and J. Perdew, J. Chem. Phys. **119** (2003).

26. M. J. G. Peach, M. J. Williamson, and D. J. Tozer, J. Chem. Theory Comput. **7** (2011).