



Cite this: *Phys. Chem. Chem. Phys.*,  
2022, **24**, 3761

## Diffusion Monte Carlo evaluation of disiloxane linearisation barrier†

Adie Tri Hanindriyo, <sup>a</sup> Amit Kumar Singh Yadav, <sup>b</sup> Tom Ichibha, <sup>c</sup> Ryo Maezono, <sup>d</sup> Kousuke Nakano <sup>de</sup> and Kenta Hongo <sup>\*f</sup>

The disiloxane molecule is a prime example of silicate compounds containing the Si–O–Si bridge. The molecule is of significant interest within the field of quantum chemistry, owing to the difficulty in theoretically predicting its properties. Herein, the linearisation barrier of disiloxane is investigated using a fixed-node diffusion Monte Carlo (FNDMC) approach, which is one of the most reliable *ab initio* methods in accounting for the electronic correlation. Calculations utilizing the density functional theory (DFT) and the coupled cluster method with single and double substitutions, including noniterative triples (CCSD(T)) are carried out alongside FNDMC for comparison. It is concluded that FNDMC successfully predicts the disiloxane linearisation barrier and does not depend on the completeness of the basis-set as much as DFT or CCSD(T), thus establishing its suitability.

Received 5th April 2021,  
Accepted 17th December 2021

DOI: 10.1039/d1cp01471d

rsc.li/pccp

## 1 Introduction

The simplest molecule containing the Si–O–Si bond is disiloxane or  $\text{Si}_2\text{H}_6\text{O}$ . Also called disilyl ether, its structure is a single Si–O–Si bond terminated by three H atoms at each end ( $\text{H}_3\text{Si}-\text{O}-\text{SiH}_3$ ). There have been numerous studies investigating the Si–O–Si bond,<sup>1–5</sup> particularly owing to its importance in the modelling of silica compounds, which are the most abundant constituent of the Earth's crust. Most importantly, silica compounds range in function from glasses to quartz crystals, both of which comprise large sectors in industry.<sup>6–8</sup> In some studies, disiloxane has been used as a sealant and as a component in cosmetics,<sup>9</sup> or as a prototype region of a zeolite or clay substrate for applications ranging from catalysis to prebiotic synthesis.<sup>10</sup>

Experimental evaluations of the Si–O–Si bond indicated an anharmonic bending potential with a low linearisation barrier, which makes it quite difficult to attain sufficient accuracy in such measurements.<sup>11,12</sup> Despite the significant volume of

previous studies dedicated to the Si–O–Si bond,<sup>10,13–19</sup> the properties of Si–O–Si bond obtained in most of these studies are not consistent with each other. In *ab initio* studies, in particular, multiple calculation methods resulted in different values for the bond angle and length,<sup>18,20,21</sup> as well as the linearisation energy<sup>13,14,17</sup> and Si–O–Si potential energy surface,<sup>10,18</sup> among other factors. These properties and the bond geometry itself were shown to be sensitive to the choice of basis-set and level of theory (the former more than the latter) according to at least one previous study.<sup>17</sup>

To narrow down the possibilities, it would be ideal to employ the most reliable methods at our disposal. Quantum Monte Carlo (QMC), one of the most reliable many-body calculation methods, is expected to provide a reliable result.<sup>22</sup> We used the fixed-node diffusion Monte Carlo (FNDMC) method,<sup>22</sup> which has been widely and successfully applied to several molecular systems.<sup>23–30</sup> Although FNDMC results are also affected by the choice of basis-set,<sup>27,31</sup> we note that the basis-set dependency is considerably different from that for a self-consistent-field (SCF)-based method, such as the density functional theory (DFT) and molecular orbital (MO) methods. In SCF-based approaches, the choice of the basis-set affects both the amplitude and the nodal positions of the corresponding many-body wavefunctions (although the methods do not explicitly employ a many-body wavefunction condition). With FNDMC, by contrast, the choice only affects the nodal positions. The amplitude can be automatically adjusted such that its shape approaches that of the exact solution as closely as possible under a restriction with fixed nodal positions.<sup>22,32</sup> Its typical example is the description of electron–nucleus cusps.<sup>33</sup> Even without explicit inclusion of the singular

<sup>a</sup> School of Materials Science, JAIST, Asahidai 1-1, Nomi, Ishikawa, 923-1292, Japan. E-mail: adietri@icloud.com

<sup>b</sup> Department of Electrical Engineering, Indian Institute of Technology Gandhinagar, Palaj 382355, Gujarat, India

<sup>c</sup> Materials Science and Technology Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA

<sup>d</sup> School of Information Science, JAIST, Asahidai 1-1, Nomi, Ishikawa, 923-1292, Japan

<sup>e</sup> Scuola Internazionale Superiore di Studi Avanzati (SISSA), via Bonomea, 265-34136 Trieste, Italy

<sup>f</sup> Research Center for Advanced Computing Infrastructure, JAIST, Asahidai 1-1, Nomi, Ishikawa 923-1292, Japan. E-mail: kenta\_hongo@mac.com

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d1cp01471d



electron–electron cusp analytical form in the wavefunction, numerical evolution driven by FNDMC will still produce an appropriate singular cusp amplitudes on electron–electron interaction.<sup>22,32</sup> (However, practically speaking, such poor description of the wavefunction frequently causes singularity of the local energy due to electron–nucleus cusps and numerical instability within the algorithm. This is easily remedied by introducing cusp correction proposed by Ma *et al.*<sup>33</sup>) The self-healing property of the amplitude works to considerably reduce the basis-set dependency: the bias arises from using an incomplete basis-set is generally significantly lower than SCF-based methods.<sup>17</sup> This bias in FNDMC is dependent on the fixed nodal positions (the nodal surface), which can be characterised by the total energy due to the fixed-node variational principle.<sup>22</sup> This principle allows for the comparison of qualities of nodal surfaces based on the comparison of the total energies; the lower the total energy, the higher the quality of the nodal surface. In turn, a higher nodal surface quality means the closer it is to the true ground state nodal surface, which translates to a more reliable FNDMC calculation result.

In this study, we apply single determinant FNDMC to investigate the basis-set dependence of the linearisation energy of a disiloxane molecule in comparison with other methods including DFT and CCSD(T), as well as empirical measurements from earlier studies.<sup>11,12,34</sup> The FNDMC depends on the basis-set implicitly through its fixed nodal surface provided by the Slater determinant whose orbitals are expanded in terms of the basis-set, whereas DFT and CCSD(T) strongly depend on the choice of basis-set. The cc-pVxZ<sup>35–37</sup> and cc-pCVxZ<sup>38</sup> basis-sets ( $x = D, T, Q$ ), which are usually applied to account for electron correlation, are examined in the present study and then applied to all the *ab initio* and semi empirical calculations to investigate the basis-set dependency of linearisation energy evaluation.

## 2 Model and methodology

Our target property is simply a linearisation barrier of disiloxane  $\text{Si}_2\text{H}_6\text{O}$  between “linear” and “nonlinear” (bent) structures. As can be seen from Fig. 1, the linear structure possesses an eclipsed  $D_{3h}$  symmetry<sup>16</sup> and the nonlinear structure is a bent conformation of the  $C_{2v}$  symmetry.<sup>39</sup> For the two fixed structures, the barrier,  $\Delta E_{\text{barrier}}$ , is defined as the energy difference between the linear and nonlinear structures.

The accuracy of computed  $\Delta E_{\text{barrier}}$  values can be calibrated by referring to experimentally observed  $\Delta E_{\text{barrier}}$  values. To compare our computational results with the experimental results, however, we note that the experimental  $\Delta E_{\text{barrier}}$  involves the zero-point

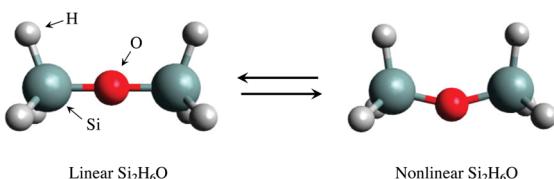


Fig. 1 Linear and nonlinear molecular structures of disiloxane,  $\text{Si}_2\text{H}_6\text{O}$ .

energy (ZPE) contribution, whereas our computational  $\Delta E_{\text{barrier}}$  value considers only the electronic contribution. The ZPE contributions can also be evaluated at each level of theory, although their accuracy depends on methods adopted and the basis-set. To investigate the basis-set dependence of  $\Delta E_{\text{barrier}}$ , we simply consider the electronic  $\Delta E_{\text{barrier}}$ . Thus, the experimental  $\Delta E_{\text{barrier}}$  values are corrected by adopting a theoretical estimate from the work of Koput.<sup>15</sup> We address this issue in more detail later.

In the present study, the energies of the two structures  $E_{\text{linear}}$  and  $E_{\text{nonlinear}}$  are computed through various combinations of (1) *ab initio* and semi empirical methods and (2) basis-sets.

(1) Calculation methods applied are DFT with B3LYP exchange–correlation functional (DFT-B3LYP),<sup>40</sup> CCSD(T),<sup>41</sup> and FNDMC<sup>22</sup> with trial wavefunctions generated from DFT-B3LYP. Within the framework of DFT, DFT-B3LYP is a standard method for covalent systems. Within the correlated methods, CCSD(T) is known to be the “gold standard” in quantum chemistry. A variety of methods and basis-sets are chosen in line with the previous results<sup>17</sup> indicating dependence of the Si–O–Si bond description on such choice. FNDMC is known to be comparable with CCSD(T); however, for the first time, we apply FNDMC to an evaluation of  $\Delta E_{\text{barrier}}$ . In order to account for relativistic effects, some calculations at the CCSD(T) level have been performed using the Douglas–Kroll–Hess (DKH) second-order Hamiltonian. We compare these results with regular CCSD(T) calculations to gauge the significance of relativistic effects on disiloxane. To our knowledge, this is the first such investigation of relativistic effects for the disiloxane molecule.

(2) Our basis-sets are a family of Pople triple zeta basis-set from 6-311G, as well as polarisation function added 6-311G\*\* and 6-311G(3df), and accounting for diffuse functions as well in 6-311+G, 6-311+G\*\*, and 6-311+G(3df) sets.<sup>42,43</sup> Also studied is the family of Dunning's correlation-consistent basis-set (cc-pVxZ;  $x = D, T, Q$ )<sup>35–37</sup> and their core-valence correlated variants (cc-pCVxZ;  $x = D, T, Q$ ).<sup>38</sup> The cc-pVxZ and cc-pCVxZ basis-sets were originally developed for correlated methods such as CCSD(T), but have been found to be appropriate even for DFT-B3LYP and FNDMC in properly reproducing the dynamic electron correlation.<sup>35</sup> In particular, the polarisation functions included implicitly in cc-pVxZ are essential for properly describing the Si–O–Si bond. To describe the correlation effects more precisely, the present study considers the dynamic correlation between core and valence electrons by applying the cc-pCVxZ basis-set, which can reproduce the core-valence correlation by minimising the difference in the correlation energies between all-electron and valence-only (using pseudopotentials) calculations.<sup>44</sup> In order to compare with state-of-the-art approaches in quantum chemistry, CCSD(T) calculations are performed using even more robust basis-sets, including the quintuple zeta sets (cc-pV5Z and cc-pCV5Z), as well as their aug-cc-pVxZ and aug-cc-pCVxZ counterparts, which possess added diffuse functions. The cc-pV(x + d) basis-sets are also used, which possess additional *d* functions, to better describe Si atoms.<sup>35,37,45</sup> In order to extrapolate to the complete basis-set (CBS) limit, both quadruple and quintuple



zeta sets are included. Extrapolation to the CBS limit are performed using the scheme of Halkier *et al.*,<sup>46</sup> with extrapolation coefficients from the optimisation of Truhlar.<sup>47</sup>

In the present study we obtain the linear and nonlinear structures of disiloxane using the second-order Møller-Plesset perturbation theory (MP2)<sup>48</sup> with the cc-pVQZ basis-set (MP2/cc-pVQZ level), and previous studies conducted at the MP2/cc-pVQZ level of theory<sup>16,20</sup> have established a reliable favourability toward a nonlinear structure and provide good agreement for the Si–O–Si angle in accordance with experimental results.<sup>39</sup> Modelling of the Si–O–Si angle in particular has been heavily emphasised in previous theoretical studies for both disiloxane<sup>17</sup> and pyrosilic acid.<sup>18,19</sup> While geometry optimization by CCSD(T) is desirable, the computational cost associated is unfortunately much higher, especially for the more robust basis-set used in this work. For example, the cc-pVQZ basis-set for disiloxane translates to over 350 basis functions. With 20 degrees of freedom in the disiloxane molecule, optimization with CCSD(T) involved steep computational costs, and optimization at the MP2/cc-pVQZ level is chosen instead. A comparison between the MP2 and CCSD(T) geometry optimization with the cc-pVTZ basis-set is shown in the ESI.† With a difference of 0.5 degrees for the Si–O–Si angle between the two calculation methods at the cc-pVTZ level, and a linearization barrier difference of less than 0.001 kcal mol<sup>−1</sup> calculated for both geometries at MP2 and CCSD(T) level, geometry optimization at the MP2 level of theory does not seem to significantly impact the linearization barrier calculated at the CCSD(T) level. The two optimised structures at the MP2/cc-pVQZ level of theory are used in this work to calculate all  $\Delta E_{\text{barrier}}$  values, *i.e.*, common to all levels of theory, where the linear structure has a Si–O–Si angle of 180° and the nonlinear structure has an optimised Si–O–Si angle of 146.8° comparable with the experimental value of 144.1°.<sup>39</sup> The other structural parameters are given in the ESI,† and are also in good agreement with experiments.

The present study adopted no pseudopotential calculations, but rather all-electron calculations with a total of 42 electrons for a single disiloxane molecule. This molecular system is not so large that the CCSD(T) calculation within the frozen core approximation is feasible, despite the computational cost of CCSD(T) scaling as  $N^7$ , where  $N$  is the number of electrons in the system. By contrast, the DFT cost scales with  $N^3$  (or less) and is therefore unimportant. Similar to DFT, the FNDMC cost scales as  $N^3$ , although the prefactor of FNDMC is much larger than that of DFT. This is because a vast number of random sampling points are required to obtain a sufficiently small error bar (sub-chemical accuracy of  $\sim 0.1$  kcal mol<sup>−1</sup>) to calibrate a small  $\Delta E_{\text{barrier}}$  ( $\sim 0.5$  kcal mol<sup>−1</sup>). Recent parallel computers, however, enable us to apply FNDMC to evaluate such a tiny  $\Delta E_{\text{barrier}}$ , because its algorithm is intrinsically parallel.

In our FNDMC calculations, we adopted Slater–Jastrow type wavefunctions as their fixed-node trial wavefunctions.<sup>49</sup> Molecular orbitals entering the single Slater determinants were generated by DFT-B3LYP with various types of basis-set. The Jastrow factor consists of one-, two-, and three-body terms<sup>50</sup> including 88 variational parameters in total. These parameters

are optimised through a variance minimisation scheme,<sup>51</sup> and only the two-body term holds the electron–electron cusp condition.<sup>52</sup> The electron–nucleus cusp condition, which is a short-range one-body correlation effect, is satisfied by the cusp correction scheme applied to the Gaussian basis-set<sup>33</sup> instead of imposing the cusp condition on the one-body term. Note that the Jastrow factor does not change the (fixed) nodal surfaces and is responsible for the numerical stability in FNDMC. However, the quality of the nodal surfaces determines the accuracy of the FNDMC energies in terms of the fixed-node variational principles.<sup>53</sup> Accordingly, the FNDMC accuracy depends implicitly on the basis-set adopted, which is used to expand the molecular orbitals entering the Slater determinant. In addition to the fixed-node error, another source of bias in actual FNDMC calculations arises from the short-time approximation with finite (small) timesteps.<sup>54</sup> To remove this bias, it is common to use multiple timesteps to make the linear regression and obtain calculation results for  $\delta t \rightarrow 0$ . This regression is an approximation of the theoretical  $\delta t = 0$  result. The present study considers both linear and quadratic extrapolations. It is commonly understood that the size of the error bar from FNDMC calculations scale inversely proportional with the square root of the number of samples in the calculations.<sup>22</sup> In order to halve the error bar, for example, the computational cost is increased fourfold. This results in an effective limit on the size of the error bar due to feasible computational cost of FNDMC calculations.

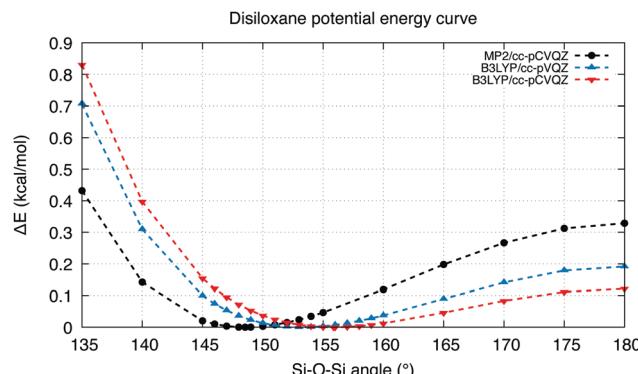
The software package Gaussian09<sup>55</sup> was used to conduct the MP2 (geometry optimisation), DFT-B3LYP, and CCSD(T) calculations, whereas the CASINO<sup>56</sup> code was used to perform FNDMC calculations. An electron–nucleus cusp correction scheme<sup>33</sup> for Gaussian orbitals was utilised in the all-electron FNDMC calculation in CASINO. Most of the basis-sets utilised in this work are available in the Gaussian09 software package. Basis-sets of the cc-pCVxZ, cc-pV(x + d)Z and aug-cc-pCVxZ families are obtained from the online Basis-Set Exchange library.<sup>57–59</sup> In FNDMC for both structures, the number of target population was 11 520, and the number of steps in the imaginary-time evolution was set to 2000 and 500 000 for equilibrated and accumulated phases, respectively. In addition, we carried out FNDMC calculations with different timesteps of 0.01, 0.005, and 0.001 a.u.<sup>−1</sup> to remove the short-time bias (see ESI† for more details).

## 3 Results and discussion

### 3.1 Geometry and ZPE contribution

Several studies have been conducted on the geometry of the disiloxane molecule, although most *ab initio* methods tested did not manage to replicate the available experimental results. Rather than the Si–O bond length, the Si–O–Si bond angle and linearisation barrier have been found to be greatly dependent on the choice of basis-set used to represent the wavefunction.<sup>17</sup> This study focuses on the linearisation barrier of disiloxane, taking the same optimised geometries for all calculations,





**Fig. 2** Potential energy curve of disiloxane molecule calculated by MP2/cc-pCVQZ, B3LYP/cc-pVQZ, and B3LYP/cc-pCVQZ combinations of calculation method and basis-set. Dashed lines serve to better illustrate the trends present in the data.

partly owing to the difficulty to optimise the geometries by FNDMC.<sup>‡</sup>

The bending potential energy of the disiloxane molecule is a shallow one, as shown in Fig. 2. In comparison with the water molecule for example, which has a relatively harmonic bending potential curve with a potential difference of 1.0 kcal mol<sup>-1</sup> at deviation of H-O-H bending angle of around 8°,<sup>71</sup> the disiloxane molecule bending potential curve is anharmonic, with a potential difference of 1.0 kcal mol<sup>-1</sup> at deviation of Si-O-Si bending angle of around 15° to 20° less than the optimal, and a small linearisation barrier of less than 0.5 kcal mol<sup>-1</sup>. Although the linearisation barriers shown in Fig. 2 differ depending on the calculation method and basis-set used, all potential curves share these similar properties with one another.

The computational and experimental linearisation barriers of disiloxane,  $\Delta E_{\text{barrier}}$ , are listed in Table 1 and plotted in Fig. 3. It has been commonly observed experimentally that the nonlinear (bent) structure of disiloxane is energetically more favourable than a linear structure, which translates to a positive linearisation barrier. All three experimental results seem to agree on this point, and the work of Aronson *et al.*<sup>34</sup> in particular reports a higher linearisation barrier (at 1.1 to 1.4 kcal mol<sup>-1</sup>) than the other two results,<sup>11,12</sup> which report a barrier of approximately 0.3 kcal mol<sup>-1</sup>.

The experimental work of Durig, Flanagan, and Kalasinsky obtained Raman (10–3500 cm<sup>-1</sup>) and infrared spectra of gaseous (4000–30 cm<sup>-1</sup>) and solid (4000–450 cm<sup>-1</sup>) disiloxane and deuterated disiloxane.<sup>11</sup> Most significantly, their assessment of the Raman (10–200 cm<sup>-1</sup>) spectra produced potential energy functions which results in a disiloxane linearisation barrier of 112 ± 5 cm<sup>-1</sup> (0.32(1) kcal mol<sup>-1</sup>), while the analysis of Koput and Wierzbicki of the low frequency Raman spectra (10–200 cm<sup>-1</sup>) produced a similar value of 103.9 ± 2.6 cm<sup>-1</sup>.

<sup>‡</sup> Geometry optimisation by FNDMC is a long-standing issue due to the existence of the systematic bias mainly depending on the fixed-nodes quality<sup>60</sup> and the infinite variance problem.<sup>61,62</sup> Although the continuous efforts are overcoming the problems,<sup>60–70</sup> the methods are still in development level and its application is practically accompanied by many difficulties.

**Table 1** Linearisation barrier values obtained from theoretical calculations and experiments. Its ZPE correction and ZPE-corrected Raman value are also given. Computational values ( $\Delta \varepsilon_0$ ) are to be compared with a ZPE-corrected experimental value ( $\Delta \varepsilon_1 - \Delta \text{ZPE}$ ); see the text for the definition of notation, sign, etc.

Basis-set	$\Delta \varepsilon_0$ [kcal mol <sup>-1</sup> ]		
	DFT-B3LYP	CCSD(T)	FNDMC
6-311G	-2.20	-2.45	-0.12 ± 0.27
6-311+G	-2.46	-1.98	-0.35 ± 0.16
6-311G**	-0.43	0.06	0.07 ± 0.15
6-311+G**	-0.53	0.24	-0.32 ± 0.14
6-311G(3df)	0.10	0.47	0.45 ± 0.24
6-311+G(3df)	0.15	0.58	-0.42 ± 0.23
cc-pVQZ	0.80	1.44	0.40 ± 0.15
cc-pVTZ	0.18	0.46	0.47 ± 0.14
cc-pVQZ	0.18	0.44	0.26 ± 0.10
cc-pV5Z	—	0.49	—
cc-pCVQZ	0.76	1.42	0.54 ± 0.12
cc-pCVTZ	-0.01	0.32	0.31 ± 0.11
cc-pCVQZ	0.06	0.35	0.36 ± 0.10
cc-pCV5Z	—	0.48	—
aug-cc-pVQZ	—	0.50	—
aug-cc-pV5Z	—	0.48	—
aug-cc-pCVQZ	—	0.46	—
aug-cc-pCV5Z	—	0.47	—
cc-pV(Q + d)Z	—	0.45	—
cc-pV(5 + d)Z	—	0.48	—

Far IR spectrum <sup>34</sup>	IR-Raman (solid) <sup>11</sup>	$\Delta \varepsilon_1 - \Delta \text{ZPE}$ [kcal mol <sup>-1</sup> ]	
		Raman <sup>12</sup>	ZPE-corrected Raman <sup>15</sup>
1.1–1.4	0.32	0.30	0.36

(0.297(7) kcal mol<sup>-1</sup>).<sup>12</sup> These results were reported more recently than the first and achieve good consilience with the DFT predictions from the highest quality basis-set.<sup>17</sup> Therefore, it is reasonable to infer that a linearisation barrier of approximately 0.3 kcal mol<sup>-1</sup> is a reliable value for the disiloxane molecule.

It needs to be stressed, however, that the experimentally measured linearisation barrier may not be comparable to the ground state values calculated from the first principles. The ground state energy obtained through *ab initio* calculations are physically unobtainable experimentally because of the ZPE, *i.e.* the difference between the ground state and the lowest energy vibrational state. Even at absolute zero temperature, the lowest energy level achievable is a vibrational state  $\varepsilon_1$  instead of the electronic ground state  $\varepsilon_0$ :

$$\Delta \varepsilon_1 = \Delta \varepsilon_0 + \Delta \text{ZPE} \quad (1)$$

Consequently, any energetic barriers measured in the experiments is at best the difference between the lowest vibrational states  $\Delta \varepsilon_1$ , whereas energetic barriers calculated through *ab initio* methods are from the electronic ground states  $\Delta \varepsilon_0$ . Therefore, a comparison between energetic barriers obtained from the theoretical calculations and experimental measurements must also account for the difference in ZPE.

The preceding theoretical study by Koput<sup>15</sup> estimates a  $\Delta \text{ZPE}$  value of  $-20 \text{ cm}^{-1} \approx -0.06 \text{ kcal mol}^{-1}$ , which in line with eqn (1) should result in a higher ground state linearisation



# Si<sub>2</sub>H<sub>6</sub>O Linearisation Barrier

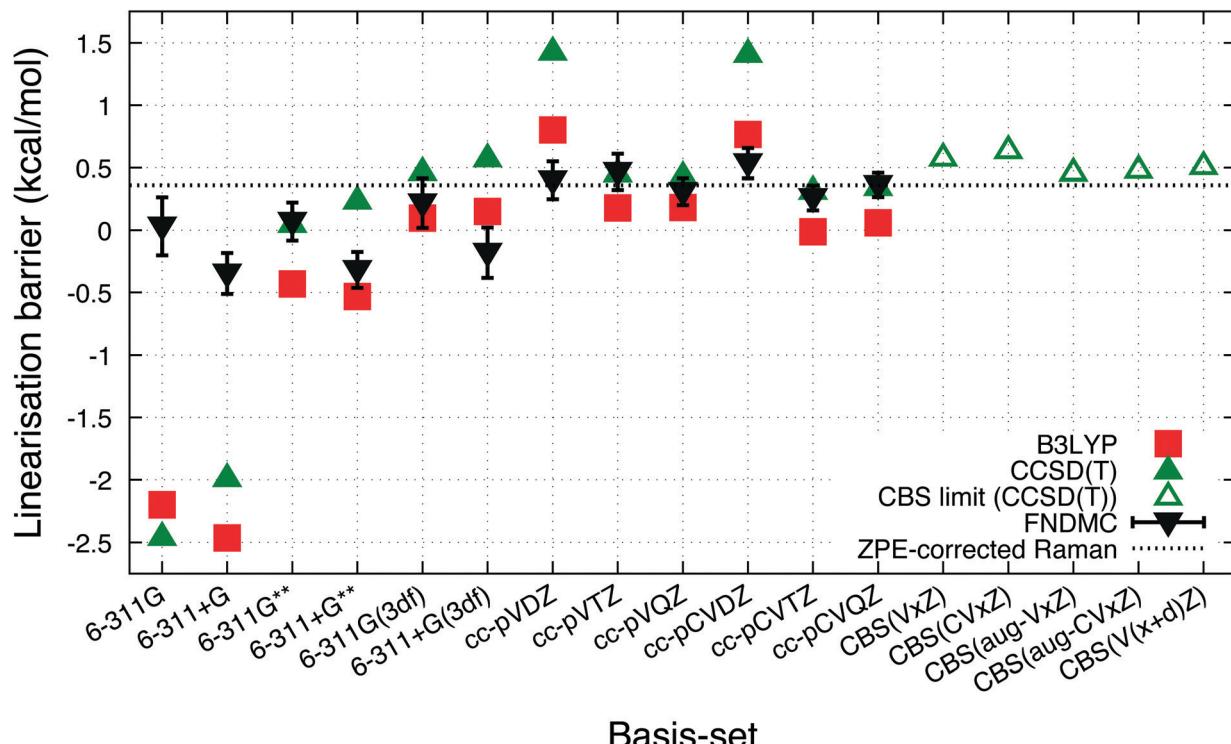


Fig. 3 Linearisation barrier of disiloxane calculated in this work.

barrier than the experimentally measured result. The study considered separately the large amplitude motion (Si–O–Si bending, which is the imaginary frequency mode in linear disiloxane) and the small amplitude vibrations, leading to a more refined estimation of the zero-point vibrational energy. By constructing the harmonic force constants at the MP2 level of theory using D/T basis-set, the energy levels of the small amplitude vibrations of the disiloxane molecule were estimated. The upper bound of the contribution was estimated at 20 cm<sup>-1</sup> from Si–O–Si bending angles of 180° to 150°. Therefore, this work treats the ground state linearisation barrier of 0.36 kcal mol<sup>-1</sup> as a reasonably accurate “exact” linearisation barrier for a point of comparison with *ab initio* calculations. This value is referred to as “ZPE-corrected Raman” in Table 1 and based on a dotted line in the figures herein, as a point of comparison. ZPE correction has not been considered in *ab initio* comparisons or based on experimental results, which have cited 0.3 kcal mol<sup>-1</sup> as the point of comparison.<sup>17</sup>

## 3.2 DFT-B3LYP results

The results for DFT-B3LYP in this study, particularly the results for the quadruple zeta basis-set cc-pVQZ and cc-pCVQZ, seem to indicate that previous agreement between the B3LYP results and experimental data<sup>17</sup> is due to coincidence instead of convergence toward the complete basis-set limit. Indeed, for the largest basis-set used in this study, the linearisation barrier calculated by B3LYP is at best approximately 0.1 kcal mol<sup>-1</sup>

smaller than the ZPE-corrected measurement. It does not seem that the B3LYP exchange–correlation functional provides adequate accounting of the electron correlation for properly describing a disiloxane molecule.

## 3.3 CCSD(T) results

CCSD(T) calculations were conducted with the same geometries and basis-sets as the B3LYP calculations. The results of these calculations are shown in Table 1. The trends of the linearization barrier agree almost perfectly with B3LYP calculations, with a universal shift in the energetic favourability toward the non-linear conformer denoted by the larger linearisation barriers. We achieve good agreement with another CCSD(T) calculation in an earlier study<sup>17</sup> reporting a linearisation barrier of 0.48 kcal mol<sup>-1</sup> using the cc-pVTZ basis-set. These results show that accounting for the electron correlation is indeed necessary to properly model the Si–O–Si bond in disiloxane, and supports the conclusion of earlier studies citing cc-pVTZ and CCSD(T) as the minimum reliable level of description for the disiloxane molecule.<sup>16,17</sup>

Calculations with more state-of-the-art basis-sets have also been performed on the disiloxane linearisation barrier. Quintuple zeta basis-set cc-pV5Z and cc-pCV5Z are used, as well as quadruple and quintuple zeta sets for diffusion function augmented aug-gcc-pVxZ family of basis-set. Extra d functions present in the cc-pV(x + d)Z family of basis-set are also taken into account. Extrapolations to the CBS limit are performed

according to the scheme in eqn (2):

$$E_{\infty}^{\text{tot}} = \frac{X^{\alpha}}{X^{\alpha} - Y^{\alpha}} E_X^{\text{HF}} - \frac{Y^{\alpha}}{X^{\alpha} - Y^{\alpha}} E_Y^{\text{HF}} + \frac{X^{\beta}}{X^{\beta} - Y^{\beta}} E_X^{\text{corr}} - \frac{Y^{\beta}}{X^{\beta} - Y^{\beta}} E_Y^{\text{corr}} \quad (2)$$

for quadruple and quintuple basis-sets ( $X = 5$ ,  $Y = 4$ ), using coefficients optimised in the work of Truhlar<sup>47</sup> with  $\alpha = 3.4$  and  $\beta = 2.4$ . Since electronic correlation energy converges less rapidly than electronic energy to the CBS limit,<sup>46</sup> it is necessary to separate contributions to the linearisation barrier due to electronic energy (Hartree–Fock or HF) and correlation energy (corr). The results from the CCSD(T) calculations performed using these quadruple and quintuple zeta basis-sets, as well as the extrapolated CBS limit, is shown in Table 3 and in Fig. 4.

The trends of CCSD(T) results seem to converge to linearisation barrier values around 0.1–0.3 kcal mol<sup>−1</sup> higher than the ZPE-corrected Raman results of 0.36 kcal mol<sup>−1</sup>, with the extrapolated CBS limit for various basis-set families showing linearisation barriers of 0.45–0.65 kcal mol<sup>−1</sup>. CCSD(T) calculation shows then that previous works underestimated the linearisation barrier of disiloxane, though it must be noted that the calculations are performed with MP2/cc-pVQZ geometries owing to the high computational cost of CCSD(T) geometry optimisations with equivalent basis-sets.

Comparison is also made to measure the significance of relativistic effects between the non-relativistic and scalar relativistic Hamiltonian at the CCSD(T) level. The basis-sets used are of the cc-pV<sub>x</sub>Z and cc-pCV<sub>x</sub>Z families ( $x = D, T, Q$ ). These results are shown in Table 2. At the higher level basis-sets, inclusion of relativistic effects seem to contribute strengthening of the linearization barrier in the order of 0.05–0.1 kcal mol<sup>-1</sup>. As relativistic effects serve to energetically stabilize the s- and p-orbitals,<sup>72,73</sup> it follows that the lone pair repulsion at the oxygen atom is strengthened due to radially contracted p orbitals, leading to an energetic stabilization of the non-linear structure. While usually considered negligible, these effects prove somewhat significant in the context of the disiloxane linearization barrier (0.36 kcal mol<sup>-1</sup>).

**Table 2** Disiloxane linearisation barrier values from CCSD(T) calculations with non-relativistic and scalar relativistic Hamiltonians

Basis-set	$\Delta\epsilon_0$ [kcal mol $^{-1}$ ]	
	$\Delta E_{\text{barrier}}^{\text{non-relativistic}}$	$\Delta E_{\text{barrier}}^{\text{scalar relativistic}}$
cc-pVDZ	1.44	1.24
cc-pVTZ	0.46	0.50
cc-pVQZ	0.44	0.50
cc-pCVDZ	1.42	1.23
cc-pCVTZ	0.32	0.42
cc-pCVQZ	0.35	0.46

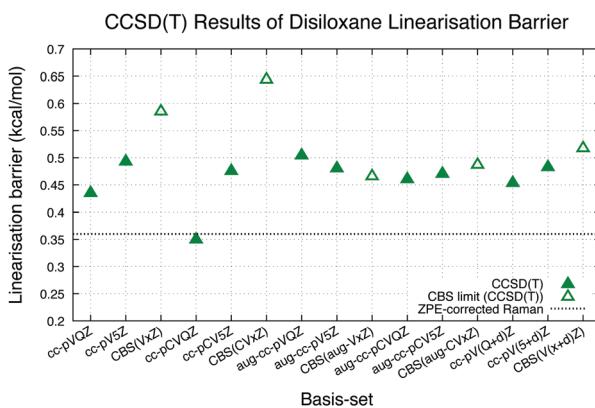
### 3.4 FNDMC results

The linearisation barrier calculated by FNDMC is also shown in Table 1 and Fig. 3. These values from FNDMC converge to a value close to the ZPE-corrected Raman benchmark adopted in this study of 0.36 kcal mol<sup>-1</sup> for the largest basis-sets. Also observed is the relative independence of the linearisation barrier calculated from the basis-set used to form the initial trial wavefunction using DFT-B3LYP relative to the other two calculation methods. The influence of the different basis-sets on the end result of the FNDMC calculations directly translates into how they affect the trial wavefunction nodal surfaces. The amplitudes of the trial wavefunction, meanwhile, do not affect the end result at ( $\tau \rightarrow \infty$ ), which, in turn, limits the dependence of the end result on the basis-sets used to form the trial wavefunction. Pople basis-sets in particular affects the FNDMC calculations relative to the correlation consistent basis-sets, with relatively lower linearisation barriers and particularly for basis-sets with diffuse functions (6-311+G, 6-311+G\*\*, and 6-311+G(3df)) even energetically favouring linear structures of disiloxane. This dependence is clearly visible in Fig. 3, though the magnitude is lesser than DFT-B3LYP and CCSD(T).

As mentioned previously, the fixed-node variational principle allows for characterization of nodal surface quality by way of comparing the absolute values of total energies. These absolute values are shown in Table 4, sorted in accord with the quality of the nodal surface (from low quality, high total energy, to high quality, low total energy). It can be observed that the cc-pCVTZ produces a nodal surface of slightly better quality than cc-pVQZ, indicating that the electron correlation between the core and valence electrons can be a significant factor for improvement beyond the triple zeta level of atomic orbital description.

### 3.5 Effects of basis-sets

Disiloxane linearisation barrier dependence on the basis-set was observed for all cases. In agreement with previous studies,<sup>17</sup> this dependence is more significant than the methodologies used in the *ab initio* calculations, particularly considering the double zeta-level basis-sets cc-pVDZ and cc-pCVDZ. Triple zeta basis-sets seem to offer the minimum level of description to reliably predict the linearisation barrier, whereas quadruple zeta basis-sets result in an extremely good prediction, especially for the CCSD(T) and FNDMC calculations. Fig. 3 clearly shows very similar trends for each calculation method with a substantial dependence on the basis-set indicated, particularly



**Fig. 4** Linearisation barrier of disiloxane calculated by CCSD(T) using state-of-the-art basis-set.

**Table 3** Disiloxane linearisation barrier values from CCSD(T) calculations with more state-of-the-art basis-sets, contribution from HF and electronic correlation, and extrapolation to the complete basis set (CBS) limit

Basis-set	$\Delta\epsilon_0$ [kcal mol <sup>-1</sup> ]		
	$\Delta E_{\text{barrier}}^{\text{HF}}$	$\Delta E_{\text{barrier}}^{\text{corr}}$	$\Delta E_{\text{barrier}}$
cc-pVQZ	-0.00	0.44	0.44
cc-pV5Z	-0.02	0.52	0.49
CBS(VxZ)			0.59
cc-pCVQZ	-0.05	0.40	0.35
cc-pCV5Z	-0.04	0.51	0.48
CBS(CVxZ)			0.65
aug-cc-pVQZ	0.01	0.50	0.51
aug-cc-pV5Z	-0.03	0.51	0.48
CBS(aug-VxZ)			0.47
aug-cc-pCVQZ	-0.03	0.49	0.46
aug-cc-pCV5Z	-0.04	0.51	0.47
CBS(aug-CVxZ)			0.49
cc-pV(Q + d)Z	-0.05	0.50	0.45
cc-pV(5 + d)Z	-0.03	0.52	0.48
CBS(V(x + d)Z)			0.52

**Table 4** Total energies from FNDMC calculations, sorted from highest to lowest

Basis-set	$E_{\text{total}}$ [Hartree]	
	Nonlinear	Linear
cc-pVDZ	-657.7994(2)	-657.7988(2)
cc-pCVDZ	-657.8062(1)	-657.8053(1)
cc-pVTZ	-657.8270(2)	-657.8263(2)
cc-pVQZ	-657.8366(1)	-657.8362(1)
cc-pCVTZ	-657.8376(1)	-657.8371(1)
cc-pCVQZ	-657.8437(1)	-657.8431(1)

towards the smaller sets. Convergence of the disiloxane linearisation barrier is generally observed for all three calculation methods, albeit not necessarily converging to the same value.

Previous theoretical studies suggest that this converging trend is attributed to the increasing addition of polarisation functions<sup>17</sup> within the basis-sets used. Adding polarisation functions serves to better reproduce dynamical correlations in the system. This is in line with previous theoretical works with semi-empirical methods implying that calculations without electron correlation favour the linear structure, thereby resulting in negative values of linearisation barrier. This is also reflected in preceding geometry optimisation calculations without polarisation functions, resulting in Si–O–Si angles close to 170°.<sup>13,14</sup> Therefore, it is expected that both cc-pVxZ and cc-pCVxZ basis-sets should converge to a reliable predicted linearisation barrier value because polarisation functions are systematically included with an increasing description of the atomic orbitals.

FNDMC is overall less dependent on the basis-set used compared to both B3LYP and CCSD(T). Although previous studies have recommended treating disiloxane with the cc-pVTZ basis-set at a minimum, cc-pVDZ is shown to generate a sufficiently high quality nodal surface for use in DMC calculations, giving a linearisation barrier in good agreement with the experimental values. Even for the smallest basis-sets

tested in this study, FNDMC shows relatively more accurate values of the linearisation barrier, and therefore less dependence on the basis-set used to form the trial wavefunction per the initial expectations.

Comparing the FNDMC results to the CCSD(T) ones using more robust basis-sets show that in general there are improvements that can be significant to the nodal surface by employing these more robust basis-sets, as CCSD(T) results at the CBS limit clearly show more energetic preference toward the nonlinear structure of disiloxane. The effects of using these more robust basis-sets to the nodal surface may be intriguing to explore, in order to find basis-sets more suited for FNDMC application in disiloxane.

### 3.6 Effects of methodologies

Previous expectations on the trend in methodologies are derived from previous studies,<sup>13,74</sup> particularly the study by Koput in 1990,<sup>13</sup> in which the inclusion of the electron correlation proved vital to predicting the energetic favourability of the nonlinear structure of disiloxane because the SCF calculation produced a near-linear structure of disiloxane. This and other studies<sup>74</sup> gave rise to the general expectation that the inclusion of the electron correlation is important to properly model the structure of disiloxane. As previously mentioned, it can be observed that the B3LYP exchange–correlation functional is insufficient to properly recover the electron correlation and thereby predict the disiloxane linearisation barrier. Our CCSD(T) calculation results shown in Table 3 reinforce the significance of accounting for electron correlation, as most of the energy difference between the nonlinear and linear structures of disiloxane come from electron correlation. Increasingly accounting for correlation, therefore, leads to calculation results favouring the nonlinear structure, and thus resulting in a higher linearisation barrier.

FNDMC has the added advantage of being able to reliably predict the energetics of disiloxane with smaller basis-sets. While improvements to the nodal surface is clearly possible by accounting for electron correlation, the difference in scalability between CCSD(T) and FNDMC means that this advantage only grows more significant when considering larger systems. With better scalability for application in high performance computing (HPC) systems and the availability of pseudopotentials, FNDMC is a promising alternative method to CCSD(T) for describing the Si–O–Si bonds, particularly for larger systems.

## 4 Conclusion

The Si<sub>2</sub>H<sub>6</sub>O linearisation barrier was calculated using three separate methods, DFT-B3LYP, CCSD(T), and FNDMC, with six different basis-set choices in line with expectations derived from previous theoretical studies on disiloxane. The disiloxane molecule was optimized for both the non-linear and linear structures at the MP2/cc-pVQZ level. Geometry optimization at the MP2 level of theory does not seem to significantly impact



the linearization barrier calculated at the CCSD(T) level, with differences of linearization barrier at less than 0.001 kcal mol<sup>-1</sup> with the cc-pVTZ basis-set. Similar with previous studies, we observed that the systematic inclusion of polarisation functions, along with an increasing level of description for atomic orbitals, eventually result in a reliable prediction of the disiloxane linearisation barrier. All calculation methods eventually produced converged values with increasing level of basis-sets for the linearisation barrier, with 0.1 kcal mol<sup>-1</sup> for DFT-B3LYP and 0.45–0.65 kcal mol<sup>-1</sup> for CCSD(T), respectively. Inclusion of relativistic effects by scalar relativistic calculations slightly increases the linearization barrier by 0.05–0.1 kcal mol<sup>-1</sup> due to the increased lone pair repulsion from the contracted p-orbitals. The agreement between the experimental measurements and the DFT-B3LYP results at 0.3 kcal mol<sup>-1</sup> are shown to be likely accidental. ZPE-corrected experimental measurements are in good agreement with FNDMC results, with the ground state linearisation barrier taken at 0.36 kcal mol<sup>-1</sup>. The CCSD(T) results using state-of-the-art basis-sets, however, show that the linearisation barrier might be underestimated in these results. FNDMC is shown to be least dependent on the choice of basis-set among the three calculation methods applied, in line with initial expectations owing to the nature of FNDMC.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

The computations in this work have been performed using the facilities of Research Center for Advanced Computing Infrastructure at JAIST. T. I. acknowledges support by the US Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division and support from Grant-in-Aid for JSPS Research Fellow (18J12653). R. M. is grateful for financial supports from MEXT-KAKENHI (JP16KK0097, JP19H04692, and JP21K03400), FLAGSHIP2020 (project no. hp190169 and hp190167 at K-computer), the Air Force Office of Scientific Research (AFOSR-AOARD/FA2386-17-1-4049; FA2386-19-1-4015), and JSPS Bilateral Joint Projects (with India DST). K. H. is grateful for financial support from the HPCI System Research Project (Project ID: hp210019, hp210131, and jh210045), MEXT-KAKENHI (JP16H06439, JP17K17762, JP19K05029, JP19H05169, and JP21K03400), and the Air Force Office of Scientific Research (Award Numbers: FA2386-20-1-4036). K. N. acknowledges a support from the JSPS Overseas Research Fellowships, that from Grant-in-Aid for Early-Career Scientists (Grant Number JP21K17752), and that from Grant-in-Aid for Scientific Research(C) (Grant Number JP21K03400).

## Notes and references

1 E. Dupree and R. F. Pettifer, *Nature*, 1984, **308**, 523–525.

- 2 U. Selvaray, K. J. Rao, C. N. R. Rao, J. Klinowski and J. M. Thomas, *Chem. Phys. Lett.*, 1985, **114**, 24–27.
- 3 R. F. Pettifer, E. Dupree, I. Farnan and U. Sternberg, *J. Non-Cryst. Solids*, 1988, **106**, 408–412.
- 4 F. Mauri, A. Pasquarello, B. G. Pfrommer, Y. G. Yoon and S. G. Louie, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2000, **62**, R4786.
- 5 W. J. Malfait, W. E. Halter and R. Verel, *Chem. Geol.*, 2008, **256**, 269–277.
- 6 M. F. M. Santos, E. Fujiwara, E. A. Schenkel, J. Enzweiler and C. K. Suzuki, *Int. J. Miner. Process.*, 2015, **135**, 65–70.
- 7 K. I. Vatalis, G. Charalambides and N. P. Benetis, *Procedia Economics and Finance*, 2015, **24**, 734–742.
- 8 E. C. Peres, J. C. Slaviero, A. M. Cunha, A. H. Bandegharaei and G. L. Dotto, *J. Environ. Chem. Eng.*, 2018, **6**, 649–659.
- 9 K. M. Ryan, A. D. Drumm, C. E. Martin, A. K. Krumpfer and J. W. Krumpfer, *Reactive and Functional Polymers Volume One*, Springer, 2020, pp. 301–328.
- 10 B. T. Luke, *J. Phys. Chem.*, 1993, **97**, 7505–7510.
- 11 J. R. Durig, M. J. Flanagan and V. F. Kalasinsky, *J. Chem. Phys.*, 1977, **66**, 2775.
- 12 J. Koput, *J. Mol. Spec.*, 1983, **99**, 116–132.
- 13 J. Koput, *J. Chem. Phys.*, 1990, **148**, 299–308.
- 14 G. I. Csonka and J. Réffy, *Chem. Phys. Lett.*, 1994, **229**, 191–197.
- 15 J. Koput, *J. Phys. Chem.*, 1995, **99**, 15874–15880.
- 16 C. Carteret, A. Labrosse and X. Assfeld, *Spectrochim. Acta, Part A*, 2007, **67**, 1421–1429.
- 17 A. R. A. Derži, A. Gregušová, K. Runge and R. J. Bartlett, *Int. J. Quantum Chem.*, 2008, **108**, 2088–2096.
- 18 F. Noritake and K. Kawamura, *J. Comput. Chem. Jpn.*, 2015, **14**, 124–130.
- 19 F. Noritake, *J. Comput. Chem. Jpn.-Int. Ed.*, 2019, **5**, 2018–0016.
- 20 J. B. Nicholas and M. Feyereisen, *J. Chem. Phys.*, 1995, **103**, 8031.
- 21 X. Yuan and A. N. Cormack, *J. Non-Cryst. Solids*, 2003, **319**, 31–43.
- 22 W. M. C. Foulkes, L. Mitas, R. J. Needs and G. Rajagopal, *Rev. Mod. Phys.*, 2001, **73**, 33.
- 23 K. Hongo, M. A. Watson, R. S. Sánchez-Carrera, T. Iitaka and A. Aspuru-Guzik, *J. Phys. Chem. Lett.*, 2010, **1**, 1789–1794.
- 24 M. A. Watson, K. Hongo, T. Iitaka and A. Aspuru-Guzik, *A Benchmark Quantum Monte Carlo Study of Molecular Crystal Polymorphism: A Challenging Case for Density-Functional Theory*, 2012, ch. 10, pp. 101–117.
- 25 K. Hongo, N. T. Cuong and R. Maezono, *J. Chem. Theory Comput.*, 2013, **9**, 1081–1086.
- 26 K. Hongo, T. Iitaka, M. Watson, A. Aspuru-Guzik and R. Maezono, *J. Chem. Theory Comput.*, 2015, **11**, 907–917.
- 27 J. Koseki, R. Maezono, M. Tachikawa, M. D. Towler and R. J. Needs, *J. Chem. Phys.*, 2008, **129**, 085103.
- 28 K. Hongo and R. Maezono, *J. Chem. Theory Comput.*, 2017, **13**, 5217–5230.
- 29 H. Takagishi, T. Matsuda, T. Shimoda, R. Maezono and K. Hongo, *J. Phys. Chem. A*, 2019, **123**, 8726–8733.
- 30 T. Ichibha, Z. Hou, K. Hongo and R. Maezono, *Sci. Rep.*, 2017, **7**, 2011.

31 K. Nakano, R. Maezono and S. Sorella, *J. Chem. Theory Comput.*, 2019, **15**, 4044–4055.

32 R. Maezono, *J. Comput. Theor. Nanosci.*, 2009, **6**, 2474–2482.

33 A. Ma, M. D. Towler, N. D. Drummond and R. J. Needs, *J. Chem. Phys.*, 2005, **122**, 224322.

34 J. R. Aronson, R. C. Lord and D. W. Robinson, *J. Chem. Phys.*, 1960, **33**, 1004.

35 T. H. Dunning Jr., *J. Chem. Phys.*, 1989, **90**, 1007.

36 R. A. Kendall and T. H. Dunning Jr., *J. Chem. Phys.*, 1992, **96**, 6796.

37 D. E. Woon and T. H. Dunning Jr., *J. Chem. Phys.*, 1993, **98**, 1358.

38 K. A. Peterson and T. H. Dunning Jr., *J. Chem. Phys.*, 2002, **117**, 10548.

39 A. Almenningen, O. Bastiansen, V. Ewing, K. Hedberg and M. Traetteberg, *Acta Chem. Scand.*, 1963, **17**, 2455–2460.

40 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 1372.

41 H. J. Monkhorst, *Int. J. Quantum Chem.*, 1977, **12**, 421–432.

42 R. Krishnan, J. S. Binkley, R. Seeger and J. A. Pople, *J. Chem. Phys.*, 1980, **72**, 650.

43 A. D. McLean and G. S. Chandler, *J. Chem. Phys.*, 1980, **72**, 5639.

44 D. E. Woon and T. H. Dunning Jr., *J. Chem. Phys.*, 1995, **103**, 4572.

45 T. H. Dunning Jr., K. A. Peterson and A. K. Wilson, *J. Chem. Phys.*, 2001, **114**, 9244–9253.

46 A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen and A. K. Wilson, *Chem. Phys. Lett.*, 1998, **286**, 243–252.

47 D. G. Truhlar, *Chem. Phys. Lett.*, 1998, **294**, 45–48.

48 C. Møller and M. S. Plesset, *Phys. Rev.*, 1934, **46**, 618.

49 R. Jastrow, *Phys. Rev.*, 1955, **98**, 1479–1484.

50 N. D. Drummond, M. D. Towler and R. J. Needs, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2004, **70**, 235119.

51 N. D. Drummond and R. J. Needs, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2005, **72**, 085124.

52 T. Kato, *Commun. Pure Appl. Math.*, 1957, **10**, 151–177.

53 P. J. Reynolds and D. M. Ceperley, *J. Chem. Phys.*, 1982, **77**, 5593.

54 C. J. Umrigar, M. P. Nightingale and K. J. Runge, *J. Chem. Phys.*, 1993, **99**, 2865.

55 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09*, Gaussian Inc., Wallingford CT, 2009.

56 R. J. Needs, M. D. Towler, N. D. Drummond and P. L. Ríos, *J. Phys.: Condens. Matter*, 2010, **22**, 023201.

57 B. P. Pritchard, D. Altarawy, B. Didier, T. D. Gibson and T. L. Windus, *J. Chem. Inf. Model.*, 2019, **59**, 4814.

58 D. Feller, *J. Comput. Chem.*, 1996, **17**, 1571–1586.

59 K. L. Schuchardt, B. T. Didier, T. Elsethagen, L. Sun, V. Gurumoorthi, J. Chase, J. Li and T. L. Windus, *J. Chem. Inf. Model.*, 2007, **47**, 1045–1052.

60 R. Assaraf and M. Caffarel, *J. Chem. Phys.*, 2003, **119**, 10536–10552.

61 R. Assaraf and M. Caffarel, *J. Chem. Phys.*, 2000, **113**, 4028–4034.

62 P. L. Ríos and G. J. Conduit, *Phys. Rev. E*, 2019, **99**, 063312.

63 J. Tiihonen, R. C. Clay and J. T. Krogel, *J. Chem. Phys.*, 2021, **154**, 204111.

64 A. Badinski and R. J. Needs, *Phys. Rev. E*, 2007, **76**, 036707.

65 A. Badinski and R. J. Needs, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2008, **78**, 035134.

66 C. Filippi and C. J. Umrigar, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2000, **61**, R16291–R16294.

67 C. Attaccalite and S. Sorella, *Phys. Rev. Lett.*, 2008, **100**, 114501.

68 S. Sorella and L. Capriotti, *J. Chem. Phys.*, 2010, **133**, 234111.

69 C. Filippi, R. Assaraf and S. Moroni, *J. Chem. Phys.*, 2016, **144**, 194105.

70 R. Assaraf, S. Moroni and C. Filippi, *J. Chem. Theory Comput.*, 2017, **13**, 5273–5281.

71 M. R. Milovanović, J. M. Živković, D. B. Ninković, I. M. Stanković and S. D. Zarić, *Phys. Chem. Chem. Phys.*, 2020, **22**, 4138.

72 K. S. Pitzer, *Acc. Chem. Res.*, 1979, **12**, 271–276.

73 W. Liu, *Natl. Sci. Rev.*, 2016, **3**, 204–221.

74 S. Shambayati, S. L. Schreiber, J. F. Blake, S. G. Wierschke and W. L. Jorgensen, *J. Am. Chem. Soc.*, 1990, **112**, 697–703.

