Charge Distribution and Spectroscopic Properties

Partial atomic charges

A very old concept in chemistry – molecular polarity is associated with charge built-up or depletion on the individual atoms. Hydrogen bond in water: the oxygen is ‘negative’ and the hydrogen atoms are ‘positive’. This concept allows us to conveniently ignore the wave character of the electrons and deal only with atoms – these atoms will reflect electronic distribution by the degree to which they carry positive or negative charge. Critical to the efficiency of most force-field methods is that they compute electrical interactions as atom-centered charge-charge interactions – the development of methods for assigning accurate partial charges to atoms in molecules has seen considerable research.

However, the concept of a partial atomic charge is ill-defined and there is no universally agreed upon ‘best’ procedure for computing partial atomic charges, because they are used in different ways within the context of different quantitative and qualitative models in chemistry. It is helpful to categorize different partial charge methodologies into four classes.

Class I charges

A Class I charge is not determined from quantum mechanics, but through some arbitrary or intuitive approach. For instance, in a neutral diatomic molecule having a known dipole moment, one atomic charge must be $+q$ and the other $-q$. One obvious choice for $q$ will be that value that, in conjunction with the experimental bond length $r_{eq}$, causes the classical dipole moment $qr_{eq}$ to be equal to the experimental value.
A less limited Class I charge model – the electronegativity equalization model: charge is distributed from atoms to their bonded partners in a manner that is dependent on their Pauling electronegativities and bond orders. This model is convenient for molecular mechanics simulations where it is desirable that the partial atomic charges be able to change as a function of a variable local environment.

Class II charges

Class II charge models involve a direct partitioning of the molecular wave function into atomic contributions following some arbitrary, orbital-based scheme. The first such scheme – proposed by Mulliken (1955) – Mulliken population analysis. Electrons are divided up amongst the atoms according to the degree to which different atomic AO basis functions contribute to the overall wave function. We can write the expression for the total number of electrons as

\[ N = \sum_{j} \int \psi_j(r_j) \psi_j(r_j) dr_j = \sum_{j} \sum c_{jr} \varphi(r_j) \sum c_{js} \varphi(s_j) dr_j = \sum_j \left( \sum_r c_{jr}^2 + \sum_{r \neq s} c_{jr} c_{js} S_{rs} \right) \]

\( r \) and \( s \) – index AO basis functions \( \varphi \); \( c_{jr} \) – the coefficient of basis function \( r \) in MO \( j \); \( S \) – the overlap matrix element.

From the last line of the equation we see that we may divide the total number of electrons up into two sums; one includes only squares of single AO basis functions and the other includes products of two basis functions. The electrons associated with only a single basis function should be assigned entirely to the atom on which that basis function resides.
The second term represents the electrons ‘shared’ between different basis functions. Mulliken suggested that one might divide them up evenly between the two atoms on which basis functions $r$ and $s$ reside. Then, we can compute the atomic population $N_k$ as

$$N_k = \sum^\text{electrons}_j \left( \sum_{r \in k} c_{jr}^2 + \sum_{r,s \in k, r \neq s} c_{jr} c_{js} S_{rs} + \frac{1}{2} \sum_{r \in k, s \notin k} c_{jr} c_{js} S_{rs} \right)$$

Note that the orthonormality of basis functions of different angular momentum both residing on the same atom $k$ causes many terms in the second sum of the above equation to vanish. The Mulliken atomic charge is then defined as

$$q_k = Z_k - N_k$$

With minimal or small split-valence basis sets, Mulliken charges tend to be reasonably intuitive, at least in sign. Analysis of changes in charge as a function of substitution or geometric change – the best use for Mulliken charges.

The problems with Mulliken population analysis: the use of a non-orthogonal basis set, which can lead to physically non-meaningful results – it is possible that for individual basis function can have occupation numbers greater than 2 or less than 0. Second, the rule that all shared electrons should be divided up equally between the atoms on which sharing basis functions reside ignores the possibly very different electronegativities of these atoms. Finally, Mulliken partial charges are very sensitive to basis-set size – with very complete basis sets, Mulliken charges have a tendency to become unphysically large.
To overcome these problems – the atomic basis set has to be orthonormalized: the Natural Population Analysis (NPA) scheme by Reed, Weinstock, and Weinhold (1985). Orthogonalization takes place in a four-step process. The electron density around each atom is initially rendered as compact as possible, and further diagonalization is carried out so as to preserve the shape of strongly occupied atomic orbitals to as large extent as possible. Following orthogonalization, a Mulliken-like analysis in the new basis set gives the atomic populations with no contributions from off-diagonal terms. In NPA, each atomic partial charge effectively converges to a stable value with increasing basis-set size.

Class III charges

Rather than being determined from an arbitrary analysis of the wave function itself, Class III charges are computed based on the analysis of some physical observable that is calculated from the wave function. There is an obvious relationship between atomic partial charges and dipole moments in diatomics. Cioslowsky (1989) has generalized this idea for polyatomic molecules, defining the generalized atomic polar tensor (GAPT) charge as

\[
q_k = \frac{1}{3} \left( \frac{\partial \mu_x}{\partial x_k} + \frac{\partial \mu_y}{\partial y_k} + \frac{\partial \mu_z}{\partial z_k} \right)
\]

\(\mu\) - the molecular dipole moment as a function of moving atom \(k\) in each of the three Cartesian directions. The dipole moment itself can be computed when the wave function is available as the expectation value of the dipole moment operator:

\[
\langle x \rangle = \sum_{i}^{\text{atoms}} Z_i x_i - \int \Psi(r) \left( \sum_{j}^{\text{electrons}} x_j \right) \Psi(r) dr
\]

and similarly for the \(y\) and \(z\) components.
GAPT charges converge quickly with respect to basis-set size. However, if a level of theory fails to give a good dipole moment (e.g., HF), it is then unlikely to give useful charges. In addition, GAPT charges are relatively expensive to compute – equivalent to the cost of a vibrational frequency calculations.

An alternative physical observable used to define partial atomic charges – the electron density. In X-ray crystallography, the electron density is directly measured, and by comparison to spherically symmetric neutral atoms, atomic partial charges may be defined experimentally, following some decisions what to do with respect to partitioning space between atoms. Bader and coworkers – a particular partitioning scheme for use with electronic structure calculations – the atoms-in-molecules (AIM) method (1990). An atomic volume is defined as that region of space including the nucleus that lies within all zero-flux surfaces surrounding the nucleus. If we move from one nucleus to another, density drops first, then we will come to a point where the density gradient reaches zero, because it passes from a negative value (falling back towards the nucleus we started from) to a positive value (falling towards the second nucleus). The points of zero gradients (zero-flux) are called ‘bond critical points’ and their union gives the zero-flux surface. Within the AIM theory, partial atomic charge is defined as nuclear charge less the total number of electrons residing within the atomic basin:

$$q_k = Z_k - \int_{\Omega_k} \rho(r)dr$$

$\Omega_k$ – the spatial volume encompassed by the zero-flux surface of $k$. As the necessary determinations of atomic volumes and the integration over them are time-consuming, AIM analysis is not a routine procedure.
Class IV charges

A Class IV charge is derived by a semiempirical mapping of a precursor charge (either from a Class II or Class III model), in order to reproduce an experimentally determined observable. Charge Model 1 (CM1) – the first method developed to compute Class IV charges (1995). In this model, the input charges are Mulliken charges determined at a semiempirical level (AM1 or PM3) and the experimental observable for which the mappings were optimized is the molecular dipole moment as computed from the atomic partial charges:

$$\mu = \left( \sum_k q_k x_k \right)^2 + \left( \sum_k q_k y_k \right)^2 + \left( \sum_k q_k z_k \right)^2$$

The CM1 charged is defined as

$$q^{CM1}_k = q^{(0)}_k + B_k \Delta q_k - \sum_{k' \neq k} B_{kk'} \Delta q_{k'}$$

$q^{(0)}_k$ - the input Mulliken charges; $B_{kk'}$ – the bond order between atoms $k$ and $k'$; $B_k$ – the sum of the bond orders of atom $k$ to all other atoms. The quantity $\Delta q_k$ is where the mapping comes in:

$$\Delta q_k = c_k q^{(0)}_k + d_k$$

Parameters $c$ and $d$ are optimized for each atom $k$ so as to minimize errors in the predicted dipole moments. The bond order is determined, as originally suggested by Mulliken, from the population analysis:

$$B_{kk'} = \sum_{\mu \in k} \sum_{\nu \in k'} P_{\mu \nu}^2$$

The CM1 models for AM1 and PM3 give root-mean-square errors of 0.30 and 0.26 D, respectively, in the dipole moments of 195 neutral molecules. The CM1-AM1 charges are very useful for arbitrary solutes in force-field simulations.
Spectroscopy of nuclear motion

Within the Born-Oppenheimer approximation, we can define the potential energy surface and then it is possible to formulate and solve Schrödinger equation for nuclear motion (as opposed to electronic motion):

\[- \sum_i^N \frac{1}{2m_i} \nabla_i^2 + V(q) \Xi(q) = E \Xi(q)\]

$N$ – the number of atoms, $m$ – atomic mass, $V$ – the potential energy from the PES as a function of $3N$ nuclear coordinates $q$, $\Xi$ – the nuclear wave function. Solution of this equation allows one to calculate rotational and vibrational spectra.

Rotational spectroscopy

The simplest approach to modeling rotational spectroscopy is the so-called ‘rigid-rotor’ approximation – the geometry of the molecule is assumed to be constant at the equilibrium geometry $q_{eq}$. Then, $V(q_{eq})$ becomes simply a multiplicative constant, so that we may write the rigid-rotor rotational Schrödinger equation

\[- \sum_i^N \frac{1}{2m_i} \nabla_i^2 \Xi(q) = E \Xi(q)\]

$E_0$, the eigenvalue for this equation corresponding to the lowest-energy rotational state, is taken to be the electronic energy for the equilibrium geometry. If restricted to only two particles, the equation is identical in form to the radial component of the electronic Schrödinger equation for the hydrogen atom expressed in polar coordinates about the system’s center of mass.
In rotational spectroscopy of polyatomic molecules, the kinetic energy operator is much more complex, but here we consider only two simple examples. The simplest possible case is a non-homonuclear diatomic (non-homonuclear because a dipole moment is required for a rotational spectrum to be observed). In that case, solution of the equation is entirely analogous to solution of the corresponding hydrogen atom problem, and the eigenfunctions $\Xi$ will the spherical harmonics $Y^m_j(\theta,\phi)$, with eigenvalues given by

$$E_j = \frac{J(J+1)\hbar^2}{2I}$$

$I$ – the moment of inertia:

$$I = \sum_{k} m_k r_k^2$$

For heteronuclear diatomic, rotation occurs exclusively about a single axis passing through the center of mass and perpendicular to the bond:

$$I = \mu r^2_{eq}, \quad \mu - \text{the reduced mass: } \mu = \frac{m_1 m_2}{m_1 + m_2}$$

Because the wave functions are the spherical harmonics, each rotational level is $(2J + 1)$-fold degenerate (over the quantum number $m$); the lowest rotational level has a rotational energy of zero, consistent with the earlier assumption that the total energy associated with this level is just the electronic energy of the equilibrium structure. Selection rule – transitions occur only between adjacent levels: $\Delta J = \pm 1$. The energy change observed for transition from level $J$ to level $J + 1$ is

$$\Delta E = \frac{\{(J+1)[(J+1)+1] - J(J+1)\} \hbar^2}{2I} = \frac{2(J+1)\hbar^2}{2I}$$
When probed spectroscopically, the absorption frequency $\nu$ is

$$\nu_J = \frac{\Delta E}{\hbar} = \frac{2(J+1)\hbar^2}{2hI} = 2(J+1)B$$

$B$ – the molecular rotational constant:

$$B = \frac{\hbar}{8\pi^2 I}$$

Non-linear molecules are more complicated than linear – they are characterized by three separate moments of inertia. For instance, in molecules possessing an axis of rotation that is three-fold or higher in symmetry, the two moments of inertia for rotation about two axis perpendicular to the high symmetry axis will be equal. Example: fluoromethane ($C_{3v}$ symmetry) – there is one moment of inertia, $I_A$, about the symmetry axis A, and there are two equal moments of inertia, $I_B$ and $I_C$, about the axes perpendicular to axis A. In this particular case, $I_B > I_A$ – the molecule is called a prolate top. In the case of a prolate top, the rotational eigenvalues are given by

$$E_J^K = \frac{J(J+1)\hbar^2}{2I_B} + K^2\left(\frac{1}{I_A} - \frac{1}{I_B}\right)\frac{\hbar^2}{2}$$

$K$ – the quantum number, running over $-J$, $-J + 1$, ..., $J - 1$, $J$, expressing the component of the angular momentum along the highest symmetry axis. The selection rules for a rotational transition are $\Delta J = \pm 1$, and $\Delta K = 0$, therefore, again, we can write

$$\nu_J = 2(J+1)B \quad B = \frac{\hbar}{8\pi^2 I_B}$$
Less symmetric molecules require a considerably more complicated treatment, but in the end their spectral transitions are functions of their three moments of inertia – prediction of rotational spectral lines depends only on the moments of inertia, and hence only on the molecular geometry. Thus, any method that provides good geometries will permit an accurate prediction of rotational spectra within the regime where the rigid-rotor approximation is valid. Since even very low levels of theory can give fairly accurate geometries, rotational spectra are quite simple to address computationally, at least over low rotational quantum numbers. For higher-energy rotational levels, molecular centrifugal distortion becomes an issue – more sophisticated solution of the rotational Schrödinger equation is required.

In GAUSSIAN output file, rotational constants are printed after Cartesian coordinates:

Standard orientation:

<table>
<thead>
<tr>
<th>Center Number</th>
<th>Atomic Number</th>
<th>Atomic Type</th>
<th>Coordinates (Angstroms)</th>
<th>Rotational constants (GHZ):</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>0</td>
<td>0.393029 -1.689341 0.000000</td>
<td>232.8871195 4.7233220 4.6294298</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>0</td>
<td>-0.257142 -0.613540 0.000000</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>0</td>
<td>0.000000 0.613670 0.000000</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>0</td>
<td>-0.168030 1.765936 0.000000</td>
<td></td>
</tr>
</tbody>
</table>
If frequency calculations are carried out using the ‘freq’ command, the output also contains moments of inertia:
Principal axes and moments of inertia in atomic units:

\[
\begin{array}{ccc}
1 & 2 & 3 \\
\end{array}
\]

EIGENVALUES -- 7.74942 382.09149 389.84092

**Vibrational spectroscopy**

When we think about chemical thermodynamics and kinetics, it is convenient to picture a molecule as a ball rolling on a potential energy surface. In this simple model, the exact position of the ball determines the molecular geometry and the potential energy, and its speed as it rolls in a frictionless way determines its kinetic energy. Quantum mechanical particles are different from classical ones. One of the most important differences – they are subject to the uncertainty principle.

One consequence of the uncertainty principle – polyatomic molecules, even at absolute zero, must vibrate – within the simple ball and surface picture, the ball must always be moving, with a sum of potential and kinetic energy that exceeds the energy of the nearest minimum by some non-zero amount. This energy is contained in molecular vibrations.

Transitions in molecular vibrational energy levels typically occur within the IR range of frequency spectrum. IR spectroscopy has a long history of use in structure determination.

Vibrational frequencies also have other important uses, for example in kinetics and computational geometry optimization, so their accurate determination has been a long-standing computational goal.
One-dimensional Schrödinger equation

\[
\left[ -\sum_{i}^{N} \frac{1}{2m_i} \nabla_i^2 + V(q) \right] \Xi(q) = E \Xi(q)
\]

Let’s take again the simplest possible case, the diatomic molecule. When restricted to the vibrational motion alone, the equation is a function of only a single variable, the interatomic distance \( r \). Solutions of differential equations of only a single variable are typically straightforward and the only challenge here is that we do not know exactly the potential energy function \( V \). However, we can compute \( V \) point by point and those points may then be fit to any convenient analytic function, polynomial, Morse, etc., and the one-dimensional Schrödinger equation can be solved to yield eigenfunctions and eigenvalues.

In a more complicated polyatomic molecule, the potential is a function of \( 3N \) variables. If \( N \) is not small (> 3, 4), it becomes tedious to generate \( V \) pointwise. Moreover, neither the fitting of \( V \) to an analytic form nor the solution of the resulting multidimensional differential equation is at all trivial – the approach described for diatomic is rarely used to compute vibrational data for larger molecules.

To simplify the problem: we are primarily concerned only with regions of the PES relatively near to the minimum, so we may take advantage of Taylor expansion to simplify our construction of \( V \).
Harmonic oscillator approximation

For our simple diatomic case, using Taylor expansion for the potential energy truncated at second order, the vibrational Schrödinger equation transformed to internal coordinates becomes

\[
\left[ -\frac{1}{2\mu} \frac{\partial^2}{\partial r^2} + \frac{1}{2} k (r - r_{eq})^2 \right] \Xi(r) = E \Xi(r)
\]

\(\mu\) - the reduced mass; \(r\) – the bond length; \(k\) – the bond force constant, i.e., the second derivative of the energy with respect to \(r\) at \(r_{eq}\). This equation is the quantum mechanical harmonic oscillator equation. Its eigenfunctions are products of Hermite polynomials and Gaussian functions, and its eigenvalues are

\[
E = \left(n + \frac{1}{2}\right) \hbar \omega \quad \omega = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}}
\]

\(n\) – the vibrational quantum number. The selection rules for the QM harmonic oscillator permit transitions only for \(\Delta n = \pm 1\) – the energy separation between any two adjacent levels is always \(\hbar \omega\), the predicted frequency for any allowed absorption is simply \(\nu = \omega\). So, in order to predict the stretching frequency within the harmonic oscillator equation, all that is needed is the second derivative of the energy with respect to bond stretching at \(r_{eq}\), \(k\).

The importance of \(k\) has led to considerable effort to derive analytical expressions for second derivatives, and they are now available for HF, MP2, DFT, QCISD, CCSD, MCSCF and some other levels of theory, although they can be quite expensive at some of the more highly correlated levels of theory.
It is important to address the errors introduced by the harmonic approximation. These errors are intrinsic to the truncation of the Taylor expansion, and will remain even for an exact level of electronic structure theory. The most critical difference – real bonds dissociate as they are stretched to longer and longer values. Thus, the separation between vibrational levels narrows with increasing vibrational quantum number, and the total number of levels is finite. By contrast, the harmonic oscillator has an infinite number of levels, all equally spaced.

An ‘exact’ harmonic frequency will always be greater than the true frequency.

For the polyatomic case, we must carry out a multi-dimensional Taylor expansion:

$$\left[ -\sum_{i} \frac{1}{2m_i} \frac{\partial^2}{\partial x_i^2} + \frac{1}{2} (x - x_{eq})^T H(x - x_{eq}) \right] \Xi(x) = E \Xi(x)$$

\(x\) – the vector of atomic coordinates, \(x_{eq}\) defines the equilibrium structure, \(H\) – the Hessian matrix.

It is quite difficult to solve this equation in its original form, but by a transformation into a unique set of mass-dependent spatial coordinates \(q\), it is possible to separate the \(3N\)-dimensional equation into \(3N\) one-dimensional Schrödinger equations, identical to the one-dimensional harmonic oscillator equations – force constants and reduced masses in those equations are defined by the action of the transformation process on the original coordinates.

Each component of \(q\) – molecular vibration referred to as a ‘normal mode’ for the system. With each component there is an associated set of harmonic oscillator wave functions and eigenvalues that can be written entirely in terms of square roots of the force constants found in the Hessian matrix and the atomic masses.
Because the equation is over the full $3N$ coordinates, the transformed coordinate system $q$ includes three translational and three rotational (two for linear molecules) ‘modes’. The eigenvalues associated with these modes are typically very close to zero; the degree to which they are close to zero can be regarded as diagnostic of how well optimized the structure is in terms of being at the local minimum geometry.

An important technical point: the first derivatives in the Taylor expansion disappear only when the potential is expanded about a critical point on the PES (since then the gradients are all zero) – there is little or no value in a frequency calculation for a particular geometry under the harmonic oscillator approximation unless the geometry in question was optimized at the same level of theory.

The harmonic oscillator equation is valid not only for minima but also for other stationary points. However, then there will be one or more normal mode force constants that will be negative, corresponding to motion along the modes that lead to energy lowering. As the frequencies are computed from the square roots of the force constants that will be negative, this leads to an imaginary frequency (often these are called negative frequencies but that is not correct). Frequency calculations – diagnostic of the nature of stationary points: all real frequencies – local minimum; one imaginary frequency – a transition state.

Let’s now consider accuracy of different methods for calculations of vibrational frequencies.
<table>
<thead>
<tr>
<th>Level of theory</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MO theoretical methods</strong></td>
<td></td>
</tr>
<tr>
<td>HF/6-311G(3df,2p)</td>
<td>144</td>
</tr>
<tr>
<td>MP2/6-31G(d,p)</td>
<td>99</td>
</tr>
<tr>
<td>CCSD(T)/6-311G(3df,2p)</td>
<td>31</td>
</tr>
<tr>
<td><strong>LSDA functionals</strong></td>
<td></td>
</tr>
<tr>
<td>SVWN/6-31G(d,p)</td>
<td>75</td>
</tr>
<tr>
<td><strong>GGA functionals</strong></td>
<td></td>
</tr>
<tr>
<td>BLYP/6-311G(d,p)</td>
<td>59</td>
</tr>
<tr>
<td>BPW91/6-311G(d,p)</td>
<td>69</td>
</tr>
<tr>
<td>PWPW91/6-311G(d,p)</td>
<td>66</td>
</tr>
<tr>
<td>mPWPW91/6-311G(d,p)</td>
<td>66</td>
</tr>
<tr>
<td><strong>Hybrid functionals</strong></td>
<td></td>
</tr>
<tr>
<td>BH&amp;HLYP/6-311G(d,p)</td>
<td>100</td>
</tr>
<tr>
<td>B1LYP/6-311G(d,p)</td>
<td>33</td>
</tr>
<tr>
<td>B1PW91/6-311G(d,p)</td>
<td>48</td>
</tr>
<tr>
<td>mPW1PW91/6-311G(d,p)</td>
<td>39</td>
</tr>
<tr>
<td>B3LYP/6-311G(d,p)</td>
<td>31</td>
</tr>
<tr>
<td>B3PW91/6-311G(d,p)</td>
<td>45</td>
</tr>
<tr>
<td>mPW3PW91/6-311G(d,p)</td>
<td>37</td>
</tr>
</tbody>
</table>
HF theory shows the poorest performance (AM1 and PM3 are in general somewhat worse than HF with a moderate basis set). MP2 shows a significant improvement over HF, but substantial error remains. CCSD(T) and some of the hybrid levels of DFT show the highest accuracies. In general, the BLYP combination is more accurate than BPW91, whether pure or hybrid in formulation, but PWPW91 is nearly as accurate as BLYP, again whether pure or hybrid in formulation.

Of some interest in the error analysis is the degree to which the error is systematic. Although HF errors are large, they are very systematic. HF overemphasizes bonding, so all force constants are too large, and thus so are all frequencies – application of a constant scaling factor to the HF frequencies improves their accuracy enormously. Scott and Radom (1996) studied this issue in detail for eight different levels of theory using a database of 122 molecules and 1066 fundamentals (i.e., measured, anharmonic vibrational frequencies).

### Scale factors and post-scaling errors in vibrational frequencies

<table>
<thead>
<tr>
<th>Level of theory</th>
<th>Scale factor</th>
<th>RMS error (cm$^{-1}$)</th>
<th>Outliers %</th>
</tr>
</thead>
<tbody>
<tr>
<td>AM1</td>
<td>0.9532</td>
<td>126</td>
<td>15</td>
</tr>
<tr>
<td>PM3</td>
<td>0.9761</td>
<td>159</td>
<td>17</td>
</tr>
<tr>
<td>HF/3-21G</td>
<td>0.9085</td>
<td>87</td>
<td>9</td>
</tr>
<tr>
<td>HF/6-31G(d)</td>
<td>0.8953</td>
<td>50</td>
<td>2</td>
</tr>
<tr>
<td>HF/6-31G(d,p)</td>
<td>0.8992</td>
<td>53</td>
<td>3</td>
</tr>
<tr>
<td>HF/6-311G(d,p)</td>
<td>0.9051</td>
<td>54</td>
<td>3</td>
</tr>
<tr>
<td>MP2/6-31G(d)</td>
<td>0.9434</td>
<td>63</td>
<td>4</td>
</tr>
<tr>
<td>QCISD/6-31G(d)</td>
<td>0.9537</td>
<td><strong>37</strong></td>
<td>2</td>
</tr>
<tr>
<td>BLYP/6-31G(d)</td>
<td>0.9945</td>
<td>45</td>
<td>2</td>
</tr>
<tr>
<td>BP86/6-31G(d)</td>
<td>0.9914</td>
<td>41</td>
<td>2</td>
</tr>
<tr>
<td>B3LYP/6-31G(d)</td>
<td>0.9614</td>
<td><strong>34</strong></td>
<td>1</td>
</tr>
<tr>
<td>B3PW91/6-31G(d)</td>
<td>0.9573</td>
<td><strong>34</strong></td>
<td>2</td>
</tr>
</tbody>
</table>
Outliers – number of frequencies still in error by more than 20% of the experimental value after scaling.

Even though the scaling factor required for the HF/6-31G(d) level is substantial, reducing every frequency by more than 10%, the final accuracy is quite high – better than considerably more expensive MP2. The pure DFT functional BLYP requires essentially no scaling.

Including HF character in hybrid DFT results in proportionately too high predictions in vibrational frequencies, although the scaling procedure is very effective. The errors in the semiempirical levels are quite high, and scaling is only modestly helpful. Results from molecular mechanics can also be of reasonable accuracy if the molecules addressed contain only functionality well represented in the force-field training set. A comparison of MM3 and MMFF94 over a test set of 157 frequencies from organic molecules gave RMS errors of 57 and 60 cm$^{-1}$, respectively.

In comparing complete theoretical spectra to complete experimental spectra, there can be a large number of lines – it is helpful to compare not only the absorption frequencies themselves but also the intensities of the absorptions. In experiment, they are usually reported as strong, medium, or weak, although in careful experiments absorption cross-sections can be measured accurately. In computation, the prediction of IR intensities can be accomplished using the mixed second derivatives of the energy with respect to geometric motion and an external electric field – estimation of the changes in the dipole moment as a function of the vibrations, which is what IR intensities are proportional to. The actual computed values tend to be no better than qualitative in the absence of using a very complete basis set and accounting for electron correlation.
An alternative experiment that measures the same vibrational fundamentals subject to different selection rules – Raman spectroscopy. Raman intensities are more difficult to compute than IR intensities – a mixed third derivative is required to approximate the change in molecular polarizability with respect to the vibration that is measured in experiment – more sensitive to the level of theory.

**NMR spectral properties**

NMR is probably the mostly widely applied spectroscopic technique in modern chemical research used for structure determination and kinetics measurements. NMR measurements assess the energy difference between a system in the presence and absence of an external magnetic field. For a chemical shift measurement on a given nucleus, there are two magnetic fields of interest: the external field of the instrument and the internal field of nucleus. The chemical shift is proportional to the second derivative of the energy with respect to these two fields, and it can be computed using the following equation:

$$\frac{\partial^2}{\partial X^2} \langle \Psi | H | \Psi \rangle = \langle \Psi | \frac{\partial^2 H}{\partial X^2} | \Psi \rangle$$

$X$ – the magnetic field contribution to the Hamiltonian, $H$ – the complete Hamiltonian which includes magnetic field terms.

The integrals that need to be computed in this case are more complex because, unlike the electric field, which perturbs the potential energy term of the Hamiltonian, the magnetic field perturbs the kinetic energy term (it is the motion of the electrons that generates electronic magnetic moments). The nature of the perturbed kinetic energy operator is such that an origin must be specified defining a coordinate system for the calculation – ‘gauge origin’.
The magnetic field is independent of the choice of the gauge origin and so are the computed magnetic properties if the wave function used is exact. To reduce artifacts associated with the gauge origin, there are two different approaches. The older method employs gauge-including atomic orbitals (GIAOs) as a basis set (London 1937). By an incorporation of the gauge origin into the basis functions themselves, all matrix elements involving the basis functions can be arranged to be independent of it. The second method is the ‘individual gauge for localized orbitals’ (IGLO), where different gauge origins are used for each localized MO (Schindle and Kutzelnigg, 1992). Modern implementations of GIAO are somewhat more robust, but it is possible to obtain good results with both methods.

GAUSSIAN command lines for NMR GIAO calculations:

# hf/6-311+g** nmr=giao
# mp2/6-311+g** nmr=giao
# b3lyp/6-311+g** nmr=giao

Of course, geometry has to be optimized before NMR calculations. It is recommended to use at least triple-ζ basis sets with diffuse and polarization functions. For molecules including heavy atoms, if the core electrons of the heavy atoms are represented by an ECP, it is not in general possible to predict the chemical shift. However, ECPs may be an efficient choice if the only chemical shifts of interest are computed for other nuclei. Another issue – the influence of relativistic effects. In terms of computing absolute chemical shifts, relativistic effects can be very large in heavy elements. For relative chemical shifts, since relativistic effects are primarily associated with core orbitals that do not depend much on chemical environment, the effect is typically much smaller. Nevertheless, accurate calculations involving atoms beyond the first row transition metals are still a challenge.
Chemical shifts

Experimental chemical shifts are reported in parts per million (ppm) to make them independent of the external magnetic field strength. They are usually reported as values relative to some standard compound, e.g., tetramethylsilane, which is often used for $^1$H and $^{13}$C. To compute a relative chemical shift, one must carry out a separate calculation for the reference species.

Errors in absolute and relative chemical shifts at various levels

<table>
<thead>
<tr>
<th>Level of theory</th>
<th>$\text{MAE}_{\text{abs}}$</th>
<th>$\text{MAE}_{\text{rel}}$</th>
<th>in ppm</th>
</tr>
</thead>
<tbody>
<tr>
<td>HF</td>
<td>20.1</td>
<td>17.4</td>
<td></td>
</tr>
<tr>
<td>MP2</td>
<td>7.8</td>
<td>4.1</td>
<td></td>
</tr>
<tr>
<td>LDA</td>
<td>27.0</td>
<td>24.7</td>
<td></td>
</tr>
<tr>
<td>BLYP</td>
<td>23.3</td>
<td>14.9</td>
<td></td>
</tr>
<tr>
<td>BP86</td>
<td>12.8</td>
<td>10.3</td>
<td></td>
</tr>
<tr>
<td>B3LYP</td>
<td>31.3</td>
<td>24.3</td>
<td></td>
</tr>
<tr>
<td>PBE1PBE</td>
<td>10.4</td>
<td>7.5</td>
<td></td>
</tr>
</tbody>
</table>

For molecules composed of only first-row atoms, heavy-atom chemical shifts can be computed with a fair degree of accuracy. Even HF theory gives acceptable accuracy in most instances and some improvements occur in favorable instances from DFT (but neither LDA nor B3LYP) and MP2. MP2 is quite accurate, but at relatively high cost. The mean absolute errors for absolute chemical shifts are larger than those for relative chemical shifts.
The high anisotropy of multiple bonds makes the chemical shifts of the atoms involved quite sensitive to the level of theory, especially for nitrogen and oxygen atoms. $^1$H NMR chemical shifts are more difficult to work with because they span a fairly modest spectral range, ~15 ppm in typical chemical environments. For a test set of 80 organic molecules, reasonable results for $^1$H chemical shifts were demonstrated and scaling improves agreement with experiment.

The calculation of spin-spin coupling is less routine than the calculation of chemical shift – additional complications are due to two local magnetic moments as opposed to one moment. The most commonly reported couplings in experiment are proton-proton couplings and this is the most complicated case for theory. Accurate prediction requires very flexible basis sets – spin-spin coupling calculation is still a challenging task.