

## EDF2: A Density Functional for Predicting Molecular Vibrational Frequencies

Ching Yeh Lin,<sup>A</sup> Michael W. George,<sup>A</sup> and Peter M. W. Gill<sup>A,B</sup>

<sup>A</sup> School of Chemistry, University of Nottingham, Nottingham NG7 2RD, UK.

<sup>B</sup> Author to whom correspondence should be addressed (e-mail: peter.gill@nott.ac.uk).

The majority of calculations of molecular vibrational spectra are based on the harmonic approximation but are compared (usually after empirical scaling) with experimental anharmonic frequencies. Any agreement that is observed in such cases must be attributable to fortuitous cancellation of errors and it would certainly be preferable to develop a more rigorous computational approach. In this paper, we introduce a new density functional model (EDF2) that is explicitly designed to yield accurate harmonic frequencies, and we present numerical results for a wide variety of molecules whose experimental harmonic frequencies are known. The EDF2 model is found to be significantly more accurate than other DFT models and competitive with the computationally expensive CCSD(T) method.

Manuscript received: 2 October 2003.

Final version: 3 February 2004.

Spectroscopy has had an enormous influence on the understanding of all aspects of chemistry. In particular, by providing direct information about the curvature of the potential energy surface (PES) near the minimum in which a molecule resides, vibrational spectroscopy provides a powerful probe of the structural and bonding characteristics of solids, liquids, and gases.

The infrared or Raman spectrum of a medium-size molecule contains a wealth of information but, regrettably, this treasure trove is normally regarded as a liability, not an asset. Normally, analysis focuses on a small subset of familiar bands (e.g. carbonyl stretches) and largely neglects the rest of the information. Why is such a profligate approach the established norm? Simply because it is difficult to assign the majority of the peaks in a vibrational spectrum unless one performs a sophisticated normal-mode analysis requiring the synthesis of several isotopomers. Sadly, the time required for such synthetic excursions is usually prohibitive.

There is, however, another way forward. As the speed of desktop computers continues to rise exponentially, it is becoming increasingly practical to apply quantum mechanics to make first-principles predictions of chemical behaviour, and it is now possible to calculate the vibrational spectra of twenty-atom molecules in a few hours using a standard PC and an efficient quantum chemistry package. Armed with this combination of hardware and software, it is straightforward to predict the infrared spectrum of a newly synthesized molecule and gain significant insights by comparing the result to an experimental measurement.

Performing theoretical calculations of vibrational spectra is much more cost-effective than the synthesis and

measurement of isotopic variants but, of course, there is a catch. Although quantum mechanics provides a formally exact theoretical basis for the whole of chemistry, all of the computationally tractable quantum chemical models involve approximations that introduce errors into the resulting predictions. Whether such errors are acceptably small depends on the model employed and the nature of the chemical question being asked. The most accurate models usually agree very well with experiment but, unfortunately, consume impractically large amounts of computer time. Less accurate, less time-consuming models, such as the various flavours of density functional theory (DFT), mirror reality rather less faithfully but have nonetheless found widespread use in all branches of chemistry.

The ideal theoretical procedure is to solve the vibrational Schrödinger equation directly in a fully coupled, anharmonic basis.<sup>[1]</sup> This can be done but the approach is limited by its complexity to small systems. Alternatively, one can first calculate normal-mode (harmonic) frequencies and then apply the necessary corrections for anharmonicity but, whereas the first of these steps is now routine, current approaches to the second step remain computationally demanding.<sup>[2]</sup> Much more commonly, harmonic calculations are performed and the resulting frequencies scaled by empirical scale factors<sup>[3]</sup> which account (it is hoped) for the neglect of anharmonicity. However, when this works, it obviously gives the right answers for the wrong reasons. Moreover, calculated harmonic frequencies are often found to deviate significantly from experimental harmonic frequencies, in cases where the latter are known, implying that the theoretical method is yielding a PES with the wrong curvature (namely, second derivatives) at the bottom of the well.

This led us to pursue a two-stage approach. First, build a DFT model that yields PESs with correct curvatures and, thus, correct harmonic frequencies. Second, develop computationally efficient schemes for the inclusion of anharmonic corrections. The present manuscript describes our progress on the first stage.

We began by assembling a large set of experimental harmonic frequency data. These are available for a wide variety of diatomics and a smaller number of larger systems. The diatomic data come primarily from Huber and Herzberg<sup>[4]</sup> but the polyatomic data that we have used are drawn from a variety of sources.<sup>[5–19]</sup> The complete set of 315 molecules is listed in Table 1, and they possess 612 distinct normal modes. As Fig. 1 indicates, the dataset covers approximately two-thirds of the elements.

We then calculated the harmonic vibrational frequencies of each molecule at a variety of levels of theory. These included four popular DFT methods:

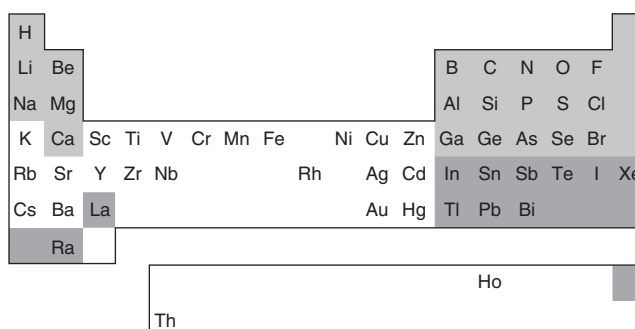
- (1) LSDA,<sup>[20,21]</sup> a combination of the Dirac exchange functional  $E_X^{D30}$  with the Vosko–Wilk–Nusair correlation functional  $E_C^{VWN}$ ;
- (2) BLYP,<sup>[22,23]</sup> a combination of the Becke exchange functional  $E_X^{B88}$  with the Lee–Yang–Parr correlation functional  $E_C^{LYP}$ ;
- (3) B3LYP,<sup>[24,25]</sup> the three-parameter hybrid functional of Stephens et al.;

- (4) EDF1,<sup>[26]</sup> an empirical functional that includes the modified Becke exchange functional  $E_X^{EDF1}$  and a modified Lee–Yang–Parr correlation functional  $E_C^{EDF1}$ ,

and three widely used wavefunction methods:

- (1) HF,<sup>[27,28]</sup> which uses the Fock exchange functional  $E_X^{F30}$ ;
- (2) MP2,<sup>[29]</sup> which uses second-order Møller–Plesset perturbation theory;
- (3) CCSD(T),<sup>[30]</sup> which uses coupled-cluster theory with single, double, and perturbative triple substitutions.

Because coupled-cluster calculations are computationally expensive for large systems, we split the full dataset into 296



**Fig. 1.** Elements in dataset (dark symbols) and the basis sets used (shading; cc-pVTZ light grey, CRENBL dark grey, SRSC white).

**Table 1.** The complete set of molecules used in this study

Diatomics <sup>A</sup>											
Ag <sub>2</sub>	Au <sub>2</sub>	BeH	Cl <sub>2</sub>	FO	InF	MnF	PbF	ScF	SrH	Triatomics	
AgAl	AuAl	BeI	ClF	GaBr	InH	MnH	PbH	ScO	SrI	CO <sub>2</sub> <sup>E</sup>	HCN <sup>B</sup>
AgBi	AuBa	BeO	ClO	GaCl	InI	MnI	PbI	ScS	SrO	CS <sub>2</sub> <sup>D</sup>	HNC <sup>F</sup>
AgBr	AuBe	BeS	CN	GaF	InO	MnO	PbO	SeBr	Te <sub>2</sub>	H <sub>2</sub> O <sup>B</sup>	HNO <sup>F</sup>
AgCl	AuCa	BF	CO	GaH	IO	MnS	PbS	SeO	TeBr	H <sub>2</sub> S <sup>B</sup>	SO <sub>2</sub> <sup>F</sup>
AgCu	AuCl	BH	CrH	GaI	K <sub>2</sub>	N <sub>2</sub>	PbTe	SeS	TeI	Tetraatomics	
AgF	AuCu	BiBr	CrO	GaO	KBr	Na <sub>2</sub>	PCl	SH	TeS	H <sub>2</sub> CO <sup>B</sup>	NH <sub>3</sub> <sup>F</sup>
AgGa	AuGa	BiCl	CrS	GeBr	KCl	NaBr	PF	Si <sub>2</sub>	TeSe	HCCF <sup>E</sup>	PH <sub>3</sub> <sup>F</sup>
AgH	AuGe	BiH	CS	GeCl	KF	NaCl	PH	SiCl	ThO	HCCH <sup>C</sup>	SO <sub>3</sub> <sup>F</sup>
AgI	AuH	BiI	Cs <sub>2</sub>	GeH	KH	NaF	PI	SiF	TiO	Pentatomics	
AgIn	AuMg	BN	CsBr	GeI	KI	NaH	PN	SiH	TiBr	CH <sub>2</sub> Cl <sub>2</sub> <sup>B</sup>	SiH <sub>4</sub> <sup>F</sup>
AgTe	AuSi	BO	CSe	GeO	LaO	NaI	PO	SiI	TiCl	CH <sub>3</sub> Cl <sup>B</sup>	CH <sub>3</sub> Br <sup>B,Q</sup>
Al <sub>2</sub>	AuSn	Br <sub>2</sub>	CsF	GeS	LaS	NaK	PS	SiN	TiF	CH <sub>3</sub> F <sup>F</sup>	CH <sub>3</sub> I <sup>B,Q</sup>
AlBr	AuSr	BrCl	CsH	GeSe	Li <sub>2</sub>	NaRb	RaCl	SiO	TiH	CH <sub>4</sub> <sup>F</sup>	HCOOH <sup>N,Q</sup>
AlCl	AuTe	BrF	CsI	GeTe	LiBr	NbO	Rb <sub>2</sub>	SiS	VO	Larger molecules <sup>Q</sup>	
AlF	B <sub>2</sub>	BS	Cu <sub>2</sub>	H <sub>2</sub>	LiCl	NBr	RbBr	SiSe	XeCl	CH <sub>2</sub> CH <sub>2</sub> <sup>B</sup>	SF <sub>6</sub> <sup>N</sup>
AlH	BaBr	C <sub>2</sub>	CuBr	HBr	LiF	NCl	RbCl	SiTe	YCl	CH <sub>3</sub> CH <sub>3</sub> <sup>B</sup>	C <sub>3</sub> H <sub>8</sub> <sup>B</sup>
AlI	BaCl	CaBr	CuCl	HCl	LiH	NF	RbH	SnBr	YF	C <sub>6</sub> H <sub>6</sub> <sup>G</sup>	C <sub>6</sub> D <sub>6</sub> <sup>H</sup>
AlO	BaF	CaCl	CuF	HF	LiI	NH	RbI	SnCl	YO	<sup>13</sup> C <sub>6</sub> H <sub>6</sub> <sup>I</sup>	<sup>13</sup> C <sub>6</sub> D <sub>6</sub> <sup>I</sup>
AlS	BaH	CaF	CuGa	HgCl	LiNa	NiH	RhC	SnF	ZnBr	<i>n</i> -butane <sup>O</sup>	3-hexyne <sup>P</sup>
AlSe	BaI	CaH	CuH	HgI	MgBr	NO	S <sub>2</sub>	SnI	ZnCl	<i>n</i> -pentane <sup>P</sup>	
As <sub>2</sub>	BaO	CaI	CuI	HI	MgCl	NS	Sb <sub>2</sub>	SnO	ZnF	Metal carbonyls <sup>Q</sup>	
AsCl	BaS	CaS	CuO	HoF	MgF	NSe	SbBi	SnS	ZnH	Ni(CO) <sub>2</sub> <sup>J</sup>	Fe(CO) <sub>5</sub> <sup>K</sup>
AsF	BBr	CCl	CuSe	I <sub>2</sub>	MgH	O <sub>2</sub>	SbBr	SnTe	ZnI	Ni(CO) <sub>3</sub> <sup>J</sup>	Cr(CO) <sub>6</sub> <sup>L</sup>
AsN	BCl	CdCl	CuTe	IBr	MgI	OH	SbCl	SO	ZrO	Ni(CO) <sub>4</sub> <sup>J</sup>	
AsO	BeBr	CdI	F <sub>2</sub>	ICl	MgO	P <sub>2</sub>	SbF	SrBr			
AsP	BeCl	CF	FeCl	IF	MgS	PbBr	SbP	SrCl			
AsS	BeF	CH	FeF	InCl	MnBr	PbCl	ScCl	SrF			

<sup>A</sup> Ref. [4]. <sup>B</sup> Ref. [5]. <sup>C</sup> Ref. [6]. <sup>D</sup> Ref. [7]. <sup>E</sup> Ref. [8]. <sup>F</sup> Ref. [9]. <sup>G</sup> Ref. [10]. <sup>H</sup> Ref. [11]. <sup>I</sup> Ref. [12]. <sup>J</sup> Ref. [13]. <sup>K</sup> Ref. [14]. <sup>L</sup> Ref. [15]. <sup>M</sup> Ref. [16]. <sup>N</sup> Ref. [17]. <sup>O</sup> Ref. [18]. <sup>P</sup> Ref. [19]. <sup>Q</sup> CCSD(T) calculations not performed.

‘small’ and 19 ‘large’ molecules (see footnote Q of Table 1) and did not attempt CCSD(T) calculations on any of the ‘large’ molecules.

All of the DFT, HF, and MP2 geometry optimization and frequency calculations were performed using *Q-Chem 2.0*.<sup>[31]</sup> Coupled-cluster calculations were carried out using *Molpro*.<sup>[32]</sup> The Dunning cc-pVTZ basis set<sup>[33]</sup> was used for light atoms, and CRENBL<sup>[34]</sup> and SRSC<sup>[35]</sup> pseudopotential basis sets for heavy atoms (Fig. 1). In the DFT calculations, we employed an Euler–Maclaurin–Lebedev quadrature scheme<sup>[36–38]</sup> to integrate the exchange–correlation functional and used 100 radial and 194 angular grid points on each atom to ensure that low-frequency modes are treated satisfactorily.<sup>[39]</sup> Every frequency calculation was preceded by an optimization—it is not physically meaningful to discuss vibrations about non-stationary points on a PES.

Following the construction philosophy of the EDF1 functional<sup>[26]</sup> (the acronym stands for ‘Empirical Density Functional 1’), we sought to develop a new DFT functional by linearly combining several existing functionals. However, whereas EDF1 was optimized to yield accurate *thermochemistry* when used with the 6–31+G\* basis set, our new target (EDF2) was optimized to give accurate *harmonic frequencies* when used with the basis sets indicated in Fig. 1.

To begin, we wrote an initial approximation to the new functional as Equation (1)

$$E^{\text{EDF2}} = a_1 E^{\text{HF}} + a_2 E^{\text{LSDA}} + a_3 E^{\text{BLYP}} + a_4 E^{\text{B3LYP}} + a_5 E^{\text{EDF1}} \quad (1)$$

and determined the five unknown  $a_i$  coefficients by a simple least-squares fit of the HF, LSDA, BLYP, B3LYP, and EDF1 harmonic frequencies to the experimental ones. This strategy was termed ‘external optimization’ by Adamson et al.<sup>[26]</sup> Then, since each of the five functionals is itself a sum of simpler components, we could write the exchange–correlation part of the new functional as Equation (2)

$$E_{\text{XC}}^{\text{EDF2}} = b_1 E_{\text{X}}^{\text{F30}} + b_2 E_{\text{X}}^{\text{D30}} + b_3 E_{\text{X}}^{\text{B88}} + b_4 E_{\text{X}}^{\text{EDF1}} + b_5 E_{\text{C}}^{\text{VWN}} + b_6 E_{\text{C}}^{\text{LYP}} + b_7 E_{\text{C}}^{\text{EDF1}} \quad (2)$$

Since external optimization leads to a linear combination of model chemistries, not a linear combination of functionals, the  $b_i$  coefficients at this stage were not optimal. Therefore, in a second step, we performed an internal optimization of the coefficients using the downhill simplex minimization algorithm.<sup>[40]</sup> On each iteration of this (each time the coefficients were refined), we re-optimized the geometries of all of the molecules in Table 1, re-computed their harmonic frequencies at these geometries, and determined the resulting root-mean-square (RMS) error between these and the experimental values. After a few months of this, the optimal coefficients were found to be  $b_1 = 0.1695$ ,  $b_2 = 0.2811$ ,  $b_3 = 0.6227$ ,  $b_4 = -0.0551$ ,  $b_5 = 0.3029$ ,  $b_6 = 0.5998$ , and  $b_7 = -0.0053$ .

There are several different ways in which the accuracies of different functionals can be compared and Tables 2–4 and Fig. 2 provide different perspectives on this. Table 2 shows the

**Table 2. Errors [ $\text{cm}^{-1}$ ] in calculated harmonic frequencies for some small molecules**

Molecule	Expt	BLYP	EDF1	B3LYP	EDF2	HF	MP2	CCSD(T)
H <sub>2</sub>	4401	-54	-26	19	2	186	125	8
HF	4138	-206	-120	-48	-33	344	57	39
CO <sub>2</sub>	2397	-69	-18	19	42	167	42	-2
	1354	-48	-25	18	26	157	-11	-8
	673	-40	-30	-2	-1	100	-12	-13
H <sub>2</sub> O	3943	-188	-107	-42	-24	284	50	2
	3832	-176	-98	-31	-15	295	40	8
	1649	-38	-25	-10	-18	104	2	19
SO <sub>2</sub>	1381	-135	-91	-30	-20	195	-45	-28
	1167	-98	-60	-2	6	212	-47	-19
	526	-45	-33	-9	-8	82	-22	-13
NH <sub>3</sub>	3577	-118	-53	1	20	230	99	20
	3506	-158	-102	-44	-30	179	31	-35
	1691	-46	-38	-13	-20	104	1	-3
	1022	46	54	46	31	105	55	87
SO <sub>3</sub>	1409	-131	-80	-34	-19	168	6	-14
	1048	-82	-42	3	13	193	6	9
	539	-60	-48	-26	-24	60	-21	-21
	504	-61	-45	-24	-21	79	-14	-17
CH <sub>2</sub> Cl <sub>2</sub>	3182	-58	-30	17	18	182	56	24
	3123	-75	-47	-1	1	162	54	7
	1464	-46	-38	-2	-7	131	31	10
	1295	-53	-40	-8	-12	119	14	-2
	1177	-51	-33	-7	-8	106	26	3
	917	-46	-36	-17	-17	64	2	-11
	771	-109	-72	-53	-41	55	33	3
	724	-70	-34	-24	-15	40	22	-4
	284	-19	-12	-6	-4	20	7	-2
CH <sub>3</sub> Cl	3166	-74	-40	1	7	149	62	15
	3074	-73	-45	-2	2	143	60	6
	1482	-36	-32	1	-5	120	34	12
	1383	-49	-38	-6	-10	119	22	1
	1038	-48	-35	-14	-15	77	10	-6
	740	-74	-34	-31	-17	29	37	2
CH <sub>4</sub>	3158	-97	-48	-24	-11	88	60	-5
	3026	-63	-32	3	9	122	-17	8
	1567	-36	-32	-3	-9	99	33	4
	1357	-43	-46	-13	-21	97	-174	-14

**Table 3. RMS errors [ $\text{cm}^{-1}$ ] of the harmonic frequencies of small and large molecules**

Molecules	BLYP	EDF1	B3LYP	EDF2	HF	MP2	CCSD(T)
‘Small’	65	53	40	34	108	67	35
‘Large’	67	54	34	33	133	55	–
All	66	54	38	34	117	64	35

**Table 4. RMS errors [%] of the harmonic frequencies of small and large molecules**

Molecules	BLYP	EDF1	B3LYP	EDF2	HF	MP2	CCSD(T)
‘Small’	7.0	5.8	4.5	3.9	9.8	6.4	4.1
‘Large’	4.7	4.4	2.4	2.3	9.0	3.6	–
All	6.3	5.4	3.7	3.3	9.5	5.6	4.1

errors in the harmonic frequencies of ten familiar molecules, as calculated by four DFT methods and three wavefunction methods. The HF method is clearly the least accurate, followed by BLYP and then EDF1. The most accurate methods are B3LYP, EDF2, and CCSD(T). Our results reveal that the poor accuracy of B3LYP observed in previous studies<sup>[41,42]</sup> for the H<sub>2</sub> and HF molecules is largely an artefact of the neglect of anharmonicity: the *harmonic* frequencies from B3LYP agree well with the *harmonic* experimental results. Moreover, the new EDF2 functional is even better in this respect.

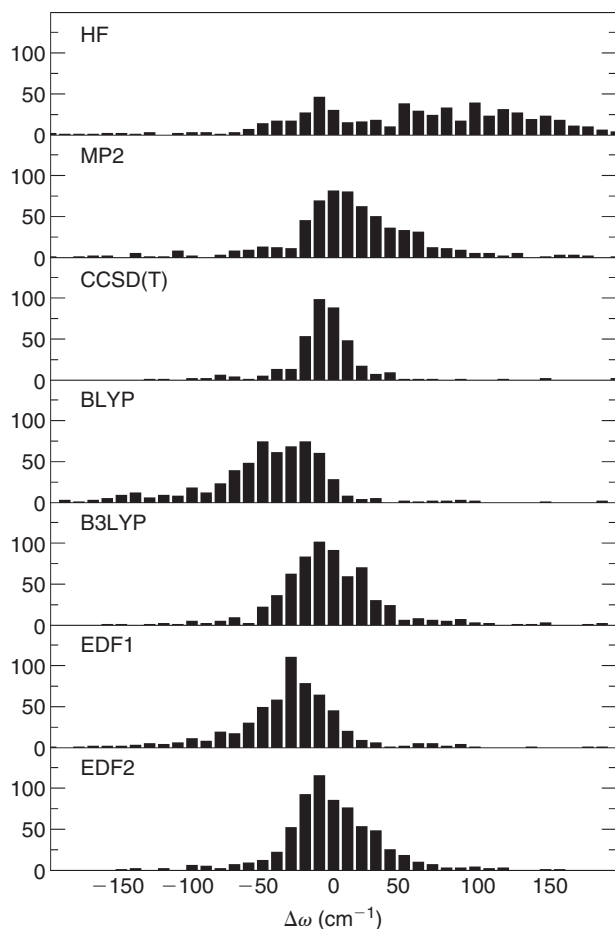


Fig. 2. Distributions of error  $\Delta\omega = \omega_{\text{calc}} - \omega_{\text{exp}}$  [ $\text{cm}^{-1}$ ] in calculated harmonic frequencies.

The histograms in Fig. 2 show the distribution of errors for the seven methods and normal modes of the full set of molecules. (As noted above, CCSD(T) was applied only to the small molecules.) It is well known that HF usually overestimates harmonic frequencies, but the tendency of BLYP and EDF1 to underestimate is less well documented. Bearing in mind that anharmonic corrections are often negative, this helps to explain the curious observation that unscaled BLYP and EDF1 harmonic frequencies often agree well with experimental anharmonic frequencies.<sup>[3,43,44]</sup> The most accurate predictions come from B3LYP, CCSD(T), and EDF2, for which errors greater than  $100 \text{ cm}^{-1}$  are rare.

Table 3 shows the RMS errors of the harmonic frequencies predicted by the seven methods. Since CCSD(T) was not applied to the ‘large’ molecules, the results for the ‘small’ molecules, the ‘large’ molecules, and the full dataset are presented separately. For the large molecules, which are predominantly organic, the new EDF2 functional is comparable to B3LYP. For the small molecules, most of which are inorganic, EDF2 is superior to B3LYP and, remarkably, appears comparable to the much more expensive CCSD(T) method.

Because the harmonic frequencies span two orders of magnitude (from 26 to  $4401 \text{ cm}^{-1}$ ), it is more appropriate to discuss relative, rather than absolute, errors. Table 4 gives the RMS percentage errors of the predicted frequencies and

Table 5. Errors [ $\text{cm}^{-1}$ ] in harmonic frequencies using EDF2 with various basis sets

Molecule	Expt	6-31+G*	6-311G**	cc-pVDZ	cc-pVTZ	cc-pVQZ <sup>A</sup>
B <sub>2</sub>	1051	-31	-33	-35	-31	-32
BF	1402	-28	-11	-50	19	15
BH	2367	-5	-41	-48	-24	-21
BN	1515	61	70	47	65	68
BO	1886	35	50	22	43	46
C <sub>2</sub>	1855	11	36	37	40	38
CF	1308	-18	-34	-1	14	4
CH	2859	-22	-52	-87	-38	-31
CN	2069	97	98	93	100	100
CO	2170	43	61	41	53	54
F <sub>2</sub>	917	133	84	105	149	143
FO	1029	94	92	88	108	103
H <sub>2</sub>	4401	34	2	-45	2	-2
HF	4138	-187	-5	-97	-33	-36
I <sub>2</sub>	351	-11	-9	-10	-10	-8
LiF	910	-15	58	84	13	15
LiH	1406	-10	5	-20	9	7
N <sub>2</sub>	2359	105	98	104	101	98
NF	1141	18	34	39	33	27
NH	3282	-21	-42	-102	-27	-21
NO	1904	90	98	103	87	84
O <sub>2</sub>	1580	56	76	85	64	70
OH	3738	-84	-23	-100	-29	-26
MAE		53	48	63	48	46

<sup>A</sup> All *g*-functions were removed.

confirms that the EDF2 functional represents a substantial improvement over B3LYP across the board. For the inorganic molecules, EDF2 is slightly more accurate than even CCSD(T). Behind these, in order of decreasing accuracy, lie EDF1 and MP2 (which appear broadly comparable), BLYP, and HF.

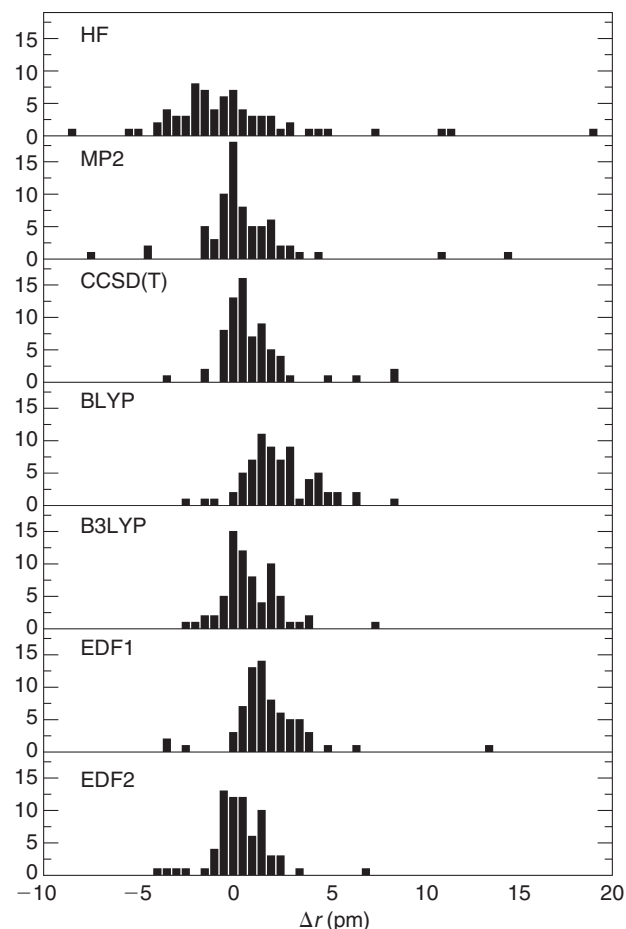
Does EDF2 yield equally satisfactory harmonic frequencies when combined with other basis sets? To test this, we calculated the EDF2 frequencies of a selection of first-row diatomics using the 6-31+G\*, 6-311G\*\*, cc-pVDZ, cc-pVTZ, and cc-pVQZ basis sets and these are shown in Table 5. We see that, though individual frequencies vary considerably as the basis is changed, the mean absolute errors (MAE) are surprisingly constant. The double-zeta Dunning basis (cc-pVDZ) gives the largest mean error ( $63 \text{ cm}^{-1}$ ) but the errors of the four other bases all lie between 46 and  $53 \text{ cm}^{-1}$ . The most difficult molecule, across the board, is F<sub>2</sub>. Judging by this small set of molecules, the EDF2/6-31+G\* model seems to offer a promising route to fairly accurate harmonic vibrational frequencies at a modest computational cost.

The EDF2 functional was explicitly constructed to yield PES curvatures, and it is obviously important to enquire whether our demand for accuracy in the associated second derivatives has entailed serious sacrifices in the quality of the first derivatives (which affect structure) and/or the energy itself (which affects thermochemistry). We have addressed both of these questions.

To assess the accuracy of EDF2 for structural prediction, we have examined the diatomics in our set for which (a) the cc-pVTZ basis set is available and (b) the bond length is known experimentally. The results are shown numerically in Table 6 and graphically in Fig. 3 and reveal that, for bond length predictions, EDF2 is more accurate than any of the

**Table 6. Errors [pm] in calculated bond lengths using the cc-pVTZ basis set**

Molecule	Expt	BLYP	EDF1	B3LYP	EDF2	HF	MP2	CCSD(T)
AlF	165.4	4.6	4.1	2.8	2.2	0.0	2.4	2.1
AlH	164.8	3.1	2.9	1.8	1.6	0.5	0.1	0.7
AlO	161.8	3.0	2.1	1.7	0.9	7.5	1.0	2.0
BeCl	179.7	1.9	1.5	1.0	0.4	1.7	0.2	0.3
BeF	136.1	1.7	1.6	0.6	0.1	-0.5	0.9	0.6
BeH	134.3	0.5	1.0	0.0	-0.1	0.1	-1.0	-0.4
BeO	133.1	0.9	0.6	-0.8	-1.1	-3.7	1.8	0.9
BeS	174.2	1.5	0.9	0.0	-0.5	-1.3	0.7	0.8
B <sub>2</sub>	159.0	2.8	2.7	2.2	1.6	4.8	-0.2	-0.3
BCl	171.6	2.4	2.0	0.9	0.2	0.7	-0.7	0.5
BF	126.3	1.4	1.3	0.2	-0.3	-1.4	0.0	0.3
BH	123.2	0.9	1.4	0.1	0.1	-1.0	-1.3	-0.6
BN	128.1	5.4	5.2	3.8	3.4	0.9	3.0	4.8
BO	120.5	1.1	0.9	-0.1	-0.5	-2.2	0.3	0.4
C <sub>2</sub>	124.3	1.3	1.0	0.5	0.1	11.3	14.3	0.3
CaF	196.7	-1.1	-2.4	-1.5	-2.8	2.2	0.1	2.0
CaH	200.3	-2.3	-3.6	-2.2	-3.3	0.2	0.0	2.3
CCl	164.5	4.1	2.5	2.2	1.3	2.2	-0.3	1.4
CF	127.2	2.2	1.5	0.4	-0.2	-1.8	-0.3	0.2
CH	112.0	1.4	1.4	0.4	0.3	-1.4	-1.5	-0.6
Cl <sub>2</sub>	198.8	6.6	3.1	3.5	2.4	-0.4	0.5	2.5
ClF	162.8	5.2	2.8	1.9	1.0	-3.8	0.6	1.4
ClO	157.0	4.9	2.5	2.4	1.3	2.5	-0.2	2.3
CN	117.2	0.3	0.1	-0.9	-1.2	-2.1	-4.7	6.3
CO	112.8	1.0	0.7	-0.2	-0.5	-2.4	0.7	0.4
CS	153.5	1.9	1.3	0.2	-0.2	-1.9	0.4	1.0
CSe	167.6	2.4	1.4	0.4	-0.2	-2.1	0.0	0.7
F <sub>2</sub>	141.2	2.0	0.0	-1.5	-2.3	-8.3	-1.6	0.2
FO	132.6	5.4	3.4	2.4	1.6	-1.3	0.3	2.7
GaCl	220.2	6.3	4.2	4.0	2.6	3.1	0.0	-1.4
GaF	177.4	4.4	3.3	2.2	1.3	-0.8	0.0	-0.3
GaH	166.3	4.0	3.2	2.3	1.8	1.0	0.0	-1.6
GaO	174.4	-1.7	-3.4	-2.4	-3.8	4.3	0.0	-3.6
GeH	158.8	1.9	1.9	1.9	1.3	1.9	1.9	-0.7
GeO	162.5	2.8	1.7	0.5	0.0	-3.4	0.0	0.7
GeS	201.2	4.2	2.5	1.8	1.0	-1.3	0.0	1.0
H <sub>2</sub>	74.1	0.5	0.4	0.1	0.3	-0.7	-0.4	0.1
HBr	141.4	2.2	1.3	1.1	0.8	-0.6	0.0	-0.5
HCl	127.5	1.9	1.1	0.9	0.7	-0.7	-0.3	0.1
HF	91.7	1.6	0.9	0.5	0.4	-1.9	0.0	-0.1
Li <sub>2</sub>	267.3	3.3	6.4	2.5	2.5	11.0	4.5	-0.6
LiCl	202.1	1.2	1.0	0.3	-0.4	1.7	0.8	0.7
LiF	156.4	1.7	1.8	0.6	0.1	0.0	1.6	1.2
LiH	159.6	0.1	0.6	-0.5	-0.6	1.2	-0.4	0.0
LiNa	281	8.4	13.5	7.4	6.9	18.8	11.0	8.6
MgF	175.0	3.8	3.4	2.0	1.4	-0.2	1.7	1.3
MgH	173.0	2.8	2.6	1.4	1.0	0.7	-0.7	0.3
MgO	174.9	1.1	0.4	-0.6	-1.3	-2.1	-1.3	0.5
N <sub>2</sub>	109.8	0.6	0.3	-0.6	-0.9	-3.1	1.2	0.3
NaF	192.6	3.0	3.3	1.3	0.5	0.4	2.3	1.8
NaH	188.7	0.6	1.8	-0.2	-0.7	3.2	1.9	2.8
NF	131.7	2.6	1.1	0.1	-0.6	-2.5	-0.9	-0.1
NH	103.6	1.6	1.2	0.5	0.4	-1.5	-0.8	0.0
NO	115.1	1.2	0.6	-0.5	-0.8	-3.5	-1.5	0.2
NS	149.4	2.7	1.8	0.7	0.2	4.2	-7.4	1.4
O <sub>2</sub>	120.8	2.4	1.1	-0.2	-0.7	-4.9	1.3	0.1
OH	97.0	1.6	1.0	0.5	0.4	-1.9	-0.4	0.0
P <sub>2</sub>	189.3	2.5	1.3	0.2	-0.3	-3.6	2.8	1.9
PF	159.0	4.5	3.3	2.2	1.5	-1.3	1.2	1.3
PH	142.2	2.0	1.5	0.9	0.7	-0.8	-0.5	0.3
PN	149.1	1.5	1.0	-0.3	-0.6	-3.8	3.3	1.2
PO	147.6	3.1	2.4	1.1	0.7	-2.9	1.9	1.6
SeO	164.8	3.0	1.3	0.0	-0.6	-5.5	0.0	1.2
SH	134.1	1.9	1.3	0.9	0.7	-0.8	-0.5	0.1
SiF	160.1	4.7	3.9	2.7	2.1	-0.5	1.9	1.7
SiH	152.0	2.5	2.2	1.3	1.1	-0.2	1.7	0.6
SiN	157.2	1.7	1.0	0.1	-0.3	1.4	-4.3	8.3
SiO	151.0	2.8	2.2	0.9	0.5	-2.5	2.2	1.4
SO	148.1	4.4	3.2	1.9	1.4	-3.0	1.7	1.7
MAE		2.6	2.1	1.3	1.1	2.6	1.5	1.3

**Fig. 3.** Distributions of error  $\Delta r = r_{\text{calc}} - r_{\text{exp}}$  [pm] in calculated bond lengths.

other methods tested, eclipsing both B3LYP and CCSD(T). Its worst failure (6.9 pm) occurs for the weakly bound LiNa dimer but its mean absolute error for the 69 molecules tested is only 1.1 pm.

To assess the accuracy of EDF2 for thermochemical prediction, we have applied it to the G2 dataset of Pople et al.<sup>[45]</sup> which includes 56 atomization energies, 40 ionization energies, 25 electron affinities, and 8 proton affinities. RMS errors for the four DFT methods are compared with HF, MP2, and CCSD(T) in Tables 7 and 8.

Table 7 shows thermochemical results obtained using the 6-31+G\* basis set, the set for which the original EDF1 functional was designed to be optimal. It is not surprising to find that, when this modest basis is used, EDF1 is the most accurate of the functionals tested. It is surprising, however, to discover that CCSD(T) is no better than MP2 (because of the small basis set) and EDF2 is comparable to B3LYP.

Table 8 shows thermochemical results obtained using the aug-cc-pVTZ basis set, which is much larger than 6-31+G\*. Overall, we find that EDF2 is the most accurate of the four DFT methods tested, bettering even the venerable B3LYP method in all tests except proton affinities. Of course, none of the DFT methods can approach the thermochemical accuracy of CCSD(T) with a large basis set, a combination that Dunning has called ‘the gold standard of quantum chemistry’.

**Table 7. RMS errors [kcal mol<sup>-1</sup>] of the thermochemical properties of the molecules in the G2 dataset using the 6-31+G\* basis set**

	BLYP	EDF1	B3LYP	EDF2	HF	MP2	CCSD(T)
Atomization energies	5.75	4.41	8.10	6.17	91.25	27.60	28.89
Ionization potentials	5.14	4.34	5.05	4.77	24.65	9.60	9.71
Electron affinities	4.39	3.79	4.40	4.38	29.64	11.47	11.44
Proton affinities	5.23	3.77	5.51	6.38	7.20	6.02	5.16

**Table 8. RMS errors [kcal mol<sup>-1</sup>] of the thermochemical properties of the molecules in the G2 dataset using the aug-cc-pVTZ basis set<sup>A</sup>**

	BLYP	EDF1	B3LYP	EDF2	HF	MP2	CCSD(T)
Atomization energies	7.67	7.50	6.12	5.94	86.34	6.33	3.84
Ionization potentials	5.14	4.47	5.28	4.96	25.41	5.07	2.35
Electron affinities	3.81	2.80	3.80	2.97	29.70	4.98	2.21
Proton affinities	2.50	1.63	1.91	2.03	3.96	1.25	1.20

<sup>A</sup> Except for Li, Be, Na, and Mg for which cc-pVTZ was used.

The EDF2 functional was constructed to yield accurate harmonic vibrational frequencies but it appears also to afford accurate structures and thermochemistry. When used with a large basis set it yields harmonic frequencies that are comparable to those of CCSD(T) and thermochemical predictions that are generally superior to those of B3LYP, at least for the G2 molecular dataset. We are encouraged by these results and look forward to applying EDF2 to a range of chemical problems.

## Acknowledgements

We thank the referees for several very helpful comments. C.Y.L. gratefully acknowledges an ORS Award from the DfES. M.W.G. thanks the Royal Society of Chemistry for the support of the Sir Edward Frankland Fellowship. P.M.W.G. thanks the organizers for the opportunity to present this work at the 19th RACI Organic Conference in Lorne, Australia. This research was supported by a JREI computer hardware grant (GR/R62052) from the Engineering and Physical Sciences Research Council.

## References

- [1] S. Carter, J. M. Bowman, *J. Chem. Phys.* **1998**, *108*, 4397. doi:10.1063/1.475852
- [2] J. M. Bowman, *Acc. Chem. Res.* **1986**, *19*, 202.
- [3] A. P. Scott, L. Radom, *J. Phys. Chem.* **1996**, *100*, 16502. doi:10.1021/JP960976R
- [4] K. P. Huber, G. Herzberg, *Molecular Spectra and Molecular Structure* **1979** (Van Nostrand Reinhold: New York, NY).
- [5] K. J. Miller, F. S. Ganda-Kesuma, *J. Mol. Spec.* **1991**, *145*, 429. doi:10.1016/0022-2852(91)90130-3
- [6] B. G. Johnson, P. M. W. Gill, J. A. Pople, *J. Chem. Phys.* **1993**, *98*, 5612. doi:10.1063/1.464906
- [7] N. J. Harris, *J. Phys. Chem.* **1995**, *99*, 14689.
- [8] J. R. Thomas, B. J. DeLeeuw, G. Vacek, T. D. Crawford, Y. Yamaguchi, H. F. Schaefer III, *J. Chem. Phys.* **1993**, *99*, 403. doi:10.1063/1.465764
- [9] W. J. Hehre, L. Radom, P. von R. Schleyer, J. A. Pople, *Ab Initio Molecular Orbital Theory* **1986** (Wiley: New York, NY).
- [10] A. Miani, E. Cane, P. Palmieri, A. Trombetti, N. C. Handy, *J. Chem. Phys.* **2000**, *112*, 248. doi:10.1063/1.480577
- [11] N. C. Handy, P. E. Maslen, R. D. Amos, J. S. Andrews, C. W. Murray, G. J. Laming, *Chem. Phys. Lett.* **1992**, *197*, 506. doi:10.1016/0009-2614(92)85808-N
- [12] L. Goodman, A. G. Ozkabak, S. N. Thakur, *J. Phys. Chem.* **1991**, *95*, 9044.
- [13] L. Manceron, M. E. Alikhani, *Chem. Phys.* **1999**, *244*, 215. doi:10.1016/S0301-0104(99)00110-X
- [14] R. S. Weiner, *Ph.D. Thesis* **1970** (Purdue University: West Lafayette, IN).
- [15] L. Jones, R. S. McDowell, M. Goldblatt, *Inorg. Chem.* **1969**, *8*, 2349.
- [16] M. Freytes, D. Hurtmans, S. Kassi, J. Liévin, J. Vander Auwera, A. Campargue, M. Herman, *Chem. Phys.* **2002**, *283*, 47. doi:10.1016/S0301-0104(02)00507-4
- [17] R. S. McDowell, J. P. Aldridge, R. F. Holland, *J. Phys. Chem.* **1976**, *80*, 1203.
- [18] R. M. Levy, O. d. l. L. Rojas, R. A. Friesner, *J. Phys. Chem.* **1984**, *88*, 4233.
- [19] B. R. Henry, D. M. Turnbull, D. P. Schofield, H. G. Kjaergaard, *J. Phys. Chem. A* **2003**, *107*, 3236. doi:10.1021/JP021650N
- [20] P. A. M. Dirac, *Proc. Camb. Phil. Soc.* **1930**, *26*, 376.
- [21] S. J. Vosko, L. Wilk, M. Nusair, *Can. J. Phys.* **1980**, *58*, 1200.
- [22] A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098. doi:10.1103/PHYSREVA.38.3098
- [23] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785. doi:10.1103/PHYSREVB.37.785
- [24] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648. doi:10.1063/1.464913
- [25] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, *J. Phys. Chem.* **1994**, *98*, 11623.
- [26] R. D. Adamson, P. M. W. Gill, J. A. Pople, *Chem. Phys. Lett.* **1998**, *284*, 6. doi:10.1016/S0009-2614(97)01282-7
- [27] D. R. Hartree, *Proc. Camb. Phil. Soc.* **1928**, *24*, 89.
- [28] V. Fock, *Z. Physik* **1930**, *61*, 126.
- [29] C. Möller, M. S. Plesset, *Phys. Rev.* **1934**, *46*, 618. doi:10.1103/PHYSREV.46.618
- [30] J. Cizek, J. Paldus, *Phys. Scripta* **1980**, *21*, 251.
- [31] J. Kong, C. A. White, A. I. Krylov, D. Sherrill, R. D. Adamson, T. R. Furlani, M. S. Lee, A. M. Lee, et al., *J. Comput. Chem.* **2000**, *21*, 1532. doi:10.1002/1096-987X(200012)21:16<1532::AID-JCC10>3.0.CO;2-W
- [32] H.-J. Werner, P. J. Knowles, *Users' Manual for MOLPRO, ver. 2000.1* **2000** (University of Birmingham: Birmingham).
- [33] D. E. Woon, T. H. Dunning, *J. Chem. Phys.* **1993**, *98*, 1358. doi:10.1063/1.464303
- [34] M. M. Hurley, L. Fernandez Pacios, P. A. Christiansen, R. B. Ross, W. C. Ermler, *J. Chem. Phys.* **1986**, *84*, 6840. doi:10.1063/1.450689
- [35] M. Dolg, U. Wedig, H. Stoll, H. Preuss, *J. Chem. Phys.* **1987**, *86*, 866. doi:10.1063/1.452288
- [36] V. I. Lebedev, *Sibirsk. Mat. Zh.* **1977**, *18*, 132.
- [37] C. W. Murray, N. C. Handy, G. J. Laming, *Mol. Phys.* **1993**, *78*, 997.
- [38] P. M. W. Gill, B. G. Johnson, J. A. Pople, *Chem. Phys. Lett.* **1993**, *209*, 506. doi:10.1016/0009-2614(93)80125-9
- [39] J. M. L. Martin, C. W. Bauschlicher, A. Ricca, *Comput. Phys. Commun.* **2001**, *133*, 189. doi:10.1016/S0010-4655(00)00174-0
- [40] J. A. Nelder, R. Mead, *Comput. J.* **1965**, *7*, 308.
- [41] M. W. Wong, *Chem. Phys. Lett.* **1996**, *256*, 391. doi:10.1016/0009-2614(96)00483-6
- [42] I. Bytheway, M. W. Wong, *Chem. Phys. Lett.* **1998**, *282*, 219. doi:10.1016/S0009-2614(97)01281-5
- [43] B. G. Johnson, P. M. W. Gill, J. A. Pople, *J. Chem. Phys.* **1993**, *98*, 5612. doi:10.1063/1.464906
- [44] T. M. Watson, J. D. Hirst, *J. Phys. Chem. A* **2002**, *106*, 7858. doi:10.1021/JP025551L
- [45] L. A. Curtiss, K. Raghavachari, G. W. Trucks, J. A. Pople, *J. Chem. Phys.* **1991**, *94*, 7221. doi:10.1063/1.460205