Invited Comment

Molecular simulation by knowledgeable quantum atoms

Paul L A Popelier
Manchester Institute of Biotechnology (MIB), 131 Princess Street, Manchester, M1 7DN, Great Britain, UK
School of Chemistry, University of Manchester, Oxford Road, Manchester M13 9PL, Great Britain, UK
E-mail: pla@manchester.ac.uk

Received 26 December 2015
Accepted for publication 12 January 2016
Published 24 February 2016

Abstract
We are at the dawn of molecular simulations being carried out, literally, by atoms endowed by knowledge of how to behave quantum mechanically in the vicinity of other atoms. The ‘next–next-generation’ force field that aims to achieve this is called QCTFF, for now, although a more pronounceable name will be suggested in the conclusion. Classical force fields such as AMBER mimic the interatomic energy experienced by atoms during a molecular simulation, with simple expressions capturing a relationship between energy and nuclear position. Such force fields neither see the electron density nor exchange-delocalization itself, or exact electrostatic interaction; they only contain simple equations and elementary parameters such as point charges to imitate the energies between atoms. Next-generation force fields, such as AMOEBA, go further and make the electrostatics more accurate by introducing multipole moments and dipolar polarization. However, QCTFF goes even further and abolishes all traditional force field expressions (e.g. Hooke’s law and extensions, Lennard-Jones) in favor of atomistic kriging models. These machine learning models learn how fundamental energy quantities, as well as high-rank multipole moments, all associated with an atom of interest, vary with the precise positions of atomic neighbors. As a result, all structural phenomena can be rapidly calculated as an interplay of intra-atomic energy, exchange-delocalization energy, electrostatic energy and dynamic correlation energy. The final QCTFF force field will generate a wealth of localized quantum information while being faster than a Car–Parrinello simulation (which does not generate local information).

Isn’t it enough to see that a garden is beautiful without having to believe that there are fairies at the bottom of it too? (Douglas Adams).

Keywords: QTAIM, quantum chemical topology (QCT), force field, multipole moments, protein, IQA, kriging

(Some figures may appear in colour only in the online journal)

1. Introduction

The scope of this journal is that of a general reader, which spurred the author to write this contribution in a lighter style, perhaps even chattier, than used in any of his previous perspectives [1–3] or book chapters [4–9]. While such a light
style can take the pronounced form [10] of musing on a prize just won, interwoven with a personal scientific journey, the present article will adopt another type of pronounced form: the subsections will be very short and almost self-contained. This style hopefully improves readability and the conveyance of a point or opinion. Rest assured, the interested reader can still consult the original references mentioned above, which contain technically precise statements peppered with equations. Also, some material of quantum chemical topology (QCT) published [11] in an earlier volume of this journal is complementary to the current article. Finally, a drastic decision caused all equations to be omitted.

The main sections of this article, however, are simple and standard: after this introduction there is the main body containing the ‘stream of consciousness’ explaining the title of this article and then there is the conclusion. The middle section contains future prospects as well as brief descriptions of work in progress.

2. Molecular simulation by knowledgeable quantum atoms

2.1. Molecular simulation

Molecular simulation is a huge field and probably the ultimate destiny of the whole of computational chemistry, even in connection with the development of algorithms and concepts. Condensed matter simulations are challenging: they demand the rapid yet accurate evaluation of the energy of a system of many thousands of atoms, varying over long time scales. Calculating the quantum mechanical energies during the simulation puts a grave restriction on the length and size of the simulations. Only so-called force fields overcome this huge computational demand but only at the cost of parameterization and loss of accuracy.

Any problem, ranging from protein folding to the atomistic workings of a solar cell or a more efficient battery, is investigated by molecular simulation. The latter should perhaps be more correctly called the ‘atomistic simulation of condensed matter’ because the simulation may contain non-molecular species, such as (solvated) atomic ions, or ionic crystals. Small charged species are a challenge for force fields because of the ions’ strong electric fields. These fields cause substantial polarization, which can be anisotropic, that is, of varying strength depending on direction. The nature and range of species that a molecular simulation tackles gives an early and important clue about how to calculate the required energies. We will discuss this problem later in section 2.10 on perturbation theory. For now, we only point out that it is conceptually cleaner to think of condensed matter as an aggregation of atoms rather than one of molecular entities (that is, in the widest sense, so including molecular ions and inorganic complexes). In other words, it is the atoms are that the well-defined basic units of chemical matter; they are chemically indestructible.

Capturing the nature and behavior of atoms in the presence of other atoms is a strategy for energy evaluation that surely cannot go wrong. Yet another way of putting this idea is that thinking of chemical matter in terms of Lewis diagrams appears natural and innocent but comes with a price. This price is the actually artificial distinction, and hence computational treatment, between bonded atoms and non-bonded atoms. This distinction has a huge imprint on the architecture of familiar force fields such as AMBER. We will argue below that it is conceptually leaner, safer, more streamlined and future-proof to think of atoms interacting with other atoms through physical well-defined energy contributions, regardless of where these atoms are situated in the total aggregate. Again, this idea will be worked out in quantitative detail in later subsections.

2.2. Bonded versus non-bonded interactions

Interatomic interactions are typically classified as either bonded or non-bonded, which is a natural consequence of the omnipresent Lewis diagram. Let us first focus on the bonded part, which is described by its own type of energy terms.

In force field parlance, bonded atoms participate in a so-called 1,2 interaction. Atoms that are separated by two bonds are said to interact by a 1,3 interaction. In general, 1,n interactions occur between atoms that are separated by n−1 bonds. In classic force fields, all 1,2; 1,3 and 1,4 interactions are treated by ‘bonded potentials’. This architecture is inspired by the molecular mechanics view that bonds can be viewed as little springs with stiffness roughly proportional to the bond order. This structural design sufficed to treat alkanes as embodied in the MMn (n = 2, 3, 4) family of force fields [12, 13]. A question that is not obvious at first but turns out to be very important is this: Why do bonded atoms not interact electrostatically within the ansatz explained above?

It is only when we consider 1,4 interactions (and higher, i.e. n > 4) that, suddenly, interatomic electrostatic interaction appears. Indeed, classical force fields describe atomic electron densities as mere point charges and calculate their electrostatic interaction via the familiar Coulomb law. We will discuss the crudeness and limited accuracy [14] of the point charge approach at short range in section 2.12. Here we want to emphasize that there is a non-physical dichotomy in the treatment of bonded and non-bonded interactions. The cause is most likely the Lewis diagram.

The Lennard-Jones potential, which in the context of interatomic forces in force fields combines interatomic dispersion and Pauli-like repulsion, adds to this dichotomy. Both type of interactions are universal to chemical matter, and physically act between any two atoms, whether bonded or not. However, in classical force fields, the Lennard-Jones potential only acts between non-bonded atoms. This dichotomy is not a pedantic detail. The (arbitrary) switching off of interactions that actually exist, means that other energy terms (i.e. the bonded ones) are burdened with expressing the missing energy contributions. Even worse, these bonded terms do not even pretend to take care of dispersion, as nothing in their name or mathematical shape reveals this fact.

The main point of this subsection is that the imprint of the Lewis diagram onto the architecture of the force field is
large. This imprint is also vulnerable to hydrogen bonding, which spoils the artificial division of energy contributions into bonded and non-bonded terms. Hydrogen bonding is disruptive, both in an intermolecular and intramolecular context. Hydrogen bonding can take place when a linear molecule curls up and forms a ring containing a hydrogen bond, which is common in carbohydrate chemistry, for example. The covalent character associated with a hydrogen bond then suddenly turns the otherwise non-bonded 1,2 interaction into a weak 1,2 interaction. Equally, a water molecule, for example, can unexpectedly appear near a solute, from its liquid phase, and form an intermolecular hydrogen bond that strongly influences the preferred conformation of the molecule that this water interacts with. Hydrogen bonding is a major effect in condensed matter simulation, especially in life science.

The issues discussed above invite one to think of condensed matter (and even a single molecule in the gas phase) as an aggregate of atoms, with ever fluctuating interactions, and of waxing or waning covalent character. If one adds reactions to the simulation problem to be addressed, then the invitation is even more prominent. The question is now how to define such an atom while it sits inside a system. Should one hold to the idea that a free atom can be spherical and should remain so while part of a system?

2.3. How to define an atom in a system then?

The problems caused by the Lewis diagram urges one to ‘think atom’ and establish a definition of an atom inside a system, which can be a single molecule, a crystal, a liquid, a solution, or any aggregate of atoms. There is no unique solution to this problem but we will follow an elegant route that has withstood the test of time since 1972, when it was first proposed [15]. Originally, the atoms proposed at that time were called virial fragments, but now they are called topological atoms. The name virial fragment refers to the (quantum mechanical) virial theorem [16] of Slater. This theorem shows that the potential energy of a molecule is trivially related to its kinetic energy, after correction for the residual forces acting on the molecule’s nuclei. Similarly, in 1972 it was shown that the potential energy of a topological atom is also trivially related to the kinetic energy of that atom, provided the atom is in a stationary point on the potential energy surface, i.e. all the forces on the nuclei vanish. Thus the virial fragment was born, but it turned out that these fragments had peculiar, non-trivial shapes.

Figure 1 provides a quick peek at the virial fragments or topological atoms in the amino acid threonine (capped by peptide bonds at both termini). For now, we just recognize that these atoms are ‘bubbles’ or ‘boxes’, and that together they form the original unpartitioned system (e.g. molecule). Later we will briefly explain how these topological atoms are constructed but it is fair to say already now that the atoms actually construct themselves. The prescription for the partitioning is so minimal that the molecule naturally falls apart into topological atoms.

At the root of this virial fragment approach lies the rigorous partitioning of energy into atomic contributions. Being able to do this rigorously offers a solid foundation to the development of a novel force field but the virial fragments had to wait more than three decades before this idea was finally worked out. This development culminated in QCTFF [1], the current and rather bland name for the protein force field under construction that uses these virial fragments. The acronym QCTFF consists of two acronyms actually: QCT and FF. The latter of course refers to ‘force field’ and is used in other force fields such as GAFF [17] or ReaxFF [18]. The partial acronym QCT refers to quantum chemical topology, which is a collective name for a number of approaches that all share one common idea: the spatial partitioning of a system by means of a gradient vector field.

2.4. The gradient vector field

The idea of a gradient vector field goes back to Faraday who in 1851 visualized, in his well-known experiment, the magnetic lines of forces by means of iron filings that form a pattern of curves. Each iron filing is like a little needle that locally points in the direction of the magnetic field. The type of gradient vector field we encounter in QCT consists of a set of curves in three-dimensions (3D), the vast majority of which terminate at a so-called attractor. This idea is illustrated in figure 2, which shows two gradient paths terminating in the red attractor. Clearly, a gradient path has a direction: it always moves from a lower value of a 3D scalar function to a higher value. In each point of the gradient path, this direction is determined by the gradient vector. This vector is marked as ∇S in figure 2, where S refers to the scalar function of interest. This function can be the (quantum mechanical) electron density ρ, its Laplacian [19], or the electron density of a molecule, for example.
localization function [20, 21], just to name a few; for an extended list of functions see box 8.1 in [5].

The gradient vector is always tangent to the gradient path. Similarly, one can see the gradient path as a succession of very short gradient vectors. Another property of a gradient path is that it pierces through all 3D contour surfaces of constant value in \( S \) in an orthogonal way. This orthogonality is a consequence of the same property of the gradient vector \( \nabla S \), which always travels through the 3D scalar function \( S \) in the direction of steepest ascent.

The application of the gradient vector field idea to the electron density is at the heart of the quantum theory of atoms in molecules (QTAIM) [22–25]. The gradient vector is the parameter-free instrument that automatically partitions a system into (topological) atoms. Figure 3 shows how one of the simplest molecules, that is, the diatomic LiH, naturally falls apart into a hydrogen atom and a lithium atom, while both inside the diatomic. The gradient vector field is shown only inside the lithium atom in order not to clutter the picture. The field’s pattern is reminiscent of a spider web, imaging the spider at the nucleus.

2.5. The relationship between the gradient vector field and contour lines

A glance at figure 3 confirms that the gradient paths are everywhere orthogonal to the constant electron density contour lines. In fact, the gradient vector field can be regarded as a mathematical dual to the set of contours. Both this set and the vector field contain the same information and each can be re-constructed from the other. However, it may be easier to observe something in one object of the duality than in the other. The gradient vector field clearly shows how a nucleus dominates a portion of space by attracting to the nucleus all gradient paths present in that portion. If one was allowed to travel only along a gradient path then one would be trapped inside the lithium atom when starting the journey anywhere in the middle right hand side of figure 3.

It is also possible to understand how QCT partitions LiH by looking at the topological nature of the contour lines, although this route is harder than via the gradient vector field. The outer contour lines all encompass the whole molecule, while the inner contour lines encircle individual nuclei, that is, either hydrogen or lithium but not both. There must be a contour line that is situated exactly at the boundary between the outer and inner contours. Indeed, there is such a contour and part of it is marked by the red lines in figure 3. The black square is positioned exactly where the two inner contour lines just touch. This can be seen as a point where two topological atoms touch. This special point serves as the terminus of a bundle of gradient paths, originating at infinity. This bundle is called the interatomic surface and it forms the boundary between the two atoms. Only two gradient paths are shown in figure 3 though, which form the intersection of the interatomic surface with the plotting plane. In the case of a diatomic molecule with a cylindrically symmetric electron density the interatomic surface is a surface of revolution: it suffices to rotate the shown intersection around the molecular axis to obtain the interatomic surface. In general, interatomic surfaces have non-trivial local (Gaussian) curvature [26], as
studied by differential geometry, but one can obtain an analytical expression [27] for them.

Finally, imagine the electron density in a plane through the Li–H axis to be plotted as a height, and that one could walk in this landscape from Li to H along the ridge that connects the two peaks, each peak associated with a nucleus. The black square along this path corresponds to the point of lowest height along this ridge. If one looked at the two peaks from a distance then this point would naturally mark the boundary between the two peaks. However, drawing a plane through this point, perpendicular to the ridge (i.e. molecular axis) is not the right QCT partitioning. This plane would be a naïve extrapolation of the correct boundary point marked by the little square. The correct interatomic surface is a more sophisticated curved object, one that reflects the variation of the electron density away from the black square. This non-planar surface bounds a subspace that has a well-defined kinetic energy, which one could call a quantum atom [5]. This atom is also a topological atom here.

2.6. The character of the topological partitioning

We pause to emphasize important consequences of the spatial arrangement of topological atoms. From figures 1 and 3 it is clear that there is no empty space or gap between two atoms sharing an interatomic surface, obviously; the atoms literally touch. As a result, each point in space must belong to (at least one) atom. The full repercussions of this fact will be worked out in our lab in connection with protein-ligand docking [2] while using topological atoms. Typical images of this research field show atoms, represented by some van der Waals-like surfaces, often separated by vast amounts of empty space. Such image gives the impression that there are sizeable portions of space that cannot or even should not be allocated to any atom. This view is incompatible with the QCT approach. This dislocation between the traditional view and the QCT view has consequences for the partitioning of energy as well, which can potentially be serious. Indeed, electrostatic energy is mathematically connected to electron density. Thus, if electron density does not belong to an atom then the associated electrostatic energy does not either. This means that energy is unaccounted for, which upsets the overall energy balance, and hence any predictions that energy is a quantum field could make. Finally we should also make clear that the sharp boundary between atoms also means that they do not overlap. This means any (non-boundary) point in space belongs to one atom at most. When combined with the earlier statement of any point belonging to at least one atom, the final result is that any point in space belongs to exactly one atom (unless it lies exactly on the interatomic surface). Overall, we can say that topological atoms exhaust space: their properties are additive. So, if one needs the net charge of a functional group, for example, then one simply adds the charges of the atoms involved; no need to correct for any double counting.

As a prelude to further discussions below on the character of the topological partitioning, we compare the topological atom to a stationary pattern in a flowing river. Figure 4 shows an example of how water can form a (more or less) fixed shape while still in motion. The rock formation determines this shape but also the amount of water (and its speed). The rock corresponds to the nuclear potential in the Hamiltonian. More precisely, each separate stone can be associated with an atomic number Z to mark its size and its position can be described by a nuclear position vector \( \mathbf{R} \). The water embodies the electrons; if their number were reduced then the water pattern would change in response. This change corresponds to ionizing a molecule, which results in the topological atoms looking different to those in the neutral molecule.

2.7. A broader mathematical language

The gradient vector field is part of the mathematical language of bifurcation theory and dynamical systems theory, which are rooted in the huge and rich mathematical branch of topology. Dynamical systems occur in the study of classical mechanical problems such as the pendulum. The area of dynamical systems has its own terminology, which applies to QCT. We have encountered the term attractor and gradient path (or phase curve), but we could have called the interatomic surface a separatrix. This is a sharp boundary between two regimes; in the case of the pendulum this boundary divides the regime of swinging and that of full rotation. We also could have called the attractor (or maximum electron density at the nucleus, or very near it) a critical point. Four possible types of critical point can occur in 3D and the attractor is one type (i.e. maximum). There is also a minimum of course, and two types of saddle points. A saddle point is a critical point where the function is either a minimum or a maximum depending on the direction in which the saddle point is approached. One saddle type is called bond critical point and is shown as a black square in figure 3. More examples of bond critical points appear as little purple spheres in figure 1. The other type of saddle point, dual to the bond
critical point, is the so-called ring critical point. The full classification of critical points and more details on them are given in box 8.4 of [5].

We should mention that each gradient path originates at a critical point and terminates at another one. Interestingly, one can also classify gradient paths, by inspecting which two types of critical points the gradient path links. This classification was achieved [28] in 2003, for the first time.

In mathematics, catastrophe theory [29] is a branch of bifurcation theory in the study of dynamical systems. As a result it also applies to the topological analysis of electron densities, which was first investigated by Collard and Hall in Nottingham, UK, in their 1977 seminal paper [30].

Finally, this subsection is the appropriate place to mention that atomic properties are obtained by integrating a density function of interest over the 3D volume of an atom. There are various independent algorithms that achieve this but they are complex and often delicate. Reviewing them would take up a whole subsection but the interested reader can consult the rather exhaustive introduction of [31] for details.

2.8. The full consequence of no overlap

Having sharp boundaries between atoms is not a mainstream view, yet, one may argue optimistically. The question is why there is resistance to this idea. After all, an atom is just another object in nature and, of course, there are many objects in the world that are delineated by sharp boundaries. For example, the phases in a typical phase diagram have sharp (i.e. non fuzzy) boundaries: the phases do not overlap each other. As another example, thermodynamics upholds a sharp distinction between the system and the surroundings. Also, life itself could not have emerged were it not for well-defined and localized boundaries. In human society too, there are plenty of examples of sharp boundaries: the land masses of France and Germany do not overlap, and the law defines someone as dead or alive, or married or not. These examples do not compel one to adopt the image of a limited system yet, one may argue optimistically. The question is why is it necessary to adopt the image of a finite-boundary atom but they make it easier to embrace this concept.

Yet the chemical establishment continues to think of atoms as fuzzy clouds of electrons that penetrate when forming a molecule. The recovery of the original atoms in a molecular electron density, newly formed from the atomic densities, is then a challenge. It is all very well to hold on to the mental image of two originally isolated atoms and then simply superimpose these two original atoms but then one faces an assignment problem: Which part of the final system belongs to which original atom? The philosophical nature of this problem has been discussed ([1], especially figures 11 and 12) at length about 10 years ago, using a metaphor of colliding galaxies and star density. Now we illustrate the main point again but this time with clouds in the sky.

Figure 5 shows two snapshots of moving clouds. The left frame occurred before the right frame. Note that the dark gray cloud kept its color while going from one frame to the other. This fact is crucial in order to interpret the right picture as two overlapping clouds, i.e. the gray cloud being superimposed on the white cloud. Had we be given only the right frame and with a more uniform white color, then the interpretation of two overlapping clouds would be harder to maintain or even think of in the first place. However, we can think of a gradient vector field emerging from spatially varying water vapor density, in analogy with an electron density gradient vector field (as in figure 3). This vector field allows the separation of the gray and the white cloud. We do not even have to know the colors of the clouds. Thanks to this vector field partitioning, we do not need to know the history of the formation of the whole cloud system. The partitioning does not rely on a mental picture of two original clouds that overlap.

The above insight also applies to electron clouds. Let us regard a molecular electron density as a cloud. The question is then how to interpret a van der Waals complex of two molecules in terms of their original molecules. Which part of the complex’s electron density belongs to one molecule and which part to the other molecule? The traditional interpretation sees the complex as a superposition of two overlapping molecules, which necessarily penetrate each other. Again, one can protest by pointing out that one can only conceive this overlap if one knows what the original isolated molecules looked like. When using QCT, this condition is not necessary: molecule A and molecule B can be identified, when inside the van der Waals complex A...B, without reference to the original state of the isolated molecules. In summary, QCT provides an interpretation of a van der Waals complex that is reference-state-free. The popular energy decomposition analysis (EDA) [32] does not enjoy this property. An exhaustive comparison [33] between EDA (and NBO) and QCT shows the repercussions of this.

Actually, there is a stronger argument to proceed with QCT rather than with an overlap model. This argument will be worked out in the section 2.10, which briefly discusses perturbation theory. For now, we remind ourselves that electrons are indistinguishable. This means that, at short medium and especially short range, one cannot allocate 10 electrons to water molecule A, and allocate another 10 electrons to molecule B, for example. Electrons cannot be labeled, which is equivalent to saying that there are no gray or white clouds in quantum mechanics. The electrons in the van der Waals complex cannot be assigned to either participating molecule.

2.9. No orbitals needed for QCT partitioning

We already observed that QCT is parameter-free, as explained in section 2.4. In section 2.8 we concluded that QCT is also reference-state-free, a point that will be elaborated in section 2.12 where a non-QCT population analysis is discussed. We do not need to know what the original free atoms looked like in order to recover the atoms inside the molecular system. Here we show that QCT is also orbital-free.

A third strength of QCT, in the spirit of Occam’s razor, is that it is orbital-free, as far as the shape of the atom is concerned, to be more precise. Indeed, one can still introduce orbitals after the atoms have been obtained without using orbitals. For example, in the calculation of an (interatomic)
bond order one performs an integration over the atomic volumes of an integrand that contains orbitals. However, we emphasize again that the atoms themselves directly depend on the electron density, no matter how it was obtained, that is, by (i) the LCAO-SCF-MO route using Gaussian, Slater or plane-wave basis functions, (ii) quantum Monte Carlo producing the electron density in a grid, or (iii) high-resolution crystallography. To be clear, it is the topological atom itself that is defined without any reference to orbitals. Of course one can still use this spatial partitioning in combination with orbitals in order to calculate exchange energies between atoms, to give another example.

Many non-QCT partitioning methods do not enjoy the ‘route independence’ detailed above because they do not operate on the electron density itself. A prime example of such a non-QCT method, suffering from orbital dependence, is the popular and mature partitioning protocol of distributed multipole analysis (DMA) \[34\]. This fact causes DMA atomic charges firstly to be highly unstable and unreliable when diffuse gaussian primitives are used, and secondly even evaporate if plane waves were used. The former problem has been overcome \[35\] by essentially introducing less fuzzy boundaries to the DMA atoms. In other words, DMA’s originator eventually realized that there is a natural need to truncate the electron density belonging to an atom. This procedure spoils the computational advantage of the original DMA method. The second problem, however, cannot be overcome.

Finally, we note that DMA lies at the heart of SIBFA \[36\], XED \[37\], EFP \[38\], AMOEBA \[39\], NEMO \[40\], DMACRYS \[41\], which are all sustained and admirable efforts to establish a next-generation multipolar force field, which will hopefully one day replace classical force fields such as AMBER, CHARMM, OPLS, MM3 or GROMOS.

2.10. No perturbation theory

More than a decade ago, it was decided that the topological force field was not to be built nor developed within the context of long-range Rayleigh–Schrödinger perturbation theory \[42\]. This formalism is the cornerstone of the traditional view on intermolecular forces at long range. The main idea is that if molecules are far enough apart, the overlap between their wavefunctions can be ignored. This idea goes back to London in the 1930s. This ansatz has a stronger imprint on next-generation force fields (listed just above) than one would expect at first. In practice this overlap is never exactly zero, but it can be sufficiently small such that the corresponding perturbation series converges (although textbooks never quantify this qualifier ‘small’). In the so-called polarization approximation \[43\] the energy of two interacting ground-state molecules A and B is written, to second order, as the sum of (i) the individual molecules’ energy (0th order), (ii) the purely electrostatic energy between A and B (1st order), and (iii) the induction energy of A and of B and the dispersion energy between A and B (2nd order).

All these energy contributions are well-defined, *within this particular formalism*, and can be straightforwardly calculated with the appropriate formulae. However, one can ask if dispersion actually exists, (of the same sense that one can legitimately ask if electron correlation (of electronic structure theory) actually exists. Formally, both concepts are well-defined and can hence be quantified but one should not forget that their existence critically depends on the formalism that
generated them in the first place (and hence depends on the underlying idea or world view). To demonstrate the difficulty with this formalism, one can blur the distinction between intra- and intermolecular interaction. One can ask, for example, what the dispersion energy is between two close-by atoms that lie at either end of a curled chain made up of a single covalent molecule. The trouble here is that the unperturbed Hamiltonian cannot be written as a sum of two Hamiltonians, for each isolated fragment A or B, because there are no two fragments; there is a single covalent molecule. Yet, surely there must be dynamical correlation between these two end-chain atoms.

Moreover, at short-range, Rayleigh–Schrödinger perturbation theory actually breaks down because there is no unique definition of the order of a term in the perturbation expansion. One can still proceed with exchange perturbation theories of which there are two main groups: symmetric methods (e.g. [44]) or symmetry-adapted theories (e.g. SAPT [45]). From a practical point of view, short-range perturbation theory is CPU intensive: one needs to know which molecule B a given molecule A interacts with. This fact contrasts with long-range perturbation theory where no information on the interacting partner is necessary. In other words, one can calculate the polarizability of a given molecule without having to know which other molecule causes the polarization. As short-range perturbation theory is computationally so expensive, one may wonder why not simply use supermolecular wavefunctions? This drastic decision was taken many years ago, at the outset of the QCTFF project. We abandoned perturbation theory and hence the full imprint it has on the design of a force field. We do not think of overlapping entities, entities that penetrate each other, series expansions defining energy contributions nor polarizabilities. So what do we think of?

We think of atoms interacting as malleable boxes of finite volume, wherever they are, that is, within a single molecule or across two molecules. The core of QCTFF’s architecture does not differentiate between intra- and intermolecular interactions. This ansatz is possible because, topologically, the electron density of the overall system partitions itself while dominated by the nuclei, rather than by where the molecules stop or start. This way of working is not adopted by any other next-generation force field mentioned above. Indeed, while targeting condensed matter, QCTFF is ultimately trained by supermolecular clusters in which atoms are always well defined.

2.11. No penetration, no damping functions

There is an extensive literature on damping functions and whole PhD theses are dedicated to this topic. A damping function is an artificial mathematical function that prevents an energy becoming unreasonably large (and even infinite) at short range. A very popular example of such a damping function is that [46] of Tang and Toennies, which operates on calculated dispersion energies. The familiar $C_n/R^n$ energy contributions (where $R$ is the internuclear distance and $n$ is typically 6, 8 or 10) are replaced by $f_n(R)C_n/R^n$ where the damping function $f_n(R)$ is required to tend to 1 for large $R$ and to zero as $R$ approaches 0. The cause for damping functions can be traced back to the so-called penetration energy. This energy is in turn due to a conceptual picture of electron clouds extending infinitely far and thus able to substantially overlap with each other. Note that if electron clouds do not overlap then one has no penetration energy at all, as is the case for QCT.

An illuminating exact calculation can be found in the book [42] of Stone (section 6.4.1). Suppose we are interested in the interaction of a proton and a hydrogen-like atom of nuclear charge $Z$. The electron density for such an atom can be conveniently obtained by solving the 1-electron Schrödinger equation exactly. The potential can then be exactly calculated at the position of the proton, by using the Poisson equation. The potential depends on $r$, the distance from the $Z$-nucleus, and deviates from $-1/r$ by a factor $[1 - \exp(-2\pi r)](\pi Z + 1)$, which is the (exact) damping function in this case. The exponential factor describes the correction to the potential at short distances. One can then calculate the corresponding penetration energy, which yields a negative correction to the interaction energy.

We keep in mind that both the damping function and penetration energy arise from the fact that the point where we want to know the potential at, resides inside the electronic charge density that generates this potential. However, with QCT atoms it is perfectly possible to take a point that is rigorously outside the atom (i.e. the finite object that generates the potential) and calculate [47] the potential there.

2.12. Point charges, population analysis and charge transfer

In polar systems, which are the majority of systems, the electrostatic interaction dominates, and therefore one should try to model it correctly. The standard way of representing the electrostatic interaction is still via point charges and the simple Coulomb law. This would only be accurate (within 0.1 kJ mol$^{-1}$ of the exact answer) if the charges were far apart [48], say 25 Å, depending on the interacting elements actually. Unfortunately, many atoms interact at distances shorter than this one. To model this medium to short range correctly one has two choices: either one adds more point charges, away from the nuclear position, or one adds point multipole moments centered on the nucleus. In fact, one can work with a mixture of these two extremes. However, the third option, which is to search for that elusive point charge that will magically turn out to be correct at short range, is fruitless. This quest has gone on for a long time, however. The spirit of this ill-informed quest leads to the occasional false acquisition of a point charge not being up to the job of generating the correct electrostatic potential. Such mischief has happened to QCT’s atomic monopole moments (i.e. point charges). We argue that QCT monopole moments measure charge transfer and nothing more. If the electrostatic potential [47, 49] or interaction [50] is to be modeled exactly then QCT atomic multipole moments need to be added.

A naïve solution to the problem of assigning electronic charge to a given atom, while within a system, was proposed [51] by Hirshfeld in the late 1970s. He suggested that a given
atom inside a molecule should obtain a share proportional to the molecular function, which is valid in a larger system. For example, a carbon in a sp² hybridized state, surrounded by functional groups would not exist. The question, however, is how to quantify transferability. Curiously, this question is under-researched, certainly in connection with the classical force fields (e.g. AMBER). Here the concept of transferability is embodied in the so-called atom type, a packet of information associated with an atom in a specific environment. For example, a carbon in a sp² hybridized state, surrounded by other carbons will have such and such a charge, and \( C_6 \) coefficient (to be used in a combination rule to parameterize the dispersion part in a Lennard-Jones potential). Atom types

2.13. Transferability

Transferability is the capacity of a small system to provide information that is valid in a larger system. For example, imagine that one calculates the atomic charge of oxygen in propanol and finds out that it differs by only 1% to that in octanol. If one is happy with this difference, then the propanol value will be used to represent the oxygen’s charge in octanol and most likely in higher alkanols (nonanol etc) too. It is tempting to push the practical quantitative boundary of transferability: one could save time by calculating the oxygen’s charge in an even smaller molecule, such as methanol, but then find that the percentile error increases to say 7%, which may be deemed too high. As a result one settles the slightly larger system of ethanol, which strikes the right balance with the still acceptable error of say 3%.

It is fair to call the concept and realization of transferability the zeroth cornerstone of any force field. Transferability is the essence of any force field because without it, one would have to build a force field for each large molecule targeted. This obviously defeats the object because then one would need at least 50 000 force fields, one for each of the proteins known in the human body, for example. A first cornerstone would determine if the force field had point charge electrostatics or multipolar electrostatics. A second cornerstone would focus on the way of implementing polarization. Before any of these cornerstones can be put in place, one has to guarantee first a quantitatively sufficient degree of transferability. In summary, before any force field design strategy can be worked out, one has to verify that the transferability indeed operates.

Chemistry teaches us that transferability exists; if not, functional groups would not exist. The question, however, is how to quantify transferability. Curiously, this question is under-researched, certainly in connection with the classical force fields (e.g. AMBER). Here the concept of transferability is embodied in the so-called atom type, a packet of information associated with an atom in a specific environment. For example, a carbon in a sp² hybridized state, surrounded by other carbons will have such and such a charge, and \( C_6 \) coefficient (to be used in a combination rule to parameterize the dispersion part in a Lennard-Jones potential). Atom types
are determined a priori in classical force fields, by chemical intuition or just ad hoc fixing of a potential. Atom types should emerge from a transferability study, which would be the correct and future-proof way forward. To the best of our knowledge, atom types where for the first time computed, in our lab [62–64]. In this work, we looked at atomic transferability from the point of view of intrinsic (non-interacting) atomic properties such as atomic energy, volume, population and the magnitude of the dipole, quadrupole, octopole and hexadecapole moment. This systematic study was based on a cluster analysis of 760 atoms, all drawn from the 20 natural amino acids. This systematic work [64] exposed inconsistencies in the designation of AMBER atom types, such as over-differentiated nitrogen atom types.

An important question is which physical quantity needs to be monitored for transferability. The example at the beginning of this subsection focused on atomic charge, which is important, but we already know that nucleus-centered atomic charge alone cannot accurately represent electrostatic interaction. The work of Nalewajski et al has shown [65] that one obtains the best atomic transferability, in the information-theoretical sense, when going from the isolated state to the molecule itself, when ρ(r) is given in terms of Hirshfeld atomic densities. However, preliminary results using the QCT energy decomposition scheme fed with Hirshfeld atoms shows [58], that, in an energetic sense, these atoms are considerably less transferable than QCT atoms. Is energy not more important to be transferable than any other physical quantity? Ultimately, it is energy (and its gradient or atomic forces) that controls the structure and dynamics of any system being simulated, as well as its thermodynamic and kinetic properties. If so, then maximum transferability based on information theory, while failing to satisfy energetic transferability, is a red herring.

Acknowledging the importance of energy, we started monitoring the atomic electrostatic potential [66], which is a half-way house between multipole moments and interatomic electrostatic energy. Indeed, the potential is a special case of interaction energy, in which the interaction partner is fixed as a proton. Next, transferability was investigated [67] in terms of electrostatic interaction energy, which invokes a ‘probe’. Focusing on the water trimer and serine...(H2O)3 as pilot systems, it turned out that the atomic multipole moments in serine are more transferable than those of the water cluster, with the exception of the hydroxy1 group.

Moving closer to ultimate transferability means moving beyond the interatomic Coulomb energy and focusing on the intra-atomic energy. This is a sensitive probe, which includes the atomic kinetic energy, the Coulomb and exchange energy of the electrons within the atom interacting with themselves, and the intra-atomic electrons interacting with the nucleus inside the atom. The typical magnitude of the intra-atomic energies for the five possible elements, are huge as shown by the following ballpark figures: nitrogen −140 000 kJ mol−1, oxygen −195 000 kJ mol−1, carbon −100 000 kJ mol−1, hydrogen −1200 kJ mol−1 and sulphur −1042 000 kJ mol−1, respectively. Very recently, transferability was systematically tested [68] on seven homo-oligopeptides (glycine, alanine, serine, threonine, cysteine, valine and leucine). As usual in transferability tests, the value of a physical quantity of a smaller system is compared with that in the larger system, keeping the geometries the same for the nuclei in common between the two systems. Think of a single amino acid as the smallest system (capped by peptide bonds at both termini), which then appears flanked by two amino acids, one at the C-terminus and one at the N-terminus. This system serves as the larger system, leading to deviations ranging from 2.6 kJ mol−1 (for glycine) to 39.2 kJ mol−1 (for alanine) for the peptidic nitrogen. Given the other large deviations found in this range, a single amino acid does not suffice to accurately model a tripeptide, and hence a protein. We thus extend the tri-peptide by two more amino acids, again one amino acid flanking each terminus. Now we find that the Cα, Hα, N, O and S atoms in a tri-peptide are energetically comparable to those in their penta-peptide configurations, within about 2 kJ mol−1 in absolute value. Across all five elements, this energy difference is on average ∼0.3 kJ mol−1. On average, the tri-peptide sample systems represent a ∼8.2 Å ‘atomic horizon’ around the central atoms of interest.

The ultimate transferability test has not been conducted yet, as it would involve all four possible fundamental energy types mentioned in the abstract: intra-atomic energy, exchange-delocalization energy, electrostatic energy and dynamic correlation energy. The latter energy is only available for very small pilot systems, both in our lab [69] and elsewhere [70]. However, the first three energies are available and can be summed such as to provide a single number for each atom. Such a study is in progress in our lab, starting with large water clusters (up to about 10 Å in radius). The conclusion that single amino acids are not accurate enough as providers of (transferable) atomic information is also being acted upon, by systematically investigating tripeptides.

### 2.14. Multipolar electrostatics and convergence

This subsection is quite short, not because it lacks importance but because much on this topic has been divulged at great length in other publications [4, 6, 9]. Multipolar electrostatics is at the heart of QCTFF, and was therefore given much attention from the outset. The mathematical origin of multipole moments lies in the Laplace expansion of the electrostatic energy. The formula expressing the latter contains a factor 1/|r1 − r2|, where r1 and r2 describe the position vectors of two interacting electron densities. This factor is an ‘entangled’ expression, which means that both r1 and r2 are needed at the same time in the evaluation of this mathematical expression. However, factorization of this expression undoes this entanglement. The multipolar expansion offers this factorization and thereby enables the calculation of quantities involving only r1 and quantities involving only r2. The advantage of this procedure is that the interaction between two electron densities can be calculated in two stages: first one calculates the multipole moments of each electron density (expressing its characteristics) and then one calculates the interaction between these multipole moments. The price paid
for this convenience is that the latter calculation, that is, the multipole expansion itself, may not converge.

Figure 6 illustrates what is meant by convergence or lack thereof. Formal convergence means that the multipolar expansion indeed reaches the exact interatomic Coulomb interaction, in a stable manner, as more and more higher order terms are added to the expansion. Note that the exact energy is obtained by six-dimensional integration and hence without multiple moments. It can be shown that this formal convergence is elusive if the respective atoms are in contact. The multipole expansion itself, may not converge.

This is the case for DMA but it is not the case for QCT. In other words, QCT can offer perfect convergence. An example of this desirable behavior is shown in the left panel of figure 6, where two hydrogens (engaged in a 1, 4 interaction) converge to the exact interaction energy of \(-9.2 \text{ kJ mol}^{-1}\). The x-axis represents the interaction rank \(L\), which is the sum of the ranks of the interacting multipole moments incremented with unity. For example, a dipole moment (rank 1) interacting with an octopole moment (rank 3) corresponds to \(L = 5\). In the next example (figure 6, right panel) convergence breaks down for very high \(L\) values: the energy diverges to unusable values. This phenomenon is sometimes called asymptotic convergence. However, the good news is that there is a stable plateau between \(L = 4\) and \(L = 12\), where the multipolar energy differs very little from the exact value. This pseudo-convergence behavior is practically useful in molecular dynamics simulations where hydrogen bond interactions need to be modeled in an accurate and stable manner.

The topic of multipolar convergence has been thoroughly researched [47, 49, 50, 72–79] within QCTFF but will not be discussed in detail here. However, it would be nice if all this work was finally taken note of in classic reference works such as [42], the second edition of which keeps mentioning the same erroneous statement on the poor convergence of topological atoms. The message to take away is that 1,2 and 1,3 interactions do not converge, in general, but 1,5 and higher 1, \(n\) interactions invariably do. The 1,4 interactions are a gray zone.

Systematic findings [49] harvested while studying the small protein crambin, provide a wealth of information on universal convergence behavior between the most common elements in life science (C, H, N, O and S). One of five key questions answered in that paper is this: for a given convergence energy error \(\Delta E\) and a given pair of atoms A and B, how does the interaction rank \(L\) at which the energy has converged, change with increasing internuclear distance \(R\)? This is actually a five-dimensional function, involving the quantities \(R, L, \Delta E\), and the qualifiers A and B. This work serves as a guide for QCTFF’s practical implementation in terms of multipolar electrostatics. Short-range electrostatics cannot be achieved by multipole moments, but is obtained exactly, via six-dimensional integration over two interacting atoms.

2.15. Non-electrostatic energy contributions

Clearly, not all interactions in a molecule are electrostatic in nature. For example, this type of interaction cannot explain why two neutral atoms are strongly held together in a covalent molecule. In the world of molecular quantum mechanics, the electrostatic interaction is the most classical in nature and therefore easier to explain than non-electrostatic interactions. It is important to represent the electrostatic interaction accurately but if this were the only interaction present, then matter would collapse.

The covalent bond can be explained by the so-called exchange delocalization. Curiously, this very strong effect stems from the fact that electrons are indistinguishable. Strictly speaking, one cannot label an electron and then put it in an orbital. However, one strangely does exactly that but then corrects for this labeling mistake by putting that same electron in all other possible orbitals. While exhausting all possible permutations, one must ensure that the total wavefunction is antisymmetric. It is then possible to calculate an exchange energy between two atoms and use this energy in QCTFF. The most documented and applied branch of QCT that defines non-electrostatic energies is called interacting quantum atoms (IQA) [80].

Figure 6. The exact Coulomb interaction energy between two (topological) atoms (dashed vertical line) compared with the energy obtained by multipolar expansion according to interaction rank \(L\) (as explained in the main text) for (left) H\(_5\) and H\(_6\) in glycine [71], and (right) O\(_5\) and H\(_2\) in a water dimer [72].
The roots of the IQA energy decomposition lie in a paper [74] published in 2001, which calculated the Coulomb potential energy between two atoms without using the atomic virial theorem. This means that, at no stage, the atomic kinetic energy was invoked (which is obtained after integration over the 3D atomic volume). Instead, the interatomic Coulomb energy was obtained by a 6D integration, essentially carried out simultaneously over two atoms. The atomic virial theorem is only practically valid, and therefore useful, when the system it belongs to is a stationary point on a potential energy surface. Hence, equilibrium geometries qualify but non-equilibrium geometries do not. In other words, the orthodox QTAIM version can only provide atomic energies for equilibrium geometries, which is a severe shortcoming. The '6D two-electron algorithm' lifted this important restriction. This new avenue was further exploited in our lab [81–83] but, in parallel, worked out to perfection by an excellent lab in Oviedo, Spain, EU, who established IQA and wrote its original concomitant program [84] called PROMOLDEN. The incorporation of 6D integration in the highly optimized and well maintained program [85] AIMAll made IQA popular, which is now being applied to an increasing number of important chemical questions. Very recently, IQA has also been adopted as the vehicle to provide the non-electrostatic energy contributions to QCTFF, whereas before this strategy was not clear [86].

IQA calls the interatomic exchange energy \( V_X(A, B) \), which has been studied [87] systematically for a raft of simple but varied compounds. Much can be learnt from plotting \( V_X(A, B) \) versus the internuclear distance \( d(A, B) \). In these plots, \( 1n \) interactions typically appear as isolated islands, roughly an order of magnitude apart (when in kJ mol\(^{-1}\)) for non-delocalized materials (so excluding metals). There is also the overall trend of \( V_X(A, B) \) decreasing with increasing internuclear distance. The interactions that defy this broad trend or the island formation deserve attention. For example, hydrogen bonding in a D–H...A systems (D = donor, A = acceptor), turns out to involve the three atoms D, H and A, rather than just H and A. Indeed, both \( V_X(D, A) \) and \( V_X(H, A) \) values are anomalous in that they do not appear in the expected place in the \( V_X(A, B) \) versus \( d(A, B) \) plot: \( V_X(A, B) \) is about an order of magnitude too strong given its \( d(A, B) \) value. Another example follows from a study of saturated alkanes, where the weakest \( V_X(H, H) \) interactions are found when two hydrogen atoms are in gauche–gauche conformations, whereas the strongest are found in trans-planar arrangements. The larger than expected values in staggered ethane can be understood as the real-space analogue of hyperconjugation; in fact, as a modern localized quantum chemical measure, the \( V_X(H, H) \) 'trans planar' effect could even replace hyperconjugation. Finally, we mention two more properties of \( V_X(A, B) \): (i) it is transferable to a high degree (for example, \( V_X(C, H) = 749 \pm 4 \) kJ mol\(^{-1}\) for saturated alkanes), and (ii) it is a measure of covalent energy because of its proportionality to bond order (divided by the internuclear distance) as proven in [7].

IQA also defines an intra-atomic energy, which lump together three contributions: (i) the Coulomb interactions of electrons with each other, provided they are within the same atom, (ii) the corresponding exchange energy \( V_X(A, A) \), and (iii) the kinetic energy. This quantity was already discussed in section 2.13 in connection with its transferability properties. Although it is of the order of a hundred thousand of kJ mol\(^{-1}\) for second period elements, we are only interested in 'wobbles' in this energy, i.e. much smaller energy changes (order of a hundred kJ mol\(^{-1}\)) in response to changes in molecular geometry or substituent effects. A dramatic but perhaps helpful metaphor is to think of the total energy of the Sun as representing the enormous values of the intra-atomic energy, while the Sun’s protuberances (or solar winds) are the wobbles. The intra-atomic energy plays an important role in stereo-electronic effects such as the rotation barriers in biphenyl [88] or the gauche effect [89].

Finally, IQA defines a dynamic correlation energy term \( V_c(A, B) \), where A and B can be equal. As mentioned in section 2.13 this energy is today only available for very small pilot systems such as H\(_2\), LiH, or H\(_2\)O, because it is so expensive to compute. Only in 2015, energies have been obtained [70] for the coupled cluster method, for MP2 wavefunctions in our lab [69] (drawing the second-order density matrix from the program GAUSSIAN) and for full configuration interaction and complete active space multi-configuration calculations (CAS[n, m]), already a decade ago [80]. In H\(_2\), \( V_c(H, H) \) is only 6% of the value of \( V_X(H, H) \) and amounts to about 40 kJ mol\(^{-1}\). The quantity \( V_c(A, B) \) may be relatively small but it is the key to calculating and understanding (intermolecular or interfragment) dispersion energy. This energy is part of the ubiquitous van der Waals interactions, which are important between aromatic amino acids in proteins (so called \( \pi \ldots \pi \) stacking). The dream is to be able to make \( V_c(A, B) \) calculations feasible for systems of the size of a phenol-dimer, for example, and thereby avoid having to resort to 'bolt on' procedures such as that of Grimme (and coworkers) [90]. The latter strategy would not fit well with a streamlined (and hence rather elegant, rigorous and future-proof) training protocol for QCTFF, as outlined in the next subsection.

2.16. The machine learning method kriging

This subsection is the longest of all because it is at the core of QCTFF. Machine learning is pivotal to how QCTFF is trained and how it is used in production mode, i.e. for geometry optimization and molecular simulation. In spite of this length a whole host of 'details' have been omitted because interested readers can find everything back in the references given. Our research all started with the handling of polarization.

Polarizability is the intrinsic propensity of an electron density to deform itself in response to an external electric field. In the simplest description of polarization, this field is homogeneous, while the polarizability is isotropic (i.e. the same in all directions), and measured by the change in the dipole moment. The reality of polarization is more complex than this description: the field is typically generated by another molecule close by (and produces strongly curved field lines), the polarizability may differ dramatically depending on
the direction, and higher-order multipole moments (quadrupole moment, etc) can be significant. These realities can be handled by introducing anisotropic polarizabilities for moments beyond the dipole moment and details can be found elsewhere [91]. Traditional long-range perturbation theory has two conceptual drawbacks: it handles charge transfer in a rather clumsy ‘bolt on’ manner and it introduces the polarization catastrophe. Finally, a polarization scheme governed by a polarizability introduces computational overhead during a molecular simulation, having to self-consistently compute on-the-fly the new multipole moments resulting from each polarization process.

In order to deal with all this in a conceptually more minimal and streamlined way, a drastically different approach was proposed [92] by this lab almost a decade ago. This new approach did not focus on the polarizability itself but on the effect that it has, after the polarization process had done its work. This idea was adopted by QCTFF and caused it to predict the new multipole moments an atom must adopt when in the presence of a new given environment. The question is then twofold: (i) Which procedure can predict these new multipole moments, and (ii) What is the input to this procedure?

The answers turn out to be minimal but effective: only a machine learning method can cope with the complexity of the atomic environment and the input to this method is no more than the nuclear coordinates of the environment. Machine learning methods are more complicated than simple linear (or even nonlinear) least-squares regression. Essentially they are mappings between inputs and outputs, and need to be trained to achieve a successful mapping. This success is measured by an objective function, which is essentially the difference between a real output and a predicted output. The smaller this difference (or prediction error) the better the machine learning engine has done. It is best to test the engine with an external test set, i.e. with data that have not been seen before by the engine (i.e. not part of the training set).

The inspiration for machine learning methods is bewildering because it can be an ant colony, a human brain, a swarm of birds, natural selection, or a geostatistical method to find a precious commodity in a landscape. The latter method is called kriging [93–95]; it is able to predict geographical regions of high concentration of some precious material, based on a relatively small number of measurements. QCTFF adopted kriging in order to predict the effect of polarization. However, the original method used [96] was an artificial neural network but this approach was abandoned [97] in favor of kriging because this method proved to be more accurate although more CPU intensive. An important advantage of kriging is that it handles a high-dimensional input space better than neural networks. This is important because QCTFF aims at serving the simulation of condensed matter and therefore sufficiently large clusters of atoms need to be included in the training set. QCTFF has adopted kriging as a sophisticated way to handle polarization (of all multipole moments) and charge transfer. The latter is just the special case of the monopole moment being treated as ‘polarization’. Hence, unlike in perturbation theory, QCTFF handles charge transfer and multipolar polarization in a unified [98] and streamlined manner.

The basic idea behind kriging is to predict the value of a function at a given point by computing a weighted average of the known values of the function in the neighborhood of the point. The technical details of kriging, as used within the QCTFF context, are given elsewhere [99]. Here we just mention that one of the central ideas of kriging is that it maximizes the so-called likelihood $L$. The likelihood is a function returning the probability of observed outcomes. A detailed discussion illustrating this concept is given in [7] where it is illustrated by means of a biased coin, whose tossing outcomes are governed by a parameter $p_H$. If $p_H = \frac{1}{2}$ then the coin is fair. Now imagine tossing this coin three times. The likelihood $L$ that the coin was fair, given the observation of two heads being up (HHT), is $p_H p_H (1 - p_H) = 1/8$, i.e. $L = 0.125$. A quick calculation shows that $p_H = 2/3$ is the parameter value that maximizes $L$. This result is intuitively clear: a coin that is biased towards ‘heads up’ by a factor 2 over ‘tail up’, i.e. $p_H = 2/3$, maximizes the probability of the observed outcomes HHT. Kriging generalizes this principle to arbitrarily high dimension. Kriging maximizes the parameters of the likelihood function that govern the density of precious material in a landscape, such that the observations made are the most probable to have been made. Once the parameters are determined in this way, the ‘trained’ probability function captures the underlying distribution of the precious material. Not only will it return exactly the measured values but will also interpolate between those values. If the kriging model is given an input point far from its training set points, then it will return the mean output associated with its training points.

In the QCTFF context, kriging predicts a property (i.e. output) of a given atom (e.g. monopole moment or charge, dipole moment, intra-atomic self-energy, exchange energy with its full atomic environment) as a function of the nuclear coordinates of the atoms surrounding this given atom (i.e. input). The maximization procedure is technically complex [100], and involves machine-learning-aided optimization of the concentrated log-likelihood, which is in turn obtained analytically from $L$. The former optimization is achieved by so-called particle swarm optimization [101] or by differential evolution (DE) [102].

Several systems have been successfully kriged [87, 98–99, 103–109] in the development of QCTFF. The so-called $S$-curve (because of its typically sigmoidal shape) displays the full performance of a Kriging model. An $S$-curve is a validation tool and shows the error between exact and predicted property, for hundreds of examples drawn from an external test set of molecular configurations. Note that this validation is more detailed (and honest) than merely reporting an average error. An $S$-curve is a cumulative distribution function and from it one can read off which percentage (y-axis) of test configurations corresponds to an energy prediction error (x-axis) up to any desired value. For example, let us set this value to 4 kJ mol$^{-1}$, an error threshold referring to the magical 1 kcal mol$^{-1}$, an arbitrary unit still used today. The point where the $S$-curve intersects the vertical 4 kJ mol$^{-1}$ line,
one could read off the (cumulative) percentage of test configurations (which contain all local energy minima found for isoleucine) that return an error of less than 4 kJ mol$^{-1}$, which could be 70%, for example. The $S$-curves unashamedly display the worse errors obtained.

Kriging models can handle quite large systems, such as the helical alanine decapetide (103 atoms), which has [110] its total electrostatic energy predicted with a mean error of 6.4 kJ mol$^{-1}$. This work also shows that atom types can be constructed from similar atoms within the helix. Atoms can share a kriging model with a reduced number of inputs because they share a local chemical environment. Transferability tests have shown that transferable models give an error of only 6% when predicting multipole moments of an atom outside the training set. This work opens an avenue to constructing QCTFF for a whole protein, beyond training for the training set. This work only mentions the so-called concept of 'seeds and distortion' of training geometries. Originally, we obtained up to a few thousand training geometries by distorting the equilibrium geometry according to normal modes. The mathematical derivation of normal modes typically assumes a reference configuration that is situated at a stationary point on the potential energy surface. Then we discovered [109] that an analysis of a second-order inhomogeneous differential equation can generalize the derivation of normal modes to non-stationary point geometries. Importantly, this generalization enables the creation of new types of training sets, each built around so-called seeds or non-stationary points of interest. Amongst currently investigated seeds there are: (i) amino acid geometries from the Protein Data Bank, (ii) molecular dynamics snapshots of water clusters, (iii) samples along an intrinsic reaction path in an enzyme, and (iv) a manually constructed set of geometries.

Finally we should mention that kriging is also used in other force field design projects [111, 112]. However, machine learning is not used in other important next-generation force fields such as AMOEBA.

3. Conclusion

A long and large effort to establish a truly novel force field approach has been presented. The various components have been painstakingly prepared and validated over the years and are now coming together. Much upsampling and systematic testing still need to be done, which will lead to a user-friendly protocol of training and use. Soon it should be possible to carry out meaningful simulations with a methodology that literally sees matter as atoms that interact as packets of localized quantum information. It is our hope that this QCTFF methodology will be useful in many applications, including biocatalysis and disease mechanisms. QCTFF will be able to
predict accurate structure and dynamics, hand in hand with realistic chemical insight, which is not guaranteed with traditional force fields. In line with an optimistic future for QCTFF, it perhaps deserves a more pronounceable name, which we suggest to be FFLUX.

Acknowledgments

I thank the EPSRC for awarding an Established Career Fellowship and Mr Peter Maxwell for preparing figure 1.

References

[1] Popelier P L A 2015 Int. J. Quantum Chem. 115 1005–11
[2] Popelier P L A 2012 Curr. Top. Med. Chem. 12 1924–34
[3] Popelier P L A 2012 AIP Conf. Proc. 1456 261–8
[4] Popelier P L A 2012 Modern Charge-Density Analysis vol 14 ed C Gatti and P Macchi (Germany: Springer) pp 505–26
[5] Popelier P L A 2014 The Chemical Bond: Fundamental Aspects of Chemical Bonding ed G Frenking and S Shaik (New York: Wiley) ch 8 pp 271–308
[6] Popelier P L A 2012 Drug Design Strategies: Computational Techniques and Applications vol 20 ed L Banting and T Clark (Cambridge: Royal Society of Chemistry) ch 6 pp 120–63
[7] Popelier P L A 2016 The Chemical Bond—100 Years Old and Getting Stronger ed M Mings (Germany: Springer)
[8] Popelier P L A 2015 Challenges and Advances in Computational Chemistry and Physics dedicated to Applications of Topological Methods in Molecular Chemistry ed E Allikhany et al (Germany: Springer)
[9] Popelier P L A 2005 Structure and Bonding Intermolecular Forces and Clusters vol 115 ed D J Wales (Heidelberg: Springer) pp 1–56
[10] Gatti C 2013 Phys. Scr. 87 048102
[11] Pendas A M, Francisco E and Costales A 2013 Phys. Scr. 87 048106
[12] Allinger N L 1998 Encyclopedia of Computational Chemistry vol 2 ed P von Rague Schleyer (New York: Wiley) pp 1028–33
[13] Allinger N L, Chen K and Lii J H 1996 J. Comput. Chem. 17 642–68
[14] Cardamone S, Hughes T J and Popelier P L A 2014 Phys. Chem. Chem. Phys. 16 10367–87
[15] Bader R F W and Beddall P M 1972 J. Chem. Phys. 56 3320–9
[16] Slater J C 1933 J. Chem. Phys. 1 687
[17] Wang J, Wolf R M, Caldwell J W, Kollman P A and Case D A 2005 J. Comput. Chem. 15 1157–74
[18] van Dun A C T, Dasgupta S, Lorant F and Goddard W A 2001 J. Phys. Chem. A 105 9396–409
[19] Bader R F W and Essen H 1984 J. Chem. Phys. 80 1943–60
[20] Becke A D and Edgecombe K E 1990 J. Chem. Phys. 92 5397–403
[21] Silvi B and Savin A 1994 Nature 371 683–6
[22] Bader R F W 1985 Acc. Chem. Res. 18 9–15
[23] Bader R F W 1990 Atoms in Molecules A Quantum Theory (Oxford: Oxford University Press)
[24] Popelier P L A 2000 Atoms in Molecules An Introduction (London: Pearson Education)
[25] Matta C F and Boyd R J 2007 The Quantum Theory of Atoms in Molecules (New York: Wiley)
[26] Popelier P L A 1996 Can. J. Chem. 74 829–38
[27] Popelier P L A 1994 Theor. Chim. Acta 87 465–76
[28] Malcolm N O J and Popelier P L A 2003 J. Comput. Chem. 24 437–42
[29] Thom R 1975 Structural Stability and Morphogenesis English edition (Reading, MA: Benjamin)
[30] Collard K and Hall G G 1977 Int. J. Quantum Chem. 12 623–37
[31] Popelier P L A 2011 J. Phys. Chem. A 115 13169–79
[32] Bickelhaupt F M and Baerends E J 2000 Rev. Comput. Chem. 15 1
[33] Pendas A M, Blanco M A and Francisco E 2009 J. Comput. Chem. 30 98–109
[34] Stone A J 1981 Chem. Phys. Lett. 83 233–9
[35] Stone A J 2005 J. Chem. Theor. Comput. 1 1128–32
[36] Gresh N, Cisneros G A, Darden T A and Paqueual J-P 2007 J. Chem. Theory Comput. 3 1960–86
[37] Vinter J G 1994 J. Comput. Aided Mol. Des. 8 653–68
[38] Gordon M S, Slipchenko L, Li H and Jensen J H 2007 Annu. Rev. Comput. Chem. 3 177–93
[39] Ponder J W et al 2010 J. Phys. Chem. B 114 2549–64
[40] Holt A, Bostrom J, Karlstrom G and Lindh R 2010 J. Comput. Chem. 31 1583–91
[41] Price S L, Leslie M, Welch G W A, Habgood M, Price L S, Karamertzis P G and Day G M 2010 Phys. Chem. Chem. Phys. 12 8478–90
[42] Stone A 2013 The Theory of Intermolecular Forces (Oxford: Oxford University Press)
[43] Hirschfelder J O 1967 Chem. Phys. Lett. 1 363–8
[44] Hayes I C and Stone A J 1984 Mol. Phys. 53 83–105
[45] Jezierski B, Moszynski R and Szalewicz K 1994 Chem. Rev. 94 1887–930
[46] Tang K T and Toennies J P 1984 J. Chem. Phys. 80 3726–41
[47] Kosov D S and Popelier P L A 2000 J. Phys. Chem. A 104 7339–45
[48] Yuan Y, Mills M J L and Popelier P L A 2014 J. Comput. Chem. 35 343–59
[49] Kosov D S and Popelier P L A 2000 J. Chem. Phys. 113 3969–74
[50] Popelier P L A, Joubert L and Kosov D S 2001 J. Phys. Chem. A 105 8254–61
[51] Hirsfeld F L 1977 Theor. Chim. Acta 44 129–38
[52] Popelier P L A 2007 Faraday Discuss. 135 3–5
[53] Francisco E, Pendas A M, Blanco M A and Costales A 2007 J. Phys. Chem. A 105 5235–41
[54] Fonseca Guerra C, Handgraaf J W, Baerends E J and Bickelhaupt F M 2004 J. Comput. Chem. 25 189–210
[55] Bultinck P, Van Alsleben C, Ayers P W and Carbó-Dorca R 2007 J. Chem. Phys. 126 144111
[56] Verstraeten T, Ayers P W, Van Speybroeck V and Waroquier M 2013 J. Chem. Theory Comput. 9 2221–5
[57] Becke A D 1988 J. Chem. Phys. 88 2547–51
[58] Francisco E, Pendas A M and Blanco M A 2006 J. Chem. Theor. Comput. 2 90–102
[59] Francil M M, Carev C, Chirlian L E and Gange D M 1996 J. Comput. Chem. 17 367–83
[60] Breneman C M and Wiberg K B 1990 J. Comput. Chem. 11 561–73
[61] Sigfridsson E and Ryde U 1998 J. Comput. Chem. 19 377–95
[62] Popelier P L A and Aicken F M 2003 Chem. Eur. J. 9 1207–16
[63] Popelier P L A and Aicken F M 2003 J. Am. Chem. Soc. 125 1284–92
[64] Popelier P L A and Aicken F M 2003 ChemPhysChem 4 824–9
[65] Nalewajski R F and Parr R G 2000 Proc. Natl. Acad. Sci. USA 97 8879–82
[66] Popelier P L A, Devereux M and Rafat M 2004 Acta Crystallogr. 60 427–33
[67] Rafat M, Shaik M and Popelier P L A 2006 J. Phys. Chem. A 110 13578–83
