

## Improving density functional theory to study trace paramagnetic-containing solids with nuclear magnetic resonance

### Specific Aims:

Improve methods and ease of use for applying DFT (Density Functional Theory) to interpret and predict solid state NMR (Nuclear Magnetic Resonance) data of paramagnetic containing solids.

1. Create a benchmark dataset of DFT predictions of nuclear properties of solid materials containing paramagnetic metals.
2. Identify the root causes of shortcomings in current methods and theory within these calculations.
3. Develop improved theory, functionals, basis sets, and/or corrections to overcome these shortcomings.
4. Develop aids and instructions to facilitate broader use of DFT for NMR by non-expert users.

### Background:

Most of human technology is built on or out of solid inorganic materials containing paramagnetic ions, atoms with unpaired electron spins that are often present at minor or low trace “dopant” concentrations. The paramagnetic ions that introduce these unpaired spins often intentionally or unintentionally influence the properties of the materials. For example, optical properties (such as color) are controlled strongly by paramagnetic additives or defects, catalysts can be enhanced or poisoned by these species, cement must be made with iron impurities to enable its key property of hardening after being mixed with water, and almost all geologic materials can be classified as containing minor concentrations of paramagnetic ions.

At low concentrations, determining the atomic position and arrangement of paramagnetic ions is exceptionally difficult with commonly-used solid inorganic chemistry tools, akin to finding a “needle in a haystack”. During my PhD work, I demonstrated pioneering experimental and analysis techniques that utilized solid-state **NMR (Nuclear Magnetic Resonance)** to observe and quantify the atomic environments and distribution of scarce paramagnetic atoms. The quantitative precision of my work offered new insight into the fundamental chemistry of solid solutions at low concentrations and the site preference of added atoms.

The paramagnetic effects observed in NMR are the result of complex electron-nuclear interactions, and although in some settings broad implications can be determined, most often the richness of the NMR data is so difficult to interpret that only limited conclusions can be produced. An approach to interpreting NMR data is to use **DFT (Density Functional Theory)** or other computational chemistry methods to calculate the chemical system’s electronic structure and use computational predictions to decompose and assign meaning to the experimental data<sup>1</sup>. In paramagnetic systems, this approach has only been demonstrated on a select few materials, none of which contain low concentration paramagnetic atoms<sup>2</sup>. During my PhD, I attempted numerous calculations following state of the art procedures, but never produced results that could be connected to the experimental data. It seemed to indicate that in my low concentration systems, existing computational approaches do not correctly describe the system.

This begged the question, why doesn’t DFT work for low-concentration paramagnetics? And could it be fixed? While seeking expert insight, I managed to intrigue Kieron Burke with my research question, and as of July 2018 I received one year’s funding through a University of California President's Postdoctoral Fellowship to pursue exploratory work on this topic. Although I anticipate making great progress during the coming year, preliminary work already indicates additional levels of complexity that will require a significant time investment to fully characterize and solve. A stimulating aspect of this problem and the proposed research is that it spans from the fundamentals of electronic structure to observations on bulk material compounds. In this regard, the research agendas of Dr. Burke and myself sit on opposite ends of the scientific

problem, which I believe allows this collaboration to be particularly creative as we search for practical solutions.

### **Significance of the research:**

Valuable and common scientific “tools” are ones that 1) **provide informative data** consistently, 2) are **easy to operate**, and 3) provide data that are **straightforward to interpret**. For molecular chemistry and biochemistry questions, NMR meets all three criteria and is understandably a powerhouse technique in these fields. However, in the solid inorganic materials community, the technique is either unknown entirely or delegated to uncommon expert users. Solid-state NMR 1) provides insights into the arrangement of atoms in solids that are unobtainable by other methods and 2) is fairly straightforward to operate, but unfortunately, 3) the technique’s high specificity typically results in complex spectra that are difficult to interpret. A tool for simplifying the complexities of NMR spectra may well enable solid-state NMR to become a major and common technique for studying solid systems.

Recently, great strides have been made in developing and applying computational chemistry to predict NMR data by using an approach termed “GIPAW”<sup>1,3,4</sup>. However, this approach usually falls short of accurately predicting the spectra for paramagnetic solids<sup>1,5</sup>, a significant problem since many technologically-important materials (both natural and synthetic) contain transition or lanthanide metals that introduce a paramagnetic aspect. In some pioneering work with materials containing high concentrations of paramagnetic ions, computational predictions have proven instrumental in understanding the resulting spectra<sup>6,7</sup>. However, these successes are limited, cumbersome to perform, and my own efforts at replicating the work indicate that the approach is not suited for every system. To my knowledge there is no published application of any computational technique to a low concentration paramagnetic system. In a review article on computational predictions of NMR spectra the problem was well summarized: *“There is currently no methodology to predict all of the relevant interactions in paramagnetic solids at a consistent computational level.”*<sup>1</sup>

To return to the characteristics of a good scientific tool, computational chemistry may well allow 3) straightforward interpretations of paramagnetic NMR data. However, it arguably struggles with 1) providing consistently informative data and 2) being easy to operate.

**In summary:** A simple and accurate DFT approach for predicting paramagnetic effects in NMR data would give researchers a powerful new perspective into the atomic structures of Earth’s most common type of solid.

### **Proposed strategy and approach:**

There are many different computational chemistry methods; in this proposal I specifically focus on DFT, as it is the most commonly used approach for predicting spectroscopic properties in solids and is likely to be the most familiar approach for solid state community. To enable DFT to enhance NMR research, the proposed research will (a) develop a library of existing NMR observations for the purposes of testing and benchmarking DFT predictions, (b) characterize specific shortcomings of DFT methods for predicting NMR parameters, (c) develop new approaches or theory to overcome these shortcomings, and (d) develop a practical guide or tutorial to aid researchers in applying DFT approaches to predict low concentration paramagnetic NMR data.

### **Develop a benchmark library**

First, a library of paramagnetic compounds that is representative of solid state systems will be assembled. A preference will be directed towards systems with existing paramagnetic NMR data, especially oxide materials for which there is a growing body of experimental spectra<sup>8-16</sup> and the most accurate theoretical predictions to date<sup>6,7,17</sup>. A variety of materials (e.g., simple oxides,

silicates, phosphates, aluminates), coordination environments (4, 6, and 7), and donor electron cations (Mn, Fe, Co, Ni, Cu, and Lanthanides) with a diversity of bond types, angles, and lengths (all of which strongly influence the electronic structure and resulting spectra) will help generalize when and where improvements in the theory are needed. Since paramagnetic concentrations can range from dilute concentrations (0.01% Fe in  $\text{Mg}_2\text{SiO}_4$ )<sup>11</sup> to major components (NiO)<sup>13</sup>, several specific concentrations will be defined. Assembling this library will enable DFT computations on a consistent set of materials in an organized and prioritized fashion, allowing systematic comparison.

This library also offers opportunities for broader impact, as it presents an organized structure for theoretical and applied researchers to collaborate efficiently. Entries in the test library will be ranked by priority, such that researchers with limited computational resources could benchmark DFT on a simple set of systems that are most representative of broad classes of materials. Compiling the library may indicate if key representative materials are missing from the experimental literature. This may motivate experimental work in this direction, which depending on the size of the work, I could pursue directly or use to establish collaborations with lab groups specializing in the particular materials. Such a library will also provide a holding place for observed but “too complex to interpret” paramagnetic effects in solid-state NMR spectra, since the value of such work is typically minimal in the NMR and solid-state chemistry community but is exactly the type of data that is intriguing for the computational chemistry community. I envision this library to be live and accessible by internet, likely distributed and managed through the solid state NMR sharing project I started ([solidstatenmr.com](http://solidstatenmr.com)).

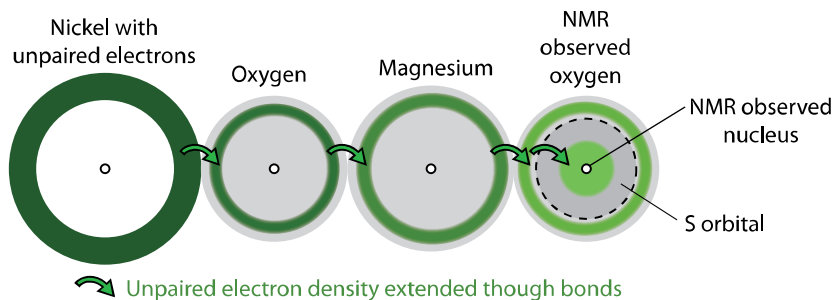
### **Characterization of shortcomings**

To address the shortcomings of DFT, initial efforts will be focused on replicating experimental NMR spectra of highly ranked compounds in the library as an initial benchmark. To date, it is unclear which circumstances produce inaccurate predictions, since publications focus almost exclusively on positive, novel results and only hint at negative results<sup>1,5</sup>. Replication will follow the approach used in the work of Pigliapochi et al.<sup>6</sup>, which splits the calculation into parts accomplished by two different software packages. For each compound, different approaches will be tested across multiple software packages, such as QuantumEspresso<sup>18</sup>, CASTEP<sup>5</sup>, CRYSTAL14<sup>19</sup>, CPK2, and possibly others. These software packages solve the same types of problems, but variations across the software packages (in either the theory used or its implementation) will provide fast analytical insight into which parts of the paramagnetic NMR calculation (which consists of multiple terms summed together) are being computed consistently and yield results most representative of real systems.

### **Improved theoretical methodologies**

Once specific shortcomings have been characterized, I will develop solutions to compensate for these problems. I describe several possible paths to resolve problems below. It is possible that the needed improvements are outside of the expertise of Dr. Burke or myself, in which case we intend to involve additional collaborators as I believe this would be the fastest route to a solution.

I suspect that at a fundamental level, part of the problem lies in DFT's tendency to overly delocalize electrons and insufficient corrections for this behavior. In the paramagnetic effects of interest, NMR is exceptionally sensitive to the transfer of electrons energy “through bonds” (termed the Fermi contact shift). The figure below illustrates these interactions, where the unpaired electron(s) originate from the paramagnetic ion's outermost orbital(s) and is then extended through bonds to a neighboring atom's S orbital where it can directly “contact” the neighboring atom's nucleus. These effects are sometimes observed to be transferred as many as 6 bonds away<sup>13</sup>. In the formulas below, the  $\delta_{\text{FCS}}$  is the Fermi contact shifts contribution to the change in



the NMR signal, a large component of which can be controlled by the isotropic hyperfine coupling constant  $A$ . In a given material and NMR experiment, the only variable in the isotropic hyperfine coupling constant (right equation written below) is the electron wave

function at the nucleus. This means that small errors in the predicted energy far away from the paramagnetic atom could result in drastically different NMR predictions. To further investigate this, a simple system such as a Hubbard dimer<sup>20</sup> or an extended Hubbard chain<sup>21</sup> provides an ideal and fast testing ground for investigating how changes in the localization of electrons around a specific nucleus might influence the resulting paramagnetic NMR data. If these interactions are found to exist, a correction involving HF-DFT, where DFT energies are evaluated on Hartree-Fock densities<sup>22</sup>, may allow for more accurate predicted energies far from the paramagnetic atom.

The Fermi contact shift

$$\delta_{\text{FCS}} = \frac{A}{\hbar} \frac{\langle S_z \rangle_j}{H_0} \frac{1}{\gamma_I}$$

$\hbar$  = Planks constant

$\gamma_I$  = Observed nuclide gyromagnetic ratio

$H_0$  = External magnetic field

$\langle S_z \rangle_j$  = Expectation value of electronic spin

Isotropic hyperfine coupling constant

$$A = -\frac{8}{3} \pi \langle \mu_n \cdot \mu_e \rangle \cdot |\psi(0)|^2$$

$\mu_n$  = Nuclear magnetic moment (observed)

$\mu_e$  = Electron magnetic moment (unpaired electron)

$\psi(0)$  = Electron wavefunction at the nucleus

Another promising route is to merge and apply existing methods that could improve DFT predictions of paramagnetic effects in oxides. For instance, Avezac et al.<sup>19</sup> incorporate accurate predictions of nuclear properties generalized for metals (a closely related phenomenon to oxide paramagnetic shifts) within the framework for predicting NMR data. By pairing this with theory developed for predicting NMR of high concentration phosphates<sup>6,7,17</sup>, it's possible the accuracy of NMR predictions may be improved sufficiently.

Optimizing the fine details of the calculations, such as the exchange and correlation interactions or basis sets, may also be necessary, especially since the best paramagnetic NMR results are calculated in highly optimized systems<sup>6,17,23</sup>. Applications of machine learning may help improve these fine details, especially for paramagnetic properties calculations. Existing application of machine learning techniques include learned functionals<sup>24-26</sup> and smaller corrections that improve overall accuracy or efficiency<sup>27-29</sup>.

### Improved practical methodologies

It is lastly possible that much of the problem in existing DFT calculations lies in knowing the "right" way to run the code. Scientists applying DFT methods—especially experimentalists seeking confirmation of their proposed interpretations—want fast, accurate, and robust results. Running a calculation involves a wide variety of ordinal and numerical parameters (basis set, functional, cutoff, k-point grid, etc.), which must be selected with the accuracy of the investigated property in mind; these ultimately influence the final output. However, selecting optimal parameters is often not straightforward, especially for complex calculations in understudied compositions and structures (the ideal workspace for novel research). As an example, Mondal and Kaupp illustrate how sensitive the NMR data predictions can be with changes to the hybrid functional<sup>30</sup>. Drawing from the improved theory this work will produce, as well as existing DFT sources, I will develop practical tutorials for the use of DFT in solid state NMR systems. The

development of practical guidance (tutorials significantly more descriptive than typical scientific publications) will help researchers gain confidence that time spent doing DFT is likely to yield useful results.

More broadly, there is growing use of online archives for collecting the calculations and results of computational chemistry work from across the globe. These databases, paired with my library of paramagnetic systems, may present an ideal dataset for applying machine learning to assist chemists in this decision making process. While assembling a robust dataset to train across multiple parameters is likely beyond the scope of this postdoc, presenting a framework for the prediction of these parameters may motivate the development of a consistent computational chemistry reporting format to assemble such a dataset.

#### **Other formal and informal aspects:**

Fellowship time to focus on research without other commitments means I can continue to establish myself in the scientific community, supporting my goal of becoming a professor. My long term scientific goal is to develop projects that scale from fundamental theory and basic experimental science to real world practical application (from atom to Earth). The strong recognition of Kieron Burke in the DFT and theoretical chemistry community will be instrumental in helping me establishing theoretical collaborators who have expertise in the area. Dr. Burke's mentorship, as well as UC Irvine's friendly research environment and collaborative nature, will help me gain a deep exposure to theoretical chemistry work in a short time.

I am also inspired by Dr. Beckman's career trajectory. While he was a career-long innovator, many of his scientific breakthroughs seem to have drawn on his early years and involvement as a college-level educator. For instance his hands on approach to tool development, and many of the specific instrument improvements seem to originate from merging a scientific lecture topic with an elegantly simple glassblowing skill, both of which he would have been teaching about. In this spirit, the dedication of the Arnold and Mable Beckman Foundation to Dr. Beckman's life would also provide me with opportunity to continue my involvement in college level education. My focuses is specifically on introducing advanced concepts to students earlier in their educational career so as to motivate further study, and efforts to increase the representation of thinkers from diverse background in the sciences. Currently I am simplify ways to introduce DFT, so community college professors can motivate introductory chemistry students about the technologically-exciting, computer-oriented aspects of chemistry. My efforts towards diversity include being a representative on the American Ceramic Society's Diversity and Inclusion sub-committee, and planned outreach activities at community colleges using my introductory DFT materials.

IP and Property Dispensation Plan: As a requirement for employment with the University of California, and consistent with State of California law, all UC-employees assign ownership of certain inventions generated under their employment with UC to UC. Any licensing decisions are made in accordance with the UC Licensing Guidelines, which provides that any licenses for IP owned by UC include a reservation of rights for other non-profit research institutions. This means that regardless of whether the postdoctoral fellow remains at UC, or moves to another non-profit institution, they will have the rights to use the invention for research and educational purposes.

Should I move to a different institution, materials will be transferred following protocols established by UBMTA.

Data and code generated by this project will be uploaded to publicly available websites, or data repositories.

**Budget.** Salary \$70,000. Benefits: \$10,000. Software and computation time: \$4,000. Conferences \$5,000. Communication \$1,000.