Check for updates

A new double-reference correction scheme for accurate and efficient computation of NMR chemical shieldings†

Deborah L. Crittenden [iD] *

Using an extensive set of benchmark geometric data and absolute NMR chemical shieldings for gas phase molecules compiled from the literature, we introduce and test a novel shift-and-scale correction scheme for bringing computed values into closer agreement with experiment. Our approach requires only one additional reference calculation relative to existing methods for computing chemical shifts, and all reference species are small molecules whose absolute nuclear shieldings can be easily obtained, both computationally and experimentally. We demonstrate that our approach is capable of correcting for errors due to choice of electronic structure method and basis set, vibrational averaging effects, and scalar relativistic effects, but cannot account for the influence that heavy atoms have on the chemical shieldings of neighbouring light atoms, *via* spin–orbit coupling. We particularly recommend using our approach in conjunction with nuclear shieldings computed at DFT optimized geometries, employing hybrid functionals with moderate-to-high quality atomic orbital basis sets, *e.g.* pc-2.

1 Introduction

Nuclear magnetic resonance spectroscopy is a widely used technique for chemical characterization and molecular structure determination. It is particularly useful for determining the molecular connectivity or two-dimensional structure of unknown or novel substances because it provides a wealth of detailed information on chemical composition (how many atoms of different types), molecular topography (which atoms neighbor one another) and atomic environment (whether atoms exist in electron-rich or electron-deficient environments with an molecule). Two-dimensional NMR spectroscopy can even give information about interatomic distances that can be used to help guide three-dimensional molecular and biomolecular structure determination.

Given the utility and widespread use of NMR spectroscopy in determining the identity of small molecules, it is surprising that spectral assignment remains by and large a manual process. This can be traced back to the fact that simple heuristic models are very effective,[1] and predicting NMR parameters to high enough accuracy for qualitative assignment of experimental spectra is computationally challenging.[2,3] Spin–spin coupling constants can predicted to sufficient accuracy (<10% relative error) using computationally affordable density

School of Physical and Chemical Sciences, University of Canterbury, Christchurch, 8140, New Zealand. E-mail: deborah.crittenden@canterbury.ac.nz
† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d2cp03992c

functional theory models with modestly sized atomic orbital basis sets,[4,5] but chemical shieldings are more challenging.[2,6]

For example, previous benchmarking studies[6] have shown that in order to obtain absolute chemical shifts for gas phase molecules to within 5 ppm (mean absolute deviation) or ∼15% (mean relative deviation) of experimental values (across 28 molecules with measured shieldings primarily available for ^{1}H, ^{13}C, ^{17}O, ^{19}F nuclei), it is necessary to employ highly correlated wavefunction models – CCSD(T) – in conjunction with large and flexible segmented-valence atomic orbital basis sets such as Dunning's correlation-consistent polarizable core–valence basis sets of at least triple zeta quality, *i.e.* cc-pCVXZ, where X = T or Q. It is also necessary to account for vibrational-averaging effects, which would otherwise contribute additional errors of ∼5 ppm or 25%. By way of comparison, DFT models predictions typically incurred average absolute errors around 20 ppm and average absolute relative errors of ∼50%.

Similarly, relativistic effects cannot be ignored,[7] even for hydrogen atoms, whose chemical shifts are influenced by neighbouring heavy atoms through spin–orbit coupling.[8] For example, relativistic corrections to ^{1}H NMR chemical shieldings range from 0.15 ppm for the hydrogen nucleus in hydrogen fluoride to 18.40 ppm for the hydrogen nucleus in hydrogen iodide.[9] For heavy atoms, scalar relativistic effects dominate and are even larger in magnitude.[9] For halogen nuclei within halogen hydrides, total relativistic corrections range from 8.9 ppm (HF) to 2060.8 ppm (HI).[9]

Of course, predicting absolute chemical shieldings is a much harder task than predicting relative chemical shifts,

which are generally the more experimentally relevant quantity. In this work, we take this idea one step further and posit that by uniformly shifting and scaling computational absolute shielding values obtained using computationally inexpensive electronic structure methods to exactly reproduce experimental values for two carefully chosen reference molecules, it may be possible to account not only for methodological errors that are constant across all molecules but also methodological errors that vary systematically between molecules.

Although this sounds simple in theory, in practice it is complicated by the vast array of computationally-affordable electronic structure models that could be tried (the density functional theory "zoo"[10]) and multiple potential sources of error:

• Electronic: incomplete, approximate or absent treatment of electron correlation, basis set incompleteness.

• Geometric: resultant errors in equilibrium geometries, vibrational-averaging effects.

• Relativistic: spin–orbit and scalar relativistic effects that cannot be fully accounted for outside a full four-component relativistic Dirac equation framework.

Therefore, it is necessary to be as systematic and selective as possible. "Computational error bars" for DFT predictions may be established using Hartree–Fock theory and Hartree–Fock–Slater density functional theory.[11] Despite both lacking any attempt at modelling electron correlation, these simple methods nonetheless exhibit systematic electron under- and over-localisation errors, respectively, and have proven effective as "bounding" functionals for molecular property estimation.[11]

An alternative strategy is to select density functionals from the literature that have been shown to be particularly accurate for predicting NMR chemical shieldings, such as the modified B3LYP variant with 5% HF exchange recommended by Handy *et al.*[12,13] Alternatively, Keal and Tozer have developed density functionals designed specifically for predicting NMR shifts and shieldings,[14,15] of which the KT3 functional has also been optimized for other molecular properties including atomization energies, ionization potentials, electron affinities, proton affinities, geometric parameters, vibrational frequencies and electronic polarizabilities.[15]

When it comes to selecting basis sets for NMR shielding calculations, it is necessary to choose high quality basis sets that can accurately describe electron densities close to nuclear centres, *i.e.* "conventional" highly-contracted basis sets optimized for computing valence properties generally do not suffice.[16] Polarization consistent (pc-n) and segmented polarization consistent (pcS-n) basis sets are particularly appropriate, because they converge rapidly with respect to n.[17,18] Even at $n = 2$, basis set incompleteness errors are substantially lower than those associated with the choice of electronic structure method.[17,18]

Geometric errors may arise from the choice of electronic structure model and neglecting to account for vibrational-averaging effects. Fortunately, these sources of error can be accounted for separately. As proposed by Vuckovic,[19] a simple way of quantifying geometric model-related errors is to compare property values computed at an "exact" reference equilibrium geometry with those computed at the optimized geometry at a given level of theory. Accounting for vibrational averaging effects at each level of theory is harder and much more computationally intensive, because it requires higher order derivatives of the energy and nuclear shieldings with respect to atomic displacements.[20] A more feasible approach is to derive anharmonic corrections at a single level of theory and then use these to "correct" the vibrationally-averaged experimental values to obtain semi-experimental equilibrium shieldings, which can then be compared directly with computed values.[6] This approach can be justified on the basis that anharmonic corrections are small in magnitude relative to overall shieldings,[6] so methodological errors are expected to be negligible, except perhaps in cases where the chosen level of theory yields qualitatively incorrect potential energy curves. For example, Hartree–Fock theory provides a qualitatively incorrect description of bonding within F_2.[3]

Relativistic effects are the most challenging to account for, because full 4-component relativistic Dirac calculations are required for quantitative prediction of absolute NMR chemical shieldings.[21,22] These are more computationally intensive than scalar and 2-component relativistic models,[23–25] and are not as widely implemented in molecular quantum chemistry programme packages.

2 Methods

2.1 Benchmark data collation

Our data set comprises 49 molecules with 34 distinct ^1H shielding values (Table 1), 29 unique ^{13}C shielding values (Table 2 and 18 different ^{19}F shielding values (Table 3). We have particularly chosen to focus on these nuclei because they are most widely used in chemical characterisation and most abundantly represented in the literature.

Benchmark equilibrium geometries and geometric parameters were compiled from both the spectroscopic[26–30] and quantum chemical[31–33] literature. Computed geometries are of

Table 1 Intermolecular-interaction-free ^1H absolute gas-phase NMR chemical shieldings measured at 300 K, σ_0 (ppm), compiled from the literature. Where there are non-equivalent hydrogen atoms present in the same molecule, the relevant nuclear centre is highlighted in bold

Molecule	Value	Ref.	Molecule	Value	Ref.
HBr	35.037	34	CH_3NH_2	28.305	34
HCl	31.124	34	CH_3Cl	27.932	34
NH_3	30.727	34	HCN	27.78	35
CH_3OH	30.671	34	SiH_4	27.625	34
CH_4	30.633	34	CH_3OH	27.350	34
H_2S	30.572	34	CH_3F	26.635	34
CH_3NH_2	30.367	34	H_2	26.293	34
H_2O	30.102	34	CH_2CCH_2	26.279	34
CH_3SH	29.888	34	LiH	25.7	36
CH_3CH_3	29.887	34	CH_2Cl_2	25.670	34
PH_3	29.761	34	CH_2CH_2	25.463	34
HCCH	29.317	34	CH_2F_2	25.291	34
CH_3CN	29.149	34	CHF_3	24.543	34
CH_3CO	28.857	34	$CHCl_3$	23.659	34
CH_3SH	28.831	34	C_6H_6	23.535	34
HF	28.51	35	CH_3COH	21.068	34
CH_3Br	28.328	34	HOF	18.511	37

Table 2 Intermolecular-interaction-free ^{13}C absolute gas-phase NMR chemical shieldings measured at 300 K, σ_0 (ppm), compiled from the literature. Where there are multiple non-equivalent carbon atoms present in the same molecule, the relevant nuclear centre is highlighted in bold

Molecule	Value	Ref.	Molecule	Value	Ref.
CH_4	195.07	38	CH_3CN	74.03	38
CH_3CN	185.05	38	CHF_3	68.74	38
CH_3Br	182.41	38	CF_3CF_3	68.49	38
CH_3SH	182.34	38	$CFCl_3$	67.21	38
CH_3CH_3	180.84	38	CF_4	64.48	38
CH_3Cl	163.84	38	CH_2CH_2	64.43	38
CH_3NH_2	158.32	38	CO_2	58.71	38
CH_3COH	156.81	38	C_6H_6	57.17	38
CH_3OH	136.51	38	OCS	30.97	39
CH_3F	116.75	38	CO	0.80	38
HCCH	116.65	38	H_2CO	−0.5	40
CH_2CCH2	115.29	38	CH_3COH	−6.66	38
CCl_4	88.99	38	CS_2	−8.11	38
HCN	82.1	41	CH_2CCH2	−29.35	38
CH_2F_2	77.73	38			

Table 3 Intermolecular-interaction-free ^{19}F absolute gas-phase NMR chemical shieldings measured at 300 K, σ_0 (ppm), compiled from the literature

Molecule	Value	Ref.	Molecule	Value	Ref.
ClF	636.8	42	CF_4	259.0	42
CH_3F	470.9	43	PF_3	228.1	42
HF	409.6	44	F_2CO	221.4	42
LiF	374.3	45	CF_2Cl_2	202.4	42
SiF_4	362.9	46	$CFCl_3$	195.7	42
CH_2F_2	338.9	47	SF_6	139.3	48
BF_3	326.9	49	NSF_3	125.7	50
C_2F_6	282.7	42	NF_3	50.1	42
CHF_3	274.0	51	F_2	−233.1	42

at least CCSD(T)/cc-pVQZ quality. Cartesian coordinates for all benchmark equilibrium geometries used in this work are reported in the ESI.†

2.2 *Ab initio* calculations

Absolute isotropic chemical shieldings were computed using the gauge-including-atomic-orbital approach[52] as implemented in DALTON2020.1,[13,53] applied to the Hartree–Fock or Kohn–Sham molecular orbitals obtained using the methods and basis sets summarised in Table 4, computed at both the benchmark equilibrium geometry and optimized geometry at each level of theory.

For the polyatomic molecules in our data set, vibrationally-averaged geometries[54] and chemical shieldings[20] are computed at HF/pc-2, following the procedure outlined in Ruud *et al.*,[55] Lutnaes *et al.*[56] and Teale *et al.*[6] In brief, the geometry is first optimized and subject to harmonic vibrational analyis. Semi-cubic force constants required to define zero-point vibrational displacements along each normal mode are computed by numerical differentiation, using a step size of 0.0075 a.u. The vibrationally-averaged geometries thus obtained are used as an expansion point for computing vibrationally-averaged shielding constants. The required cubic and quartic derivatives of the

Table 4 Electronic structure methods and basis sets employed in the present work, listed in order of decreasing % HF exchange

Category	Options
Method	HF, HnH[a], B3LYP, B3LYP$_{0.05}$[b], HFS, KT3
Basis set	pc-2, pcS-2

[a] A simple hybrid functional comprising 50% exact (HF) exchange and 50% Slater–Dirac exchange. [b] A modified variant of the B3LYP functional comprising 5% HF exchange and 95% Slater–Dirac exchange (*cf.* 20% HF and 80% Slater–Dirac exchange in the original).

energy and second derivatives of the nuclear shielding tensor with respect to displacement along normal mode coordinates are computed numerically, using a step size of 0.05 a.u. We note that these step sizes may not always be optimal in all cases, but will consider this as a potential source of error rather than attempting to optimize these values on a case-by-case basis which would be both computationally prohibitive and may introduce further uncontrolled errors.

For diatomics, accurate and precise equilibrium and vibrationally-averaged bond lengths can be determined directly from spectroscopic data. Therefore, vibrationally-averaged shieldings are computed using these coordinates rather than their HF-optimized equivalents.

Finally, anharmonic corrections are computed as:

$$\delta_{ZPVE} = \sigma_0^{HF/pc-2} - \sigma_e^{HF/pc-2} \tag{1}$$

2.3 Correction procedure

At the heart of this work is our proposal to shift and scale computed absolute shieldings to more accurately reproduce experimental values:

$$\sigma_{corr} = m\sigma_{calc} + b \tag{2}$$

The most statistically robust way of obtaining the required m and b values, that guarantees residual root-mean-squared errors will be minimised across the training data set, is linear regression analysis. However, this is not particularly practical as a general approach, because the parameters thus obtained are not transferable, and so a full calibration procedure must be performed at each intended level of theory.

However, all is not lost. The required scaling parameters may instead be estimated by comparing computed and experimental shieldings using two carefully chosen reference molecules:

$$m = \frac{\sigma_{exp}(1) - \sigma_{exp}(2)}{\sigma_{calc}(1) - \sigma_{calc}(2)} \tag{3}$$

$$b = \frac{\sigma_{calc}(1)\sigma_{exp}(2) - \sigma_{calc}(2)\sigma_{exp}(1)}{\sigma_{calc}(1) - \sigma_{calc}(2)} \tag{4}$$

Key considerations in choosing appropriate reference molecules include:

• They should be small.
• They should have significantly different shieldings.
• They should lie on or near the line of best fit one would obtain *via* a full regression analysis, *i.e.* be representative of the data set as a whole.

This journal is © the Owner Societies 2022

Phys. Chem. Chem. Phys., 2022, **24**, 27055–27063 | **27057**

Table 5 Reference molecules for our two-point correction procedure

Nucleus	^{1}H	^{13}C	^{19}F
Molecule 1	CH_4	CH_4	HF
Molecule 2	H_2	CO	CF_4

The molecules selected according to these criteria are listed in Table 5.

3 Results and discussion

The first task at hand is to assess whether computed shielding constants display systematic errors that may be amenable to linear-scaling correction. It is well known from the quantum chemistry literature that the most significant source of error in molecular property predictions comes from the underlying electronic structure model. Therefore, we will begin by comparing shieldings computed at benchmark equilibrium geometries using a range of electronic structure models (σ_e^{calc}) with semi-experimental "equilibrium" shieldings (σ_e^{expt}). To assess geometric effects, we will compare nuclear shieldings computed at benchmark and optimized geometries (σ_e^{calc} and σ_{opt}^{calc}) with vibrationally-averaged experimental shieldings (σ_0^{expt}) and semi-experimental equilibrium shieldings (σ_e^{expt}). Finally, we will apply and analyse the performance of our two-point correction model.

3.1 Electronic effects

Correlations between computed and semi-experimental equilibrium shieldings are illustrated in Fig. 1. B3LYP and B3LYP$_{0.05}$ data have been omitted from these plots because they are hard to visually distinguish from the HnH data. Similarly, computational values generated using the pcS-2 basis set are not shown, because these data points are visually indistinguishable from their pc-2 counterparts. However, key regression statistics for all 6 methods with each basis set are summarised in Table 6.

Fig. 1 shows that the KT3 functional displays qualitatively different behaviour to the hybrid functionals – it is much more accurate for predicting ^{13}C shielding constants, although it displays similar accuracy for ^{1}H and ^{19}F shieldings. Analysis of the regression parameters in Table 6 illustrates these differences even more clearly. Hybrid functionals display systematic model-dependent trends in both the slope and intercept values according to the proportion of HF exchange in the underlying functional; regression parameters for the KT3 functional do not fit these trends. KT3 functionals tend to yield more accurate predictions of shielding constants (slopes and intercepts closer to 1 and 0, respectively) but display more scatter about their lines of best fit (lower R^2 values).

These observations likely reflect differences in parameterisation strategy; hybrid functionals are parameterised to accurately model the energetics of valence electrons[57] whereas the KT functionals are designed to accurately reproduce exchange-correlation potentials, especially for electrons within 0.4–1.0 a.u. of each nuclear centre.[14] Because valence electron behaviour varies widely across our data set whereas smaller fluctuations are expected in
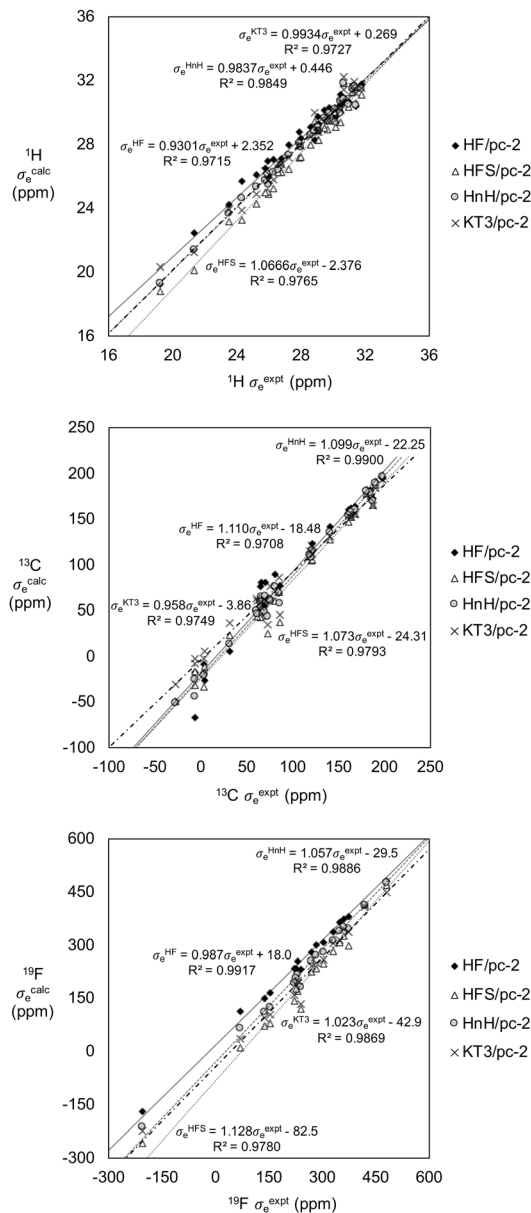

Fig. 1 Computed *vs.* semi-experimental equilibrium shieldings for ^{1}H (top), ^{13}C (middle) and ^{19}F (bottom) nuclei.

core regions, it stands to reason that hybrid functionals display larger but more systematic errors, whereas the KT3 functional exhibits smaller but less systematic errors. This suggests that hybrid functionals may be more appropriate for our proposed shift-and-scale correction procedure. In particular, it is best to choose "well-balanced" hybrid functionals in which electron under- and over-localisation errors approximately cancel and that are parameterised to model electron behaviour well in a wide range of chemical environments, which will show up as consistently high R^2 values across all nuclear centres in this context.

However, from a practical standpoint, it does not make sense to correlate DFT shieldings computed at near-exact equilibrium geometries with semi-experimental equilibrium shieldings, because both of these quantities are hard to obtain.

Table 6 Linear regression summary statistics correlating predicted NMR shieldings at benchmark equilibrium geometries σ_e^{calc} with semi-experimental equilibrium shieldings $\sigma_e^{expt} = \sigma_0^{expt} - \delta_{ZPVE}$. Zero-point vibrational corrections are computed at HF/pc-2 following the procedure of Ruud et al.[20]

	^{1}H			^{13}C			^{19}F		
Model	Slope	Int.	R^2	Slope	Int.	R^2	Slope	Int.	R^2
HF/pc-2	0.9301	2.352	0.9715	1.110	−18.48	0.9708	0.987	18.0	0.9917
HF/pcS-2	0.8890	3.085	0.9479	1.108	−17.55	0.9712	0.989	21.9	0.9923
HnH/pc-2	0.9837	0.446	0.9849	1.099	−22.25	0.9900	1.057	−29.5	0.9886
HnH/pcS-2	0.9817	0.435	0.9853	1.110	−24.74	0.9902	1.067	−29.1	0.9894
B3LYP/pc-2	1.0019	0.140	0.9866	1.057	−23.44	0.9848	1.075	−50.2	0.9917
B3LYP/pcS-2	1.0002	0.124	0.9872	1.072	−27.47	0.9864	1.088	−51.5	0.9923
B3LYP$_{0.05}$/pc-2	1.0282	−0.751	0.9832	1.048	−23.91	0.9791	1.097	−66.9	0.9888
B3LYP$_{0.05}$/pcS-2	1.0264	−0.762	0.9842	1.066	−28.95	0.9808	1.112	−69.3	0.9896
HFS/pc-2	1.0666	−2.376	0.9765	1.073	−24.31	0.9793	1.128	−82.5	0.9775
HFS/pcS-2	1.0643	−2.371	0.9781	1.094	−30.12	0.9796	1.146	−85.9	0.9789
KT3/pc-2	0.9914	0.482	0.9767	0.958	−3.86	0.9747	1.023	−42.9	0.9869
KT3/pcS-2	0.9941	0.349	0.9782	0.957	−3.33	0.9782	1.024	−39.2	0.9878

Benchmark equilibrium geometries can only be obtained by painstaking analysis of spectroscopic data or through high level *ab initio* geometry optimizations. Anharmonic corrections are computationally intensive because they require calculation of third derivatives of the energy with respect to atomic displacements and second derivatives of the nuclear shielding tensor. Therefore, it is prudent to investigate whether errors introduced by ignoring anharmonicity and/or optimizing geometries using density functional theory are systematic, and could therefore be compensated for using our proposed shift-and-scale procedure.

3.2 Geometric effects

Nuclear shieldings computed at benchmark and DFT-optimized geometries are correlated against vibrationally averaged experimental values and semi-experimental equilibrium values in Fig. 2. B3LYP/pc-2 data are illustrated as a representative example.

For ^{13}C nuclear shieldings, neither geometry optimization nor vibrational averaging effects significantly affect the correlation between computed and observed values. Therefore, the most sensible quantities to correlate are shieldings computed at the DFT-optimized geometry and vibrationally-averaged shielding constants obtained directly from experiment.

For ^{19}F nuclear shieldings, the picture becomes a bit more complicated. The most obvious outliers are due to limitations of mean-field models for describing bonding in F_2, which shows up in longer-than-expected optimized bond lengths, and therefore shieldings that are lower in magnitude than expected. In fact, perhaps the most surprising thing about this data is how well the B3LYP data correlates to experiment when the molecule is constrained to its equilibrium bond length, given the known deficiencies of the electronic structure model. Because this outlier is so extreme and also because F_2 has a very different shielding constant to other fluorine nuclei in our data set, this disproportionately affects the R^2 value. However, otherwise we observe a mix of smaller sources of scatter due to both vibrational-averaging corrections and geometry optimization.

For ^{1}H shieldings, a very different pattern emerges. In this case, geometry optimization does not seem to have much of an
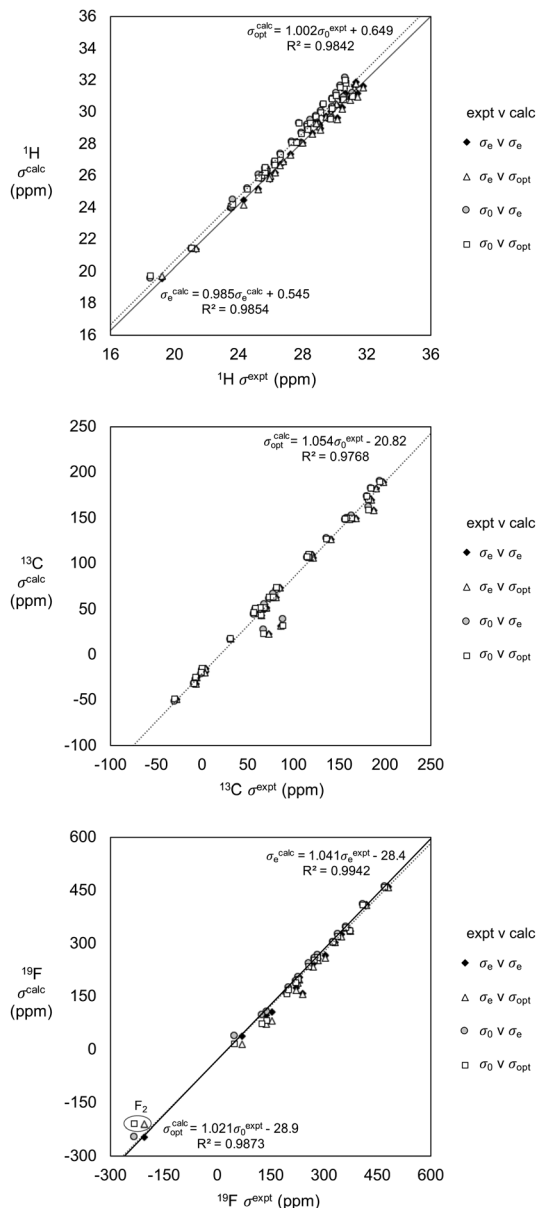

Fig. 2 Computed (B3LYP/pc-2) *vs.* experimentally-derived shieldings for ^{1}H (top), ^{13}C (middle) and ^{19}F (bottom) nuclei.

Table 7 Errors in predicted shieldings before and after application of our 2-point correction scheme. MAD = mean absolute deviation between computed (σ_{opt}^{calc}) and experimental (σ_0^{expt}) values, STD = standard deviation, MARD = median absolute relative deviation. Outliers are excluded in all cases

Model	^{1}H			^{13}C			^{19}F		
	MAD (ppm)	STD (ppm)	MARD (%)	MAD (ppm)	STD (ppm)	MARD (%)	MAD (ppm)	STD (ppm)	MARD (%)
HnH/pc-2	0.539	0.472	1.6	8.26	7.71	5.4	60.4	39.0	19.2
HnH/pc-2 (corr)	0.321	0.277	0.8	3.22	4.80	2.7	10.8	20.6	0.9
B3LYP/pc-2	0.797	0.286	2.7	12.85	5.34	14.7	29.0	18.8	9.3
B3LYP/pc-2 (corr)	0.238	0.195	0.7	2.75	3.36	2.5	7.5	9.3	1.8
B3LYP$_{0.05}$/pc-2	0.450	0.332	1.4	15.64	6.42	16.9	47.0	28.6	14.2
B3LYP$_{0.05}$/pc-2 (corr)	0.164	0.236	0.3	4.40	5.09	3.8	10.7	10.6	3.1
KT3/pc-2	1.062	0.412	3.7	4.60	5.69	4.6	34.5	16.4	10.3
KT3/pc-2 (corr)	0.235	0.358	0.6	4.60	5.60	3.9	12.8	13.6	3.9

effect, but the data clusters strongly based upon whether zero-point vibrational averaging corrections have been applied or not. This is perhaps unsurprising, as zero-point vibrational effects are expected to be strongest for lighter, peripheral atoms. What is perhaps surprising is that ^{1}H nuclear shieldings computed at DFT-optimized geometries correlate just as well with vibrationally-averaged shieldings obtained directly from experiment as computational values obtained at reference equilibrium geometries correlate to semi-experimental equilibrium data.

Therefore, we conclude that the most computationally-effective strategy will be to apply our two-point correction model to NMR shieldings computed at DFT-optimized geometries, using experimental reference data that has not been corrected for vibrational averaging effects. This approach will obviously incur significant errors if the DFT model of choice is not appropriate for a given molecule, but such errors will remain isolated provided that such molecules are not used in the calibration process.

3.3 Two-point correction model

Scaled and shifted B3LYP/pc-2 nuclear shieldings obtained using our two-point correction model are illustrated in Fig. 3, and summary statistics for corrected and uncorrected HnH/pc-2, B3LYP/pc-2 and B3LYP$_{0.05}$/pc-2 and KT3/pc-2 predictions presented in Table 7. After correction, all four models display similar error statistics, reflecting similar underlying error profiles and visually indistinguishable correlation curves.

The main take-home message is that our two-point correction procedure is extremely effective at controlling for most sources of error in equilibrium DFT predictions of vibrationally-averaged NMR shielding constants, based upon DFT-optimized geometries. This enables us to identify and characterise outliers, which fall into two categories; those that arise due to neglect of relativistic effects and those that arise due to grossly inadequate modelling of electron correlation.

Hydrogen nuclei adjacent to third- and fourth-row atoms within HBr, HCl, H$_2$S, CH$_3$SH and PH$_3$ exhibit significant spin–orbit coupling that is inherently captured experimentally but absent from non-relativistic quantum chemical models and not accounted for by the present formulation of our two-point correction procedure. Nonetheless, the residual errors appear quite systematic, and in line with expectations from relativistic calculations of spin–orbit contributions to nuclear shielding
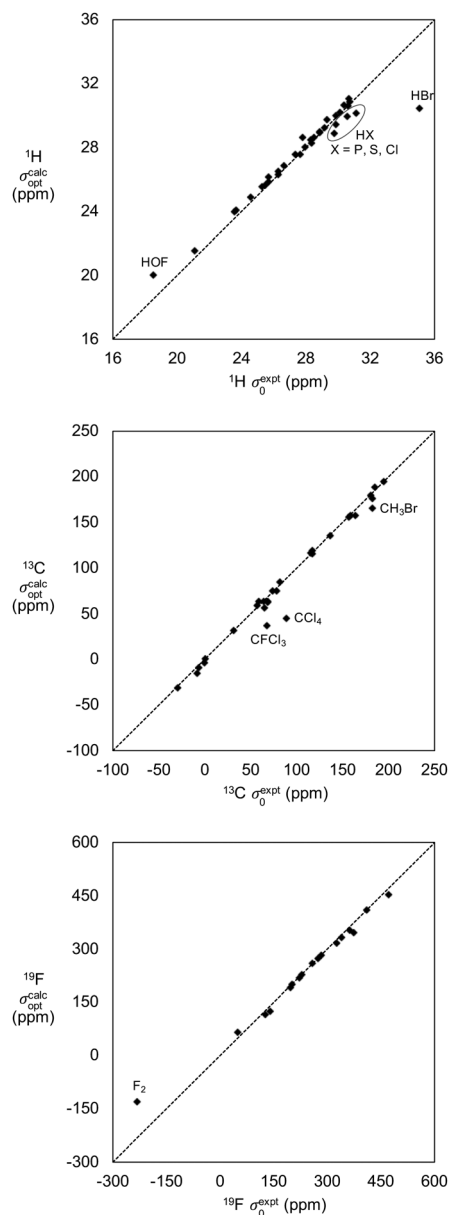

Fig. 3 Scaled-and-shifted B3LYP/pc-2 nuclear shieldings obtained using our two-point correction scheme. The dotted diagonal line represents exact equivalence between computed and experimental values.

Table 8 Aggregate mean absolute deviations (MAD) and median absolute relative deviations (MARD) for all corrected shieldings in our data set, excluding outliers

Model	MAD (ppm)	MARD (%)
HnH/pc-2 (corr)	3.8	1.4
B3LYP/pc-2 (corr)	3.1	1.0
B3LYP$_{0.05}$/pc-2 (corr)	4.4	1.3
KT3/pc-2 (corr)	4.8	1.6

constants of hydrogen halides in the literature.[9] Similarly, carbon nuclei adjacent to multiple chlorine atoms or a single, heavier bromine atom also display spin–orbit heavy-atom–light-atom effects consistent with trends reported in the literature.[8,58]

Molecular fluorine, F_2 is a canonical example of a molecule in which electron correlation effects are extremely important and must be accounted for using correlated wavefunction theories to obtain get its potential energy curve even qualitatively correct.[59] Similarly, electron correlation is known to be particularly important for molecules with O–F bonds, such as HOF.[60] Therefore it is appropriate that DFT-based models yield inaccurate predictions in these cases.

Digging a bit deeper into the data presented in Table 7 reveals that the effectiveness of our correction procedure depends on both the level of theory and the identity of the nuclear centre. The KT3 functional displays a different error profile to the hybrid functionals; it yields more accurate predictions of ^{13}C shieldings that cannot be improved using our correction scheme, but less accurate predictions of ^{1}H shieldings that show the highest fold improvement in accuracy upon correction. Overall, this supports our earlier assertion that KT functionals appear to trade off the ability to capture subtle changes in valence electron behaviour for an improved description of the near-core region.

Errors in shielding constants computed using hybrid functionals tend to be larger but more systematic, so our correction procedure improves prediction accuracy across the board. For hybrid functionals, mean absolute deviations in ^{1}H NMR shieldings are reduced by a factor of ~ 2–3, ^{13}C by ~ 2.5–4.5 and ^{19}F by ~ 4–6. In general, the less accurate the initial predictions, the more they are improved upon correction.

Finally, we note that our correction model holds up well against other benchmark computational results in the literature. To facilitate comparison, we have computed aggregate mean absolute deviations and median absolute relative deviations for all corrected shielding values obtained using hybrid functionals, excluding outliers. These values are presented in Table 8. Based on mean absolute deviations, our results are comparable in accuracy to the CCSD(T)/aug-cc-pCVnZ (n = T, Q) results of Teale *et al.*,[6] but are obtained at a fraction of the computational cost.

4 Conclusions and future work

We have introduced a novel two-point correction procedure for scaling and shifting equilibrium nuclear shielding constants computed using density functional theory to reproduce vibrationally-averaged experimental values. It relies on two empirical but non-adjustable parameters per nucleus, which are the exact gas-phase shieldings for two reference molecules. As such, it can be used in conjunction with any electronic structure model and basis set, although it is important to use high-quality basis sets that can appropriately capture changes in electron density near nuclei in nuclear shielding calculations. It is most effective when applied to shieldings computed using density functionals that prioritise accurately describing the behaviour and energetics of valence electrons.

We have demonstrated that this approach yields predictions of approximately CCSD(T)/aug-cc-pVT/QZ accuracy for a fraction of the computational cost. Further, it enables outliers that arise due to non-systematic or non-systemic effects such as spin–orbit coupling and electron correlation model failure to be clearly identified and characterised. Spin–orbit heavy-atom–light-atom effects are particularly problematic, and will need to be accounted for to ensure high accuracy predictions for all molecules, including those that contain heavy atoms adjacent to much lighter nuclei. In keeping with the spirit of this work, we propose that this too be achieved using a simple empirical correction approach, although this will require careful calibration against benchmark computational data obtained by solving the full 4-component Dirac equation.[25]

Similar correlations between computed nuclear shieldings and experimental chemical shifts have been observed for molecules in solution,[61] and so we expect our two-point correction procedure to generalise well – noting that the "shift" aspect of our procedure will naturally account for referencing when applied to predict relative experimental shifts from absolute computed shieldings. In other words, it is equally appropriate to use relative chemical shifts for reference points in our procedure as it is to use absolute shielding constants – provided the chemical shifts are all measured under the same experimental conditions. The question of which reference molecules are most appropriate – and the importance of accounting for solvation effects – will be addressed in future work.

Conflicts of interest

There are no conflicts to declare.

References

1 E. E. Kwan and S. G. Huang, *Eur. J. Org. Chem.*, 2008, 2671–2688.
2 T. Helgaker, M. Jaszunski and K. Ruud, *Chem. Rev.*, 1999, **99**, 293–352.
3 C. J. Jameson, in *Recent Advances in Nuclear Magnetic Shielding Theory and Computational Methods*, American Chemical Society; 1999, ch. 1, pp. 1–23.
4 T. Helgaker, M. Watson and N. C. Handy, *J. Chem. Phys.*, 2000, **113**, 9402–9409.

This journal is © the Owner Societies 2022

Phys. Chem. Chem. Phys., 2022, **24**, 27055–27063 | **27061**

5 J. San Fabián, J. Garca de la Vega and E. San Fabián, *J. Chem. Theory Comput.*, 2014, **10**, 4938–4949.

6 A. M. Teale, O. B. Lutnæs, T. Helgaker, D. J. Tozer and J. Gauss, *J. Chem. Phys.*, 2013, **138**, 024111.

7 J. Autschbach and S. Zheng, *Annu. Rep. NMR Spectrosc.*, 2009, **67**, 1–95.

8 J. Vícha, J. Novotný, S. Komorovsky, M. Straka, M. Kaupp and R. Marek, *Chem. Rev.*, 2020, **120**, 7065–7103.

9 L. Visscher, T. Enevoldsen, T. Saue, H. J. A. Jensen and J. Oddershede, *J. Comput. Chem.*, 1999, **20**, 1262–1273.

10 L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi and S. Grimme, *Phys. Chem. Chem. Phys.*, 2017, **19**, 32184–32215.

11 F. Peccati, R. Laplaza and J. Contreras-Garca, *J. Phys. Chem. C*, 2019, **123**, 4767–4772.

12 P. J. Wilson, R. D. Amos and N. C. Handy, *Chem. Phys. Lett.*, 1999, **312**, 475–484.

13 T. Helgaker, P. J. Wilson, R. D. Amos and N. C. Handy, *J. Chem. Phys.*, 2000, **113**, 2983–2989.

14 T. W. Keal and D. J. Tozer, *J. Chem. Phys.*, 2003, **119**, 3015–3024.

15 T. W. Keal and D. J. Tozer, *J. Chem. Phys.*, 2004, **121**, 5654–5660.

16 T. Kupka and C. Lim, *J. Phys. Chem. A*, 2007, **111**, 1927–1932.

17 T. Kupka, M. Stachów, M. Nieradka, J. Kaminsky and T. Pluta, *J. Chem. Theory Comput.*, 2010, **6**, 1580–1589.

18 T. Kupka, M. Stachów, M. Nieradka, J. Kaminsky, T. Pluta and S. P. A. Sauer, *Magn. Reson. Chem.*, 2011, **49**, 231–236.

19 S. Vuckovic, *J. Phys. Chem. A*, 2022, **126**, 1300–1311.

20 K. Ruud, P.-O. Åstrand and P. R. Taylor, *J. Chem. Phys.*, 2000, **112**, 2668–2683.

21 J. I. Melo, M. C. Ruiz de Azua, C. G. Giribet, G. A. Aucar and R. H. Romero, *J. Chem. Phys.*, 2003, **118**, 471–486.

22 J. Autschbach, *Mol. Phys.*, 2013, **111**, 2544–2554.

23 J. I. Melo, A. F. Maldonado and G. A. Aucar, *J. Chem. Inf. Model.*, 2019, **60**, 722–730.

24 L. Visscher, *J. Comput. Chem.*, 2002, **23**, 759–766.

25 M. Repisky, S. Komorovsky, M. Kadek, L. Konecny, U. Ekström, E. Malkin, M. Kaupp, K. Ruud, O. L. Malkina and V. G. Malkin, *J. Chem. Phys.*, 2020, **152**, 184101.

26 K.-P. Huber and G. Herzberg, *Molecular spectra and molecular structure: IV. Constants of diatomic molecules*, Springer Science & Business Media; 2013.

27 G. Herzberg, *Electronic spectra and electronic structure of polyatomic molecules*, van Nostrand; 1966, vol. 3.

28 K. Kuchitsu, *Structure of free polyatomic molecules: basic data*, Springer Science & Business Media; 2013.

29 J. Demaison, L. Margules and J. E. Boggs, *Struct. Chem.*, 2003, **14**, 159–174.

30 L. Halonen and T. K. Ha, *J. Chem. Phys.*, 1988, **89**, 4885–4888.

31 A. Karton, S. Daon and J. M. Martin, *Chem. Phys. Lett.*, 2011, **510**, 165–178.

32 J. Gauss and J. F. Stanton, *J. Phys. Chem. A*, 2000, **104**, 2865–2868.

33 A. R. Hoy and P. R. Bunker, *J. Mol. Spectrosc.*, 1979, **74**, 1–8.

34 P. Garbacz, K. Jackowski, W. Makulski and R. E. Wasylishen, *J. Phys. Chem. A*, 2012, **116**, 11896–11904.

35 W. T. Raynes, *Nuclear Magnetic Resonance: Volume 7*, The Royal Society of Chemistry; 1978, vol. 7, pp. 1–25.

36 L. Wharton, L. P. Gold and W. Klemperer, *J. Chem. Phys.*, 1962, **37**, 2149–2150.

37 J. Hindman, A. Svirmickas and E. Appelman, *J. Chem. Phys.*, 1972, **57**, 4542–4543.

38 W. Makulski, *Phys. Chem. Chem. Phys.*, 2022, **24**, 8950–8961.

39 K. Jackowski, M. Jaszuński, W. Makulski and J. Vaara, *J. Magn. Reson.*, 1998, **135**, 444–453.

40 C. J. Jameson, *Nuclear Magnetic Resonance: Volume 18*, The Royal Society of Chemistry; 1989, vol. 18, pp. 1–33.

41 A. K. Jameson and C. J. Jameson, *Chem. Phys. Lett.*, 1987, **134**, 461–466.

42 C. J. Jameson, A. K. Jameson and P. M. Burrell, *J. Chem. Phys.*, 1980, **73**, 6013–6020.

43 C. J. Jameson, A. K. Jameson and J. Honarbakhsh, *J. Chem. Phys.*, 1984, **81**, 5266–5267.

44 S. M. Bass, R. L. DeLeon and J. Muenter, *J. Chem. Phys.*, 1987, **86**, 4305–4312.

45 R. Braunstein and J. Trischka, *Phys. Rev.*, 1955, **98**, 1092.

46 W. Makulski, *J. Mol. Struct.*, 2013, **1036**, 168–173.

47 M. Kubiszewski, W. Makulski and K. Jackowski, *J. Mol. Struct.*, 2004, **704**, 211–214.

48 K. Jackowski and M. Jaszuński, *Concepts Magn. Reson., Part A*, 2007, **30**, 246–260.

49 K. Jackowski, W. Makulski, A. Szyprowska, A. Antušek, M. Jaszuński and J. Jusélius, *J. Chem. Phys.*, 2009, **130**, 044309.

50 O. Glemser and R. Mews, *Angew. Chem., Int. Ed. Engl.*, 1980, **19**, 883–899.

51 M. Kubiszewski, W. Makulski and K. Jackowski, *J. Mol. Struct.*, 2005, **737**, 7–10.

52 K. Wolinski, J. F. Hinton and P. Pulay, *J. Am. Chem. Soc.*, 1990, **112**, 8251–8260.

53 K. Aidas, C. Angeli, K. L. Bak, V. Bakken, R. Bast, L. Boman, O. Christiansen, R. Cimiraglia, S. Coriani, P. Dahle, E. K. Dalskov, U. Ekström, T. Enevoldsen, J. J. Eriksen, P. Ettenhuber, B. Fernández, L. Ferrighi, H. Fliegl, L. Frediani, K. Hald, A. Halkier, C. Hättig, H. Heiberg, T. Helgaker, A. C. Hennum, H. Hettema, E. Hjertenæs, S. Høst, I.-M. Høyvik, M. F. Iozzi, B. Jansík, H. J. A. Jensen, D. Jonsson, P. Jørgensen, J. Kauczor, S. Kirpekar, T. Kjærgaard, W. Klopper, S. Knecht, R. Kobayashi, H. Koch, J. Kongsted, A. Krapp, K. Kristensen, A. Ligabue, O. B. Lutnæs, J. I. Melo, K. V. Mikkelsen, R. H. Myhre, C. Neiss, C. B. Nielsen, P. Norman, J. Olsen, J. M. H. Olsen, A. Osted, M. J. Packer, F. Pawlowski, T. B. Pedersen, P. F. Provasi, S. Reine, Z. Rinkevicius, T. A. Ruden, K. Ruud, V. V. Rybkin, P. Sałek, C. C. M. Samson, A. S. de Merás, T. Saue, S. P. A. Sauer, B. Schimmelpfennig, K. Sneskov, A. H. Steindal, K. O. Sylvester-Hvid, P. R. Taylor, A. M. Teale, E. I. Tellgren, D. P. Tew, A. J. Thorvaldsen, L. Thøgersen, O. Vahtras, M. A. Watson, D. J. D. Wilson, M. Ziolkowski and H. Ågren, *Wiley Interdisc. Rev.: Comput. Mol. Sci.*, 2014, **4**, 269–284.

27062 | *Phys. Chem. Chem. Phys.*, 2022, **24**, 27055–27063

This journal is © the Owner Societies 2022

54 P.-O. Åstrand, K. Ruud and P. R. Taylor, *J. Chem. Phys.*, 2000, **112**, 2655–2667.

55 K. Ruud, P.-O. Åstrand and P. R. Taylor, *J. Am. Chem. Soc.*, 2001, **123**, 4826–4833.

56 O. B. Lutnæs, A. M. Teale, T. Helgaker, D. J. Tozer, K. Ruud and J. Gauss, *J. Chem. Phys.*, 2009, **131**, 144104.

57 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 1372–1377.

58 S. Fukawa, M. Hada, R. Fukuda, S. Tanaka and H. Nakatsuji, *J. Comput. Chem.*, 2001, **22**, 528–536.

59 L. Bytautas and K. Ruedenberg, *J. Chem. Phys.*, 2009, **130**, 204101.

60 W. Thiel, G. Scuseria, H. F. Schaefer III and W. D. Allen, *J. Chem. Phys.*, 1988, **89**, 4965–4975.

61 M. W. Lodewyk, M. R. Siebert and D. J. Tantillo, *Chem. Rev.*, 2012, **112**, 1839–1862.

This journal is © the Owner Societies 2022

Phys. Chem. Chem. Phys., 2022, **24**, 27055–27063 | **27063**