Winter 12-15-2022

Local Spectroscopy Data Infrastructure: Solid State NMR Crystallography with Experiment, First-principal Analysis and Machine learning

He Sun
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Chemistry

Dissertation Examination Committee:

Bryce Sadtler, Chair
Sophia E. Hayes
Erik Henriksen
Richard Mabbs
Courtney Reichhardt

Local Spectroscopy Data Infrastructure: Solid State NMR Crystallography with Experiment, First-Principal Analysis and Machine Learning

by

He Sun

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2022
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

Looking back at the journey that took me from a fresh college graduate to who I am now, I feel both astonished and grateful for how much I have developed, improved, and learned.

None of the stories will ever be possible without my advisor, Dr. Sophia E Hayes. It was her who opened the door of the wonderful world of physical chemistry and solid-state NMR to me and shared her knowledge and experience in such a selfless manner. I learned how to be a good scientist from her. Her passion and dedication to science, and her courage and creativity of always looking for innovative ways to solve scientific problems are inspirational. Without her trust and support, I would never be brave enough to explore the unknown area of combining computational methods with NMR, and this dissertation would never be possible. She is also an excellent teacher and caring mentor. I still remember the good time we spend together when I was the TA in her inorganic class, and I will never forget her encouragement when I was facing trouble with my first paper. Finally, she is also a great mentor and friend in my life, it is fun to hang out with her since we both love good food and tea.

I also want to send my appreciation to all my collaborators.

Dr. Shyam Dwaraknath, if Sophia is my mentor in NMR then you are like my mentor in data science and machine learning. Your guidance and wisdom in computational chemistry are invaluable to me.

Dr. Lisa F. Szczepura, I enjoyed our work together on ^{51}V NMR. I learned a lot from you about synthetic inorganic chemistry. Thank you for all the hard work and collaboration!

I also want to thank Dr. Kristin Persson and Dr. Mark Asta for our collaboration on NMR computations and thank their students Handong Ling, Shivani Srivastava, and Maxwell Venetos for our fun collaboration on DFT calculations. I want to say thank you to Dr. Phil Grandinetti and Dr. Deepansh Srivastava for our work together on Mrsimulator which inspired me to learn more about CI/CD. I want to thank Dr. Blake Hammann for his help on the ^{51}V paper, the project would not be the same without his initial experiments. I need to express my appreciation to Dr. Elijah Thimsen, Dr. Marko Bertmer and Michael Malone for our collaborations on applying computational NMR to various types of solid materials.

Many appreciations to my committee members and mentors, Dr. Bryce Sadtler, Dr. Richard Mabbs, Dr. Alexander Barns, Dr. Courtney Reichhardt, and Dr. Erik Henriksen, your questions, advice, and feedback are the source of reference for me to improve my research. I also want to thank Dr. Jacob Schaefer, he taught me how to interpret NMR beyond the vector model, learning density matrix with him was an eye-opening journey.

To all my lab mates and friends that I spend my daily life with, Dr. Robert Marti, Dr. Erika Sesti, and Dr. Chen Chia-Hsin were good tutors to me when I initially came to the lab. Dr. Shimon Daphna was so kind and generous, I asked her a lot of questions about the basic principles of NMR. Dy. Yvonne Afriyie and I spent time together learning MQMAS. Dr. Jinlei Cui and Dr. Hongyang Zou are good friends of mine, it was fun to chat with them about both science and life.

Dr. Jinlei Cui also taught me a lot about experimental NMR, he is my go-to guy when I have experiment-related questions. I also spent a lot of time filling helium during the pandemic with Dr. Ike West and Dr. Patrick Judge. My deep thanks to my close friends, Jie Chen, Ruidong Jiang, Zhiyao Zhou, Ziqi Wang and Mu Li for our great hangouts together, we keep each other company during the pandemic, good memories shall never fade.

Of course, some snacks and fish chips should be provided to my cat, Mochi, for sacrificing his soft belly when I am in desperate need of a hand warmer.

Finally, to my parents Jun Yuan, Hongbin Sun, and my fiancée Zhanyi Yao, there are no words that can describe the meaning of you guys to my life. I just hope everyone in this world has someone like you guys, like shining stars in the sky, teleporting light to my dark eyes, thousands of thousands of light-years away across the sparse and dark universe with eternal entropy increase.

The research in this dissertation is supported by the following grants and organizations. Chapter 2 is based on work supported by the National Science Foundation (NSF CHE RUI 1401686 to L.F. S.). This work is also supported (for H.S. and S.E.H.) by the National Science Foundation (NSF OAC 1640899). The work made use of computational resources and software infrastructure provided by the Washington University Center for High-Performance Computing (CHPC) at the Mallinckrodt Institute for Radiology. The authors would like to thank Katie Wentz for collecting some of the initial spectral data. Chapter 3 and 4 are supported by the U.S. National Science Foundation (NSF), Award 1640899. The research made use of computational resources and

software infrastructure provided through the Materials Project, which is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division under Contract No. DE-AC02-05-CH11231 (Materials Project program KC23MP). Additional computational resources were provided by the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the NSF (ACI-1053575), as well as the Savio computational cluster resource provided by the Berkeley Research Computing program at the University of California, Berkele.

He Sun

Washington University in St. Louis

August 2022

ABSTRACT OF THE DISSERTATION

Local Spectroscopy Data Infrastructure: Solid State NMR Crystallography with Experiment, First-principal Analysis and Machine learning

by

He Sun

Doctor of Philosophy in Physical Chemistry

Washington University in St. Louis, 2022

Research Advisor: Professor Sophia E. Hayes

Solid-state magnetic resonance (SSNMR) spectroscopy is a powerful tool for obtaining precise information about the local bonding and morphology of materials. The detailed local structure of crystalline materials cannot be easily solved by traditional experimental methods such as X-ray diffraction (XRD). SSNMR combined with first principal calculation methods such as density functional theory (DFT) can be of great use in this research area. The methodology that is called "NMR crystallography" today has been widely applied to the determination of a wide range of solid materials with an increasing amount of computationally simulated NMR spectra. The construction of a well-established computational NMR dataset is thus getting more important. The research presented in this dissertation serves the purpose of applying NMR crystallography to investigate the local structural distortion of ^{51}V in $Ag_2V_2PO_4$ (SVPO), building well-benchmarked SSNMR datasets for both spin ½ species such as ^{29}Si and quadrupolar species (spin > ½) such as ^{27}Al, and constructing machine learning mode for NMR parameters prediction utilizing computational simulated NMR database.

Chapter 1

Introduction

1.1 Basics of Solid-state NMR

In this section, we will discuss the basic principles of solid-state NMR (SSNMR), including the behavior of nuclear spin under a static magnetic field, the rise of Zeeman splitting, and NMR Hamiltonians.

1.1.1 Nuclei in an external magnetic field

Nuclear magnetic resonance (NMR) spectroscopy is a technique that detects the interactions between the nuclear spin and its surrounding environments in the presence of an external magnetic field. All the individual magnetic moments associated with each nuclear spin in the sample are summed up to yield a net magnetization **M** to represent the bulk magnetic moment.

$$\mathbf{M} = \sum_{i} \mu_i \tag{1.1}$$

Here, μ_i is the individual magnetic moment for the i^{th} nucleus in the material. The magnetic moments can also be expressed based on nuclear spin **I** and gyromagnetic ratio γ which is an intrinsic property of a nuclear isotope. For different nuclear species, the value of γ varies.

$$\mu_i = \gamma \mathbf{I_i} \tag{1.2}$$

When the material is placed under a static external magnetic field $\mathbf{B_0}$, which is traditionally designated to be along the laboratory z-direction, $\mathbf{B_0} = (0, 0, B_0)$, the magnetic dipole moment of

the nucleus interacts with the magnetic field, and the originally degenerate nuclear energy levels split into (2I+1) energy states. The energy of each state can be expressed as:

$$E_m = -m\hbar\gamma B_0 \tag{1.3}$$

Here m means the m^{th} magnetic quantum number and it can increase from -I to +I in integer steps. $\hbar$ is the reduced Planck's constant which equals h/2π. For spin ½ nuclei (I = 1/2) such as ^{1}H, ^{13}C, and ^{29}Si, there are two different energy states which are labeled as parallel and antiparallel. The energy gap between the two energy states is:

$$\Delta E = E_{-\frac{1}{2}} - E_{\frac{1}{2}} = \hbar\gamma B_0 = \hbar\omega_0 \tag{1.4}$$

This splitting of the originally degenerate nuclear energy states is called the Zeeman splitting. The ω_0 here is called the Larmor frequency, which corresponds to the frequency a which of the net magnetization **M** precesses around the external magnetic field $\mathbf{B_0}$.

The Zeeman splitting can be detected by NMR due to the small population (or probability to be thermal dynamically strict) difference between different energy states. For spin ½ species, the population differences between the parallel (m=1/2) and antiparallel (m=-1/2) state can be expressed based on the Boltzmann distribution:

$$\frac{p_{-\frac{1}{2}}}{p_{\frac{1}{2}}} = e^{\frac{\gamma\hbar B_0}{kT}} \tag{1.5}$$

Here k is the Boltzmann constant, T is the thermodynamic temperature of the sample. The population of a certain eigenstate p_i can also be derived:

$$p_i = \frac{e^{-E_i/kT}}{\sum_i e^{-E_i/kT}} \tag{1.6}$$

Here E_i is the energy of the eigenstate i.

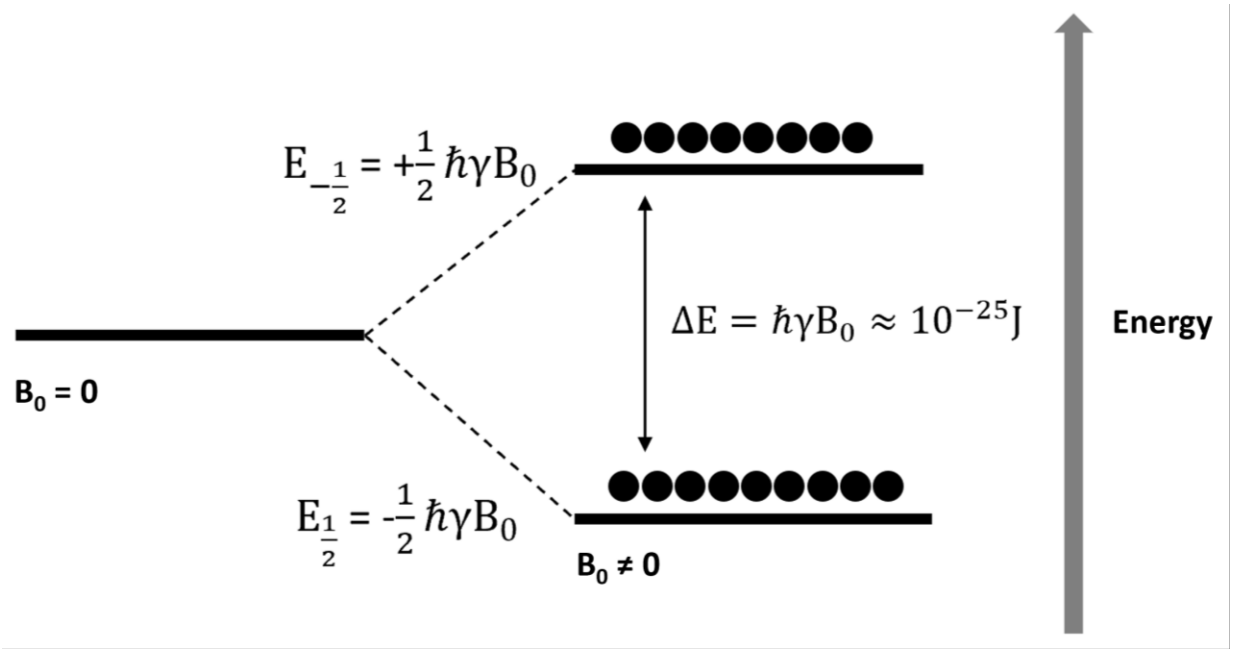


Figure 1.1 Energy scheme of Zeeman splitting for spin ½ nucleus with external magnetic field B_0 applied.

Figure 1.1 illustrates the scheme of the Zeeman splitting for spin ½ species. The energy gap between the two eigenstates is approximately 10^{-25} J[1], which is orders of magnitude smaller than the thermal energy at room temperature (298K). Under room temperature $\Delta E / kT \sim 10^{-5}$J, this indicates that the population difference between nuclear energy states is very small, thus the enhancement of signal is critical in the development of NMR spectroscopy. Researchers have been working on enhancing the signal-to-noise ratio (S/N) of NMR through either the decreasing of temperature T with low-temperature probes or the increasing of magnetic field B_0 with higher field magnets[2]. The gyromagnetic ratio γ can also be modified to enhance the S/N via techniques

such as cross-polarization (CP) and dynamic nuclear polarization (DNP) which can transfer the polarization from nuclei with higher γ or electrons to the nuclei under detection[3].

1.1.2 NMR Hamiltonians

The nuclear energy splitting is primarily determined by the Zeeman splitting, as illustrated in section 1.1.1, the magnitude of the polarization is mainly controlled through gyromagnetic ratio γ, and applied magnetic field B_0. There are, however, other weaker interactions between the nuclei of interest and its surrounding local chemical environment that could modify the energy splitting between NMR eigenstates. These weaker interactions are particularly useful for NMR because the subtle changes in the local environments can be revealed through the perturbations in the energy gap. This makes NMR one of the most powerful tools in detecting the local structural information of solid materials. For SSNMR, the total Hamiltonian ($\hat{H}_{NMR}$) can be described as a summation of multiple interactions including Zeeman.

$$\hat{H}_{NMR} = \hat{H}_{Zeeman} + \hat{H}_{cs} + \hat{H}_{Q} + \hat{H}_{D} + \hat{H}_{J} \quad (1.7)$$

The name and approximate magnitude in Hz as an expression of relative energies for each of the terms in NMR Hamiltonians could be found in **Table 1.1**.

Table 1.1 Characterization and magnitude for NMR Hamiltonians[4].

Hamiltonian	Characterization	Representative Magnitude
$\hat{H}_{Zeeman}$	Zeeman interaction	50-500 MHz
$\hat{H}_{cs}$	Chemical shielding effect	20 kHz
$\hat{H}_{Q}$	Quadrupolar interaction	1-20 MHz

$\hat{H}_D$	Dipolar interaction	50 kHz
$\hat{H}_J$	J-coupling	1-100 Hz

From **Table 1.1**, it is clear that the Zeeman splitting is the dominating part of the NMR interactions. The expression of the Zeeman splitting's Hamiltonian is:

$$\hat{H}_{Zeeman} = -\gamma B_0 \hat{I}_Z \tag{1.8}$$

Here $\hat{I}_Z$ is the Z component of the angular momentum operator. Z here is defined in the laboratory frame in which the direction of $\mathbf{B_0}$ is also along Z.

We can further derive the energy of Zeeman states (Equation 1.3) with the Schrödinger equation:

$$\hat{H}_{zeeman}\psi_m = -\gamma B_0\left(\hat{I}_Z\psi_m\right) \tag{1.9}$$

$$\hat{I}_Z\psi_m = m\hbar\psi_m \tag{1.10}$$

Here ψ_m is the wave function of Zeeman eigenstate with magnetic quantum number m.

For spin ½ species, both chemical shielding and dipolar interaction play a vital role in the determination of NMR spectra with comparable lineshape broadening (20kHz and 50kHz). The Hamiltonian of chemical shielding is:

$$\hat{H}_{CS} = -\gamma B_0 \sigma \hat{I}_Z \tag{1.11}$$

Here σ is called the shielding tensor, it is a tensor of rank 2 which describes the orientation dependencies of shielding effect relative to the external magnetic field $\mathbf{B_0}$. The shielding tensor can be further simplified to yield several chemical shielding parameters such as isotropic

shielding (σ_{iso}) which is the most widely used NMR parameter. Unfortunately, there is no consensus on the way to simplify the shielding tensor. Since the beginning of NMR, several conventions of expressing the shielding tensor had been proposed by different researchers. The choice of convention can be critical for the reliability of NMR parameters. Thus in section 1.2, we will give a detailed discussion of NMR tensor conventions.

For quadrupolar species (I>1/2), quadrupolar interaction is orders of magnitude larger than chemical shielding and dipolar interaction (MHz vs kHz). Thus, when quadrupolar interactions are present, the NMR spectrum can be dominated by quadrupolar broadening which makes an accurate determination of shielding and dipolar interactions difficult.

The quadrupolar interaction arises from the interaction of a quadrupolar (I>1/2) nucleus with the electric field gradient (EFG) at the site of the nucleus. The components of the EFG tensor are defined as a partial derivative with respect to the electrostatic potential U over the Cartesian coordinates. Here the position of electrostatic potential is the position of the nuclei. (r=0)

$$V_{\alpha\beta} = \frac{\partial U}{\partial \alpha \, \partial \beta}\Big|_{r=0} \quad (\alpha, \beta = x, y, z) \tag{1.12}$$

The EFG tensor is a second rank symmetric tensor. Under the principal axis system (PAS), the EFG tensor is a diagonal tensor with principal components, V_{XX}, V_{YY} and V_{ZZ}.

$$V = \begin{pmatrix} V_{XX} & 0 & 0 \\ 0 & V_{YY} & 0 \\ 0 & 0 & V_{ZZ} \end{pmatrix} \tag{1.13}$$

These elements are ordered as $|V_{ZZ}|>|V_{YY}|>|V_{XX}|$. Also note that matrix V is traceless.

$$V_{XX} + V_{YY} + V_{ZZ} = 0 \tag{1.14}$$

Thus, the quadrupolar Hamiltonian can be expressed in terms of the interaction between the nuclear quadrupolar moment Q and the EFG tensor.

$$\hat{H}_Q = \frac{e^2 Q}{6I(2I-1)\hbar} \hat{\mathbf{I}} \cdot V \cdot \hat{\mathbf{I}} \tag{1.15}$$

Here for a spin I, $\hat{\mathbf{I}}$ is the nuclear spin vector.

Experimentally, the EFG tensor is characterized using two parameters, the quadrupolar coupling constant, C_Q and the quadrupolar asymmetry parameter, η_Q:

$$C_Q = \frac{eQV_{ZZ}}{\hbar} \tag{1.16}$$

$$\eta_Q = \frac{V_{XX} - V_{YY}}{V_{ZZ}} \tag{1.17}$$

C_Q and η_Q are commonly used to describe the NMR lineshape of quadrupolar species.

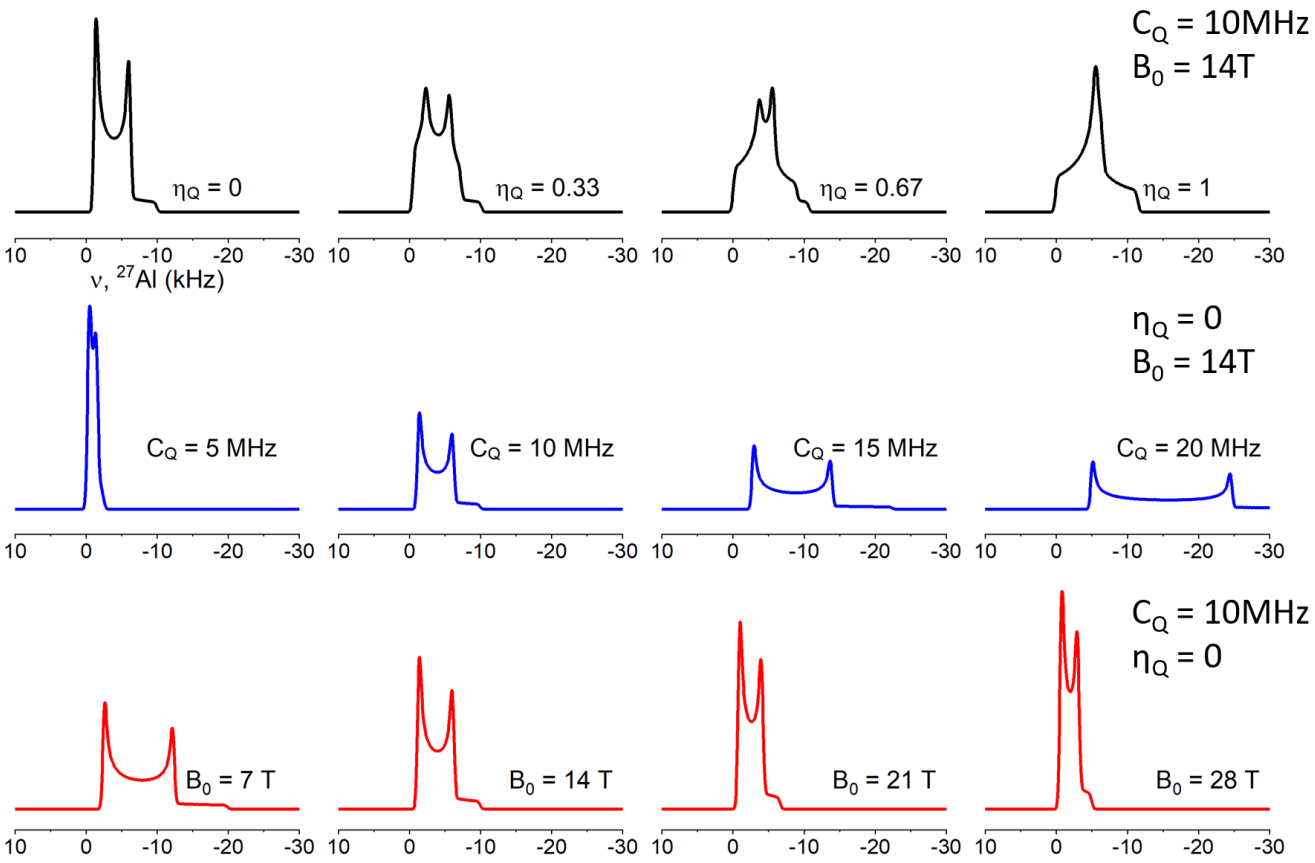


Figure 1.2 Various model lineshapes depicting the effects of quadrupolar broadening. Unless otherwise specified, $\eta_Q = 0$, $C_Q = 10$ MHz, and $B_0 = 14$ T. The first row in black shows the effect on the lineshape of a varying η_Q, which qualitatively affects the overall shape of the resonance. The second row in blue shows the effect of a varying C_Q on the width of the resonance. The third row in red shows the narrowing effect of a varying B_0 field. Each row was drawn so that the integrated intensities were constant to better illustrate the gain/loss in experimental sensitivity.

Figure 1.2 is a series of model lineshapes with varying inputs illustrating the influence of C_Q, η_Q and magnetic field B_0. which qualitatively affects the overall appearance of the resonance. Note in the third row, the effect of a varying field is shown. By definition the second-order quadrupolar broadening is inversely proportional to the external magnetic field B_0, thus, at

higher fields, the effect of the second-order quadrupolar broadening diminishes, appearing as a narrowing of the lineshape. This is one major reason why higher fields are desired when performing NMR with quadrupolar nuclei.

1.2 Conventions of NMR shielding tensor and chemical shift tensor

The chemical shielding interaction is fundamentally critical for NMR detection, almost always when presented with an NMR spectrum, the first thing to do is try to determine the *isotropic average* of the chemical *shift* tensor (**δ**) for each signal observed. The averaged value is called the isotropic chemical shift (δ_{iso}) which determines the positioning of the resonance in both liquid and solid-state NMR spectra. Here, the chemical *shift* is experimentally measured by referencing the experimental spectrum to a standard compound depending on the isotope being studied such as tetramethylsilane (TMS), whose ^{1}H singlet signal is assigned as $\delta_{iso} = 0$ ppm. Thus, in principle, the isotropic chemical shift gives the resonance frequency of the sample relative to the resonance frequency of the hydrogen atoms in TMS.

The *shielding*, on the other hand, is a measurement of the change in the resonance frequency of nuclei in different chemical environments relative to the Larmor frequency of the bare nucleus. The *shielding* tensor (**σ**) is thus a tensor that describes the orientation dependence of the shielding effect. Experimentally, reporting the shielding tensor is troublesome because the value depends on the magnitude of the external magnetic field B_0. Instead, the chemical shift has a uniform scale across different magnetic fields, thus promoting the efficiency of scientific communication.

While the chemical shift tensor is the expression of the shielding effect through measured resonances, computational simulation of NMR parameters usually reports the shielding tensor because it could be expressed as a second derivative of the electronic energy of the system. For nucleus, N, its shielding tensor can be expressed as:

$$\boldsymbol{\sigma}_{\mathbf{N}} = \frac{\partial^2 \mathrm{E}}{\partial \mathbf{B}\, \partial \mathbf{m}_{\mathbf{N}}} \tag{1.18}$$

where $\mathbf{m}_{\mathbf{N}}$ is the magnetic moment of nucleus N.

Because of chemical shift/shielding tensors' critical role in the structural determination with experimental and computational NMR, the following sections will mainly discuss three important aspects of them: derivation from shielding Hamiltonian, difference conventions of expressing tensor's principal components, and conversion between shielding and chemical shift tensor.

1.2.1 From Hamiltonian to the shielding tensor

As stated in section 1.1.2, the shielding effect can be expressed as a second-rank tensor (a 3 by 3 matrix), which determines the orientation and magnitude of the net magnetic field that can be decomposed into external magnetic field $\mathbf{B_0}$ and induced magnetic field $\mathbf{B_{ind}}$.

$$\mathbf{B_{net}} = \mathbf{B_0} + \mathbf{B_{ind}} \tag{1.19}$$

Thus the total Hamiltonian when considering only the Zeeman splitting and shielding effect is:

$$\hat{\mathrm{H}}_{\mathrm{NMR}} = -\gamma \mathbf{B_{net}} \hat{\mathbf{I}} \tag{1.20}$$

Considering equation 1.8 and 1.11, we can derive $\mathbf{B_{ind}}$:

$$\mathbf{B_{ind}} = -\boldsymbol{\sigma}\mathbf{B_0} \tag{1.21}$$

where $\boldsymbol{\sigma}$ is the shielding tensor:

$$\boldsymbol{\sigma} = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{bmatrix} \tag{1.22}$$

Thus, we could get the resonance frequency of the nuclear spin accounts modified by the shielding effect:

$$\nu = \frac{\gamma}{2\pi} B_0(1 - \sigma) \tag{1.23}$$

The shielding tensor can be further decomposed into a symmetrical and anti-symmetrical part:

$$\boldsymbol{\sigma} = \boldsymbol{\sigma}^{\text{symm}} + \boldsymbol{\sigma}^{\text{antisymm}} \tag{1.24}$$

where:

$$\boldsymbol{\sigma}^{\text{symm}} = \frac{\boldsymbol{\sigma} + \boldsymbol{\sigma}^{\text{T}}}{2}$$
$$\boldsymbol{\sigma}^{\text{antisymm}} = \frac{\boldsymbol{\sigma} - \boldsymbol{\sigma}^{\text{T}}}{2} \tag{1.25}$$

Here the $\boldsymbol{\sigma}^{\text{T}}$ is the transpose of $\boldsymbol{\sigma}$. The antisymmetric part could affect relaxation, while it does not have a significant influence on the chemical shift observed by experiments. Thus, for NMR, we usually only consider the symmetrical part of the shielding tensor[5].

The symmetric shielding tensor can be rotated to a specific set of axes (X, Y, Z) called the principal axis system (PAS). The rotation will diagonalize the tensor to yield 3 principal components of the shielding tensor. These three principal components correspond to the

orientation of the PAS. The PAS is fixed against the crystal lattice, thus useful in the determination of the local environments of the local environments that emerge from NMR analysis of sites within a material.

Alternative ways of representing the principal components are reported throughout the history of NMR. In the next section, we will discuss in detail about these conventions.

1.2.2 Conventions of the shielding tensor

Individual tensor elements have particular utility to help communicate details of the full chemical shielding anisotropy (CSA) lineshape. At issue is how best to report these tensors, since there are multiple conventions, including "Mehring" convention ("standard" convention)[6], Haeberlen[7], and the "Maryland" convention[8,9]. **Figure 1.3** below shows a graphical illustration of the definitional difference between the conventions.

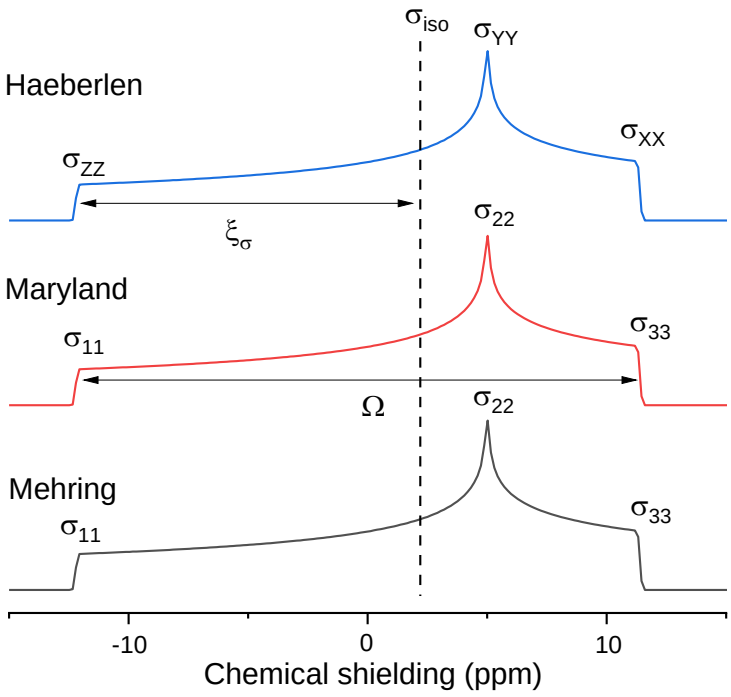


Figure 1.3 Static NMR powder patterns dominated by the shielding effects, illustrate from top to bottom, Haeberlen convention, Maryland convention and Mehring convention. Both the Maryland and Mehring convention assign the shielding tensor principal components as $\sigma_{33} \geq \sigma_{22} \geq \sigma_{11}$. Maryland convention defines span (Ω) and skew(κ) based on the principal components while Mehring convention does not have further definitions.

The Haeberlen convention is the one used by many researchers and importantly by computational programs that use these conventions to depict spectra, including the popular Dmfit[10] and SIMPSON[11] programs. The symmetric contribution can be diagonalized to yield 3 principal components of the shielding tensor, referred to in Haeberlen notation as σ_{XX}, σ_{YY} and σ_{ZZ}:

$$\boldsymbol{\sigma} = \begin{bmatrix} \sigma_{XX} & 0 & 0 \\ 0 & \sigma_{YY} & 0 \\ 0 & 0 & \sigma_{ZZ} \end{bmatrix} \tag{1.26}$$

Isotropic shielding $\sigma_{iso}$is defined as the numerical average of the principal components.

$$\sigma_{iso} = \frac{\sigma_{XX} + \sigma_{YY} + \sigma_{ZZ}}{3} \tag{1.27}$$

Most literature uses the Haeberlen convention for reporting the full chemical shielding (or shift) tensor. In Haeberlen, σ_{XX}, σ_{YY} and σ_{ZZ} are defined based on the magnitude of the frequency difference, σ_{iso}:

$$|\sigma_{ZZ} - \sigma_{iso}| \geq |\sigma_{XX} - \sigma_{iso}| \geq |\sigma_{YY} - \sigma_{iso}| \tag{1.28}$$

There are additional parameters that are often reported in the Haeberlen system reflective of the solid-state CSA lineshape. These are the shielding anisotropy, also called "reduced anisotropy", ζ_σ and "asymmetry parameter" (η_{CSA}), expressed as equation 1.29 and 1.32:

$$\zeta_\sigma = \sigma_{ZZ} - \sigma_{iso} \tag{1.29}$$

While equation 1.29 expresses the algebraic definition for this quantity, the reduced anisotropy can be visualized in terms of the relative location of the most intense point (σ_{YY}, typically) in the static lineshape – to the right or left – of the isotropic shielding, σ_{iso}. Also worth noting is that in the Haeberlen convention, there are two methods for reporting the anisotropy of the CSA: shielding anisotropy $\Delta\sigma$ and reduced shielding anisotropy ζ_σ.

$$\Delta\sigma = \sigma_{ZZ} - (\sigma_{XX} + \sigma_{YY})/2 \tag{1.30}$$

The relationship between ζ_σ and $\Delta\sigma$ can be expressed as:

$$\Delta\sigma = \frac{3}{2}\zeta_\sigma \tag{1.31}$$

Since both notations of the shielding anisotropy are commonly in use, it is advised to be careful comparing different anisotropy values from different sources.

The sign of anisotropy should also be handled carefully because when σ_{ZZ} gets closer to σ_{YY} than the frequency difference between σ_{YY} and σ_{XX}, there will be a sudden flip of the sign of $\Delta\sigma$ as well as a switch of the designation of σ_{XX} and σ_{ZZ} based on the definition of equation 1.28. This sudden change could lead to misunderstanding when comparing a series of closely related samples.

The overall shape of the line is expressed by the asymmetry parameter η_{CSA}, where ζ_σ appears in the denominator.

$$\eta_{CSA} = \frac{\sigma_{YY} - \sigma_{XX}}{\sigma_{ZZ} - \sigma_{iso}} \tag{1.32}$$

Based on the definition, the range of η_{CSA} is from zero to one. When the local geometry is axially symmetric ($\sigma_{YY} = \sigma_{XX}$), η_{CSA}=0. When $\sigma_{YY} = \sigma_{iso}$, η_{CSA}=1. This change in η_{CSA} is irrespective of the sign of ζ_σ because any change from positive to negative reduced anisotropy is canceled by a similar sign change in the numerator.

Another convention of expressing the shielding tensor is called the "Mehring notation". Instead of designating the principal components based on their distance from the isotropic values, the Mehring convention just designates the components based on their numerical order:

$$\sigma_{33} \geq \sigma_{22} \geq \sigma_{11} \tag{1.33}$$

The definition of isotropic shielding remains the same as Haeberlen:

$$\sigma_{iso} = \frac{\sigma_{11} + \sigma_{22} + \sigma_{33}}{3} \tag{1.34}$$

It is difficult to define the anisotropy and asymmetry components in a similar manner as the Haeberlen convention because anisotropy and asymmetry parameters are dependent on the relative positioning of σ_{22} between σ_{11} and σ_{33}. However, the Mehring convention also has advantages over the Haeberlen convention since it does not have the ambiguity of principal components assignment. Due to this benefit, the Mehring convention is particularly useful when benchmarking computational NMR shielding tensors against experimental ones as we will discuss in the following chapters.

There is another convention which has also been widely used called the "Maryland" notation proposed by a group of NMR scientists at a summer school in College Park, Maryland, in 1993.

The purpose of the Maryland notation is to address the discontinuity problem of anisotropy in the Haeberlen convention. The anisotropy and asymmetry parameters were replaced by two new parameters called the span (Ω) and skew(κ).

$$\Omega = \sigma_{33} - \sigma_{11}$$
$$\kappa = 3(\sigma_{iso} - \sigma_{22}) / \Omega \tag{1.35}$$

The span is always positive and describes the width of the spectrum. The skew can be both positive and negative based on the relative positioning of σ_{22} and σ_{iso}, and is conceptually analogous to the reduced shielding anisotropy.

1.2.3 Conversion between shielding and chemical shift

Experimentally, the NMR resonance frequency is converted into a referenced value, chemical shift by introducing a reference compound. [8]

$$\delta = 10^6 \frac{\nu_s - \nu_{ref}}{\nu_{ref}} \tag{1.36}$$

The values ν_s and ν_{ref} represent the NMR resonance frequency of the sample and the reference compound. Using equation 1.23, the shielding tensor can be mapped to chemical shift tensor:

$$\delta = -10^6 \frac{\sigma_s - \sigma_{ref}}{1 - \sigma_{ref}} \tag{1.37}$$

σ_s and σ_{ref} represents the shielding of the nucleus of interest in a sample and a reference compound respectively. Since for most of the cases, σ_{ref} is significantly smaller than 1, the following approximation is often used:

$$\delta = -10^6(\sigma_s - \sigma_{ref}) \tag{1.38}$$

Thus, we can calculate the chemical shift tensor with corresponding parameters isotropic chemical shift (δ_{iso}), chemical shift anisotropy (ζ_δ) and asymmetry parameter (η_{CSA}). Their relationship to the shielding parameters is shown below:

$$\delta_{iso} = \frac{\sigma_{ref} - \sigma_{iso}}{1 - \sigma_{ref}}$$
$$\zeta_\delta = \frac{-\zeta_\sigma}{1 - \sigma_{ref}} \tag{1.39}$$

Same as equation 1.37, for most cases where $\sigma_{ref} << 1$, the denominator can be ignored[12]. Since the asymmetry parameter η_{CSA} is already normalized, it is the same for both shielding and chemical shift.

1.3 NMR crystallography

The power of SSNMR lies in its capability of obtaining structural or crystallographic information from interactions such as shielding, dipolar and quadrupolar interactions, which correspond to local environments of the material and regardless of the long-range order. Thus, by careful analysis of its broadened lineshape, SSNMR has the capability of describing the local structural properties of a wide range of materials. Being one of the world's most popular tools used for structural investigation, SSNMR has been used on various types of systems such as organic/inorganic solids and biomaterials[13–24].

The term ‘NMR crystallography”, can be generally defined by using SSNMR as the central technique to determine, select or refine the materials’ structure.[15] In practice, however, different

‘versions’ of NMR crystallography research exist and can be generally categorized as the following:

1. Using SSNMR as the primary approach for the characterization of the materials’ structure[25–27].
2. A detailed characterization of materials’ structure combining SSNMR with other crystallography techniques such as X-ray diffraction, and computational simulation such as density functional theory (DFT)[28–35].
3. Using SSNMR and DFT calculation to select the right structure from a list of candidates proposed by crystal structure prediction (CSP)[36–40].

This dissertation will focus on the second aspect of NMR crystallography. As we mentioned above, SSNMR is powerful in detecting the short-range structure of materials, often regardless of the long-range order. Thus, the information obtained by NMR is complementary to XRD and is not limited to crystalline systems. Combined with XRD, SSNMR was applied to problems such as polymorph detection, hydrogen bond characterization and determination of amorphous compounds for which the ability of XRD is limited[41–46]. The use of NMR is also important for the characterization of powdered samples. Although full structure determination is possible for powdered XRD (p-XRD) using automated search/match routines, its accuracy depends on proposing suitable initial guess of the structure or structural constrains[47], SSNMR thus provides a critical verification and refinement approach for the structure proposed by p-XRD.

Experimental SSNMR could, however, still face challenges for the determination of structures with high complexity. Problems such as broadened and overlapped lineshapes with a large number of resonances, low natural abundances, or less commonly studied nuclei could greatly

increase the difficulty of signal assignment and interpretation[48]. The recent development of computer simulation methods such as DFT has been shown to be a solution. In 2001, Pickard and Mauri proposed the Gauge-Including Projector Augmented Wave (GIPAW) which enables accurate simulation of the NMR interactions for a periodic system based on plane-wave basis sets. In terms of NMR crystallography, GIPAW greatly expanded DFT's abilities for structure characterization in different ways such as fast assignment of sites for complex structures, deconvolution of overlapped spectrum lineshape, and providing information that could not be easily measured experimentally such as shielding anisotropy for quadrupolar species[12,49,50]. The introduction of DFT could also help guide experiments by providing theoretically predicted spectra parameters. Thus, the combination of SSNMR, XRD and DFT is powerful for solving materials' structures with high accuracy.

The solid systems studied in this dissertation possess translational symmetries, which enables periodic boundary conditions to be applied for DFT calculations[50]. This greatly reduces the number of atoms that need to be considered for the calculation. While the charge density of crystalline materials is fully periodic, the corresponding wave functions are only quasi-periodic, from Bloch's theorem, the single-particle wave function in a crystalline material can be expressed as:

$$\psi_{\mathbf{k}}^{n}(\mathbf{r}+\mathbf{R}) = e^{i\mathbf{k}\mathbf{r}}\psi_{\mathbf{k}}^{n}(\mathbf{r}) \tag{1.40}$$

Here n is the index of particle and **k** is the wavenumber. Thus, physical properties of a crystalline material need to be calculated as an average over all the values of wavenumber **k**. In practice, it is not feasible to go through all the possible **k**, instead DFT take an average of a physical

property over a regular spaced k-point mesh within the first Brillouin zone[51]. The choice of the k-point mesh is commonly determined based on the scheme of Monkhorst and Pack[52].

To implement DFT, the wave functions need to be expressed as a linear combination of simple mathematical functions. For periodic systems, it is convenient to use planewaves which automatically satisfy the periodic boundary conditions.

$$\psi_{\mathbf{k}}^{\mathrm{n}}(\mathbf{r}) = \sum_{\mathrm{G}} c_{\mathrm{k}}^{\mathrm{n}}(\mathbf{G})\, \mathrm{e}^{\mathrm{i}(\mathbf{k}+\mathbf{G})\mathbf{r}} \tag{1.41}$$

Here G are a set of reciprocal lattice vectors. To control the size of the basis set, the summation in equation 1.41 is limited to a set of reciprocal vectors which satisfy:

$$\frac{1}{2}|\mathbf{k}+\mathbf{G}|^2 \leq \mathrm{E}_{\mathrm{Cut}} \tag{1.42}$$

Here the E_{cut} is the cut-off energy which defines the maximum kinetic energy of the planewaves. In practices of planewaves DFT calculations, the cut-off energy is increased until a satisfactory level of convergence for the target physical property is reached. The value of cut-off energy required for convergence primarily depends on the atomic species involve in the calculation, since the highest planewaves energies are associated with the regions closest to the nucleus.

While DFT has been proven to be accurate in NMR parameter simulation, its application is limited due to the lack of efficiency, because the computational time and resources required increase cubically with the increase of the structure size[51]. One emerging solution is to build a machine learning model to map the local structure information to the NMR parameters. With more experimental/computational NMR parameter information accumulated and open to access,

it is now feasible to training machine learning models as a fast computational tool for NMR spectra prediction. More details about this aspect will be discussed in the following chapters.

1.4 Conclusion and outline of the thesis

In this chapter, we discussed the basics of SSNMR including its origin, the NMR Hamiltonian, and NMR tensors. We specifically emphasized the importance of the different conventions to express the shielding tensor, which will be shown later to be critical when assessing the accuracy of DFT calculations of NMR tensors against experiment. We also discussed the general idea of NMR crystallography and different approaches to NMR crystallography in practice. The following chapters will follow this route of discovery on different aspects of NMR crystallography. Chapter 2 is a case study on how DFT calculations combined with experimental NMR reveal information about the structural transformations of a battery material SVPO along the discharging process. The importance of NMR parameters other than isotropic chemical shift in the determination of structure is also emphasized. Chapter 3 is a study on computational NMR dataset construction by performing a benchmark of experimental spin ½ ^{29}Si NMR parameters against DFT calculated NMR parameters for 42 unique sites. Two different DFT packages, CASTEP and VASP are also compared, and a critical definitional difference between the two is revealed. Chapter 4 is a study that continues the benchmarking effort of chapter 3 but focuses on the quadrupolar nucleus ^{27}Al, then explores the possibility of predicting NMR parameter C_Q with machine learning algorithms.

Chapter 2

Structural Investigation of Sliver Vanadium Phosphorus Oxide ($Ag_2VO_2PO_4$) and its Reduction Products

The material in this chapter is adapted from Sun H, Hammann B A, Brady A B, et al. Structural Investigation of Silver Vanadium Phosphorus Oxide ($Ag_2VO_2PO_4$) and Its Reduction Products. Chemistry of Materials, 2021, 33(12): 4425-4434.

2.1 Introduction

In this chapter we study the local structural profile of a high-capacity primary lithium-ion cathode material ($Ag_2VO_2PO_4$), as well as its electrochemically reduced counterparts by NMR crystallography combining XRD, solid-state NMR(SSNMR), *Ab-initio* calculations and electrochemical impedance spectroscopy.

One of the best-known phosphonate containing cathode materials for secondary lithium ion batteries, $LiFePO_4$ has been on the market for years.[52,53] Structurally similar to the mineral olivine, this material is relatively inexpensive and has the added advantage of being more thermally and chemically stable than other lithium-based batteries.[54–58] However, $LiFePO_4$, like most phosphonate materials, displays inherently low conductivity which limits electrochemical performance. This has been addressed in commercial $LiFePO_4$ batteries by minimizing particle size, coating particles with a conductive material or doping the material to improve conductivity.[59–62] Silver vanadium oxide ($Ag_2V_4O_{11}$ or SVO) has been used for over 25 years as

a cathode material in lithium batteries most commonly found in implantable cardiac defibrillators (ICDs).[63,64] Because of the important role SVO has played in ICDs, it has been studied extensively. Of interest are the reports on the formation of silver nanoparticles and nanowires (dendrites) upon battery discharge, which leads to an increase in the conductivity of the material.[65,66] Studies involving copper vanadium oxides, such as $Cu_{1.1}V_4O_{11}$, have also shown the generation of copper metal dendrites upon reduction.[67,68] These reports led Takeuchi and coworkers to propose another approach to improving the performance of low conductivity phosphonate materials, that is, to prepare materials which are able to generate a conductive matrix *in situ*.

Based on the success of SVO, there has been interest in examining the properties of silver vanadium phosphorus oxide, $Ag_2VO_2PO_4$ (SVPO), a vanadium phosphonate material first prepared in 1993.[69] Like $LiFePO_4$, SVPO also has an inherently low conductivity. In 2009, Takeuchi and coworkers reported that reduction of lithium-based electrochemical cells utilizing $Ag_2VO_2PO_4$ as the cathode material, led to silver metal deposition, resulting in a significant increase (i.e., 15,000 fold) in the conductivity of the material.[70–72] Other studies on SVPO (pristine and reduced SVPO samples) including electrochemical studies, X-ray absorption fine structure spectroscopy and microscopy studies, have helped us to better understand the reduction process.[73,74]

An increase in the paramagnetic character of the V center during discharge has been reported using magnetic susceptibility measurements.[71] The increase in magnetic susceptibility cannot originate from a Ag center because the d^{10} electron configuration is maintained during Ag^+ reduction to Ag^0, neither of which display paramagnetic properties under these conditions.[75] V^{5+} centers in the pristine material do not display magnetic character due to the d^0 configuration.

Although there is some attenuation of magnetic susceptibilities for V^{3+} and V^{4+} in both vanadium oxides[76] and vanadium phosphates[77], a measurable increase in the magnetic susceptibility of the discharged SVPO materials has been reported. This increase in magnetic susceptibility as discharge progresses corresponds to spin-moment calculations for both the d^1 and d^2 electron configurations, which result from two reduction events: $V^{5+} \rightarrow V^{4+} \rightarrow V^{3+}$.

Due to the amorphization of $Ag_2VO_2PO_4$ during discharge, diffraction techniques become less suited for elucidating local structural changes.[64] X-ray absorption spectroscopy (XAS) techniques like X-ray Absorption Near-Edge Spectra (XANES) and Extended X-ray Absorption Fine-Edge Structure (EXAFS) are more useful for resolving changes in the ionic coordination environments. Because XAS methods are able to interrogate amorphous structures, which can provide some additional information about the mechanism of reduction.[73]

Of interest is analysis of SVPO by SSNMR spectroscopy which could give additional insight to the material's local chemical environment. The ^{51}V isotope has high natural abundance of (99.76%), making it an attractive nucleus for NMR spectroscopy. Even though it is quadrupolar (I = 7/2 for ^{51}V), the low quadrupolar moment (0.05 barn) and high receptivity make it more amenable to study.[78] The SSNMR spectroscopy (SSNMR) of vanadium-containing systems has proven valuable in exploring inorganic vanadium oxide catalysts, vanadium containing proteins, and both vanadyl and vanadate species, which are of interest in the treatment of diabetes.[79–85] Based on experimental SSNMR spectra and the XRD-proposed structure, the local environment can be further refined or confirmed with the help of *Ab-initio* methods such as density functional theory (DFT). Thus, key structural information that can be obtained from ^{51}V SSNMR plus DFT calculations includes information such as number and types of coordinated atoms and structural information such as symmetry and association of vanadium-oxygen polyhedra.[78] Here we present

the ^{51}V SSNMR spectroscopy of SVPO, $Ag_2VO_2PO_4$ as well as several electronically discharged SVPO samples from 1 electron equivalent (1ee) to 3 electron equivalents (3ee). The 1ee and 2ee discharged samples were also measured by XRD to obtain structural information. Computation of NMR tensors for three different structural models (two of which were previously published) were performed with the DFT package Cambridge Sequential Total Energy Package (CASTEP). Experimental NMR data and computed NMR parameters were compared. Structural changes of SVPO manifested by the vanadium local environment along the discharge process could be monitored with SSNMR and DFT.

2.2 Experimental Methods

2.2.1 Synthesis and Characterization

A hydrothermal method adapted from Kang et. al was used to prepare silver vanadium phosphate ($Ag_2VO_2PO_4$).[69] The vessel for the aqueous reaction was a Teflon autoclave in which Ag_2O and V_2O_5 were mixed in a 1:2.554 by mass ratio and heated at 230°C for 96 h. The yellow solid was collected via vacuum filtration. Differential Scanning Calorimetry (DSC) was conducted up to 580°C at 5°C/min to test for sample purity. A Rigaku SmartLab powder diffractometer with a DTex detector and 5° Soller-slits was used to collect an X-ray diffraction (XRD) pattern from 5° to 90° 2θ. The XRD pattern was used to determine crystal structure and was indexed to PDF#01-081-2149.

Synchrotron powder X-ray diffraction (XRD) was performed at the 28-ID beamline at NSLS-II. The powder sample was packed into a Kapton tube, and the sample distance was calibrated using an LaB_6 powder sample. The beam wavelength was calibrated to 0.2388 Å. Diffraction patterns were detected using a 16-inch CsI-scintillator fitted to an amorphous silicon flat panel. The two-

dimensional data were integrated using GSAS-II.[86] The diffraction data were analyzed by Rietveld method, using GSAS-II.[86]

Instrumental broadening was calculated using an LaB_6 powder sample. A ten-term polynomial background was used. Lattice parameters, atomic positions, and Debye-Waller factors were refined. Crystallite size and microstrain broadening were found to be negligible for the pristine sample but were fit for the discharged samples.

2.2.2 Discharge Experiments and Electronical Testing

Coin cells were assembled using a Li metal anode and $Ag_2VO_2PO_4$ cathode ($Li||Ag_2VO_2PO_4$) with electrolyte of 1 M lithium hexafluorophosphate ($LiPF_6$) in 30:70 ethylene carbonate/dimethyl carbonate. Discharge was performed at a rate of C/300 to 1 electron equivalent (1ee), 2 electron equivalent (2ee) and 3 electron equivalent (3ee) limits with a Maccor 4000 series battery-tester at 30°C using their post-assembly open circuit potential of about 3.40 – 3.55 V. The cathode material was recovered post-discharge by dismantling the cells inside an inert atmosphere glovebox, pulverizing the cathode pellets, and rinsing the cathode material with pure dimethylcarbonate.

For electrochemical testing, type 2325 coin-type cells were assembled in an argon-filled glove box. $Ag_2VO_2PO_4$ powder was pressed into pellets to construct cathodes. The anode was a disk of lithium metal and separated from the cathode using polypropylene membranes wetted with $LiPF_6$ in 3:7 ethylene carbonate and dimethyl carbonate electrolyte.

Electrochemical Impedance Spectroscopy (EIS) of the assembled cells was collected with a Biologic VSP potentiostat/galvanostat, and fitting of the impedance was performed using ZView

software. Discharge was performed at C/300 to 1ee, 2ee, and 3ee limits with a Maccor 4000 series battery-tester in a temperature-controlled chamber.

For ex-situ testing, electrodes were recovered from the coin-cells post-discharge by decrimping the cells under argon. The recovered $Ag_2VO_2PO_4$ pellet electrodes were rinsed with dimethyl carbonate and allowed to dry under argon before being packed and sealed under argon in Kapton tubes. The 28 ID-2 beamline at the National Synchrotron Light Source-II was calibrated to 0.2388 Å for collecting powder diffraction data. Measurements were collected by a CsI scintillator on an amorphous silicon detector. NIST standard LaB_6 powder was used to find instrumental broadening. GSAS-II[86] diffraction analysis software was used to refine the measured structure after integrating the 2D data as-collected to 1D data. The parameters refined were a background polynomial, scale parameters, atomic positions, Debye-Waller factors, and lattice parameters.

2.2.3 Experimental SSNMR

NMR samples were packed under an inert atmosphere (dry N_2 gas) for all NMR experiments. ^{51}V NMR data were acquired at a magnetic field strength of 13.9 T corresponding to a ^{51}V Larmor frequency of 155.1 MHz using a Tecmag Redstone spectrometer equipped with a Bruker 2.5 mm HX MAS probe. All ^{51}V MAS-NMR spectra were obtained with a small tip angle ($\pi/24$) pulse acquire sequence under magic angle spinning (MAS) conditions ν_R = 10 kHz and 25 kHz with a 1 s recycle delay. A static ^{51}V NMR spectrum was obtained with a Hahn spin echo pulse sequence (t = ms).[87] Length of the 90-pulse was set at 0.7us and the recycle delay was 20s. ^{51}V NMR spectra were referenced to a 0.1 M Na_3VO_4 solution (a secondary reference at -545 ppm relative to $VOCl_3$ at 0 ppm).[78] ^{51}V NMR data were simulated using the model “int2quad” within

the Dmfit[10] program. All the discharged samples were similar in mass, typically 23-24 mg, and measurements were repeated twice.

For the ^{7}Li MAS NMR spectral data, samples were packed and unpacked under N_2. Spectra were acquired through Bloch decay with a small tip-angle pulse (1.5 μs) at a Larmor frequency of 116.47 MHz at 7T. Samples were spun at 8 kHz with N_2 gas, and recycle delays of 2s were used.

To get a statistical analysis of the reliability and stability of the fitted NMR model, Monte Carlo simulations were performed for each of the experimental parameters within Dmfit listed in **Table 2.2** and **Table 2.5**. A small perturbation (random noise) was added to the fitted model followed by a re-fit of the spectrum with Dmfit, and the resultant fitted parameters were recorded. This process was performed iteratively 200 times for each model, and a distribution of the fitted experimental NMR parameters can be obtained. **Figure 2.1** shows an example of the resultant distribution plot of the parameters extracted from the 10-kHz MAS NMR experiment on pristine SVPO. From the plot we can see all the parameters generally exhibit Gaussian distributions. Thus, an estimation of the 95% confidence interval can be calculated by getting the standard deviation of the distribution.

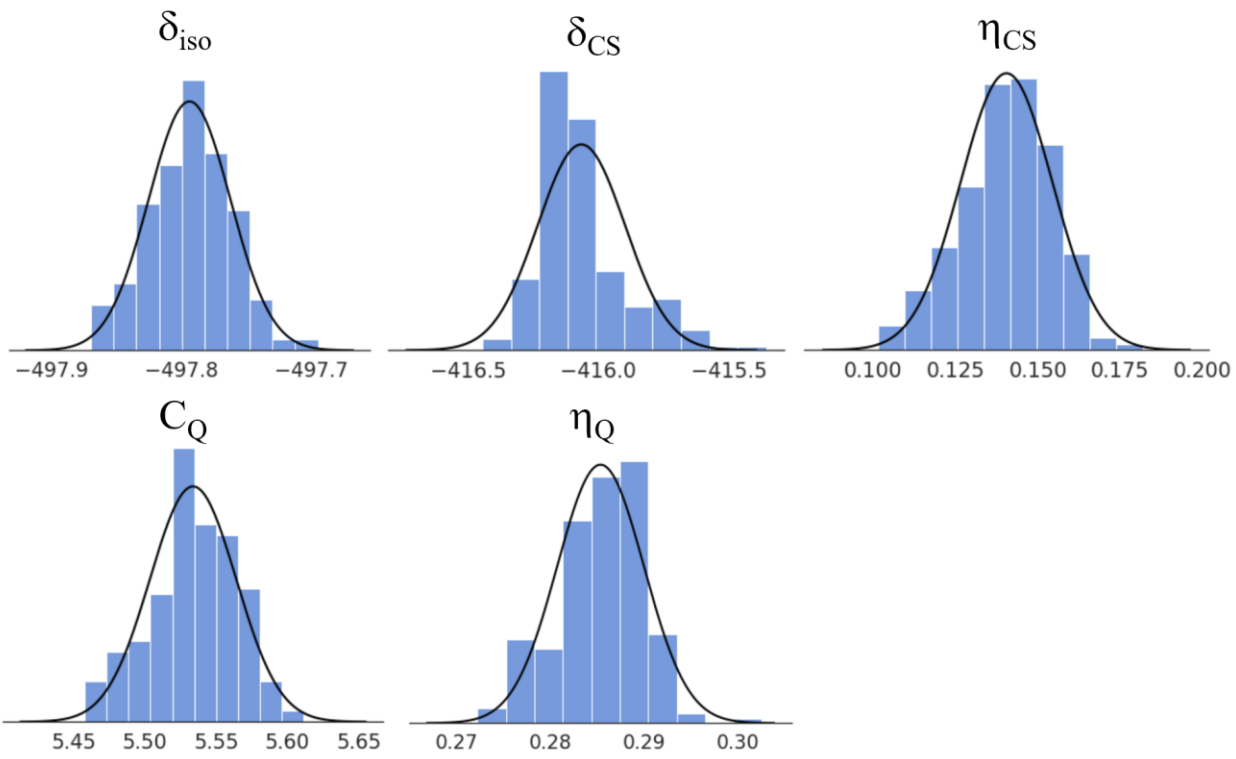


Figure 2.1 Distribution plot of experimental NMR parameters of 10kHz MAS-NMR on pristine SVPO, (representative of all Monte Carlo analyses). The black line shows the fitted Gaussian distribution based on the Monte Carlo simulation results shown as the blue histograms.

2.2.4 DFT computation of SSNMR parameters

For computational simulation of SSNMR, the package, CASTEP,[88] was used to perform density functional theory[89,90] (DFT) calculations of NMR parameters. Three SVPO structures from different sources were selected: a DFT-computed structure from the Materials Project,[91] a refined single-crystal SVPO structure by Lii and coworkers (ICSD-73580),[69] and a powder XRD SVPO structure Rietveld-refined by ourselves. The DFT computations were performed using the generalized gradient approximations (GGA) and the Perdew-Burke-Ernzerhof[92–94] (PBE) functionals. The chemical shielding tensors[30,95] (or chemical shielding anisotropy, CSA) and electric field gradient (EFG) tensors[96] were calculated under the GIPAW[30,50] method utilizing

ultrasoft-pseudo potentials.[30] All the structures were geometry optimized first to the local minimum energy. The critical parameters (cut-off energy and k-point) used in the calculations were confirmed by convergence tests before the actual calculation in order to verify the calculation method.

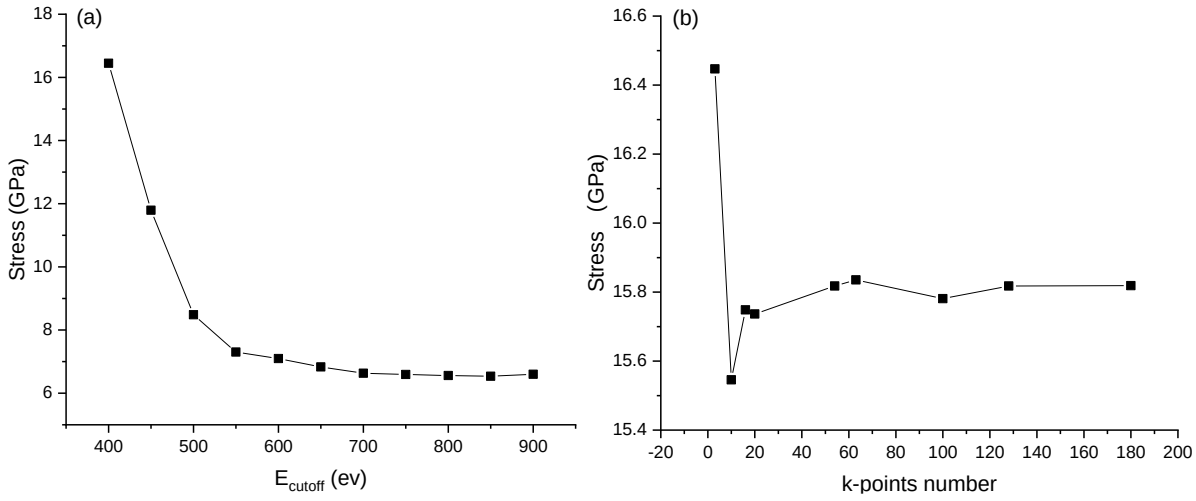


Figure 2.2 CASTEP convergence test for SVPO over a) cuf-off energy E_{cutoff} and b) number of k-points.

CASTEP is a DFT package utilizing plane wave basis set and sampling in the reciprocal space. Thus, before setting up the calculation, convergence tests should be performed with respect to cut-off energy (E_{cutoff}) and the number of k-points in the lattice. **Figure 2.2** illustrates two convergence tests that was performed on the SVPO structure model obtained from the Materials Project (MP-565532).[97] **Figure 2.2 (a)** shows the convergence test of stress with respect to the cut off energy E_{cutoff}. Here the cutoff energy is specified from 400 eV to 900 eV in increments of 50 eV. The system converged after 600 eV. **Figure 2.2 (b)** shows the convergence test of stress with increasing number of irreducible k-points. Number of k-points increases from 1 (Monkhorst-pack 1×1×1) to 180 (9x9x8). The system converged after 20 k-points (4×4×4). As a result, we chose the cut-off energy as 640 ev and 30 k-points.

2.3 Results and discussion

The structure of the synthesized material was confirmed by powder XRD. The diffraction pattern for pristine $Ag_2VO_2PO_4$ (SVPO) was found to match well with previous reports, PDF #01-081-2149, the XRD pattern and a multiple unit cell graphical rendering of the structure is shown in **Figure 2.3** and **Figure 2.4**. Results of the powder diffraction are shown in **Table 2.1**. The reported bond distances in the following paragraph are based on the results of the Rietveld refinement of powder diffraction.

Table 2.1 Rietveld refinement of pristine $Ag_2VO_2PO_4$: powder diffraction crystallographic parameters.

Chemical formula	$Ag_2VO_2PO_4$	Atomic Parameters			
Formula weight	393.7 g/mol	Atom:	x	y	z	Occupancy
Crystal system	Monoclinic	Ag_1	0.8798(6)	0.7682(7)	0.3793(5)	1.000
Space group	$C2/m$	V_1	0.626(2)	0.00000	0.085(2)	1.000
a	12.4384(1)Å	P_1	0.852(2)	0.00000	0.849(3)	1.000
b	6.29283(1)Å	O_1	0.500(3)	0.00000	0.208(7)	1.000
c	1.935(8)Å	O_2	0.895(3)	0.811(4)	0.977(2)	1.000
$beta$	90.336(3) degrees	O_3	0.724(4)	0.00000	0.801(7)	1.000
V(Å^3)	493.343704 Å^3	O_4	0.603(3)	0.00000	0.849(3)	1.000
Z	4	O_5	0.701(3)	0.00000	0.251(8)	1.000
R_{wp}(%)	3.01					

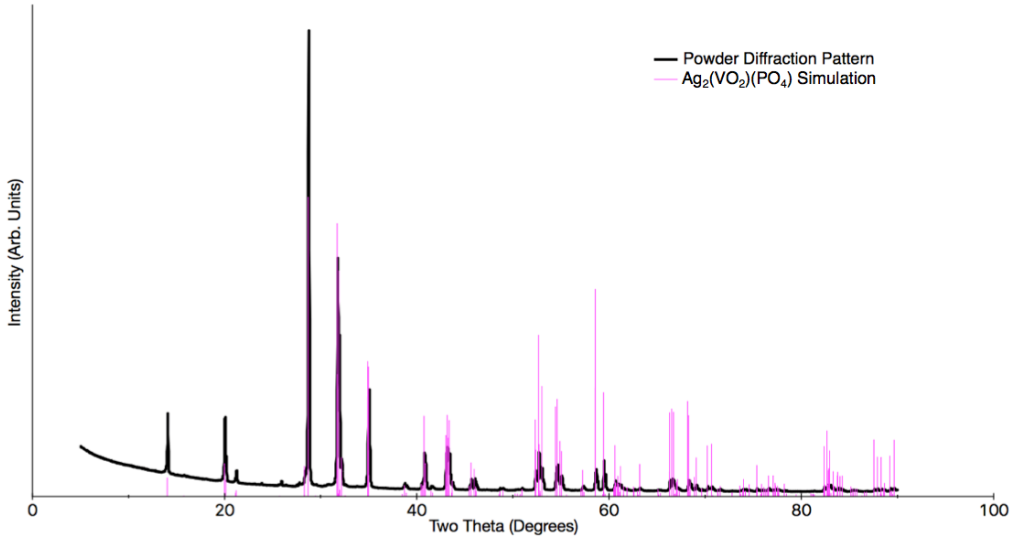


Figure 2.3 XRD pattern of $Ag_2VO_2PO_4$ overlaid with simulated peaks based on PDF Card #01-081-2149.

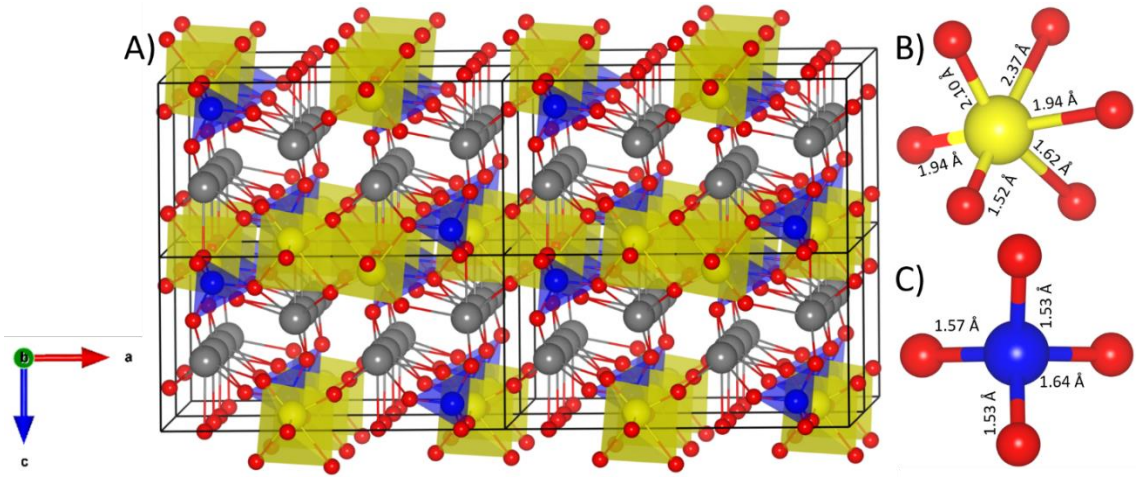


Figure 2.4 (A) Crystal structure of $Ag_2VO_2PO_4$, viewed down the *b*-axis. The vanadium octahedra (yellow) and the phosphorus tetrahedra (blue) form layers with silver (gray) interspersed in between. (B) Strongly distorted 6-coordinate vanadium. (C) Slightly distorted phosphorus tetrahedra.

$Ag_2VO_2PO_4$ forms a crystal in the *C*2/m space group as a layered structure. Pairs of 6-coordinate vanadia species are associated edge-wise, and together with phosphate groups these form a full layer perpendicular to the (001) lattice direction. The 6-coordinate vanadium sites are

significantly distorted (See **Figure 2.4**) with vanadium-oxygen (V-O) bond distances between 1.52 Å - 2.37 Å. The phosphorous tetrahedra are only slightly distorted 1.53 Å - 1.64 Å. Silver occupies the inter-layer space, and the valence is estimated at +1. These ions coordinate with multiple of oxygen atoms between 2.23 Å and 2.78 Å distances.

The purity of the synthesized material was also assessed by differential scanning calorimetry. DSC measured a single endothermic peak at 535°C, in keeping with previous reports.[74,98] The crystal structure of the starting material permits the 6-coordinate vanadium site(s) to be computed via DFT for interpretation by solid-state ^{51}V NMR. Since NMR is particularly sensitive to local electronic environments, any changes to this environment on discharge can be potentially interpreted even when the structure lacks the long-range order that is required for diffraction.

2.3.1 ^{51}V SSNMR Spectroscopy of Pristine SVPO ($Ag_2VO_2PO_4$)

^{51}V NMR spectra were acquired at 13.9 T using static NMR and magic angle spinning (MAS) at multiple spinning frequencies (ν_R). **Figure 2.5** shows two spectra of SVPO acquired at MAS rotational frequencies of 10 and 25 kHz, where symbols denote the isotropic resonances, centered at (δ_{iso} = -500.3 ppm). Accurate determination of the chemical shift anisotropy (CSA) tensor and the electronic field gradient (EFG) tensor for ^{51}V in SVPO can be accomplished through experimental modeling (using a simulation package Dmfit[10]) and calculation of the NMR tensors using a DFT package CASTEP.[99]

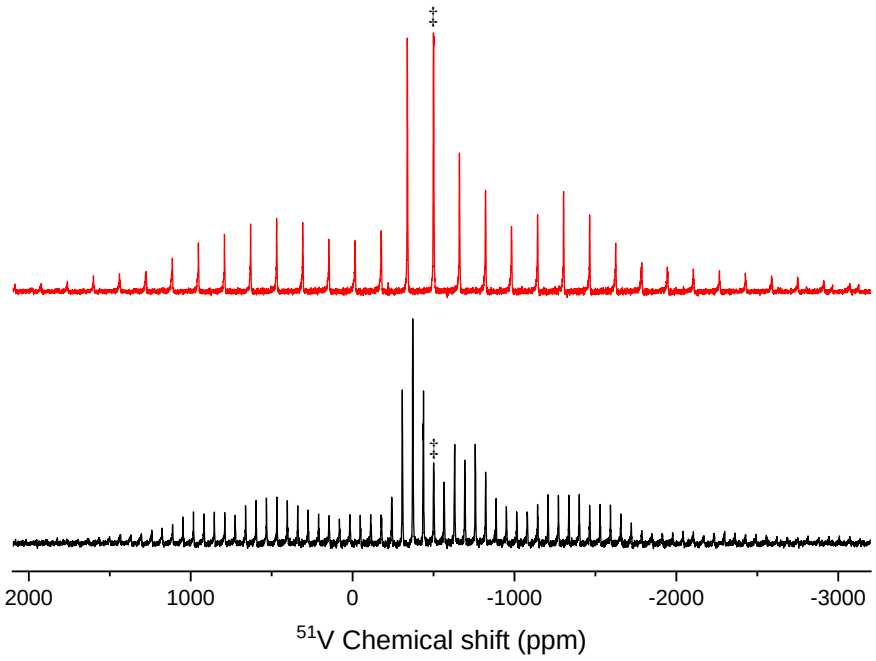


Figure 2.5 Experimental ^{51}V MAS NMR spectra of $Ag_2VO_2PO_4$ at 13.9T: (bottom) ν_R = 10 kHz and (top) ν_R = 25 kHz. Symbols (‡) indicate isotropic resonances.

Using Dmfit to deconvolute the ^{51}V NMR spectra at two different MAS rotational frequencies, one can better assess the accuracy of the NMR parameters extracted. With simulation packages such as Dmfit, choosing the right model is critical to the success and accuracy of the simulation. ^{51}V has a small quadrupolar moment such that the width of the ^{51}V central transition is relatively narrow. Thus, the line shape of ^{51}V spectra is determined by a combination of both the CSA tensor and EFG tensor. We chose the 'int2quad' model in Dmfit to perform the simulation because it takes into account first- and second-order quadrupolar effects, and the effect of chemical shielding (shift) anisotropy. We use DFT calculated NMR parameters for the single vanadium site in SVPO as a "first guess" for Dmfit. **Figure 2.6** shows the comparison between the experimental MAS NMR spectrum and the simulated spectrum from Dmfit. Excellent

agreement is found between the simulated and experimental MAS NMR spectra, and the values from the deconvolution are shown in **Table 2.2**.

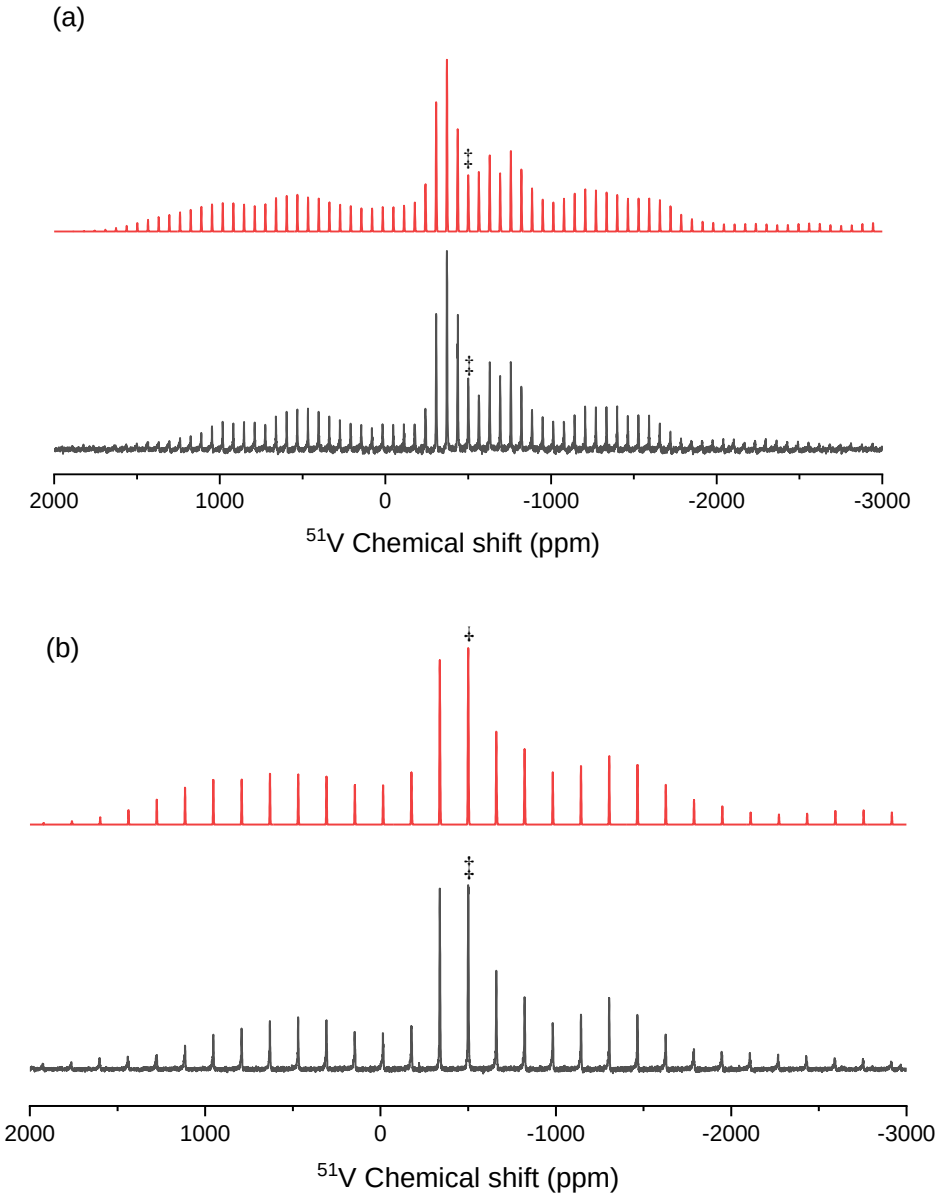


Figure 2.6 Experimental (black) and simulated (by Dmfit, red) using CASTEP-NMR tensor parameters as an "initial guess." ^{51}V MAS NMR spectra of $Ag_2VO_2PO_4$ obtained at 13.9T: (a) ν_R = 10 kHz and (b) ν_R = 25 kHz. Here the spinning speeds are chosen for the high rotor stability of the current experimental setups. Symbols (‡) indicate isotropic resonances.

Calculation of NMR tensor values using DFT serves as an important part in the process of "NMR crystallography". By comparing experimental NMR parameters with calculated NMR parameters (CSA and EFG tensors) based on structural models proposed by crystallographic methods like XRD, validation or potentially further refinement of the atomic coordinates is possible. Here, we perform DFT calculations of ^{51}V CSA and EFG tensors with three SVPO structural models with slightly different atomic coordinates from different sources (details are addressed in the Experimental Section). The calculation results are summarized in **Table 2.2**. Calculation results for individual tensor components can be found in **Table 2.3.**

Table 2.2 Experimental and DFT ^{51}V NMR parameters for $Ag_2VO_2PO_4$ (SVPO), by Haeberlen convention expressions. Note: experimental measurements are expressed in terms of chemical shift (δ) and computed values are in terms of shielding (σ).

Experimental NMR Chemical *Shift* Anisotropy Parameters (fitted by Dmfit)					
Pristine SVPO sample	δ_{iso} (ppm)	δ_{CS}	η_{CS}	C_Q (MHz)	η_Q
SVPO (MAS at 10KHz)	-497.39	-416.07	0.12	5.39	0.27
SVPO (MAS at 25KHz)	-496.9	-416.27	0.16	5.39	0.26
SVPO (static NMR)	-497.28	-416.74	0.16	5.40	0.30
DFT-Computed NMR *Shielding* Parameters*					
	σ_{iso} (ppm)	δ_{CS}	η_{CS}	C_Q (MHz)	η_Q
CASTEP SVPO#1 (computed)	-1530.36	-416.32	0.43	5.22	0.38

CASTEP SVPO#2 (single crystal XRD)	-1523.72	-413.73	0.42	5.23	0.35
CASTEP SVPO#3 (powder XRD)	-1518.66	-414.00	0.41	5.21	0.37

* CASTEP was used to calculate SVPO structures from 3 different sources: SVPO#1, computational structure from the Materials Project (MP-565532).[97] SVPO#2, a refined single-crystal SVPO structure by Lii and coworkers, and SVPO#3, a powder diffraction SVPO structure Rietveld-refined by ourselves.

Table 2.3 Principal components of the experimental and DFT calculated NMR shift/shielding tensors for pristine SVPO and discharged SVPO, and Euler angles (α, β and γ). The orientation of the principal axis system of the chemical shift tensor relative to the quadrupolar tensor is given by the three Euler angles.

Experimental NMR Chemical *Shift* Tensor Principal Components* from Dmfit Pristine SVPO						
	δ_{11} (ppm)	δ_{22} (ppm)	δ_{33} (ppm)	α	β	γ
SVPO (MAS at 10KHz)	-264.4	-314.3	-913.5	86º	161º	92º
SVPO (MAS at 25KHz)	-255.4	-322.1	-913.2	85º	161º	87º
SVPO (static NMR)	-255.6	-322.2	-914.0	81º	165º	87º
Experimental NMR Chemical *Shift* Tensor Principal Components* from Dmfit Discharged SVPO						
	δ_{11} (ppm)	δ_{22} (ppm)	δ_{33} (ppm)	α	β	γ
1ee discharged (MAS at 25 kHz)	-266.2	-343.1	-881.4	87º	164º	76º
2ee discharged (MAS at 25 kHz)	-257.6	-356.2	-875.7	89º	164º	77º
DFT-Computed NMR *Shielding* Tensor Principal Components* from CASTEP Pristine SVPO						
	σ_{11} (ppm)	σ_{22} (ppm)	σ_{33} (ppm)	α	B	γ
CASTEP SVPO#1 (computed)	-1828.0	-1649.0	-1114.0	90º	168º	90º
CASTEP SVPO#2 (single crystal XRD)	-1810.5	-1640.8	-1104.7	90º	168º	90º
CASTEP SVPO#3 (powder XRD)	-1817.5	-1643.7	-1110.0	90º	166º	90º

The table displays NMR tensors expressed in the Haeberlen convention, as well as values for chemical shift (δ, from experiment) and chemical shielding (σ, from computation) expressed in terms of the diagonalized tensor principal component values. Comparison between CASTEP shielding and experimental shift values requires referencing to a standard set of ^{51}V NMR compounds. Thus, the σ_{iso} and δ_{iso} values are not comparable between experiment and DFT. However, there is good agreement between the computed and experimental values of the other NMR parameters which are expressed in terms of the Haeberlen convention, such as the anisotropy of the chemical shift δ_{CS} (determined by the full diagonalized CSA tensor) and quadrupolar coupling constant C_Q (determined by the EFG tensor). The asymmetry parameter (η) of both CSA and EFG tensors (η_{CS} and η_Q), however, are not in as good agreement between experiment and theoretical calculations. This difficulty of accurately predicting asymmetry parameters is a familiar issue;[100] these expressions are more prone to error because the mathematical definition suffers from error propagation. The variation between CASTEP results for different structural models is small, which indicates good agreement of the structure with those published previously.

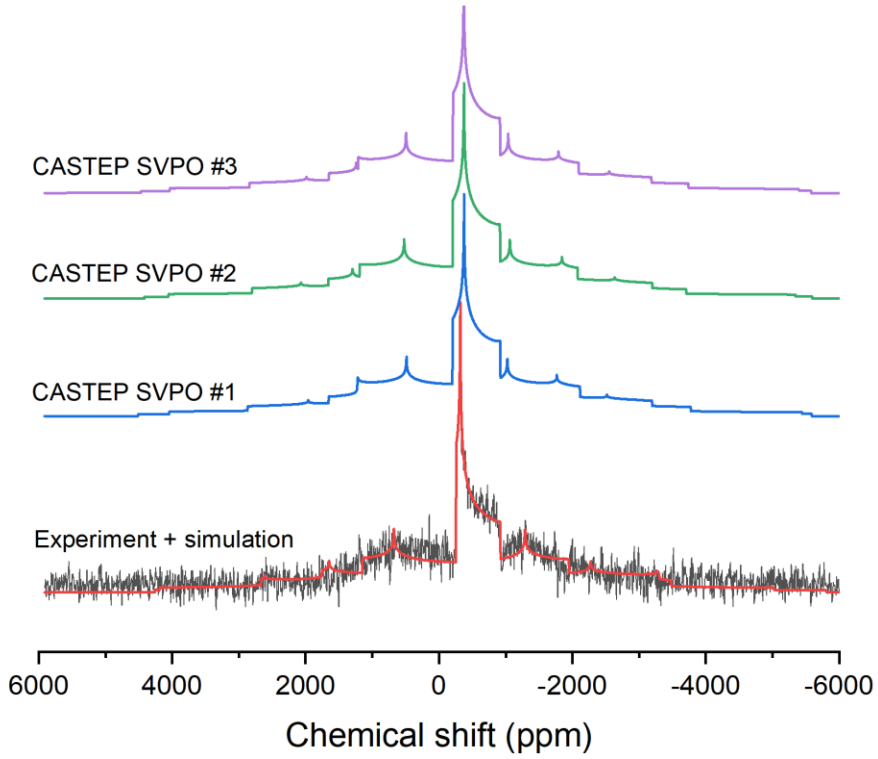


Figure 2.7 ^{51}V static SSNMR spectrum of $Ag_2VO_2PO_4$ obtained at 13.9T: experimental spectrum in black and Dmfit-simulated lineshape in red. The central transition and all satellite transitions are all depicted in this spectrum. Overlaid are the multiple CASTEP-computed models from data in **Table 2.2**.

Even with such deviations, the static NMR spectrum can illustrate the subtle differences between models. To that end, a static ^{51}V NMR spectrum of pristine SVPO was recorded, shown in **Figure 2.7**. The spectrum (black) is overlaid with the simulated static lineshape (red) from Dmfit. The lineshapes depicted represent both the central <1/2, -1/2> transition and multiple satellite transitions (for I = 7/2). The Dmfit extracted NMR parameters are also shown in **Table 2.2**. The values obtained by fitting data for the static experiment agree well with those of the MAS results.

With the NMR parameters determined, local structure at the V sites can be further confirmed. The 6-coordinate VO_6 species described here are edge-sharing dimers. Lapina et al.[78] report CSA

anisotropy (δ_{cs}) in the range of -200 to -400 ppm and CSA asymmetry parameters in the range of 0 to 0.2 for associated VO_6 which are similar to our observations for SVPO. The isotropic chemical shift found here for SVPO VO_6 edge-sharing dimers are in a range near to those reported for VO_6 dimers (-500 to -560 ppm).[101,102]

2.3.2 ^{51}V SSNMR Spectroscopy of Discharged SVPO ($Ag_2VO_2PO_4$)

Coin cells assembled using a Li metal anode and $Ag_2VO_2PO_4$ cathode ($Li||Ag_2VO_2PO_4$) were discharged from their post-assembly open circuit potential of about 3.40 – 3.55 V at a rate of C/300. Consistent with the high thermodynamic potential, yet low conductivity, of pristine SVPO, the initial potential rapidly drops to 2.3 V before recovering and levelling off at 2.6 V. The profiles observed are described by the following electrochemical processes:

$$Ag_2VO_2PO_4 + yLi \rightarrow Li_yAg_{2-y}VO_2PO_4 + yAg^0 \quad (1)$$

$$Li_2VO_2PO_4 + zLi \rightarrow Li_{2+z}VO_2PO_4 \quad (2)$$

Initial reduction of SVPO involves the reduction of the Ag^+ ions, equation (1), and corresponds to the observed potential recovery from 2.3 V to 2.6 V soon after discharge begins. As reduction proceeds, silver is displaced from the layered material as it aggregates to form a conductive Ag^0 matrix, and lithium ions intercalate in to balance charge. This reduction-displacement of silver from parent $Ag_2VO_2PO_4$, results in a drop in cell resistance and increase in conductivity. The second reaction, equation 2, represents the reduction of vanadium.

Bulk electrochemical reduction of pristine SVPO was conducted to generate 1 electron equivalent (1mol of electron discharged per molar mass of SVPO, 1ee), 2ee and 3ee discharged samples. With each reduction step comes a loss of crystallinity making these materials

challenging to characterize. Compared to pristine SVPO, 1ee reduced material is the most crystalline of the discharged samples, and a fitting analysis of the XRD data reveals that the material primarily consists of $Ag_2VO_2PO_4$ (~50%) and silver metal (~30%) along with small amounts of a (paramagnetic) lithiated vanadium phosphate phase (~10%) similar to $LiVOPO_4$ and some Li_2O (<10%) (**Table 2.4**).

Table 2.4 Refined weight fractions and estimated phase fractions from XRD fitting analysis at different levels of discharge. Relative mole estimates are based on the assumption that silver, vanadium, and phosphorous remain in the cathode. Phase fractions of $Ag_2VO_2PO_4$, Ag^0 metal, Li_2O, and $LiVOPO_4$ were determined from fits, while amorphous $Li_xVO_yPO_4$ was determined by difference such that the resulting Ag:V:P ratio was 2:1:1, consistent with $Ag_2VO_2PO_4$.

Phase	**$Ag_2VO_2PO_4$**		**$Li_1Ag_2VO_2PO_4$**		**$Li_2Ag_2VO_2PO_4$**	
	Weight %	Rel. Moles	Weight %	Rel. Moles	Weight %	Rel. Moles
$Ag_2VO_2PO_4$[7]	100%	1.0	52(7)%	0.50(9)	8(4)%	0.05(3)
Ag^0 Metal[8]	None		29(7)%	1.0(2)	82(6)%	1.90(6)
Li_2O[9]	None		6(4)%	0.8(5)	2(1)%	0.18(1)
$LiVOPO_4$[10]	None		12(3)%	0.27(8)	8(2)%	0.12(3)
Amorphous $Li_xVO_yPO_4$	None		None		Unknown	0.83(6)

As previously mentioned, reduction leads to increased conductivity as Ag^0 deposits form, and to an expansion of the inter-layer distances as lithium ions intercalate between the layers of VO_6 and PO_4 polyhedra. The amorphization of the material upon reduction makes these materials challenging to characterize by XRD. XANES data has shown that the 6-coordinate VO_6

environment is maintained throughout the discharge process despite severe distortion.[72] However, using detailed EXAFS analysis to determine local coordination environments is difficult due to beam attenuation by the electron-dense Ag atoms surrounding the V centers. SSNMR is useful as it can be used to monitor changes in the local site symmetry around vanadium upon reduction, including non-crystalline vanadium species that cannot be detected by XRD.

Geometry optimization and CASTEP-NMR computations were performed with the atomic coordinates obtained from the Rietveld-refined structure of the $Ag_2VO_2PO_4$ component of the 1ee discharged sample (shown in **Figure 2.8**). The CASTEP-computed ^{51}V EFG and chemical shielding tensors served as the starting point for Dmfit simulation of the ^{51}V MAS NMR spectrum (shown in red in **Figure 2.9**), which resolves a single vanadium site we assign to (1ee) $Ag_2VO_2PO_4$ (**Table 2.5**). When comparing ^{51}V NMR tensor values between those measured here to those of the SVPO parent compound (**Table 2.2**), it is notable that the isotropic resonance remains essentially the same, which suggests that NMR parameters other than the isotropic should be used as the probe for local structural changes here. The XRD data indicate that while the major component of the 1ee discharged SVPO sample is $Ag_2VO_2PO_4$, there are local changes about the VO_6 and PO_4 centers. Significant changes to the crystal structure can be seen through comparison of pristine SVPO to that of the 1ee discharged sample. **Figure 2.10** and **Figure 2.11** graphically depict changes to the V-O bond lengths to illustrate the shift of atoms that occurs in a pair of edge-sharing VO_6 centers upon reduction. These changes in the local structure are reflected in a significant change in the ^{51}V NMR chemical shift anisotropy (d_{CS}) and a measurable difference in the quadrupolar coupling constant (C_Q). In a study of a series of $Li_xV_2O_5$ ($0.4<x<1.4$) materials, Nakamura et al.[103] also found that as V^{5+} is reduced, the ^{51}V

NMR isotropic chemical shift stays approximately the same while d_{CS} and C_Q change. It is worth noting that for both the $Li_xV_2O_5$ study and this study, the ^{51}V NMR tensor parameters (d_{CS} and C_Q) better reflect changes to the local environment than the isotropic chemical shift.

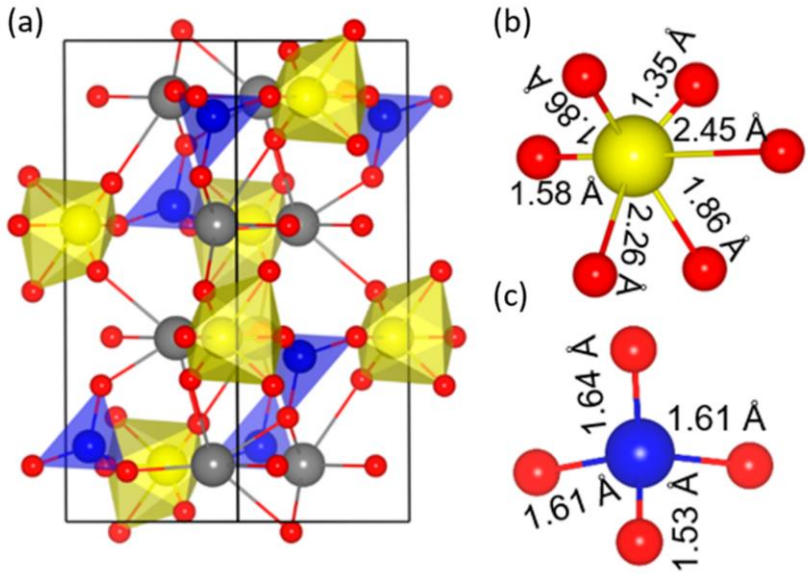


Figure 2.8. (a) Rietveld refined structure of 1ee discharged $Ag_2VO_2PO_4$ with VO_6 shown in yellow, PO_4 shown in blue and Ag atoms shown in gray. (b) Strongly distorted vanadium centers. (c) Slightly distorted phosphorus tetrahedra.

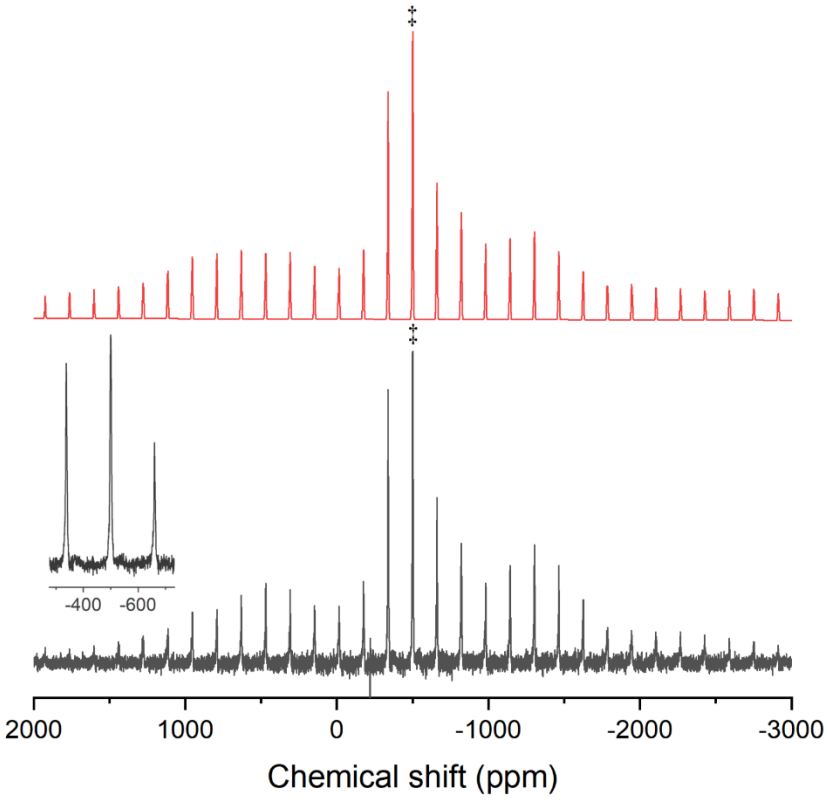


Figure 2.9 ^{51}V MAS NMR of 1ee discharged $Ag_2VO_2PO_4$, obtained at 13.9 T and ν_R = 25 kHz MAS. Experimental (black) and simulated (by Dmfit the "int2quad" model, red) using CASTEP-NMR tensor parameters as an "initial guess." Symbols (‡) indicate isotropic resonances. Inset image shows expansion of the region near the isotropic resonance.

Table 2.5 Experimental and DFT ^{51}V NMR parameters for electrochemically discharged SVPO structures, by Haeberlen convention expressions. Note: experimental measurements are expressed in terms of chemical shift (δ) and computed values in terms of shielding (σ). (See Table 2.3 for individual tensor components).

Experimental NMR Chemical *Shift* Anisotropy Parameters (fitted by Dmfit) Discharged SVPO					
Discharged SVPO sample	δ_{iso} (ppm)	δ_{CS}	η_{CS}	C_Q (MHz)	η_Q

1ee discharged (MAS at 25 kHz)	-496.90	-384.47	0.20	5.25	0.26
2ee discharged (MAS at 25 kHz)	-496.50	-379.21	0.26	5.18	0.24
3ee discharged (MAS at 25 kHz)	-495.87	--	--	--	--
DFT-Computed NMR *Shielding* Parameters Discharged SVPO					
	σ_{iso} (ppm)	δ_{CS}	η_{CS}	C_Q (MHz)	η_Q
1ee discharged SVPO (CASTEP)	-1500.76	-400.39	0.37	5.47	0.37

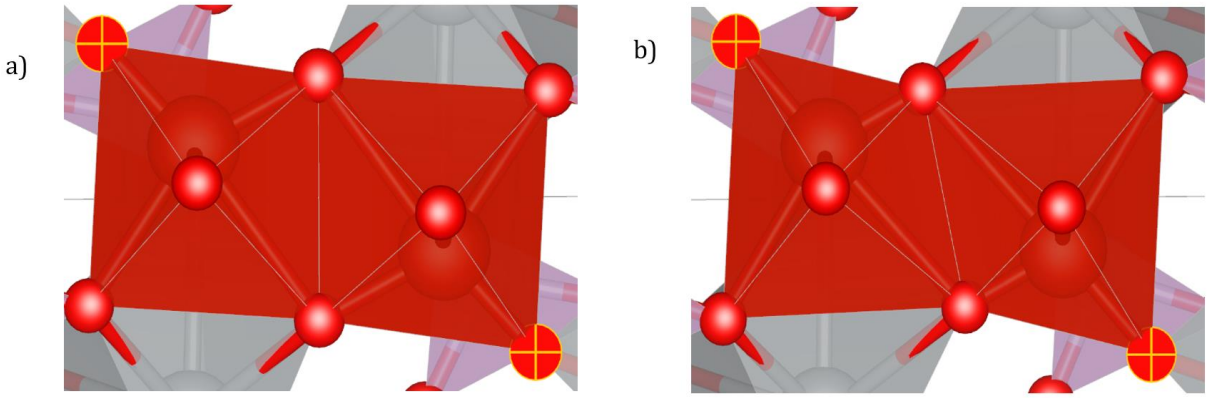


Figure 2.10 VESTA visualization of the V_2O_{10} units (made up of edge sharing VO_6 polyhedra) in a) pristine SVPO, and b) 1ee discharged SVPO.[72]

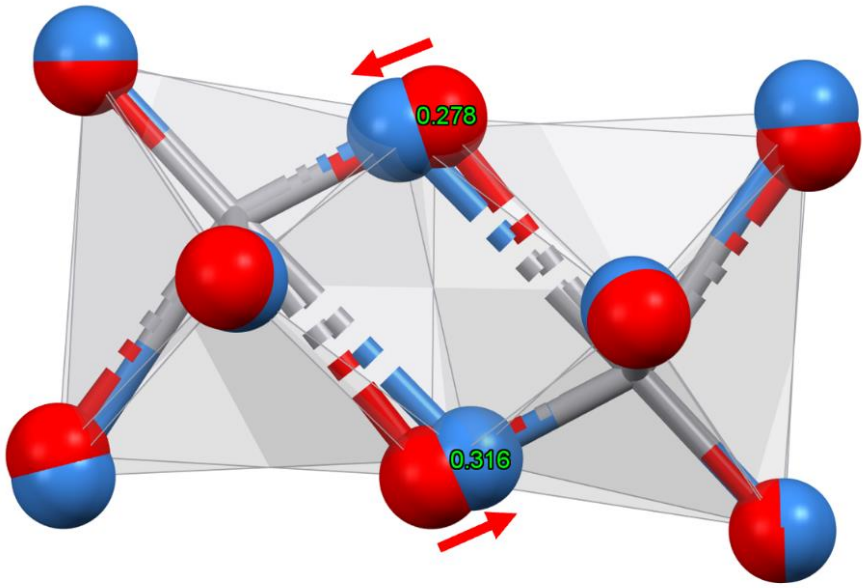


Figure 2.11 Overlay of the V_2O_{10} units (made up of edge sharing VO_6 polyhedra) in pristine SVPO (oxygens atoms in red) and 1ee discharged SVPO (oxygen atoms in blue). The most significant shifts involve the bridging O atoms (distance (Å) of the shift is indicated in green).[71]

The further reduced samples were also analyzed by ^{51}V NMR spectroscopy, and solid-state MAS spectra for the 1ee, 2ee and 3ee reduced samples are shown in **Figure 2.12**. As reduction proceeds a substantial loss of signal intensity was observed. These findings are consistent with a paramagnetic sample, as expected for V^{4+} ($Li_{2-z}VO_2PO_4$).[71] The reduction of diamagnetic V^{5+} to a lower 4+ oxidation state leads to a paramagnetic vanadium species and a concomitant increase in magnetic susceptibility.[71] The strong electron-nuclear interaction between the unpaired electron and the nucleus gives a very fast T_2 relaxation which makes paramagnetic V^{4+} and V^{3+} undetectable by NMR.[104] As expected, this increase in paramagnetic character leads to an decrease in signal/noise (S/N) ratio due to the lower abundance of detectable V^{5+}. According to the electrochemical discharge data, Ag^+ is initially reduced to Ag^0. However, Ag^+ is not completely reduced prior to reduction of V^{5+}; these occur concurrently.[73] The magnetic

susceptibility of the material increases with each reduction step, although not in a linear fashion.[71] Based on the magnetic susceptibility data, the 1ee discharged SVPO sample only contains a small amount of paramagnetic material, whereas the 3ee discharged SVPO sample contains the largest amount.[71] Notably, the magnetic susceptibility of the 2ee discharged material more closely resembles that of the 3ee discharged material. The isotropic chemical shift values for both the pristine SVPO and the discharged SVPO compounds does not change indicating that the ^{51}V NMR spectrum of V^{5+} receives limited influence from paramagnetic V^{4+} and V^{3+}. The increased paramagnetic behavior along with the increase in uncertainty of the XRD structural model introduces complications in the CASTEP calculations. With the onset of reduction, amorphization begins to occur, therefore, it is not surprising that there are some differences between the experimental and calculated parameters for the 1ee discharged sample.

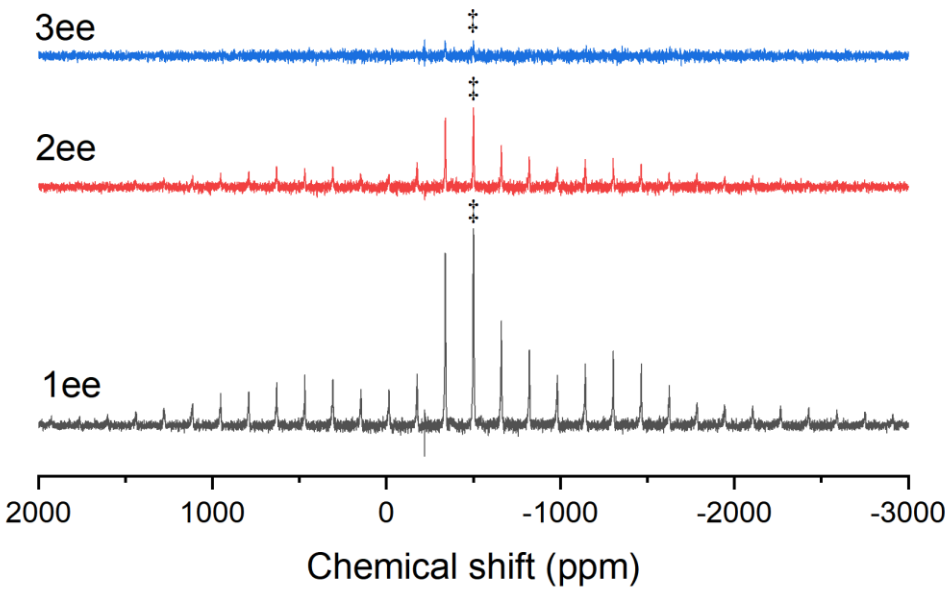


Figure 2.12 Comparison of the ^{51}V MAS NMR spectra of 1ee, 2ee and 3ee discharged $Ag_2VO_2PO_4$, obtained at 13.9 T and v_R = 25 kHz MAS.The same number of scans were collected for each spectrum. Symbols (‡) indicate isotropic resonances.

The ^{51}V MAS NMR of the 2ee discharged $Ag_2VO_2PO_4$ is shown in greater detail **Figure 2.13**. We were able to use Dmfit to deconvolute the spectrum of the 2ee discharged sample. Although the XRD analysis of the 2ee discharged sample of $Ag_2VO_2PO_4$ was performed, the amorphization of SVPO upon reduction along with the presence of multiple components severely complicates any type of detailed structure analysis through diffraction methods. In fact, the XRD analysis indicates the crystalline portion of the 2ee discharged sample consists primarily of silver metal (over 80%) along with small amounts of starting $Ag_2VO_2PO_4$ (~10%) and approximately 10% of the lithiated vanadium phosphate phase ($LiVOPO_4$) (**Table 2.4**). In the absence of a crystallographic model structure, we were not able to use CASTEP to calculate the NMR spectral data for the 2ee and 3ee discharged material with desired accuracy.

What is lacking from the XRD analysis is any information about the amorphous material that is present. Analyzing the 2ee discharged sample using ^{51}V SSNMR we note that the isotropic resonance remains the same as the pristine and the 1ee discharged SVPO. However, there are further changes to the chemical shift anisotropy and the quadrupolar coupling parameters. The general trend from pristine, to 1ee to 2ee discharged SVPO is that the anisotropy of the CSA becomes more positive (from -416 ppm in pristine SVPO to -379 ppm in the 2ee discharged sample) while the C_Q decreases slightly in magnitude (from 5.39 MHz in pristine SVPO to 5.18 MHz in the 2ee discharged sample). These data indicate that as reduction occurs the bulk material retains the basic SVPO structure, while causing greater distortions in the coordination environment surrounding the VO_6 centers. Thus, the fact that NMR spectroscopy is not impeded

by the lack of crystallinity of the sample, the ^{51}V NMR spectra of the 2ee discharged SVPO samples provide a unique window into the structure of this material.

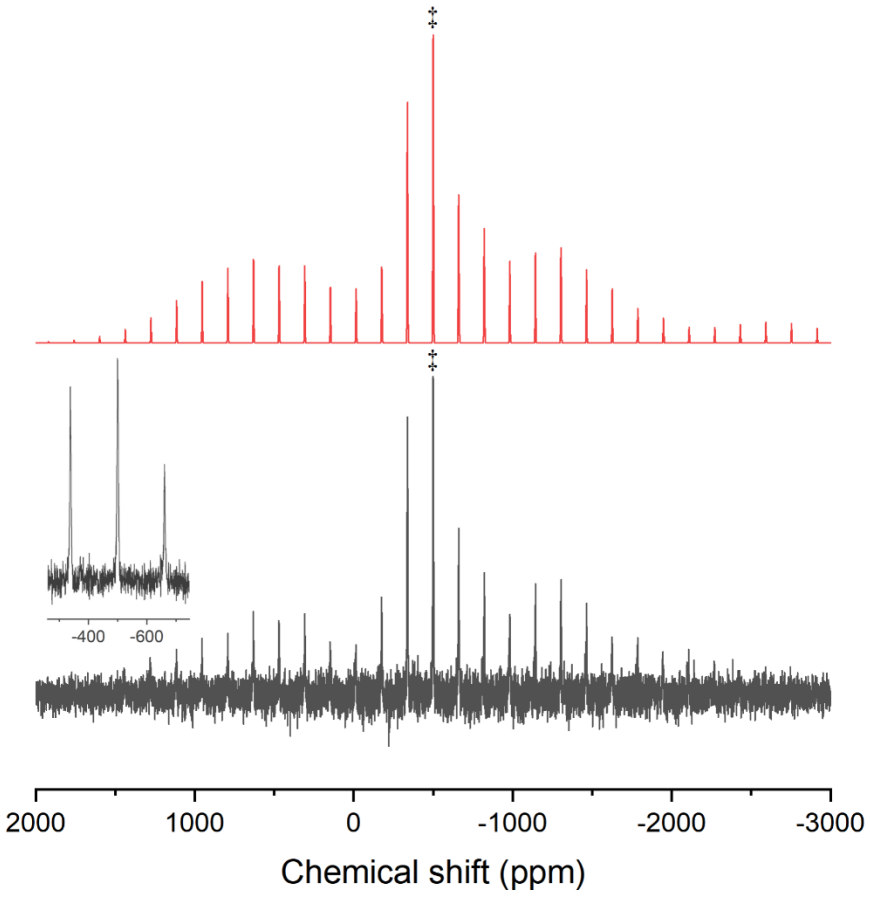


Figure 2.13 ^{51}V MAS NMR of 2ee discharged $Ag_2VO_2PO_4$, obtained at 13.9 T and ν_R = 25 kHz MAS. Experimental (black) and simulated by Dmfit the "int2quad" model, using the 1ee discharged spectrum for Site 1 (red). Symbols (‡) indicate isotropic resonances.

It is important to note that a NMR spectral study of the commercially important silver vanadium oxide ($Ag_2V_4O_{11}$, SVO) was reported in 2007 by Greenbaum and coworkers.[105] The NMR analysis focused on the ^{7}Li SSNMR data of reduced SVO (~1ee up to ~6ee discharged SVO) which showed evidence of Li^+ ions in three different environments corresponding with three separate voltage plateaus.

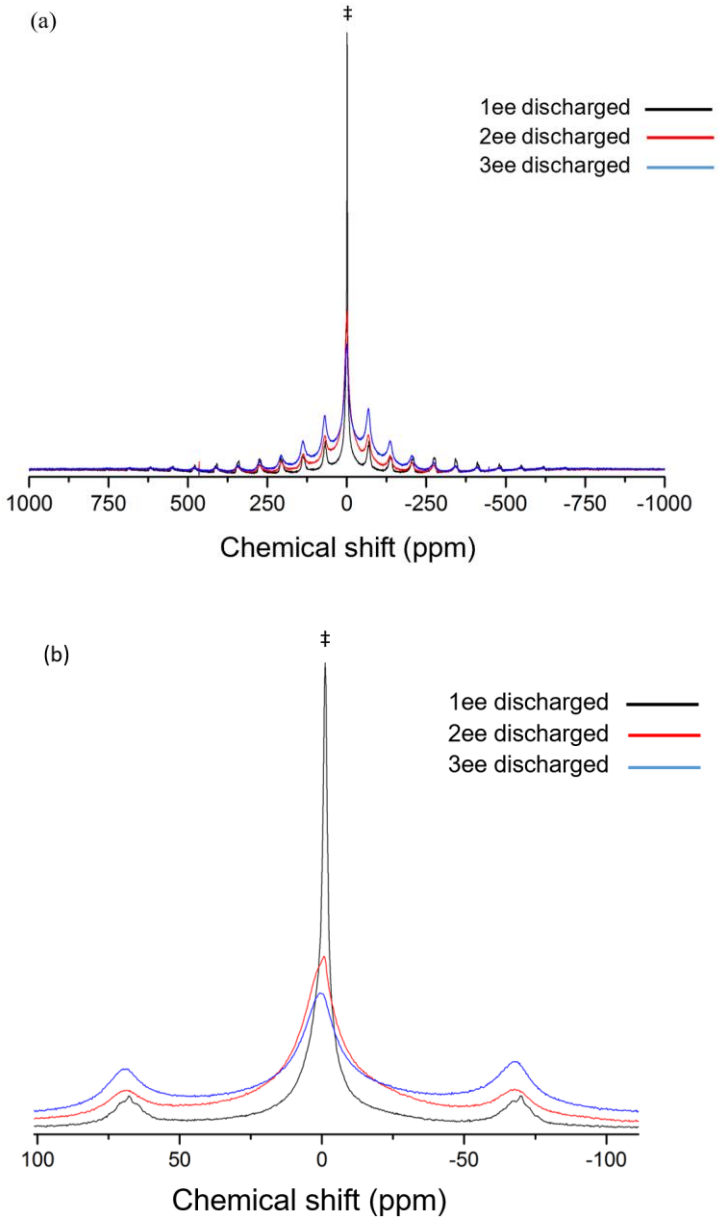


Figure 2.14 (a) Experimental ^{7}Li solid-state MAS NMR spectra of the 1ee (black), 2ee (red), and 3ee (blue) discharged reduced samples, obtained at 7 T and ν_R = 8 kHz. Symbol (‡) indicates the isotropic resonance. (b) Expanded view of the experimental ^{7}Li solid-state MAS NMR spectra of the 1ee (black), 2ee (red), and 3ee (blue) discharged reduced samples, obtained at 7 T and ν_R = 8 kHz. Symbol (‡) indicates the isotropic resonance.

Here, we provide representative ^{7}Li NMR spectral data of the discharged SVPO samples without a detailed analysis. Although there are similarities between the ^{7}Li NMR data of SVPO and SVO (evidence of Li^{+} in different environments and broadening of peaks with increased discharge), we focused our attention on the ^{51}V nucleus which is a more sensitive probe in elucidating structural changes within this material. Leifer *et al.* did report the ^{51}V NMR data (quadrupolar echo) of pristine SVO as well as some of that of the reduced samples.[105] Like the changes we observed for the 1ee reduction of SVPO, they observed minor changes in the central transition as well as the quadrupolar satellite transitions in the ^{51}V NMR spectral data of $Li_{0.72}Ag_2V_4O_{11}$ (~1ee discharged SVO) with pristine SVO. Thus, the crystallinity and local environment about vanadium are not substantially altered during the early stages of reduction for SVO as well as SVPO. The spectral data for the ~2ee discharged SVO sample displays more noticeable differences when compared with the data of pristine SVO, and further reduction (~6ee discharged SVO) displays a significant loss of signal which is expected as the paramagnetic properties of the sample increase as V^{5+} is reduced. The chemical shielding (shift) values are useful to researchers when trying to understand the components present in an amorphous material, and here we are reporting several sites that may be useful to test structural models of the discharged, amorphous SVPO. Instead of visually comparing the lineshapes, utilizing Dmfit and CASTEP to extract NMR lineshape parameters has enabled us to conduct an analysis that reflects the vanadium local environment, and we find similar observations between the two systems.

2.4 Conclusions

Spectroscopic analysis of SVPO and related discharged materials has been conducted. Dmfit was utilized to determine key NMR parameters for pristine, 1ee and 2ee discharged SVPO.

CASTEP DFT calculations were also applied to pristine and 1ee discharged SVPO samples. Good agreement was found between the DFT and experimental parameters for pristine SVPO material, while deviations of the NMR lineshape parameters were found for the 1ee discharged sample due to onset of amorphization of the material. The data reveal that while the isotropic chemical shift does not change substantially upon reduction of the material, there are significant changes in the reduced chemical shift anisotropy and the quadrupolar coupling constant. These changes, manifested as NMR lineshape differences, indicate that while the local coordination environment of VO_6 dimers is maintained throughout the reduction process, there are deviations in the local symmetry about the VO_6 center. This is information that could not be obtained from XRD data since there is a significant loss of crystallinity in the vanadium phases after more than 1ee have been added. Therefore, the sensitivity of SSNMR to local coordination environments provides a rare window into the structure, when it lacks long-range order, of the amorphous, reduced samples. Possession of NMR tensors such as these could one day enable modeling of amorphous vanadium oxide structures containing V^{5+}.

Chapter 3

Enabling Materials Informatics for ^{29}Si Solid-state NMR of Crystalline Materials

The material in this chapter is adapted from Sun H, Dwaraknath S, Ling H, et al. Enabling materials informatics for ^{29}Si solid-state NMR of crystalline materials. npj Computational Materials, 2020, 6(1): 1-7.

3.1 Introduction

This chapter will be focusing on systematically validating two density functional theory (DFT) methods (CASTEP/VASP) for the calculation of solid-state NMR (SSNMR) chemical shift/shielding tensor parameters for spin ½ species. SSNMR has revolutionized organic and biological chemistry fields, owing to its ability to provide precise structural detail through investigation of ^{1}H and ^{13}C spectra. Assignments of these spectra rely on 50+ years of comprehensive and detailed data, many of which have been catalogued in guides from Sadtler[106] and Aldrich[107] and subsequently in databases such as the AIST Spectral Database for Organic Compounds SDBS[108].

For inorganic species, there are far fewer resources, and through the Local Spectroscopy Data Infrastructure (LSDI), we have begun to develop a database of both known and predicted NMR spectra for less-commonly studied nuclei, beginning with ^{29}Si. The data infrastructure serves as a platform to compute ^{29}Si NMR tensors and generate model spectra by using crystalline compounds in The Materials Project database. ^{29}Si is attractive, because it is a nuclear spin, *I*, 1/2 species, found at moderate natural abundance (4.68%)[109], and studied as a constituent in minerals, zeolites, and amorphous glasses.

X-ray diffraction (XRD) has been the primary tool for determining the structure of crystalline materials, for nearly a century. Determination of lattice parameters, symmetry, and coordinates of moderate- to high-Z species in the lattice is relatively straightforward, making XRD a powerful and versatile analytical tool. As the demand for accuracy of atomic coordinates increases, structures proposed based only on XRD have been shown to lack accuracy for lighter elements, such as H[110–113]. In this case, other experimental techniques like neutron diffraction and recently nuclear magnetic resonance (NMR) have been employed to lend accuracy. This NMR refinement of structures is termed "NMR crystallography"[112,114–117]. Solid-state NMR is also a powerful tool to characterize the local environments of unique sites within a crystalline material, where alterations in the local environment can shift NMR resonances: small distortions to bond lengths and angles can perturb spectra in ways that are manifested in information gleaned, especially in the solid state.

The exponential increase in computational power over the past two decades enables theoretical methods to scale across structure and chemistry more easily than experimental methods. In the field of solid-state NMR, however, most of the research utilizing computational methods are focused on a handful of structures at a time. [118–120]The potential of rapidly characterizing NMR properties based on a large computational database coupled with consistent standards is still underestimated. Thus, within certain approximations necessary for tractable simulations, a dataset of simulated NMR tensors and interactive tools to visualize and explore NMR spectra has the potential to drastically increase the accuracy and efficiency of the study of solid-state materials. The LSDI is constructed with plane wave basis density functional theory (DFT) calculations using two popular codes: the Vienna Ab initio Simulation Package (VASP) and Cambridge Serial Total Energy Package (CASTEP). In this study, we seek to demonstrate that

both packages are effective at calculation of NMR shielding tensors (σ) for ^{29}Si. The isotropic chemical shift is the most familiar experimental NMR parameter to researchers (δ_{iso}); however, other lesser explored individual tensor elements from the solid-state lineshape add critical information about the local environment. Prediction of the full diagonalized tensor is useful for planning experiments, both under static solid-state or magic-angle-spinning (MAS) NMR conditions, that will enable accurate extraction of these values. It has been shown in a separate ^{13}C study[111], determination of the chemical shift tensor values enabled refinement of the H positions in a polycrystalline sample. Possessing catalogues of tensor values will ultimately accelerate "NMR crystallography"--to refine the local environment around nuclei being probed during NMR experiments.

Furthermore, this study illustrates an important aspect of cataloguing experimental data and comparing these to computations. As experimental measurements improve over time, there are often improved tools to provide more accurate interpretation of data. In this case, by examining a large set of tensors, it has been possible to identify assignment errors in tensor elements arising from the use of Haeberlen notation, described below. In addition, systematic differences between CASTEP and VASP are found, which are critical when reporting the full shielding (or shift) tensor, that are not evident when considering only the isotropic values, σ_{iso} and δ_{iso}.

3.2 Computational Methods

3.2.1 Dataset

We have identified ^{29}Si NMR of crystalline compounds to use as a benchmarking set, nearly all of which have been analyzed by solid-state magic angle spinning (MAS) NMR or static single

crystal NMR (2). This set is comprised of 31 structures[121–125], with 42 unique silicon sites primarily in minerals such as forsterite, wollastonite, zeolites, and quartz.

3.2.2 Density functional theory (DFT) calculation

CASTEP has been shown to be very effective for calculations of isotropic chemical shifts for nuclei such as ^{1}H, ^{13}C, ^{89}Y, and ^{119}Sn [114,126–129] as well as diagonalized tensor values for ^{19}F and ^{77}Se [130–132] in select systems. DFT calculations using CASTEP were performed within the Perdew-Burke-Enzerhof (PBE) Generalized Gradient Approximation (GGA) formulation of the exchange-correlation for both geometry optimization and NMR calculations. "On the fly" generated ultra-soft pseudopotentials were used to approximate the interaction between the core electrons and the nuclei.

Convergence tests on the remnant stress at the end of geometry optimization calculations with respect to cut-off energies and k-points were applied to α-quartz with CASTEP. Similarly, for NMR GIPAW calculations, calculation parameters were converged with respect to the anisotropy versus cut-off energy and number of k-points. As a result of the test, we chose the cut-off energy as 800 eV and 14 k-points. For all structures, a constant k-point density of 0.025 $Å^{-1}$ was used. As shown in **Figure 3.1**, the cut-off energy of the plane-wave basis set was 800 eV, and the separation of k-points in the reciprocal space was 0.025 $1/Å$.

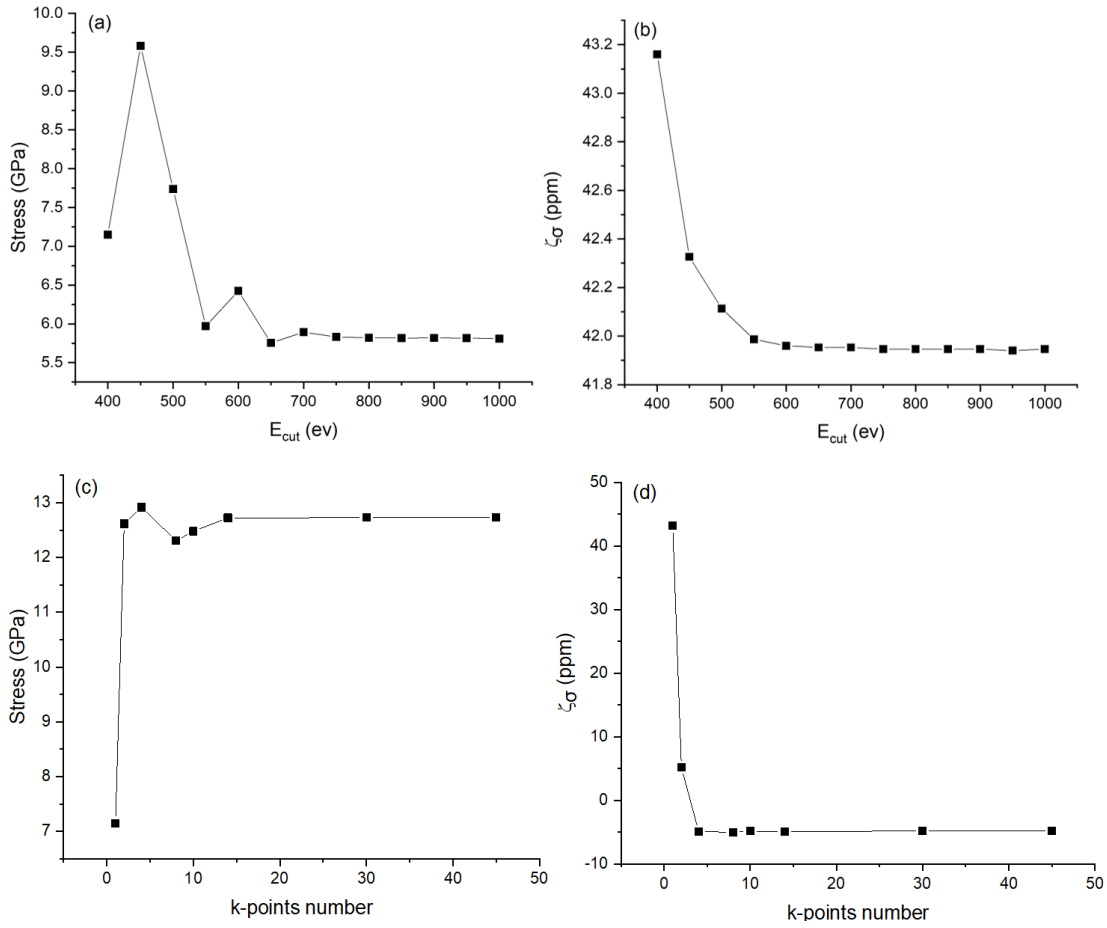


Figure 3.1 α-quartz convergence tests. (a) convergence test of stress with increasing cut off energy. Cutoff energy is specified from 400 eV to 1000 eV in increments of 50ev. The system converged after 700 eV. (b) convergence test of reduced chemical shift anisotropy ζ_σ with increasing cut off energy. The system converged after 600ev. (c) convergence test of stress with increasing number of irreducible k-points. k-points from 1 (Monkhorst-pack 1×1×1) to 155 (12×12×10). The system converged after 14 k-points (5×5×4). (d) convergence test of reduced chemical shift anisotropy ζ_σ with increasing number of irreducible k-points. The system converged after 4 k-points (3×3×2).

DFT calculations were also performed using the projector augmented wave (PAW) method[133,134] as implemented in the Vienna *Ab Initio* Simulation Package (VASP)[135–137] within the Perdew-Burke-Enzerhof (PBE) Generalized Gradient Approximation (GGA) formulation of the

exchange-correlation functional[138]. A cut-off for the plane waves of 520 eV is used and a uniform k-point density of approximately 1,000/atom is employed. We note that the computational and convergence parameters were chosen in compliance with the settings used in the Materials Project (MP)[91] to enable direct comparisons with the large set of available MP data.

CASTEP and VASP both use the Gauge Including Projector Augmented Waves (GIPAW) method[95] to reconstruct the core wavefunction and perform NMR calculations.

In this benchmarking set, we focus on species whose full CSA tensor has been reported. When possible crystalline structure coordinates accompanying the tensor values were used as the basis for DFT optimization and tensor calculation. When not explicitly specified, structures from the ICSD database were the starting point for geometry optimizations.

All the computationally-obtained parameters were subsequently used in simulations of spectra using the lineshape-generating program, Dmfit[10]. Experimentally, ^{29}Si materials are mostly measured using either static NMR or MAS NMR. For the ease of comparison between computed and experiment results, two models are used in the simulation: "CSA static" for static NMR lineshapes (CSA powder patterns), and "CSA MAS" for the NMR spectrum of the manifold of spinning sidebands found for a given MAS rotation frequency, ν_r. Since this rotation frequency is an easily-adjustable parameter, it is straightforward to simulate multiple "spinning-sideband manifolds" that essentially map onto the static CSA-broadened lineshape.

3.2.3 Chemical shift/shielding referencing through linear regression

DFT codes like CASTEP and VASP calculate the chemical shielding tensor of the target system. As discussed in chapter one, from the shielding tensor we can get the isotropic shielding σ_{iso}, which corresponds to the isotropic chemical shift δ_{iso} of the target compound when it is

referenced to the bare nucleus in vacuum[8] (see Note below). In practice, the isotropic chemical shift is usually obtained by comparing the resonance frequency between the target compound and an experimentally measured reference compound. For computational studies, it is necessary to reference the isotropic shielding to the isotropic chemical shift both the enable comparison with experimental data, and to do a better evaluation of the calculation accuracy.

There are multiple methods for referencing[139]. The method employed here is to create a correlation plot of computed σ_{iso} versus experimentally reported δ_{iso} values, shown in **Figure 3.3** in the next section. This method takes advantage of having a large number of data points from which a linear regression can determine the intercept. The advantage is to use the correlation of many such values to establish the reference where the errors from individual species may cancel. Using both values, the chemical shielding tensor can generate the computational isotropic chemical shift. The linear regression model from **Figure 3.3 (a)** between CASTEP computations and experimental data results in the following linear relation:

$$\sigma_{iso} = -1.12\ \delta_{iso} + 316.26\ \text{ppm} \qquad (3.1)$$

For VASP computations, the same linear regression method can also be applied. The resultant model is:

$$\sigma_{iso} = 1.15\ \delta_{iso} + 528.18\ \text{ppm} \qquad (S2)$$

The intercept is the reference value for isotropic shielding that is applied to all the computational data. Furthermore, the deviation of the slope from 1 (here -1.12 and +1.15) describes a systematic error[139] that could arise from a number of variables such as uncertainty in experimentally determined structure or measurement parameters, lack of correlation in the

electronic structure, or the lack of higher order terms in the full evaluation of the tensor such as J-coupling.

The results after adjusting computed σ_{iso} to computed δ_{iso} are plotted below in **Figure 3.2** for both CASTP and VASP, with the correlation to experimental δ_{iso}. The linear correlation here ideally has a slope of + 1 with an intercept of 0. It is worthwhile to note: only after the referencing process can the value of RMSE between computation and experiment be obtained. Therefore, the RMSE value shown in **Figure 3.3 (a)** and **3.3 (b)** is the same as **Figure 3.2 (a)** and **3.2 (b)**.

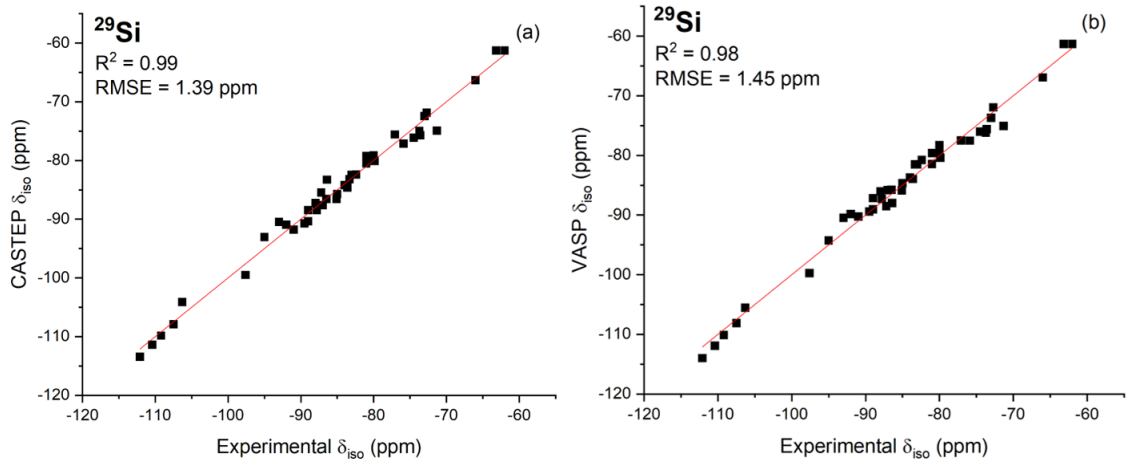


Figure 3.2 Correlation plots of ^{29}Si isotropic chemical shift/shielding between experiments and DFT calculations. a) Comparison of CASTEP computed ^{29}Si isotropic chemical shift δ_{iso} versus experimental δ_{iso} after referencing. The linear model (red) has a slope of 1.000 and an intercept of 3.07×10^{-6} ppm. b) Comparison of VASP computed ^{29}Si isotropic chemical shift δ_{iso} versus experimental δ_{iso} after referencing. The linear model (red) has a slope of 1.000 and an intercept of -3.30×10^{-6} ppm.

3.3 Results and Discussions

3.3.1 DFT benchmarking set of ^{29}Si crystalline materials

The solid-state NMR parameter known with the highest precision is the experimentally measured isotropic chemical shift, δ_{iso}. This value is the average of all 3 principal components of the diagonalized tensor. Small inaccuracies in the principal components are partially averaged when considered in their expression, as the average: $(\delta_{XX} + \delta_{YY} + \delta_{ZZ})/3$. As the most frequently reported (experimental) parameter, the comparison between experiment and computation has particular significance for researchers.

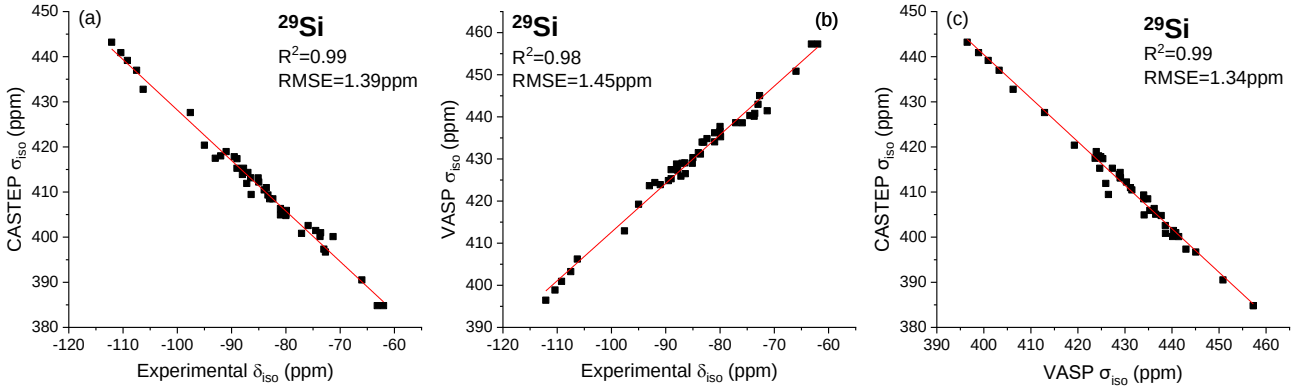


Figure 3.3 Correlation plots of isotropic chemical shift/shielding between DFT calculation and experiments. (a) Comparison of CASTEP computed ^{29}Si isotropic chemical shielding values (σ_{iso}) versus experimentally reported ^{29}Si isotropic chemical shifts (δ_{iso}), R^2=0.99 and RMSE=1.39 ppm. (b) Comparison of VASP computed ^{29}Si isotropic chemical shielding values (σ_{iso}) versus experimentally reported ^{29}Si isotropic chemical shifts (δ_{iso}), R^2=0.98 and RMSE=1.45 ppm. (c) Comparison of CASTEP-computed versus VASP-computed ^{29}Si isotropic chemical shielding (σ_{iso}) values with a linear regression, R^2=0.99 and RMSE=1.34 ppm. For all three plots, the fit to a linear regression is shown in the figure.

In the computations we extract chemical *shielding* tensors. The calculated parameters are compared with the 42 sets of experimentally reported (chemical *shift*) tensors as a benchmarking

set, and the reference isotropic chemical shift is obtained by extrapolation of a linear regression model[140] described in detail in Section 2[8].

Shown in **Figure 3.3 (a)** is the linear relationship between the CASTEP-computed ^{29}Si isotropic chemical shielding, σ_{iso}, and the experimentally measured ^{29}Si isotropic shift, δ_{iso}. **Figure 3.3 (b)** is a similar plot of VASP computed σ_{iso} versus experimental values δ_{iso}. Each data point in the plot represents a unique Si site in a crystalline material. The resultant value for reference isotropic chemical shielding within CASTEP is $\sigma_{reference} = 316.26$ ppm, and the slope of the correlation plot is -1.12. The resultant value for reference isotropic chemical shielding within VASP is $\sigma_{reference} = 528.18$ ppm, and the slope is +1.15. There is a very high degree of correlation, with an R^2 value of 0.99 and RMSE of 1.39 ppm for CASTEP, and R^2 of 0.98 and RMSE of 1.45 ppm for VASP. This strong linear correlation demonstrates the ability of DFT to compute chemical shielding with sufficient precision to match experimentally determined chemical shifts for inorganic materials. A high degree of correlation in this benchmarking set gives us confidence that additional crystalline materials will also have accurate prediction of the ^{29}Si chemical shielding/shift. Additionally, σ_{iso} of the same data set was predicted by VASP. **Figure 3.3 (c)** compares VASP and CASTEP computed σ_{iso} values demonstrating very good agreement between VASP and CASTEP that shows both platforms perform well, modeling the ^{29}Si isotropic chemical shielding. These data are all collected in tables in the Appendix A.

Beyond isotropic shift, the additional two algebraic expressions (ζ_δ and η_{CSA}) can be directly linked to the individual tensor elements that express the shape of the experimental lineshape, whether static NMR or a manifold of spinning sidebands under MAS NMR. **Figure 3.4** is a

schematic illustrating the relationship of principal components of the chemical shift tensor, as well as δ_{iso} and ζ_δ for a lineshape with a representative η_{CSA} value of 0.4.

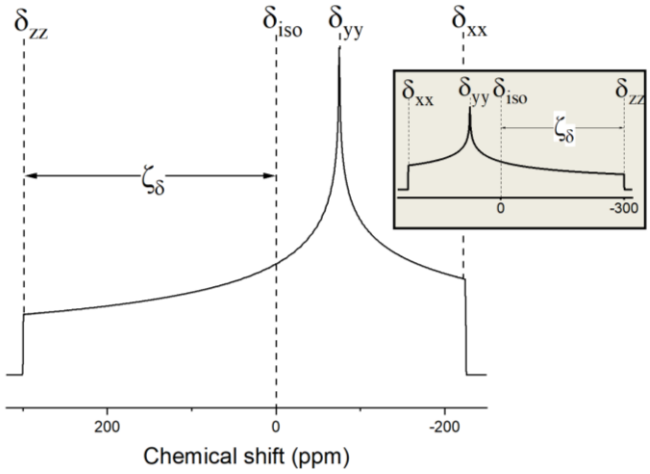


Figure 3.4 A simulated static lineshape dominated by chemical shift anisotropy (CSA) interactions. Based on this lineshape, the three principal components of the chemical shift tensor can be identified individually as δ_{XX}, δ_{YY} and δ_{ZZ} based on the notation from Haeberlen. These values are usually reported as: isotopic chemical shift (δ_{iso}), chemical shift anisotropy (ζ_δ), and the asymmetry parameter (η_{CSA}). The subplot shows an alternative version of the simulation with δ_{XX} and δ_{ZZ} switched. In this case, the spectrum has a negative value for the anisotropy.

3.3.2 Challenges for cataloguing the full shielding tensor, reduced anisotropy ζ_δ

Since most of the benchmark compounds have reported the Haeberlen quantities of "asymmetry parameter", η_{CSA}, and reduced anisotropy of CSA (ζ_δ), we examine the relationship between experimentally measured values (largely from past literature) and computations below.

We have reconciled past experimental reports of the ^{29}Si reduced anisotropy of the chemical shift (ζ_δ) and depict our findings in the following set of figures. The comparison between experimentally reported reduced anisotropy and the computed values from CASTEP (or VASP)

reveals issues faced when cataloguing data. **Figure 3.5** depicts a comparison of 42 experimentally reported reduced anisotropies from the literature with the corresponding values predicted by CASTEP. While a high degree of correlation is found for most of the data, a number of significant "outliers" are identified.

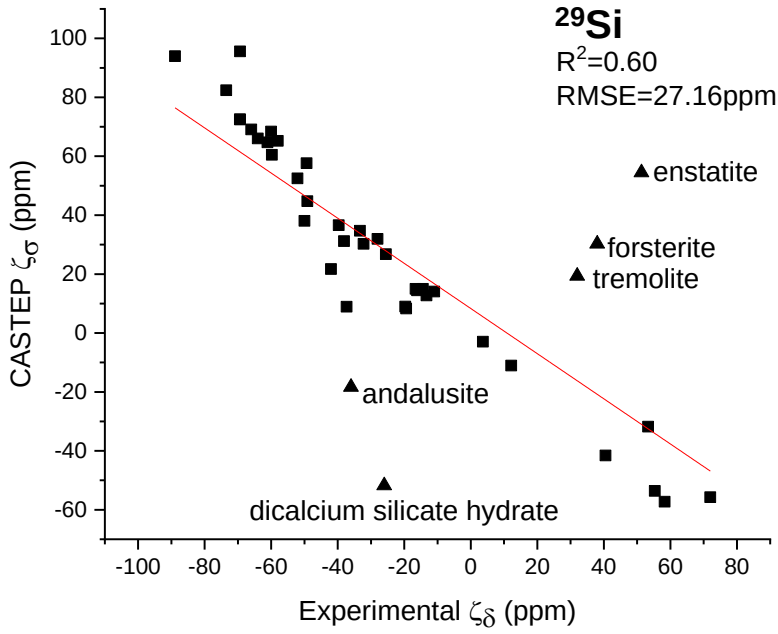


Figure 3.5 Comparison of CASTEP computed ^{29}Si reduced shielding anisotropy ζ_σ versus experimentally reported reduced chemical shift anisotropy ζ_δ. The outliers (shown by ▲symbols) were closely analyzed in terms of experimental and computational error.

Given the difficulty establishing high-quality correlations between computed and experimentally reported Haeberlen lineshape parameters, $\zeta_\sigma$and η_{CSA}, we believe it is fruitful to compare how the two different DFT packages predict these values. CASTEP utilizes ultra-soft pseudopotential to simplify the calculation of the core electron wave functions while VASP employs the Projector-Augmented-Wave (PAW) method, providing a meaningful algorithmic difference to compare the calculation and methodological robustness.

The output of DFT NMR calculations for both CASTEP and VASP performed on the same set of crystalline structures are compared with each other in **Figures 3.6**. There is a high degree of correlation, and the outliers shown in **Figure 3.5** are absent. This finding supports the argument that the outliers in **Figure 3.5** are due to the experimental mis-assignment of the δ_{XX} and the δ_{ZZ} components of the chemical shift tensor.

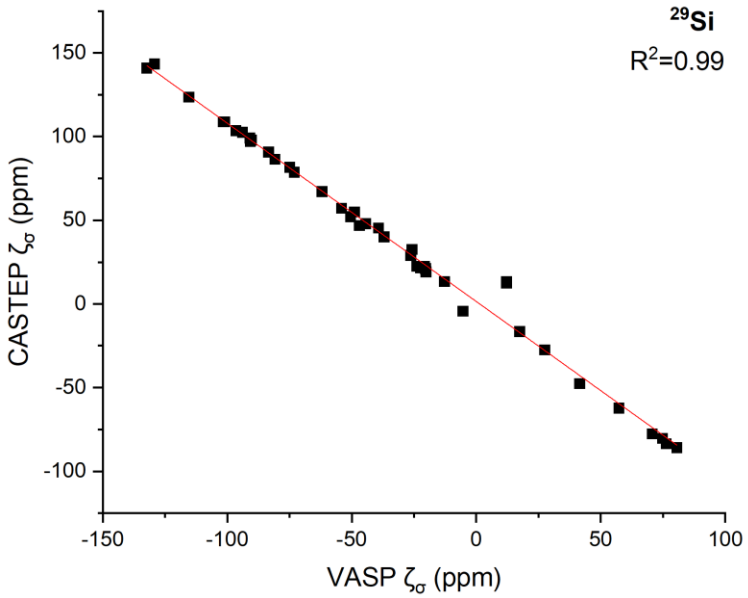


Figure 3.6 Comparison of CASTEP computed ^{29}Si reduced anisotropy versus VASP computed reduced anisotropy. The fit to a linear regression is shown in the figure (in red).

The excellent agreement in **Figure 3.6** between the computed values for reduced anisotropy for CASTEP versus VASP, gives us confidence that both programs are able to predict similar values of these tensor parameters for crystalline structures. In general, the outliers are points for which the assignment of experimentally obtained δ_{ZZ} (and hence, δ_{XX} as well) may be incorrect, as we illustrate.

The reduced anisotropy of the CSA (ζ_δ) in the Haeberlen system defines the lineshape in terms of one "extreme edge" of the static powder pattern ($\zeta_\delta = \delta_{ZZ} - \delta_{iso}$), explicitly yielding that one specific element of the tensor. This is the shoulder furthest from the isotropic chemical shift, which poses an observational challenge when examining some experimental spectra, as illustrated by **Figure 3.7**. For one manifestation of the lineshape, usually an η_{CSA} value less than about 0.7 (such as that shown in the inset image), δ_{ZZ} is unambiguous as marked.

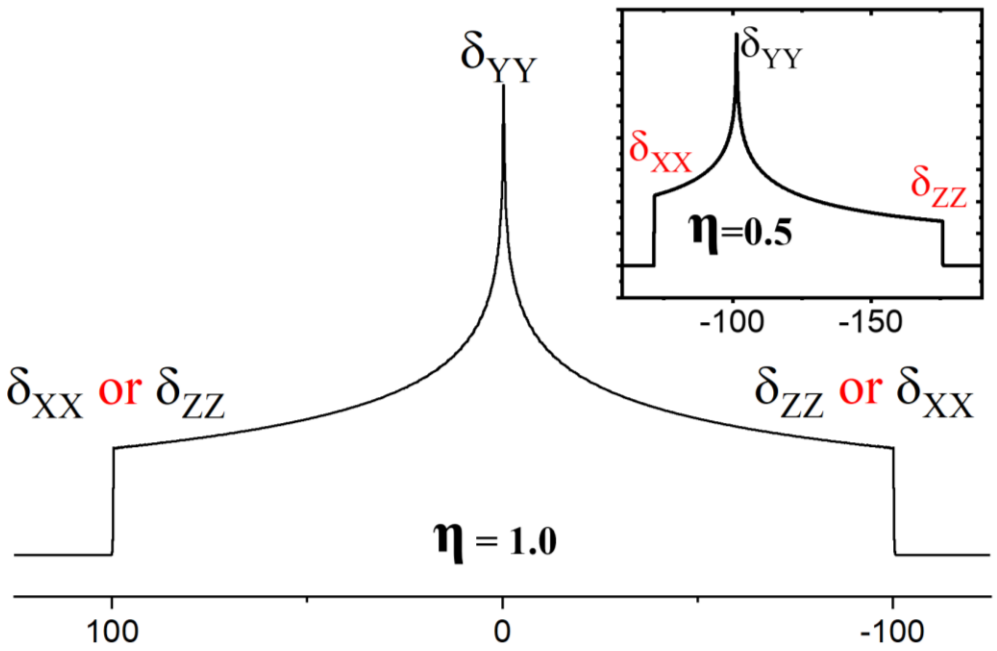


Figure 3.7 Scheme of static NMR powder pattern lineshapes for two different values of the asymmetry parameter, η_{CSA}. For values of η_{CSA} less than ~0.7, there is an unambiguous assignment of δ_{ZZ} and δ_{XX} tensor components. However, for η_{CSA} values of ~0.7 to 1.0, the determination between the tensors elements can be mis-assigned.

However, for lineshapes with large values of η_{CSA} (e.g., approaching 1.0) and for MAS NMR with few spinning side bands, the researcher must assign that shoulder to one side or another based on sparse data as illustrated schematically with the following case study.

We are showing a set of Dmfit[10] simulations of MAS NMR spectra of one of the "outliers" in the correlation plot of **Figure 3.5**, forsterite. These simulations are based on both experimentally reported and CASTEP-calculated CSA tensors, where the experimental MAS frequency was 2100 Hz. The CASTEP-computed tensor values lead to a spectrum that is nearly identical (in appearance only) to that generated from the experimentally reported spectrum. Notable from the simulations in **Figure 3.8**—the spinning sideband manifold does not contain enough information to assign the principal components, δ_{XX} and δ_{ZZ}. Hence, forsterite's reduced anisotropy ζ_δ values are similar in magnitude yet opposite in sign between the two. Given how similar the two patterns are, in the absence of a computational resource, one must make a best guess as to their values

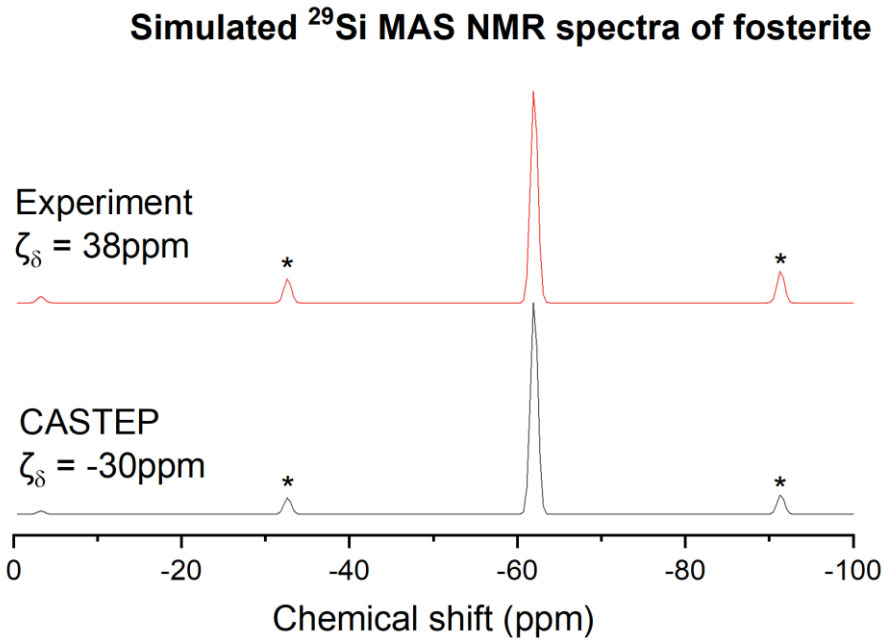


Figure 3.8 Simulation of ^{29}Si MAS-NMR spectra based on experimentally reported (rotational frequency of 2100 Hz, recorded at 8.4 Tesla) and CASTEP computed tensor values, simulated for these conditions.

The spectra are nearly identical with a different sign of chemical shift anisotropy ζ_δ., arising from the assignment/placement of the δ_{ZZ} value.

When there is sparse data, poor-signal-to-noise ratios in the experimental spectrum, or when there is a truncation of one shoulder due to radio-frequency pulse imperfections[141], the wrong value for δ_{ZZ} may be assigned—importantly, to the incorrect "side" of the lineshape. In addition to having so few spinning sidebands, this is a material where the asymmetry parameter η_{CSA} has a value approaching 1, similar to the situation illustrated schematically in **Figure 3.7**. This example further illustrates the inability to categorically assign δ_{ZZ} with certainty for situations where the asymmetry parameter values are so large.

3.3.3 Challenges for cataloguing the full shielding tensor, asymmetry parameter η_{CSA}

A consequence of such inaccurate assignments is to lead to incorrect expressions for both ζ_δ and η_{CSA}. We have also found that η_{CSA} tends to be poorly determined by observational analysis of lineshapes. This "asymmetry parameter, η_{CSA}" contains the reduced anisotropy, ζ_δ, in its denominator, as well as δ_{XX} and δ_{YY} in the numerator. Consequently, a mis-assignment of two of these tensor elements can cause this parameter to be unstable, exhibiting large fluctuations with small deviations in the direct tensor elements, resulting in a significant lack of correlation between computation and experiment as depicted in **Figure 3.9**. (Similar to what was seen for ζ_δ, there is a very good correlation between VASP and CASTEP computed values for the asymmetry parameter, shown in **Figure 3.10**.)

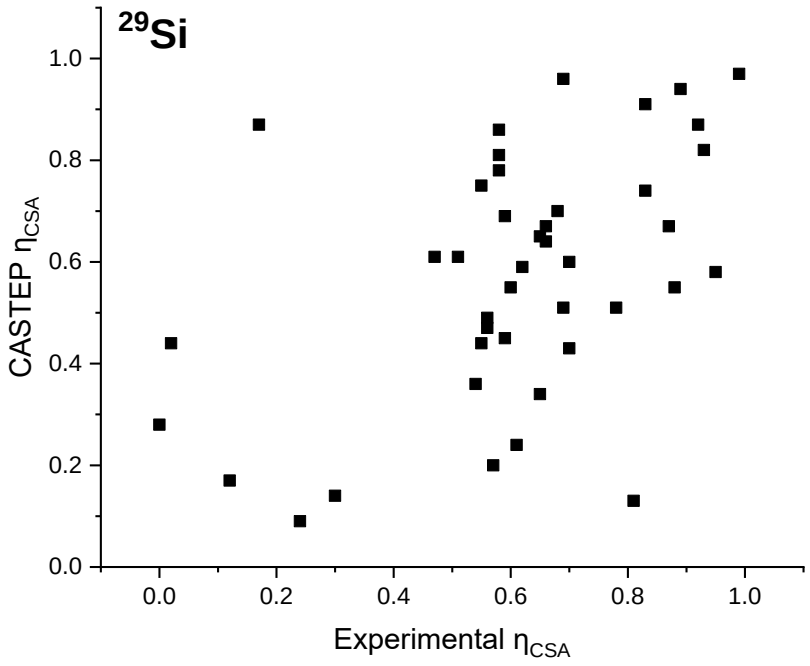


Figure 3.9 Comparison of CASTEP computed ^{29}Si η_{CSA} asymmetry parameters, versus experimentally reported values for 42 crystal structures.

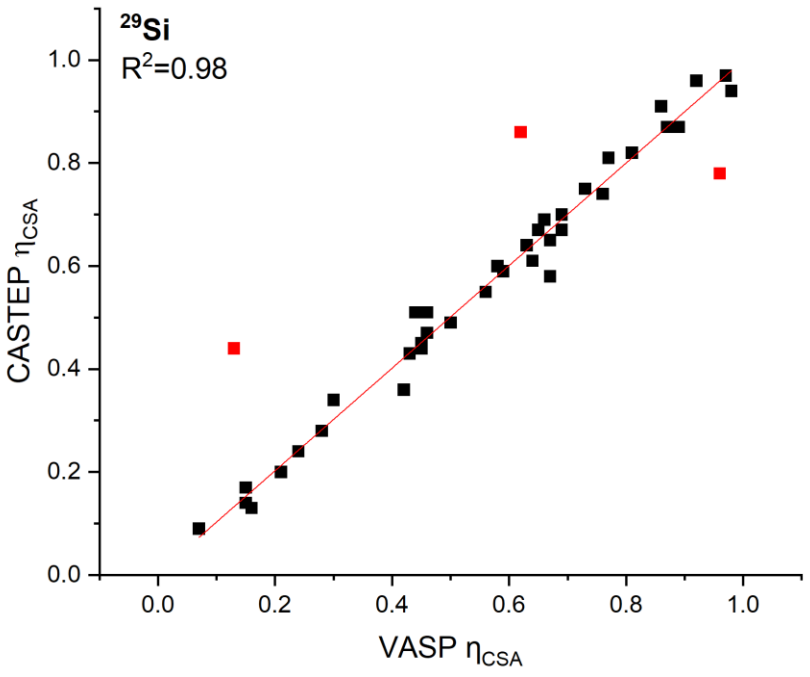


Figure 3.10 Comparison of CASTEP computed ^{29}Si asymmetry parameter η_{CSA} versus VASP computed asymmetry parameter. The fit to a linear regression is shown in the figure (in red).

3.3.4 Critical discrepancies between CASTEP and VASP

It is important to note that the shielding (σ) and the chemical shift (δ) should have a negative correlation with respect to one another. One finding in creation of this benchmarking set is the inverse correlation of tensor elements between CASTEP and VASP, which is critical to any understanding derived from comparison of experiment and computation. Both CASTEP and VASP compute chemical shielding, where the individual tensor elements are catalogued in the LSDI based on the Mehring convention, namely that the tensor elements are ordered numerically from largest to smallest. A case study is presented to illustrate this systematic difference in the shielding tensor elements, that is corrected when producing the individual chemical shift tensor elements. The LSDI catalogue will ultimately contain both computed chemical shielding and corrected chemical shift full tensors.

The inverse correlation between CASTEP and VASP data is evident from **Figure 3.11** below where the individual principal components generated by CASTEP and VASP (CASTEP 11, 22, 33 and VASP 11, 22, 33) wrongly correlated to each other (for example CASTEP 11 is correlated to VASP 33 instead of VASP 11). Such a sign difference implies a fundamental difference between the two implementations of DFT, requiring careful assignment of convention for both the tensor elements as well as the internal conventions for the theory. This is also seen in the sign difference for the local electrostatic potentials for CASTEP and VASP for instance, but can be properly accounted for when computing the work function. It's important to note that the linear regression correction *accounts for the CSA sign difference when generating the chemical shift tensor*

elements. We must address this inherent difference, when generating individual tensor elements (σ_{11}, σ_{22}, σ_{33}), that contribute to the expressions of σ_{iso} (and effectively δ_{11}, δ_{22}, δ_{33} for δ_{iso}).

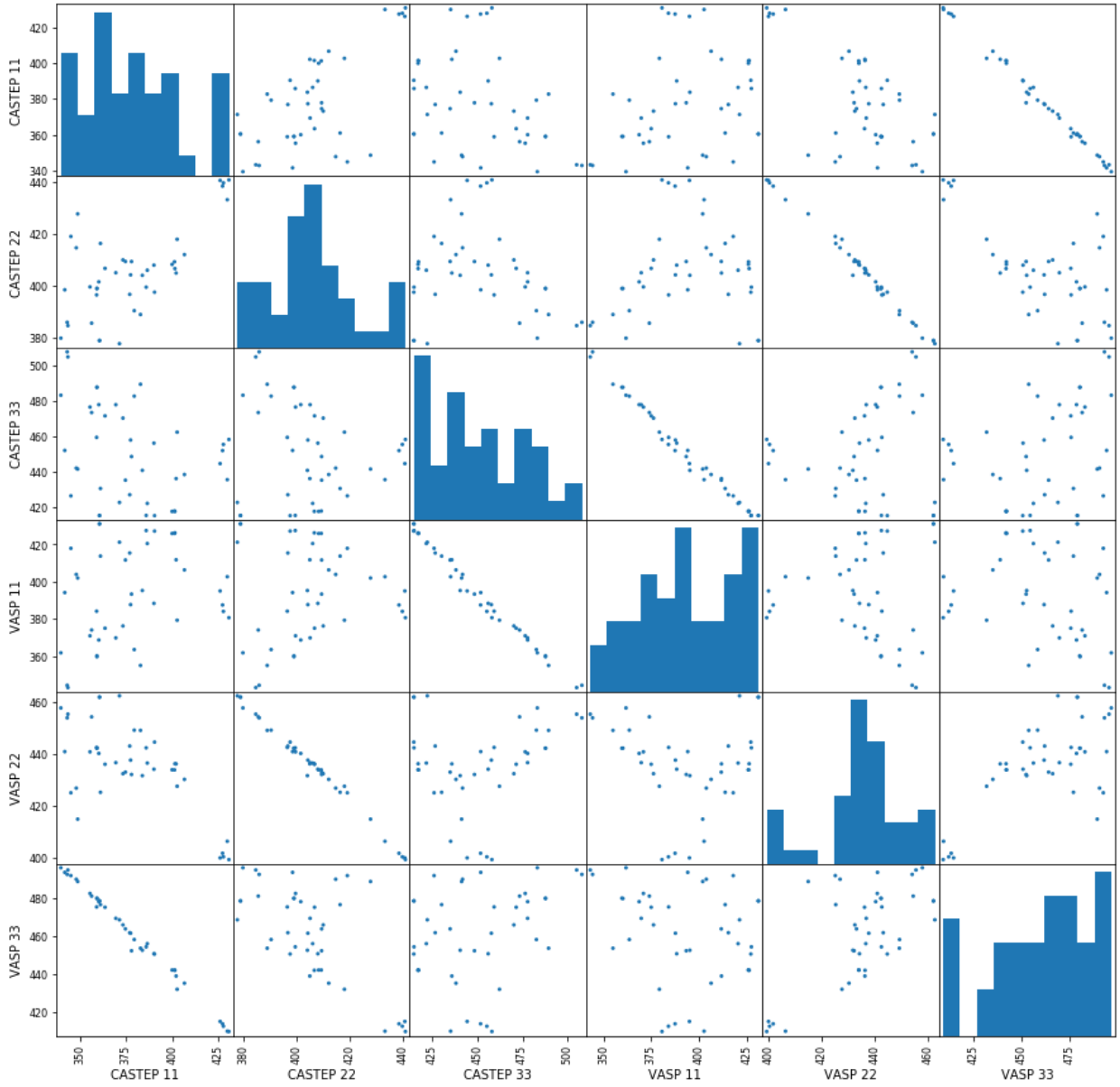


Figure 3.11 Correlation plots of CASTEP and VASP shielding tensor values σ_{11}, σ_{22}, σ_{33}, as indicated by the titles. This matrix of plots demonstrates the correlation between, for example, VASP σ_{11} elements and CASTEP σ_{33} elements. The σ_{22} tensor elements are negatively correlated. The implications of these correlations is discussed in the Case Study below.

For the purposes of clarity and understanding, we show the raw output of both VASP and CASTEP programs, of a representative compound, forsterite (Mg_2SiO_4). The change of sign and the corrections used from linear regression illustrate how the tensor elements (shielding and shift, both) are related. The matrix output of the two programs (after symmetrizing and diagonalizing) are as follows:

CASTEP	VASP
$\begin{pmatrix} 415.0589 & 0.0000 & 0.0000 \\ 0.0000 & 360.5291 & 0.0000 \\ 0.0000 & 0.0000 & 378.9098 \end{pmatrix}$	$\begin{pmatrix} 461.9618 & 0.0000 & 0.0000 \\ 0.0000 & 431.1768 & 0.0000 \\ 0.0000 & 0.0000 & 478.7577 \end{pmatrix}$

The principal components from these two diagonalized matrices are, by definition ($\sigma_{11} \leq \sigma_{22} \leq \sigma_{33}$):

CASTEP	VASP
$\sigma_{11} = 360.5291$ $\sigma_{22} = 378.9098$ $\sigma_{33} = 415.0589$	$\sigma_{11} = 431.1768$ $\sigma_{22} = 461.9618$ $\sigma_{33} = 478.7577$

However, when these are rendered, the reader will easily recognize the following (static NMR) lineshapes, based on those *shielding* values, that instead of output the same spectrum, CASTEP and VASP output spectra that mirror each other.

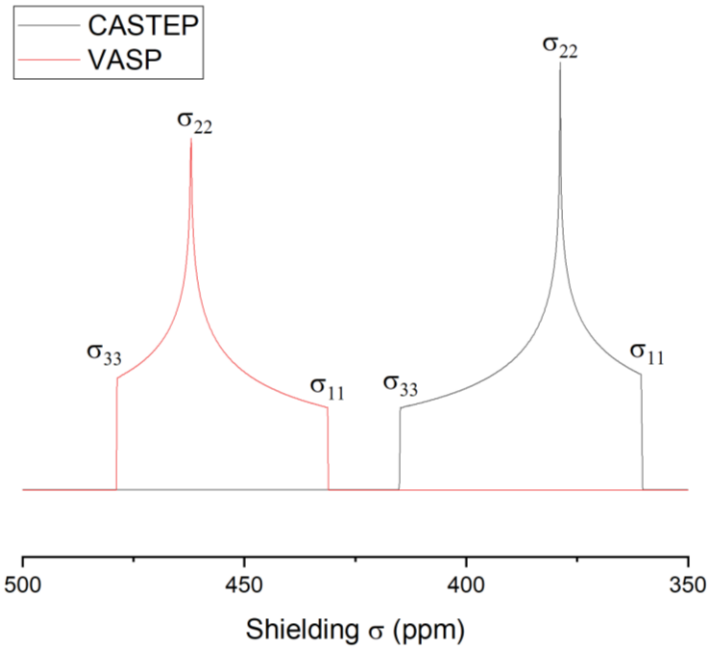


Figure 3.12 Simulation of ^{29}Si static lineshapes based on diagonalized *shielding* tensor values for CASTEP and VASP.

After applying the corrections provided by the linear regression for a reference, the diagonalized chemical shift values are obtained:

CASTEP	δ_{11} = -39.51 ppm δ_{22} = -56.00 ppm δ_{33} = -88.30 ppm	VASP	δ_{11} = -42.80 ppm δ_{22} = -57.28 ppm δ_{33} = -83.98 ppm

When these latter values of chemical shift are plotted to yield static lineshapes, the model spectra are obtained as shown below, alongside an example of the experimental ^{29}Si static NMR spectrum as shown in **Figure 3.13.**

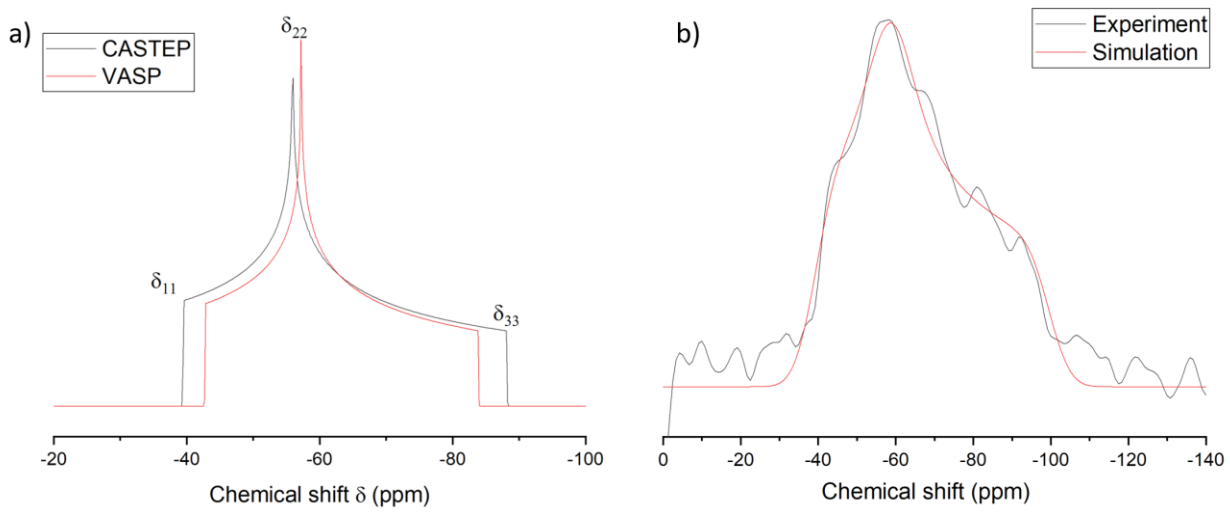


Figure 3.13 a) Simulation of ^{29}Si static lineshapes for forsterite based on diagonalized chemical shift tensor values for CASTEP and VASP. b) experimental spectrum of the solid-state ^{29}Si NMR (static) of forsterite for comparison. The fitting of the data "Simulation" give these parameters: gives δ_{iso} = -63.81 ppm, ζ_{δ} = -33.26 ppm, η_{CSA} = 0.66.

As a check, we have used the TensorView program[142] to render a graphical depiction of the shielding surface ovaloid superimposed onto a Q^3 silicon site in sodium disilicate. In **Figure 3.14**, the tensors' graphical depiction for VASP versus CASTEP is mathematically perpendicular to one another. The assigned σ_{33} CASTEP ovaloid is oriented as expected with the σ_{33} component along the single C_3 rotation axis of the Q^3 silicate site. The VASP schematic shows that σ_{33} (from VASP) is mis-identified to be at 90° from the bond along which it should lie.

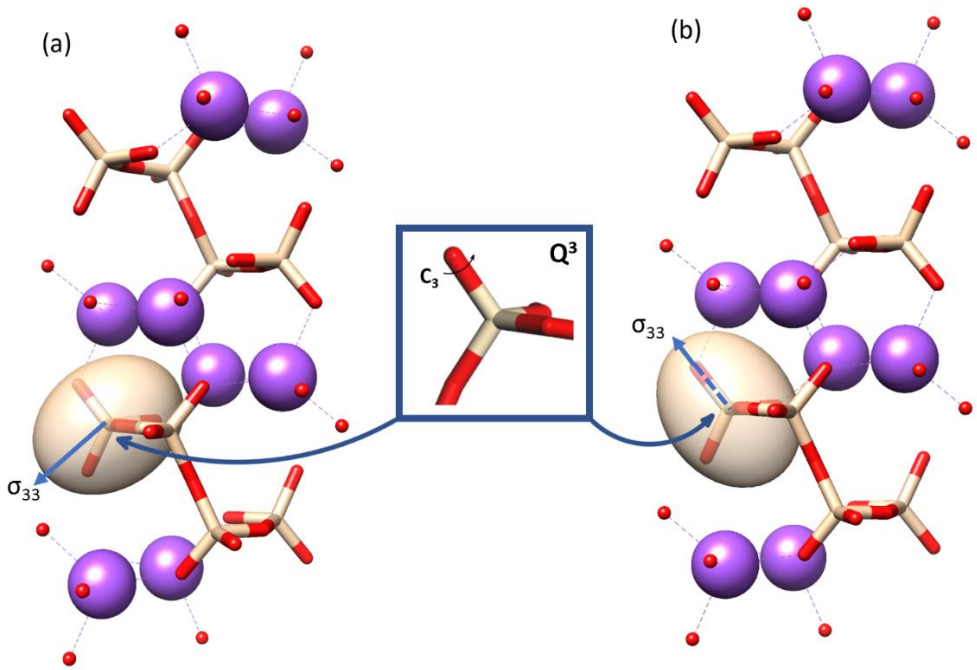


Figure 3.14 Graphical depiction of sodium disilicate with one silicon site highlighted with an ovaloid shape rendered to depict the ^{29}Si chemical shielding tensor for a Q^3 site with an asymmetry parameter of $\eta_{CSA} = 0.24$. On the left (a) is the VASP-computed tensor, and (b) the CASTEP-computed one. The long axis of the tensor is also the orientation of the σ_{33} component, which is oriented differently between the two programs. The CASTEP version is the conventional assignment for σ_{33}—along the unique Si-O bond. σ_{33} here is expressed in the "Mehring convention", $\sigma_{33} > \sigma_{22} > \sigma_{11}$.

3.3.5 The effect of convention on individual tensor elements

In light of the errors revealed in the expressions above, a strong argument can be made for reporting the individual tensor elements, and departing from the Haeberlen convention. One of the important opportunities afforded by the LSDI database is the ability to discover such systematic errors, by comparing a large number of datasets. Using the three equations from the Haeberlen convention and solving for the three unknowns (σ_{XX}, σ_{YY} and σ_{ZZ}), a correlation plot between computed and experimentally reported δ values is shown in **Figure 3.15 (a)**, with tensor elements clustered by

symbol (and color). The outliers are identified by name. Shifting to a different definition for chemical shift tensors, referred to as the "Mehring" convention[6], where σ_{11}, σ_{22} and σ_{33} are the three counterparts, organized in terms of high- to low-frequency for any lineshape, the algebraic solutions for the experimentally reported values become reconciled, as shown in **Figure 3.15 (b)**.

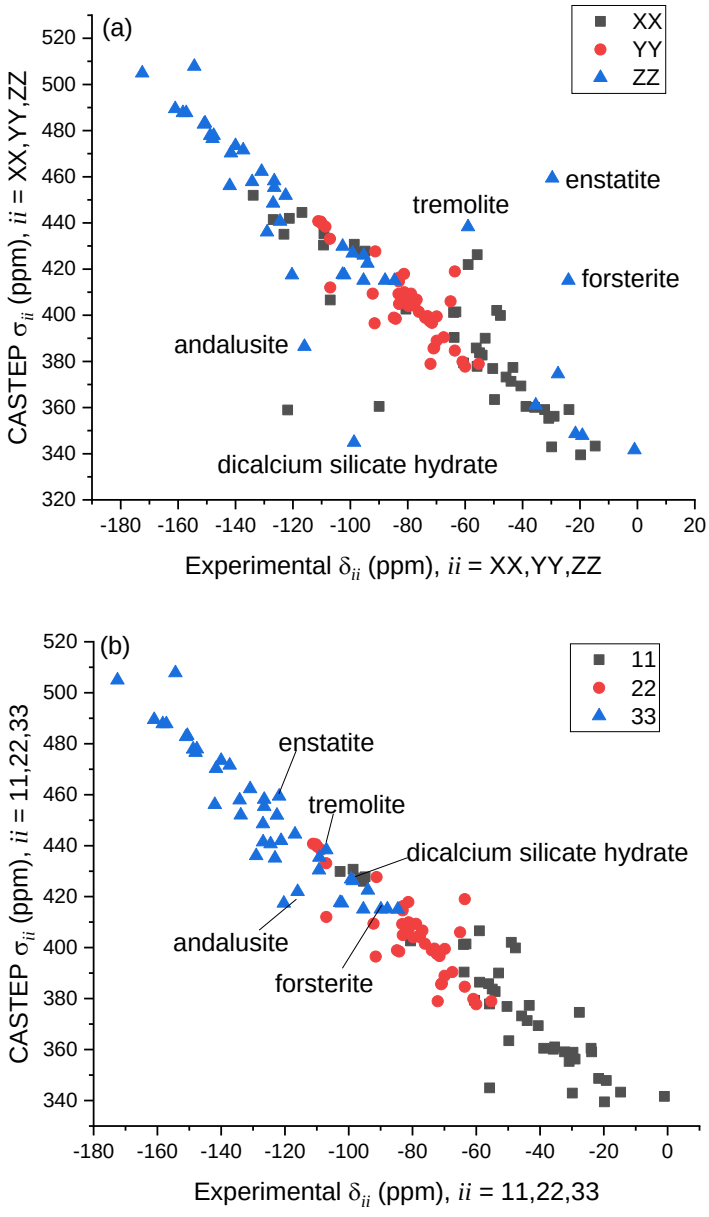


Figure 3.15 Correlation plots of CASTEP computed values (σ) versus experimentally reported values (δ). a) depicts σ_{XX}, σ_{YY} and σ_{ZZ}, and b) depicts computed values expressed instead as σ_{11}, σ_{22} and σ_{33}, versus the chemical shift (δ) tensor values from experimental reports, extracted algebraically as δ_{11}, δ_{22}, δ_{33}.

We can see that the individual tensor elements, defined in terms of their frequency using the Mehring convention, are better correlated between experiment and computation, and reporting these reduces the inaccuracies inherent in the algebraic expressions used to describe the lineshape.

3.3.6 Opportunities for applications of the "LSDI" Catalogue

CASTEP and VASP have particular strengths in the assignment of tensor elements, which will form the basis of the LSDI catalogue. The LSDI has already computed over 10,000 unique Si-sites for compounds in the Materials Project using VASP. This continually growing data set, easily accessible API (application programming interface) and collection of software tools is established as a community resource to enable easier in-silico experimentation with solid state NMR. Having such a catalogue of shift tensors allows prediction of both static and MAS lineshapes for solid-state NMR, which will aid in accurate simulation of the full lineshape and all 3 tensor elements. Furthermore, as we depict schematically in **Figure 3.16**, the ability to plan experiments (i.e., to select an ideal MAS spinning frequency, such as shown in **Fig 3.16 c** versus that in **3.16 b**) in order to accurately map out the tensor values, especially of δ_{22}, is a consequence of possessing such data.

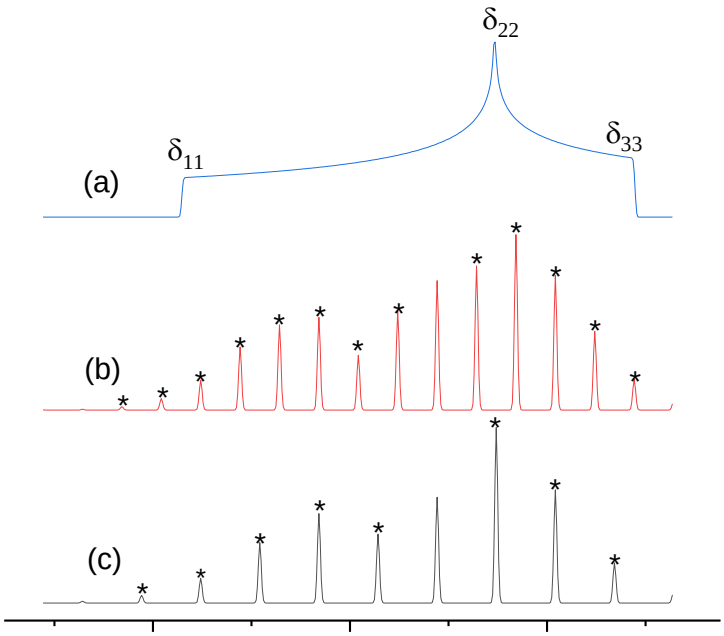


Figure 3.16 Comparison of the static CSA lineshape and corresponding NMR MAS spinning sideband manifolds at two different rotational frequencies. (a) simulated static lineshape with chemical shift principal components labeled. (b) simulated MAS lineshape with a spinning frequency of 800 Hz. (c) simulated MAS lineshape with a spinning frequency of 1200 Hz. * denote spinning side bands.

The utility of this catalogue can be demonstrated by considering the characterization of silicates by ^{29}Si solid-state NMR spectra, specifically assigning resonances to Q^n sites, a notation that reflects the local silicon environment and symmetry. Q^4 has 4 equivalent Si-O-X bonds, and X is an element that can include Si, often Si-O-Si in a network, or a species such as H (forming $Si(OH)_4$). Q^3 has 3 equivalent Si-O-X linkages and one unique Si-O-Y substituent (where in this case, Y could be a different substituent, or it could simply reflect a longer Si-O bond), and so on. Each of the Q^n sites is associated with a typical ^{29}Si chemical shift range. However, what if you have a sample with an atypical substituent? The LSDI catalogue permits a comparison of isotropic chemical shielding values for > 5000 silicate structures.

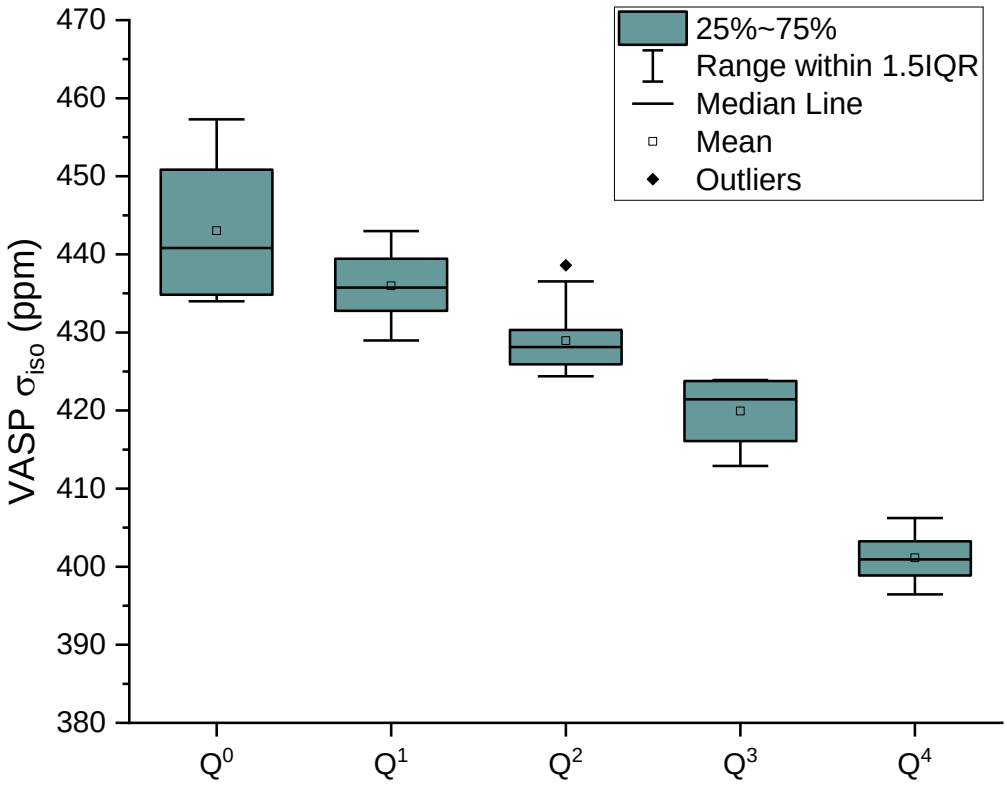


Figure 3.17 A statistical box plot illustrating the distribution of VASP calculated ^{29}Si isotropic shielding parameters (σ_{iso}) for different Q^n sites from the benchmarking set of 42 silicon sites. The symbols outside the range of 1.5 IQR are outliers (IQR = inter quartile range).

In **Figure 3.17**, a "box plot" of the VASP-computed σ_{iso} parameters from the benchmarking set shows the range of isotropic chemical shielding values predicted for different Q^n sites in silicates, with a variety of substituents. The trend as n increases is seen, as well as the range of computed values, spanning 40 – 45 ppm. A number of outliers are also found. It is possible for practitioners of ^{29}Si NMR to compare their spectra to these values in order to develop chemical insights into trends for particular bonding environments or changes of local site symmetry. What is especially helpful from such a plot is the ability to assign the chemical shifts of "less common" sites, not

based on the isotropic value alone (since these ranges overlap strongly), but through comparison to a range of compounds in the database with related chemical structures.

3.4 Conclusions

We have used 42 silicon sites as a benchmarking set to compare between CASTEP, VASP and experimentally reported expressions regarding the solid-state ^{29}Si NMR lineshapes. Through this examination, we have established a robust and systematic method for assigning the diagonalized chemical shift/shielding tensor values. Armed with confidence in this benchmarking set, over 10,000 ^{29}Si NMR shielding tensors will be publicly available via the Local Spectroscopy Data Infrastructure portion of The Materials Project. These tensors will be a guide to researchers when searching for ^{29}Si NMR assignments, as well as a platform that can assist with experimental conditions, since the appearance of spectra can be anticipated prior to measurement.

Benchmarking also revealed an unexpected systematic difference between VASP and CASTEP, where σ_{11} and σ_{33} shielding elements were interchanged, owing to a sign difference between computed tensors. This sign error is corrected when using linear regression methods (to obtain chemical shift tensor values, δ), and the final chemical shift anisotropy lineshapes that are generated are consistent with experimental measurements—from both programs. Consequently, our data tables reflect these revised values. Thus, systematic comparison of NMR properties across various methodologies, including differing computational methods or codes, should be conducted in a chemical shift basis to eliminate representation deviations that could lead to systematic error. Understandable "assignment errors" of δ_{XX} and δ_{ZZ} tensor elements have been found in the literature, owing to difficulties with the Haeberlen notation and uncertainties as the lineshapes approach large asymmetry values (η_{CSA}) closer to 1. The benchmarking set permitted discovery

of such errors, and the values are corrected in the LSDI database (and in the tables shown in the Appendix A). Consequently, the database will report all notation in the IUPAC recommended fashion using the Mehring convention of δ_{11}, δ_{22}, and δ_{33}.

The possession of such a large dataset permits comparisons of the computed parameters across a large number of structures. When NMR practitioners use the LSDI dataset, they will be permitted to compare their experimental measurements to a variety of related structures, which will ultimately facilitate assignments of those spectra. This type of dataset can open the next era in solid-state NMR spectroscopy encompassing an informatics approach to experimental design.

Chapter 4

Benchmarking ^{27}Al NMR quadrupolar and chemical shielding tensors with DFT: machine learning prediction of quadrupolar coupling constants (C_Q)

4.1 Introduction

In this chapter we will expand beyond the spin ½ ^{29}Si benchmarking set presented in chapter 3 and focus on the validation and development of computational methods of quadrupolar (nuclear spin > ½) tensor parameters of ^{27}Al. In particular, we will evaluate DFT's (CASTEP/VASP) performance for calculating the quadrupolar coupling constant (C_Q) and develop a new machine learning model for C_Q prediction based on the DFT results.

Solid-state nuclear magnetic resonance (SSNMR) parameters such as isotropic chemical shift (δ_{iso}), chemical shift anisotropy (ζ_δ) and quadrupolar coupling constants (C_Q) can accurately reflect the interactions between the nuclear spin and its surrounding charge density even when long-range order is lacking. This enables SSNMR to be widely adopted in the determination of the local structure for a variety of solid materials such as lithium battery anodes, biological solids, and zeolites.[143–148]

NMR crystallography is an emerging multi-modal characterization methodology that combines the long-range information provided by x-ray diffraction (XRD), with the local structural information of SSNMR to provide highly accurate structural determinations.[115,149,150] This technique relies on state-of-the-art first-principles calculations such as density functional theory

(DFT) to make predictions of NMR parameters and compare with experimental counterparts so that the local structure could be further refined. Despite this advantage, the computational cost of DFT is often too large for broad adoption. More importantly, the reliability of these calculations in predicting experimental parameters has to be assessed one nuclear isotope at a time, with the literature focusing on ^{1}H, ^{13}C, ^{29}Si, ^{31}P and ^{17}O in various systems.[100,151–153]

Literature benchmarks have provided large datasets of computed data available via community databases, such as in the Materials Project (materialsproject.org) and the Collaborative Computational Project for NMR (CCP-NC)[91,154] that can be utilized for advanced machine learning (ML) studies to reduce the computational cost and democratize the adoption of NMR crystallography. The cubic scaling[51] of DFT limits these datasets to focusing on small unit cells of perfect crystalline materials modeled at a temperature of 0 K. Still, appropriately trained ML algorithms have demonstrated the ability to capture local geometry to predict δ_{iso} with accuracy close to DFT while requiring only a fraction of computing time.[51,155–158] While most of the machine learning efforts were focused on the prediction of δ_{iso}, the experimentally measured isotropic chemical shift (or σ_{iso}, the DFT computed chemical shielding), there have been relatively few studies that can demonstrate the ability of machine learning algorithms in predicting expressions of the electric field gradient (EFG) tensor parameters such as C_Q for quadrupolar nuclei. These EFG parameters provide a complementary measurement of small perturbations to local environments, especially when it is hard to distinguish different sites based on isotropic chemical shift.[159,160] Thus, the development of a machine learning method for C_Q prediction can be highly informative for NMR crystallography studies.

Herein, we present a solid-state NMR ^{27}Al benchmarking set with both DFT calculated NMR parameters and their experimentally measured counterparts, reported in the literature. As noted in

chapters one and three, the most common way for these NMR tensor elements to be reported is using expressions that employ the "Haeberlen" convention. We follow that convention here, for ready comparison between computed and experimentally measured quantities, we report both computed values and their experimental complements for isotropic chemical shielding (σ_{iso}), quadrupolar coupling constant (C_Q), and quadrupolar asymmetry parameter (η_Q) in the Appendix B. The DFT calculations were performed by two popular DFT packages: Vienna Ab initio Simulation Package (VASP) and Cambridge Serial Total Energy Package (CASTEP).[88,161] The reliability of DFT predictions of σ_{iso} and C_Q for ^{27}Al materials was confirmed based on the comparison between DFT and experimental NMR parameters over the benchmarking set. We further trained a "random-forest" machine learning model to predict the quadrupolar coupling constant C_Q for compounds containing 4-, 5- and 6-coordinate ^{27}Al sites based on a larger DFT calculated dataset with 1681 aluminum-containing crystalline solid materials. To train the model, we constructed two sets of features, structural features and elemental features (sometimes termed "alchemical features" in machine learning literature), based on the crystal structure to represent the ^{27}Al local environment. We have found the ^{27}Al C_Q value is closely correlated with the geometric properties of the next-neighbor bonding environment (surprisingly, regardless of the chemical identity of the bonded species). This geometry is typically depicted as a space-filling polyhedron. Distortions to the polyhedra given by variance of bond lengths and bond angles, in combination with other features denoting elemental variance, produce a simple but effective model with a predictive accuracy of $R^2 = 0.98$ and RMSE = 0.61 MHz expressed with a correlation between VASP computed C_Q and the model predicted C_Q.

4.2 Methods

4.2.1 Data set

We have performed a literature search and collected experimental ^{27}Al NMR parameters from 56 different crystalline materials, accounting for 105 unique sites, including a few repeated structures with independent measurements. All the parameters were collected with either MAS or MQMAS experiments. All of the structures were calculated with both VASP and CASTEP.

For machine learning model training, a larger dataset of DFT computed ^{27}Al NMR parameters was constructed by VASP calculation. The dataset is composed of 1681 aluminum-containing structures which correspond to 8081 ^{27}Al sites. The coordinating environment of the ^{27}Al sites was confined to 4-coordinate (4696 sites), 5-coordinate (202 sites), or 6-coordinate aluminum (3183 sites). All the crystal structures were obtained from the Materials Project and were geometry optimized before NMR calculations.

4.2.2 DFT methods

DFT calculations with CASTEP were performed within the Perdew-Burke-Enzerhof (PBE) Generalized Gradient Approximation (GGA) formulation of the exchange-correlation functional. These were performed in two steps: an initial geometry optimization where the lattice was allowed to adjust followed by an NMR calculation on the relaxed structure. On-the-fly ultra-soft pseudopotentials were used as an approximation of nuclear and core electron interactions. Convergence tests were performed on gamma-$LiAlO_2$ to find optimal energy cutoffs and k-points. As shown in **Figure 4.1**, it was determined that 750 eV as an energy cutoff with Monkhorst-Pack of $5 \times 4 \times 4$ was enough to converge the NMR calculations to a single value.

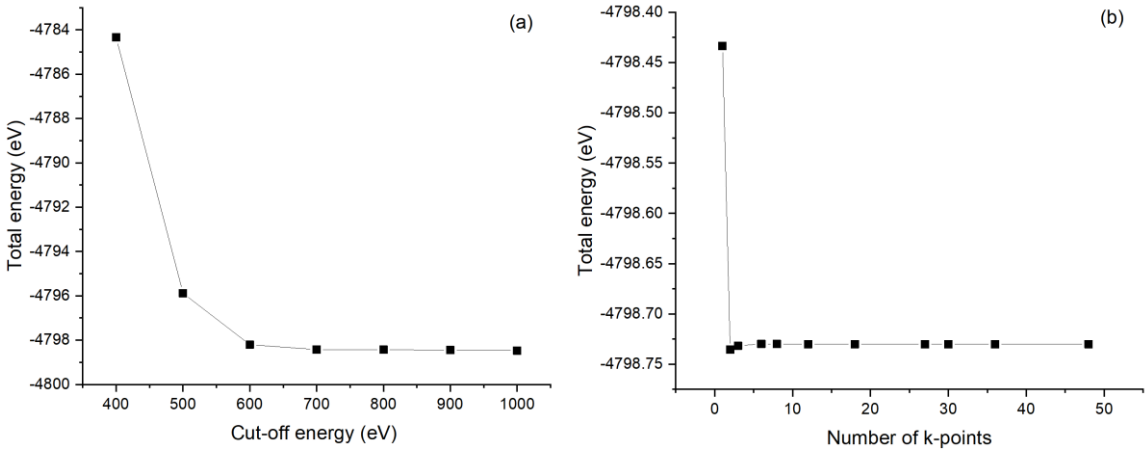


Figure 4.1 γ-$LiAlO_2$ convergence tests. (a) convergence test of total energy (eV) with increasing cut-off energy. Cutoff energy is specified from 400 eV to 1000 eV in increments of 100 eV. The system converged after 750 eV. (b) convergence test of total energy (eV) with an increasing number of irreducible k-points. The number of k-points from 1 (Monkhorst-pack 1×1×1) to 48 (8×7×6). The system converged after 12 k-points (5×4×4).

DFT calculations were also performed using the projector augmented wave (PAW) method[133,134] as implemented in the Vienna *Ab Initio* Simulation Package (VASP)[135–137] within the Perdew-Burke-Enzerhof (PBE) Generalized Gradient Approximation (GGA) formulation of the exchange-correlation functional[138]. A cut-off for the plane waves of 520 eV was used and a uniform k-point density of approximately 1,000/atom was employed. We note that the computational and convergence parameters were chosen in compliance with the settings used in the Materials Project (MP)[91] to enable direct comparisons with the large set of available MP data.

4.3 ^{27}Al DFT benchmarking

We begin by benchmarking the ability of DFT to predict chemical shielding tensors (σ_{CSA}) against experimentally compiled chemical shift tensors (δ_{CSA}). Unfortunately, the dominating influence of

quadrupolar broadening on NMR measurements makes accurately and independently specifying the chemical shift tensor difficult. In particular, many references choose to simply not report the anisotropy of the chemical shift (ζ_δ) and the asymmetry parameter (η_{CSA}) from the shielding tensor, leaving only the isotropic chemical shift (δ_{iso}) to compare with our computational dataset.

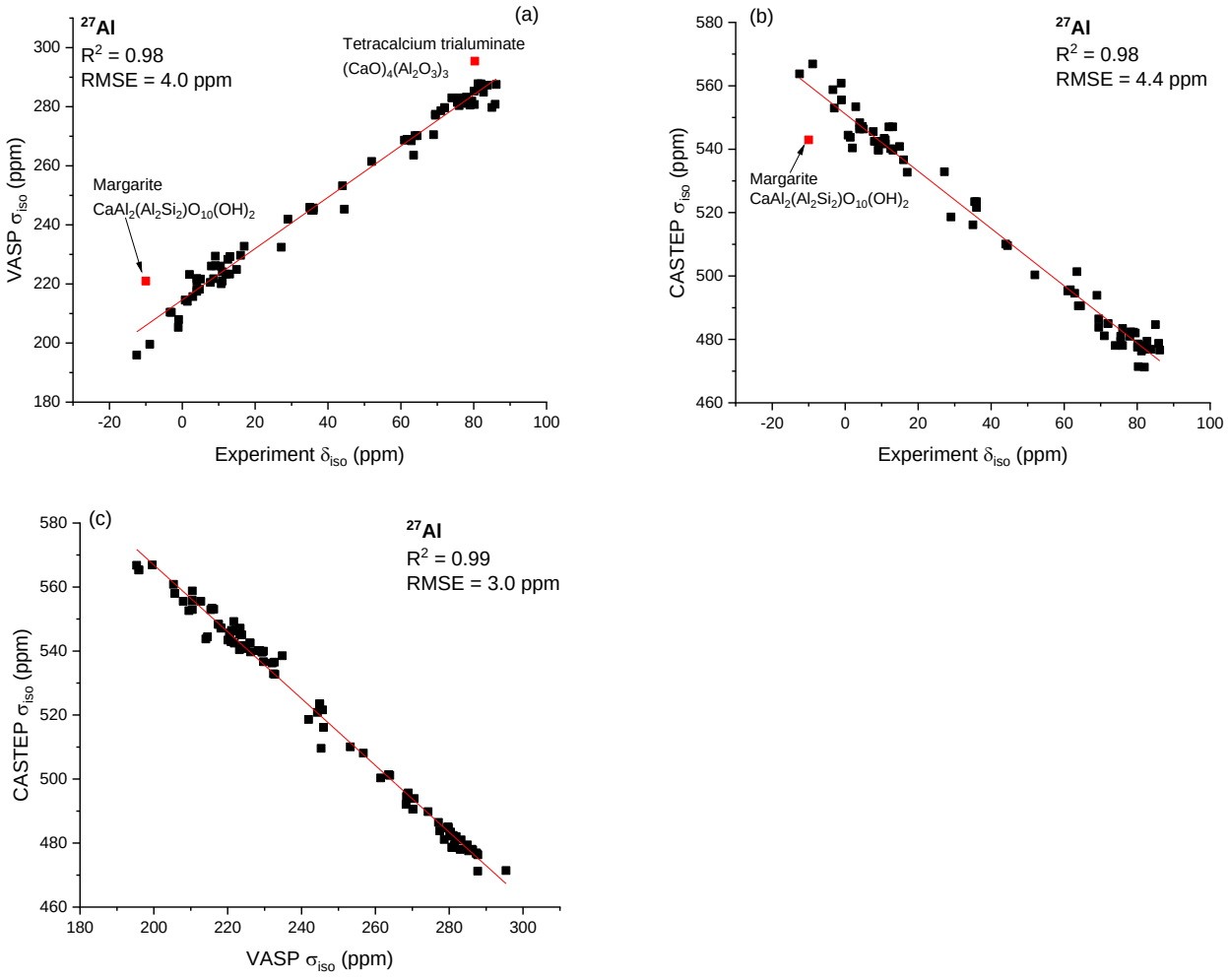
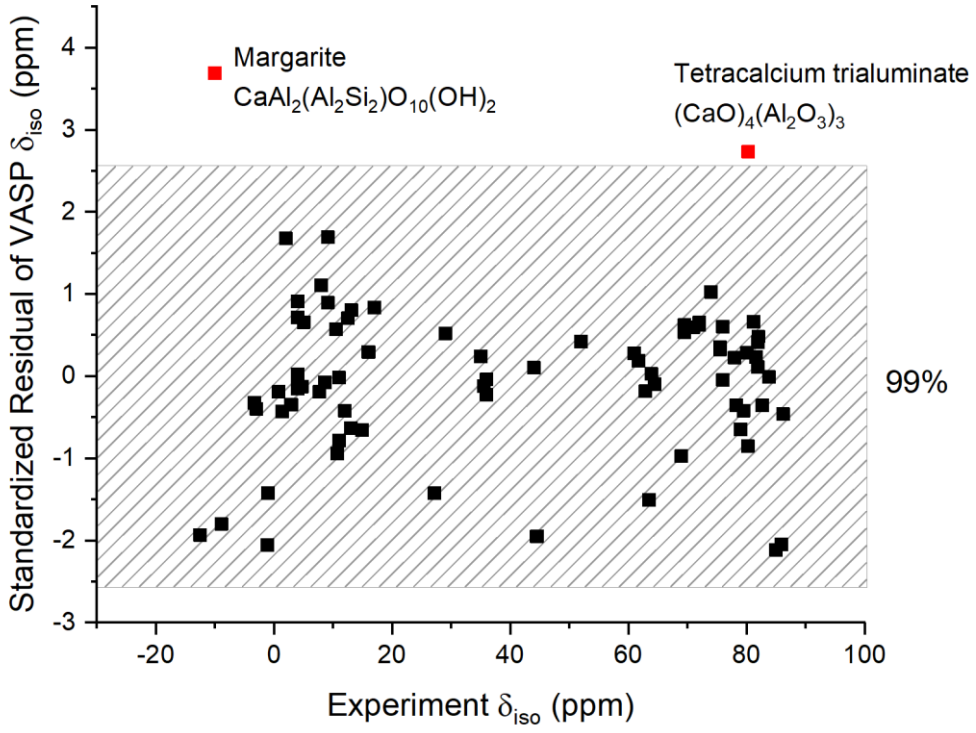


Figure 4.2 Correlation plots of ^{27}Al isotropic chemical shift/shielding between experiment and DFT calculations. a) Experimental δ_{iso} (ppm) versus VASP calculated σ_{iso} (ppm). b) Experimental δ_{iso} (ppm) versus CASTEP calculated σ_{iso} (ppm). c) VASP calculated σ_{iso} (ppm) versus CASTEP calculated σ_{iso} (ppm). In plot a) and b), the outlier species are highlighted in red. Note: the different slopes and the computed shielding values, not adjusted for reference values, reflect different embedded assumptions between VASP and CASTEP codes[100].

Figure 4.2 shows the correlation plot between the experimental isotropic chemical shift (δ_{iso}) and DFT calculated isotropic shielding (σ_{iso}) with two different packages (VASP and CASTEP). Both DFT packages demonstrate the ability to accurately predict ^{27}Al isotropic chemical shifts with R^2 = 0.98 and RMSE = 4.0 ppm and 4.4 ppm respectively. The positive correlation between VASP and experiment is due to VASP's special definition of its computed shielding tensor.[100] Further, **Figure 4.2 (c)** demonstrates a strong correlation between the two packages with R^2 = 0.99 and RMSE = 3.0 ppm, suggesting that future calculations with either of these codes should yield comparable results. Outliers can be identified by plotting the standardized residual values against each independent variable and identifying those that fall outside of a given confidence interval, which for this study was set at 99% (**Figure 4.3**).



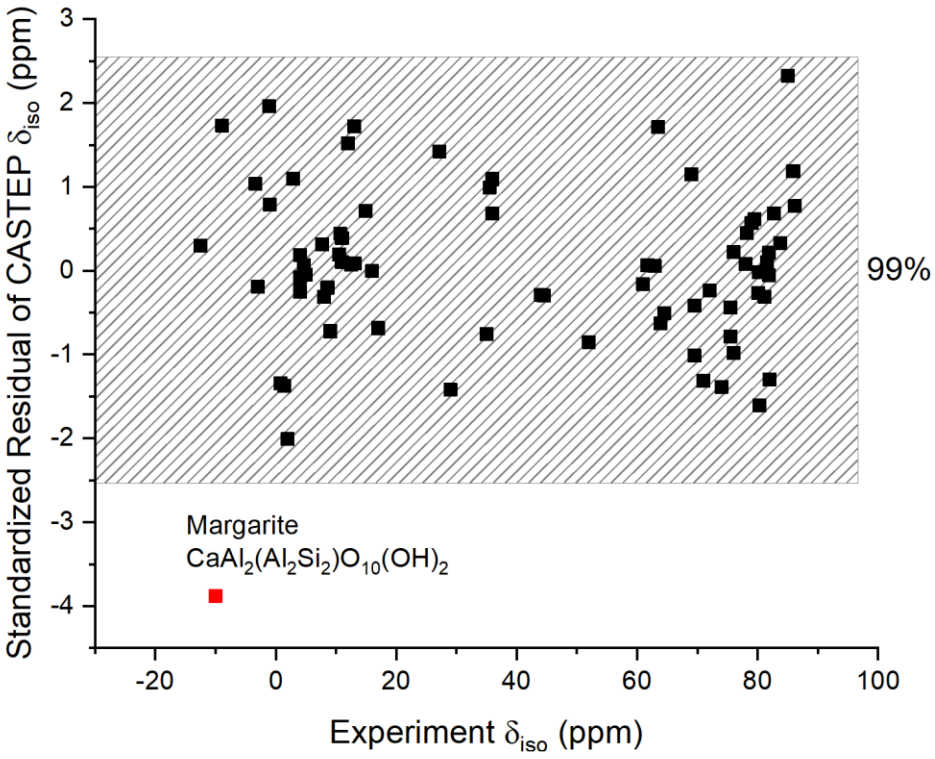


Figure 4.3 Standardized residual plot against independent variables for a) VASP versus experimental δ_{iso}; b) CASTEP versus experimental δ_{iso}. The shaded area represents the range within the 99% confidence zone for the linear regression. The red points that are outside of the 99% zone are considered outliers.

The outliers can provide useful case studies of some of the limitations of benchmarking studies, comparing calculated spectra with those measured by experiment, in some cases many decades ago. The lack of consensus on the crystalline structure of certain materials is an important source of discrepancy. One outlier corresponds to one of the two octahedral sites in the mineral, margarite ($CaAl_2(Al_2Si_2)O_{10}(OH)_2$) at -10 ppm, reported in an early solid-state NMR study from 1985.[162] Other literatures[163,164] suggest that there is only a single octahedral site, and provide a δ_{iso} value of 11 ppm for this octahedral aluminum site. Our DFT calculation identifies two octahedral aluminum sites in margarite with isotropic shielding σ_{iso} of 540.4 ppm and 542.9 ppm, which corresponds to

a chemical shift of 12.1 ppm and 9.3 ppm after linear regression referencing, respectively. These values are comparable to the literature values at 11 ppm.[164] **Figure 4.2 (a)** also has a second outlier $(CaO)_4(Al_2O_3)_3$, which is a member of a series of polycrystalline aluminates within $CaO\text{-}Al_2O_3$ system. Unfortunately, the difficulty of synthesis, the presence of defects and potentially reduced crystallinity level for the sample that was measured[165,166] can have the effect of changing the local structural motif, with consequences for the NMR measurements. As NMR is so highly sensitive to the local structure, any compositional uncertainty can result in a weak structural comparison between our DFT model and the actual structural parameters that should be used to describe the $(CaO)_4(Al_2O_3)_3$ sample that was measured.

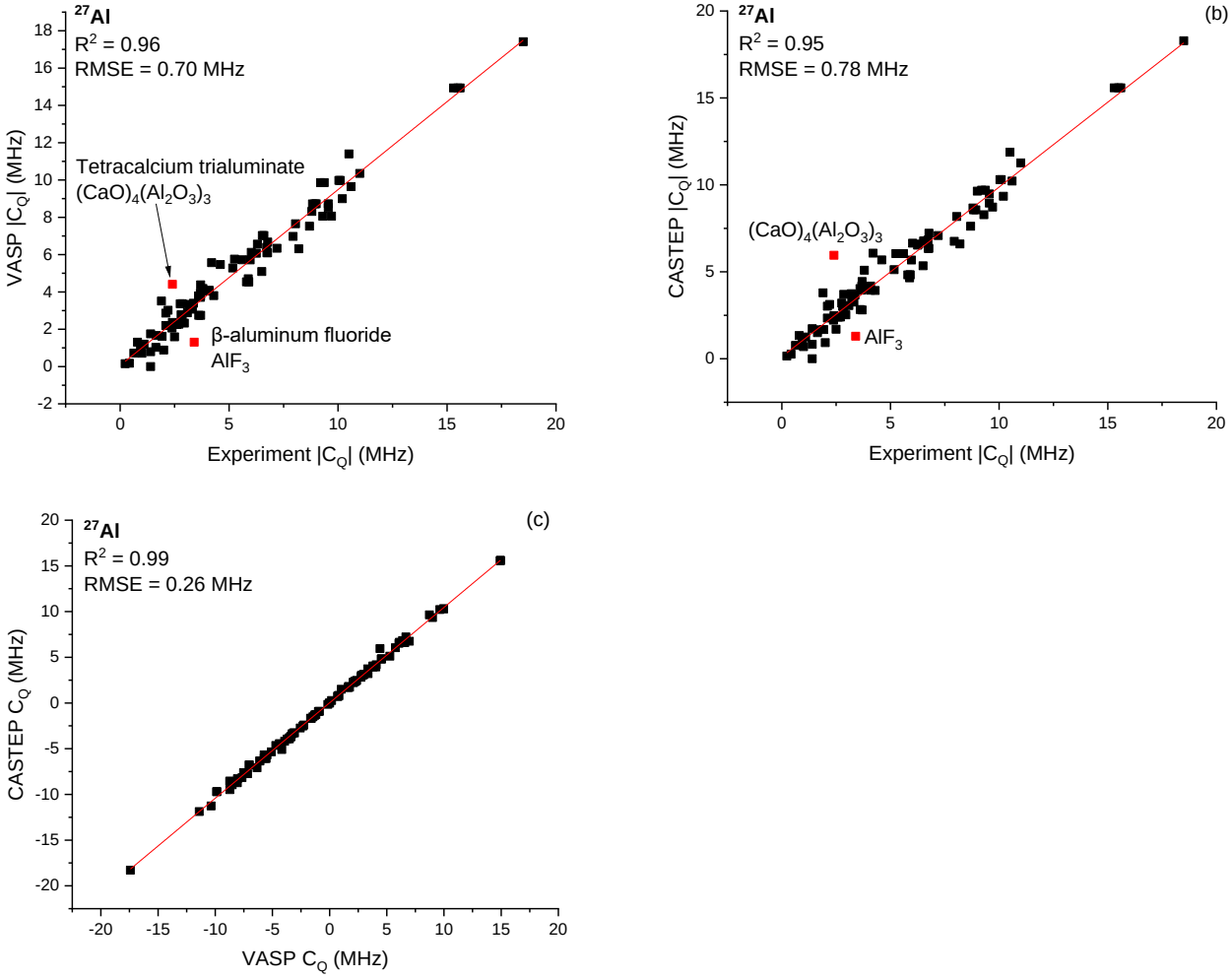


Figure 4.4 Correlation plots of the absolute value of ^{27}Al quadrupolar coupling constant $|C_Q|$ between experiment and DFT calculations. a) Experimental $|C_Q|$ (MHz) versus VASP calculated $|C_Q|$ (MHz). b) Experimental $|C_Q|$ (MHz) versus CASTEP calculated $|C_Q|$ (MHz). c) VASP calculated C_Q (MHz) versus CASTEP calculated C_Q (MHz). In plot a) and b), the outlier species are highlighted in red.

We compared the computed electric field gradient (EFG) tensor against experimentally reported values for C_Q and η_Q. (the algebraic definitions can be found in chapter one). The diagonalized EFG tensor for quadrupolar nuclei also can be translated into these convenient algebraic expressions, C_Q and η_Q, that reflect the appearance of the experimentally measured spectra. For quadrupolar species such as ^{27}Al, both C_Q and η_Q values from the EFG tensor are often compiled in the literature. **Figure 4.4** shows the correlation between experimentally measured C_Q and the corresponding values calculated by DFT packages (VASP and CASTEP). Since it is hard to measure the sign of C_Q experimentally with typical solid-state NMR, nearly all experimental papers choose to report the magnitude of C_Q. Thus, in **Figure 4.4 (a)**, **(b)**, we report the correlation between the computed C_Q from DFT with experimentally reported magnitudes: $|C_Q|$. The strong correlation between DFT and experiment for both VASP and CASTEP, with $R^2 = 0.96$ for VASP and $R^2 = 0.95$ for CASTEP, demonstrates that DFT has the ability to accurately predict the EFG tensor. We note two outliers using the same confidence interval sampling method used previously for the isotropic chemical shift (**Figure 4.5**).

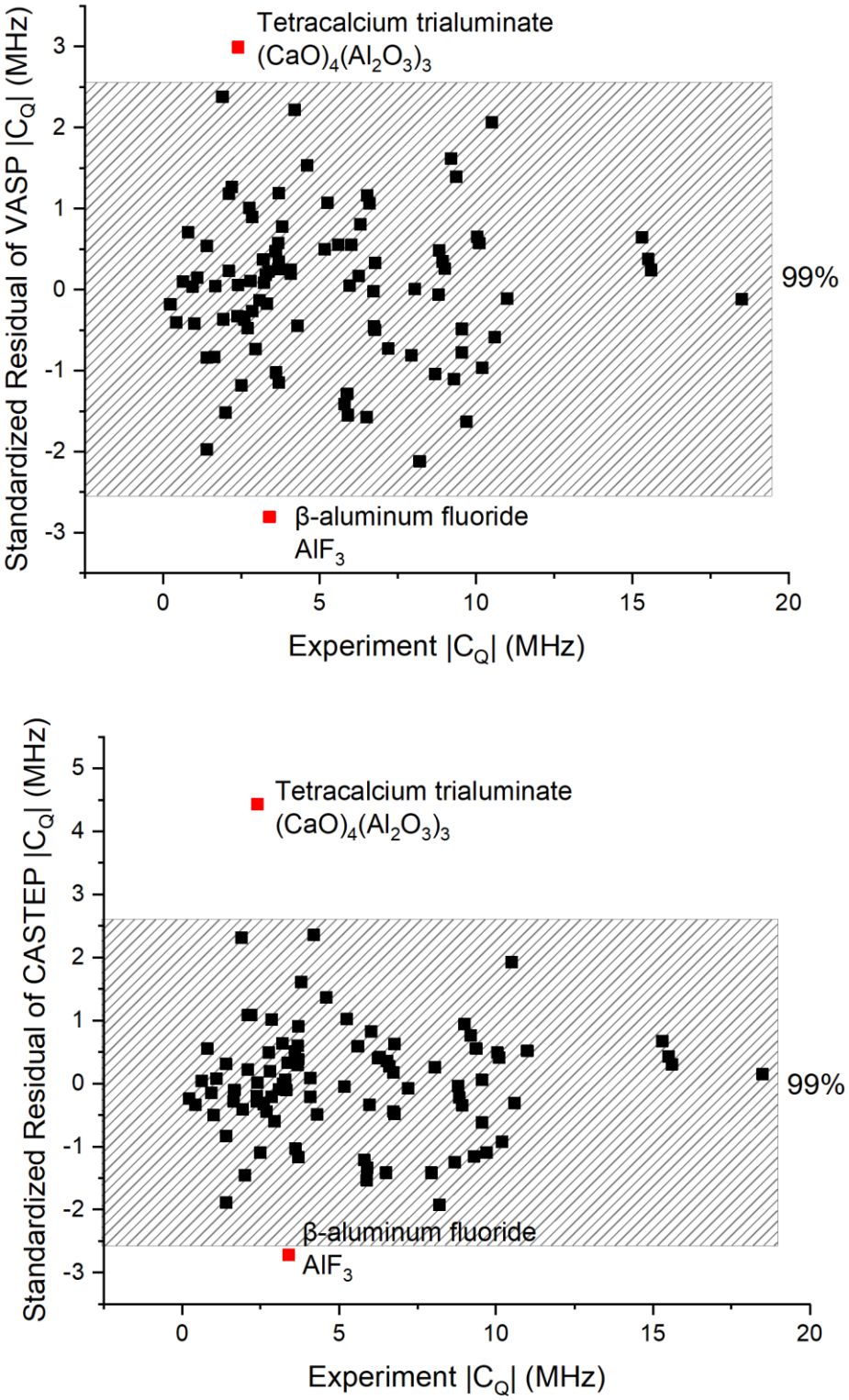


Figure 4.5 Standardized residual plot against independent variables for a) VASP versus experimental $|C_Q|$; b) CASTEP versus experimental $|C_Q|$. The shaded area represents the range within the 99%

confidence zone for the linear regression. The red points that are outside of the 99% zone are considered outliers.

Two significantly different C_Q values of β-AlF_3 were reported by previous publications with one stating the C_Q of the single ^{27}Al site in β-AlF_3 is 3.4 MHz while the other one stating a C_Q of 0.8 MHz.[167,168] Our calculation result (-1.31 MHz) suggests that 0.8 MHz appears closer to the computed value, and this result is supported by a more recent publication in 2014.[169] The second outlier is the previously noted $(CaO)_4(Al_2O_3)_3$ with experimentally reported C_Q = 2.4 MHz and VASP calculated C_Q = 4.41 MHz. As noted earlier, it is still unclear if our idealized structural model is an accurate representation of the local structural motifs in the measured sample of $(CaO)_4(Al_2O_3)_3$ resulting in an inappropriate comparison of NMR parameters. **Figure 4.4 (c)** shows a strong correlation between the two DFT packages, with $R^2 = 0.99$, for C_Q, suggesting that future calculations with either of these codes should yield comparable results.

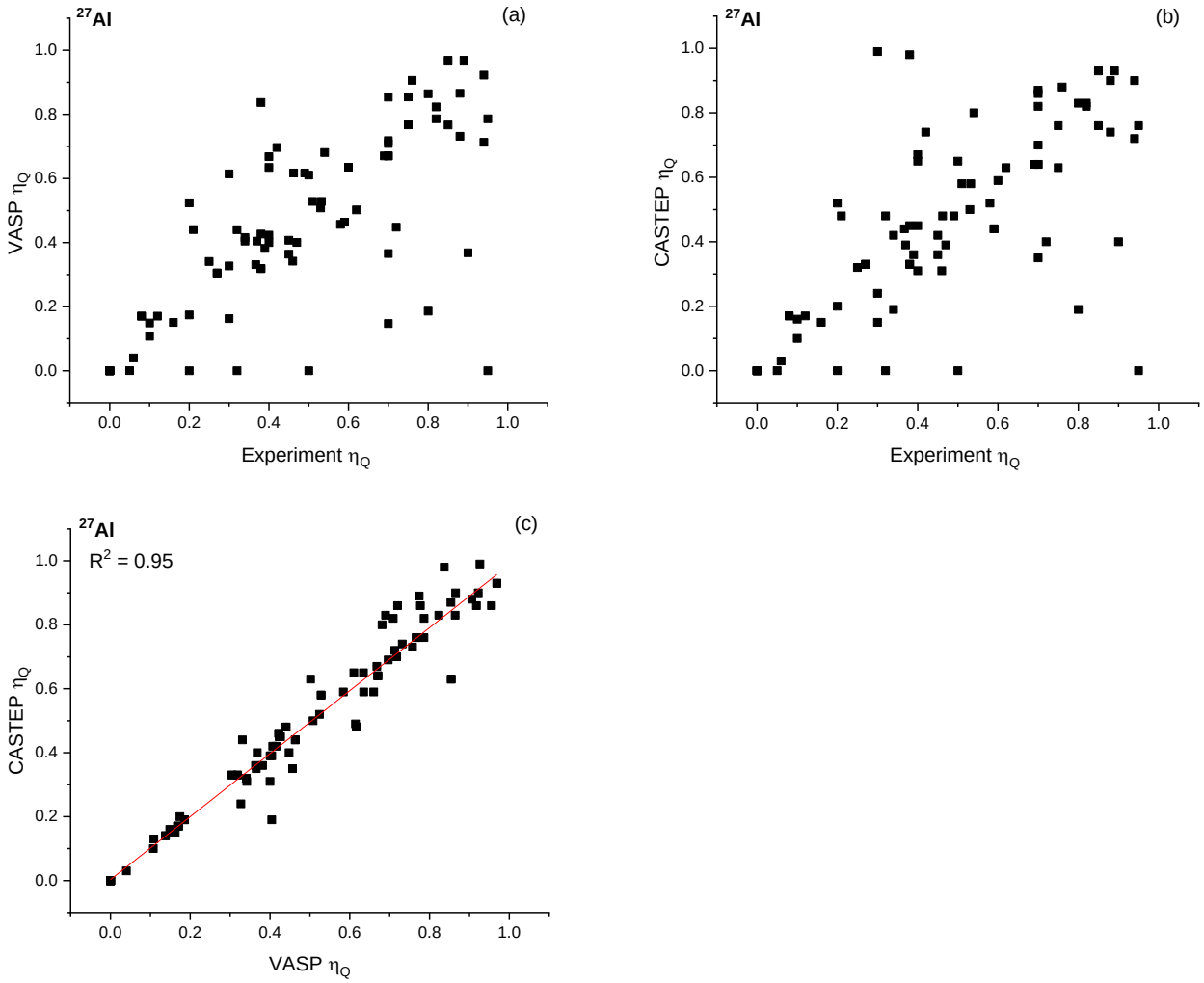


Figure 4.6 Correlation of experimentally reported η_Q with DFT calculated η_Q. a) VASP vs experiment; b) CASTEP vs experiment; c) CASTEP vs VASP.

The correlation between experimentally reported η_Q and DFT-computed values for η_Q is shown in **Figure 4.6**. Any correlation between computed and experimental values is tenuous, at best, with many outliers. It is unsurprising because asymmetry parameters are often difficult to accurately extract from an experimental spectrum for shielding and EFG tensors[170,171]. Some experimentalists resort to assuming an η_Q value based on knowledge of the crystal structures, usually as 0 or 1.[172] On the other hand, both CASTEP and VASP show a strong correlation, R^2 =0.95, which suggests that these codes remain self-consistent with respect to the full EFG tensor. Consequently, we

demonstrate that the experimentally reported asymmetry parameter may not be sufficiently robust for benchmarking comparisons.

4.4 Fast prediction of ^{27}Al C_Q with machine learning

4.4.1 DFT calculated ^{27}Al database

To predict C_Q values for ^{27}Al with machine learning, we constructed a VASP calculated database with 1681 aluminum-containing solid crystalline materials utilizing the high-throughput DFT framework of the Materials Project[91]. The sites in the database can be classified as belonging to three types of local coordination environments: 4-coordinate tetrahedron (termed "T:4"), 5-coordinate trigonal bipyramid ("T:5") and 6-coordinate octahedron ("O:6").

Unusual geometries such as 2-coordinate linear or bent geometries, 3-coordinate trigonal planar, or 4-coordinate square planar were excluded.

4.4.2 Feature engineering

One of the most critical aspects of a successful machine learning model lies in "feature engineering." In terms of materials science, features are usually properties related to the materials or values that can be derived or calculated based on materials' structural or chemical information. In terms of these chemical entities, our effort is to select features that provide a means for recognizing patterns in the data, and to correlate an NMR measurement with one or more specific chemical (or structural) properties. When successfully identified, features, either singly or combined, can form a numerical representation of the material, usually expressed in form of a 1D vector. For machine learning prediction, these numerical representations need to capture the variance of the target parameter across different materials to be successful. The process of feature engineering can be as simple as collecting the atomic numbers (i.e., the chemical identity of an

atom participating in a bond), while for many data sets, more complex constructed features are needed.

There has been considerable research on feature engineering for materials science to predict NMR parameters such as the use of smooth overlap of atomic positions (SOAP) descriptors[173,174], Coulomb matrix[175] and Behler−Parrinello symmetrical functions (BFPS)[176] . While these features are capable of describing the variance of geometries for structures with different NMR parameters such as isotropic chemical shifts/shieldings, they were designed to be very general and therefore useful in many different types of applications (beyond NMR).[177] For specific targets such as NMR spectroscopy, which aims to extract highly localized perturbations, these features may yield suboptimal results. For example, the size of a SOAP kernel scales quadratically with the number of elemental species considered, which makes it slow to process when applied to datasets with a variety of elements. Considering that NMR parameters are strongly correlated to the local environment of the target atom, the above features can be greatly simplified or replaced by customized features with better performance and efficiency.

Here we propose two types of customized features to predict the C_Q of the EFG tensor: structural features, and elemental features. Structural features are derived from the geometry of the materials' local environment without taking into consideration any difference between atomic species. Significant research in solid-state ^{27}Al NMR of aluminum-containing materials has focused on the empirical correlation between NMR measurable parameters such as C_Q and simple descriptive parameters derived from local geometry[178–181]. It appears that many of these empirical correlations are particularly useful for the recent efforts of building computational predictive models for NMR spectroscopy. For example, Ghose and Tsang[178] defined the longitudinal strain and the shear strain to quantify the distortion of the local polyhedron from the Platonic solid-like forms (i.e., with

identical faces of the geometric solid). Later Baur et al.[179] suggested a distortion index (DI) to measure the angular distortion of the local geometry. These parameters were shown to have a high level of correlation with C_Q value.

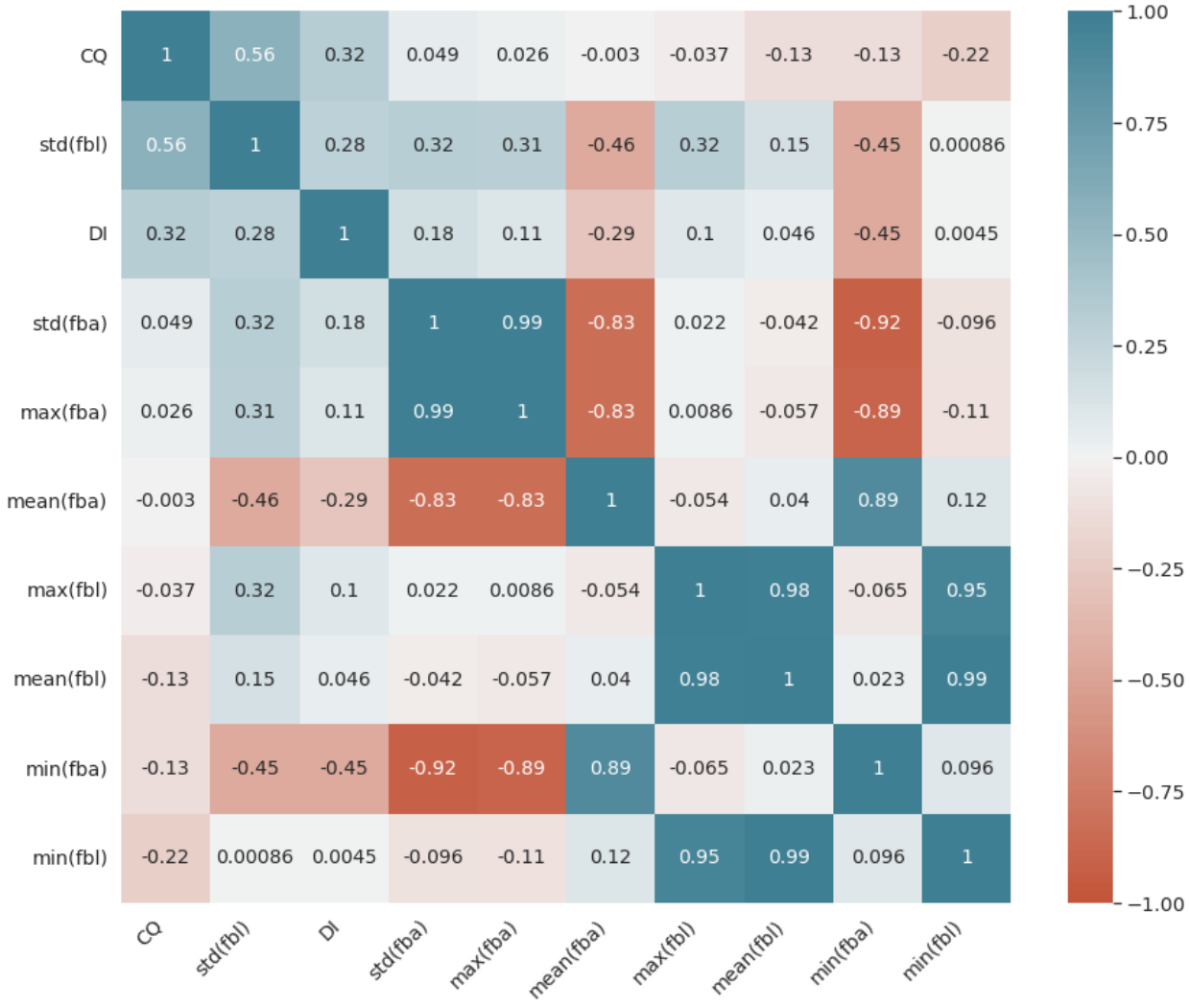


Figure 4.7 Correlation heat map between C_Q and structural features. "fbl" and "fba" here are abbreviations of the first-order bond length and the first-order bond angle. The number in each block is Pearson's correlation coefficient (PCC).

Table 4.1 List of the simple structural features and corresponding abbreviations.

Feature name	Abbreviation

First order bond length standard deviation	std(fbl)
First order bond length mean	mean(fbl)
First order bond length maximum	max(fbl)
First order bond length minimum	min(fbl)
First order bond angle standard deviation	std(fba)
First order bond angle mean	mean(fba)
First order bond angle maximum	max(fba)
First order bond angle minimum	min(fba)
Distortion index	DI

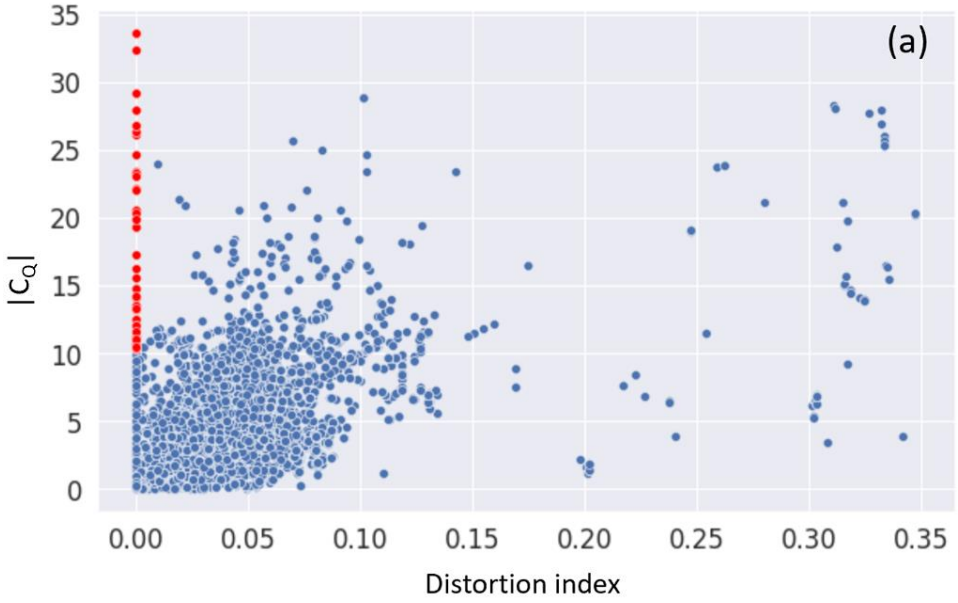
We implemented the DI parameter in python along with eight statistical averages of local polyhedron geometry in a structure: namely the maximum, minimum, standard deviation and mean of the first-order bond lengths (fbl) and bond angles (fba). A full list of structural features and their corresponding abbreviations can be found in **Table 4.1**. **Figure 4.7** shows a correlation "heat map" between the DFT-calculated NMR parameter C_Q and the structural features. The standard deviation of the first-order bond length std(fbl) has a high level of correlation with C_Q, which illustrates the power of such simple features when used for the right target. The distortion index (DI) has the second-largest correlation with C_Q.

The std(fbl) and DI characterize the distortion of the local polyhedron from its "ideal form" in terms of bond length and bond angle, respectively. We found these two features are complementary to each other in the prediction of C_Q. The distortion index (DI) is defined as follows:

$$\mathrm{DI} = \frac{\sum_{i=1}^{n} |\theta_u - \theta_0|}{\sum_{i=1}^{n} \theta_u} \tag{4.1}$$

Where θ_u represents the bond angle in an undistorted polyhedron geometry, and θ_o represents the bond angles in real cases. The DI can partially capture the distortion of the local geometry as can be seen in **Figure 4.8 (a)**; there is a correlation between the DI and C_Q, seen as a clustering of small values of |CQ| at small values of DI, and an overall trend to larger [CQ] values with

increasing DI. However, DI's prediction ability is limited when it comes to bond length variance. The red-labeled sites are those with relatively large C_Q (>10 MHz) values, whereas DI=0. These sites presumably have a small variance in bond angles but a large variance in bond lengths. Thus, DI as a parameter does not have the capacity on its own to distinguish these sites. The standard deviation of first-order bond length, however, does an excellent job of capturing the distortion of these sites. **Figure 4.8 (b)** shows a correlation between C_Q and bond length standard deviation for the red-labeled sites. DI and the first order bond length standard deviation exhibit a complementary relationship between the two features (distortion index and standard deviation of the bond lengths) for the machine learning study of C_Q.



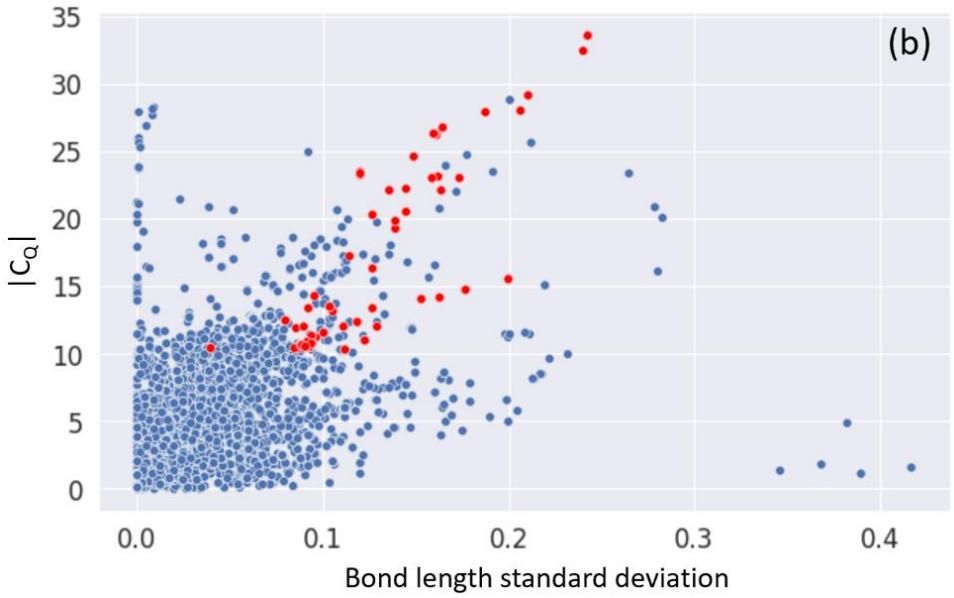


Figure 4.8 Correlation plots between a) $|C_Q|$ versus distortion index (DI); b) $|C_Q|$ versus bond length standard deviation. The ^{27}Al site with $|C_Q|$ values greater than 10 MHz and DI of 0 are highlighted in red in both plots. The red labeled sites are those with low bond angle variance and high bond length variance which makes their $|C_Q|$ more suitable to be characterized by the bond length standard deviation feature.

Two or more features having high correlations with one another is termed "multi-collinearity". While multi-collinearity does not inhibit the machine learning's ability to obtain a good prediction especially for tree-based models such as random forest, it is problematic for the model interpretation such as feature importance ranking.[182,183] After removing the features with high correlations (|Pearson's correlation coefficient|>0.8) from the structural feature set, four structural features remained: first-order bond length standard deviation (fbl_std), distortion index (DI), first-order bond angle average (fba_average) and first-order bond length average. **Figure 4.9** shows the correlation heatmap between C_Q and these features.

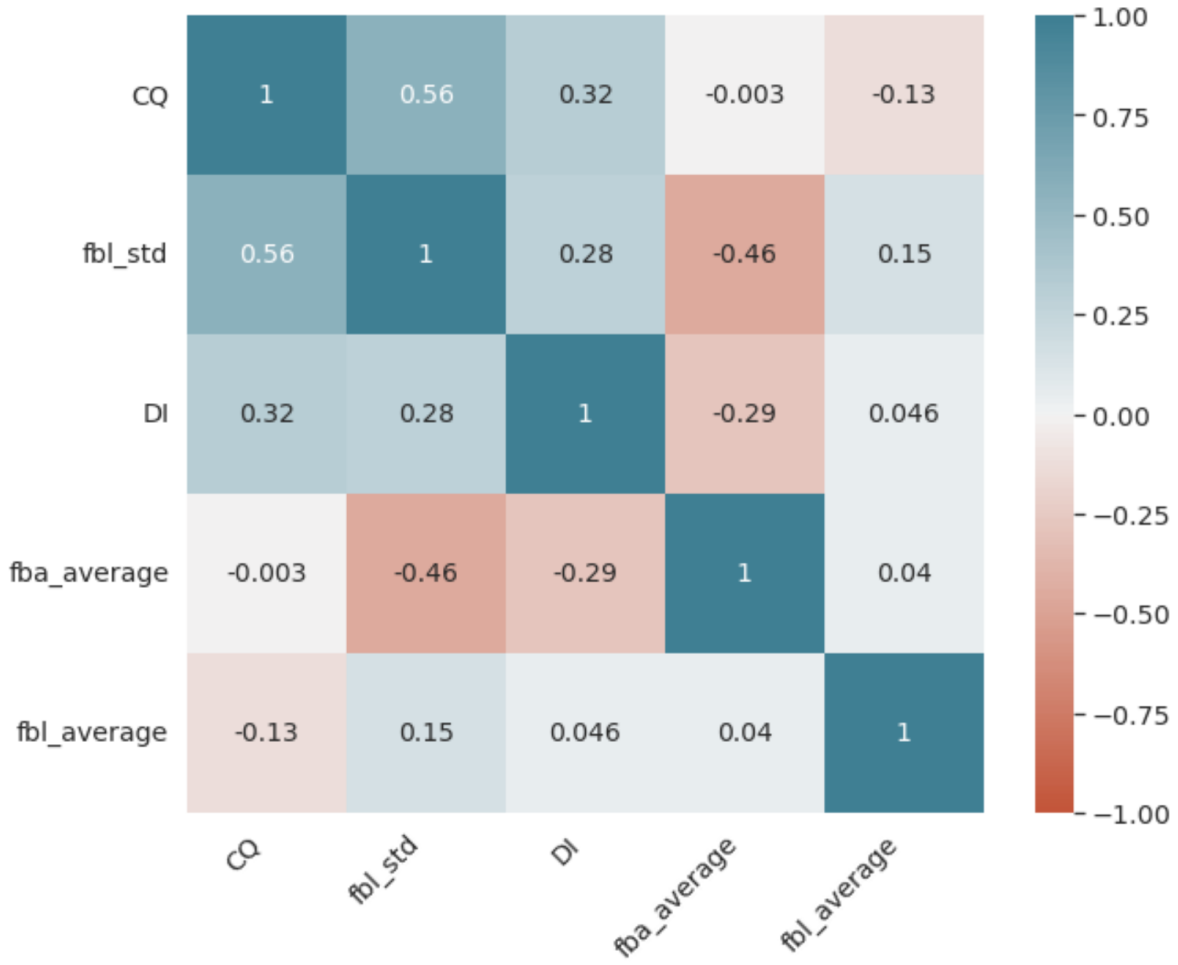


Figure 4.9 Correlation heat map between C_Q and structural features after removing the collinearity. The number in each block is Pearson's correlation coefficient (PCC).

Here we choose random forest as our machine learning model due to its stability and robustness to outliers and non-linear feature space. Using just the structural features, we trained a random forest model for ^{27}Al C_Q which derives the target value by performing data segmentation with an ensemble of decision trees.[184] **Figure 4.10** shows the correlation between the calculated DFT ^{27}Al C_Q and the model predicted C_Q. The plot shows that the set of simple structural features can already predict C_Q with a R^2 of 0.95 and RMSE of 0.77 MHz. We do note that there are still a number of

outliers suggesting characteristics other than structural features can play a significant role in dictating NMR properties.

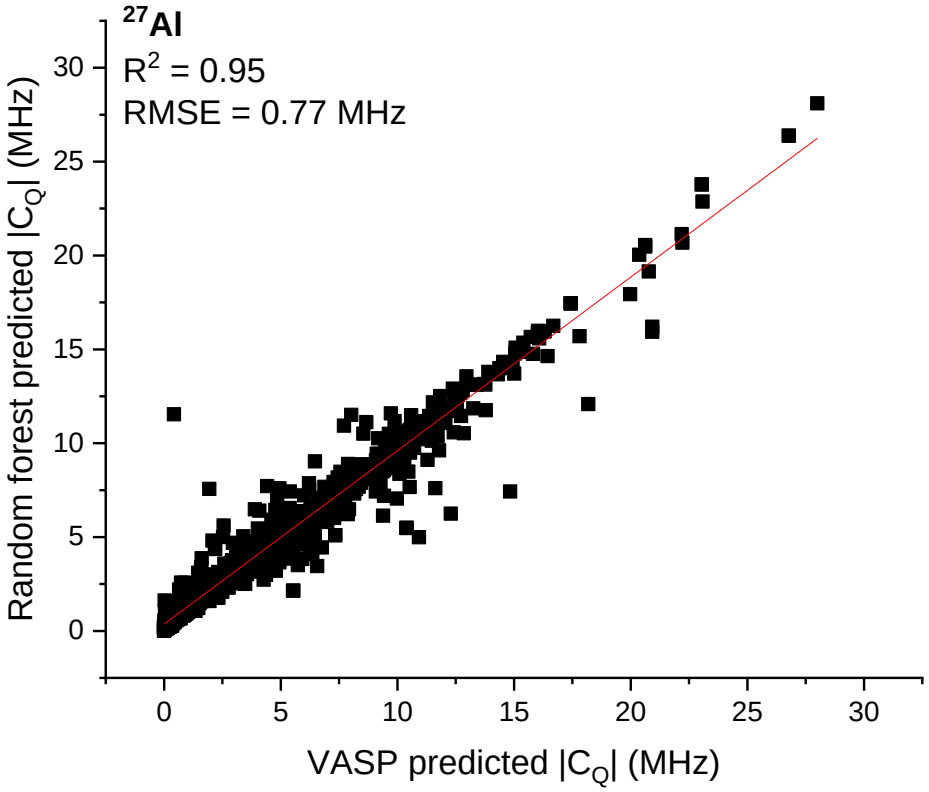


Figure 4.10 Comparison between random forest-predicted ^{27}Al C_Q with VASP-calculated ^{27}Al C_Q for aluminum-containing compounds. The random forest model was trained with structural features only (i.e., not with elemental properties). The size of the test set is 1617 individual ^{27}Al sites.

The EFG tensor is not only related to the geometry of the local environment but is also strongly influenced by the properties of surrounding atomic species because it is derived from the electronic density distribution. To further improve the prediction of C_Q, we need to represent the variation in local chemical composition into consideration. We selected twelve elemental properties such as atomic numbers, electron affinity, and other properties (**Table 4.2**) and constructed three sets of elemental features. **Figure 4.11** shows the general procedure of feature construction.

Table 4.2 List of the selected elemental properties used for feature engineering; all the values are taken from the Pymatgen library[185].

Properties Name	Abbreviation	Unit
Atomic number	Z	--
Mendeleev number ^	χ	--
Atomic mass	m_a	u (unified mass)
Electron affinity	E_{ea}	kJ/mol
Electron negativity	χ	--
Number of valence electrons	--	--
Average ionic radius‡	r_i	Å
Van der Waals radius	r_w	Å
Thermal conductivity	κ	$W/(m \cdot K)$
Melting point	--	°C
Boiling point	--	°C
Ionization energy (first)	IE_1	kJ/mol

We first get the twelve elemental properties for each atom in the first coordinate shell around the ^{27}Al sites. The first set of features is simple statistics of each of the elemental properties: its maximum, its minimum, standard deviation, and average values. The second set of features measures the differences between the neighbor atoms and the core atom (aluminum in our case).

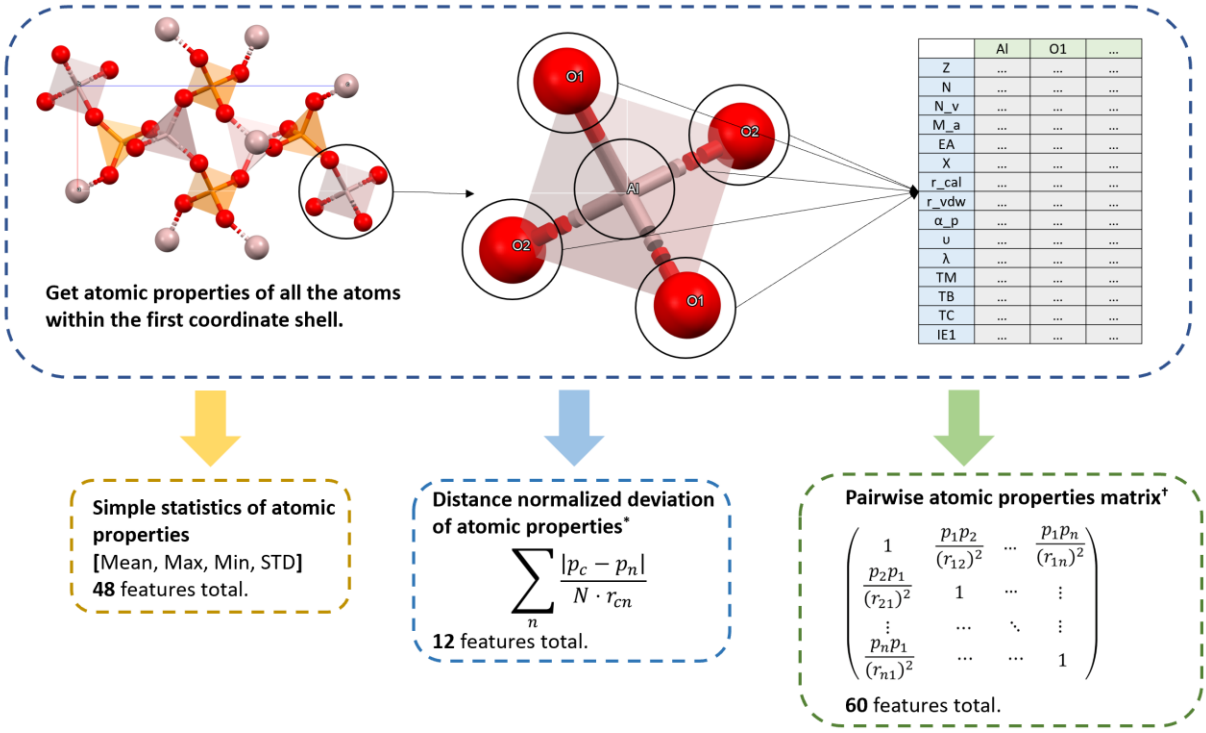
$$\sum_{\mathbf{i}} \frac{|\mathbf{p_0} - \mathbf{p_i}|}{\mathbf{N} \cdot \mathbf{r_i}} \tag{4.2}$$

Here p_0 and p_i are the atomic properties of the central atom and coordinate atoms; N is the coordination number; r_i is the corresponding bond length.

For the third set of features, we draw inspiration from the classic Coulomb matrix. For each of the twelve elemental properties, a matrix considering all the atoms within the first neighbor shell was generated.

$$\mathbf{M_{ij}} = \begin{cases} \mathbf{1}, & \mathbf{i = j} \\ \dfrac{\mathbf{p_i p_j}}{\mathbf{r_{ij}}^2}, & \mathbf{i \neq j} \end{cases} \tag{4.3}$$

Like a Coulomb matrix, this feature also considers the pairwise comparison of the selected properties between two atoms in the lattice. One challenge is that when the number of atoms considered is different, the size of the resultant matrix will also be different. In our specific case, the size of the matrix for 4-, 5- and 6-coordinated Al sites will be different. This is troublesome for machine learning prediction because most algorithms require the dimensionality of the feature space to be uniform across all the samples. To solve the problem, we decompose the matrix with singular value decomposition (SVD) and use 5 singular values, the maximum number of possible singular values for our system, as our features instead of the whole matrix.



* p_0 and p_i are the atomic properties of the central atom and coordinate atom i; N is the coordination number; r_i is the corresponding bond length.

† p_n are the atomic properties of the atoms within the first coordination shell; r_{mn} are the inter-atomic distances between atom m and atom n.

Figure 4.11 Illustration of the feature engineering process for elemental features. A list of atomic properties for each atom within the first coordination shell was collected and then transferred into 3 sets of features: simple statistics of atomic properties, distance normalized deviation of atomic properties, and pairwise atomic properties matrix.

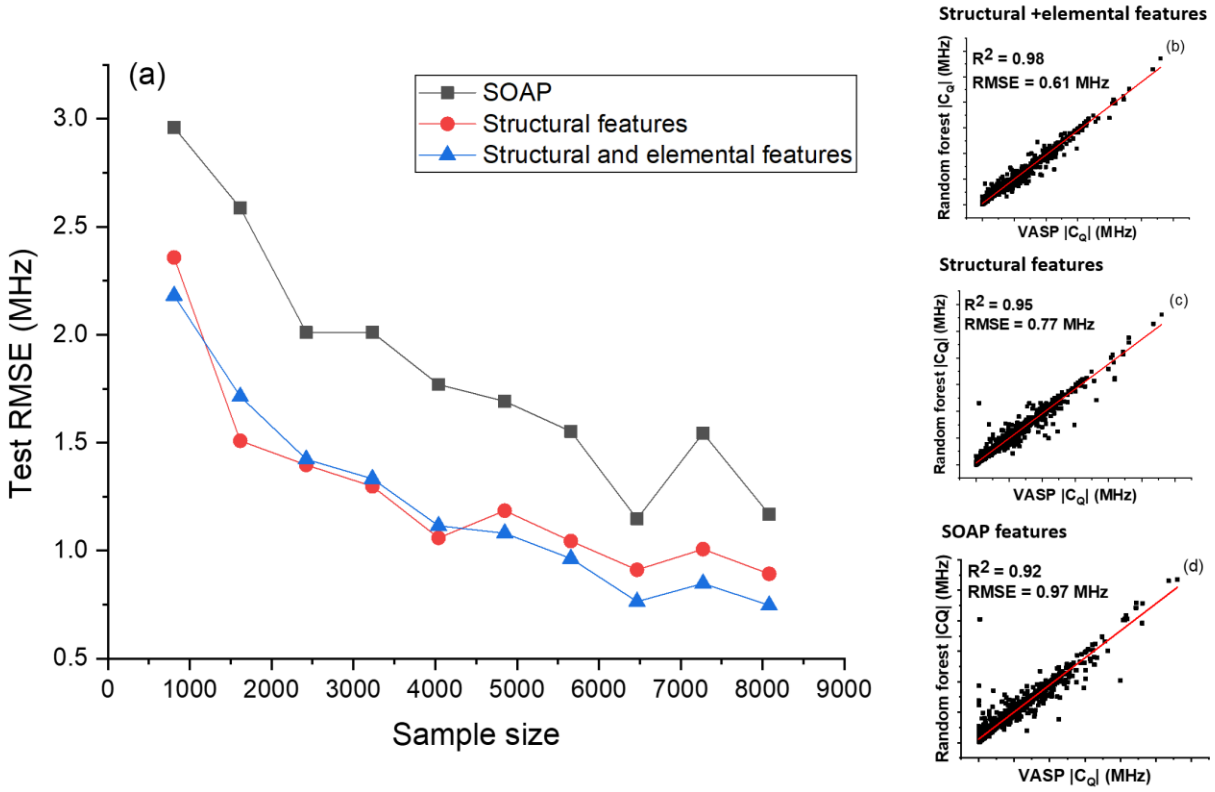


Figure 4.12 Comparison of the random forest models trained based on three different feature sets (structural+elemental, structural and SOAP). (a) The learning curve of model performance (Test RMSE) over sample size for all three models. (b) Correlation between random forest-predicted ^{27}Al C_Q with VASP-calculated ^{27}Al C_Q for aluminum-containing compounds. The random forest model was trained with structural and elemental features. (c) Correlation between random forest-predicted ^{27}Al C_Q with VASP-calculated ^{27}Al C_Q for aluminum-containing compounds. The random forest model was trained with structural features only. (d) Correlation between random forest-predicted ^{27}Al C_Q with VASP-calculated ^{27}Al C_Q for aluminum-containing compounds. The random forest model was trained with SOAP features.

We retrained the random forest model with both structural features and elemental features (structural + elemental) which improves the model accuracy to $R^2 = 0.98$ and RMSE = 0.61 MHz. To further assess the performance of the models, we also compared our models with a benchmark model using the SOAP features.[173,174] The SOAP model was also trained with a random forest algorithm based on the same set of data as our model. The only difference is the features used for training. Instead of using our structural and chemical features, we use SOAP features generated by an open-source package Dscribe,[186] the total number of SOAP features generated for our data set is 4163280. **Figure 4.12** shows a performance comparison between models based on our proposed features (structure, structural + elemental) and the model based on SOAP. As shown in **Figure 4.12 (a)**, both of our proposed features perform significantly better than SOAP no matter the size of the sample. Structural + elemental features performs better than structural features when the sample size gets larger. **Figure 4.12(b)-(d)** show the correlation plots between the VASP calculated C_Q and the machine learning predicted C_Q based on the three models. The structural + elemental model significantly reduces the number of strong outliers as seen in both **Figure 4.12 (c)** and **4.12 (d)**. The SOAP model achieves a usable performance benchmark of $R^2 = 0.92$ and RMSE = 0.97 MHz. Still, despite the significantly increased computational cost of the SOAP features, it lacks the same degree of accuracy in comparison to our simple feature set. Most importantly, the SOAP features result in strong outliers. This shows that a simple set of features that are customized for a specific problem such as NMR parameter prediction can outperform universal features because this method excludes a lot of unnecessary information that could significantly decrease the performance of the model in terms of both efficiency and accuracy.

4.5 Conclusions

By studying the correlation between experimentally measured ^{27}Al NMR parameters and DFT calculated results with a relatively large benchmarking set, we can confirm that DFT calculations are accurate in predicting the isotropic chemical shift δ_{iso} and quadrupolar coupling constant C_Q for crystalline materials that contain ^{27}Al. Similar to our previous benchmarking effort on spin ½ nuclei ^{29}Si, DFT prediction on asymmetry parameters (both η_{CS} and η_Q) is shown to be more prone to error due to the sensitivity of this parameter to slight variation in local geometry and the difficulty in accurately measuring η experimentally.

Based on our confidence in DFT's accuracy of predicting ^{27}Al NMR parameters, we built a simple machine learning model to predict ^{27}Al C_Q based on a large VASP calculated NMR dataset of 1681 aluminum-containing solid materials. The structural and elemental features that we built were proven to be effective in capturing the variation of local environments to which NMR parameters are very sensitive. It is surprising for us to find that among all the features, the pure geometrical variations such as variations in bond length and bond angles are the dominating features for C_Q prediction, which shows the possibility of building simple but effective features for the prediction of materials' properties instead of using larger universal features. Also, we can get a better understanding of the relationship between local geometry and SSNMR spectra that, specifically, SSNMR spectra for quadrupolar nuclei are determined primarily by local geometry distortions. The data are publicly available for further investigation. Our final model was proven to be effective in predicting C_Q for 4, 5 and 6 coordinated Al sites with R^2=0.98 and RMSE=0.61MHz. This accuracy is comparable with the accuracy of DFT calculations against experiment (RMSE = 0.70MHz for VASP), thus making this machine learning method a fast and lightweight complement to DFT calculations.

Chapter 5

Conclusion

In this dissertation we explored the ability of solid-state nuclear magnetic resonance (SSNMR) spectroscopy as a powerful technique for local environment refinement of solid materials with a fidelity beyond conventional diffraction-based methods. Combining both experiment and computation (such as density functional theory and machine learning) of SSNMR spectra, a methodology that is called "NMR crystallography" today has been widely applied to the determination of solid materials' structures with an increasing amount of computationally calculated NMR spectra.

We showed that while NMR crystallography has been proven to be successful on many different solid systems, two essential aspects, robustness and efficiency, remain to be tested and improved. When approaching less-commonly-studied species (i.e., isotopes such as ^{15}N, ^{27}Al, ^{29}Si, and many others throughout the periodic table), verifying the reliability of DFT computation methods becomes a challenge especially when the amount of data available is limited. The construction of a well-established computational NMR dataset is thus beneficial for both verification and guidance of computational/experimental methods. The research presented in this dissertation has created benchmarking sets for both materials with spin ½ ^{29}Si and materials with quadrupolar nuclear spin, as in ^{27}Al and ^{51}V. We compared experimentally-measured solid-state NMR tensor parameters with their first-principle calculated counterparts by two DFT packages, CASTEP and VASP. These data-driven approaches enable us to identify the source of discrepancies across a range of experimental and computational results. The information from NMR (in the form of an NMR

tensor) has been both validated, and in some cases corrected, in an effort to catalogue these for the Local Spectroscopy Database Infrastructure (LSDI), where over 10,000 NMR tensors for crystalline materials have been computed.

Based on the experimentally benchmarked DFT-SSNMR datasets, we showed that the efficiency of NMR crystallography workflow can be further improved with machine learning models trained for NMR tensor prediction. Appropriately trained ML algorithms have demonstrated the ability to capture local geometry to predict isotropic chemical shift δ_{iso} with accuracy close to DFT with only a fraction of computing time. While most of the efforts were focused on the prediction of δ_{iso}, there have been relatively few studies that can demonstrate the ability of machine learning algorithms in predicting expressions of the electric field gradient (EFG) tensor parameters such as quadrupolar coupling constant (C_Q) for quadrupolar nuclei, which provide a complimentary measurement of small perturbations to local environments, especially when it is hard to distinguish different sites based on isotropic chemical shift. In this dissertation, we developed a simple machine learning (using the random forest algorithm) model based on local structural motifs and elemental properties to predict quadrupolar tensor values for the quadrupolar nucleus ^{27}Al. This catalog permits a rapid assignment of species before validating with first-principles calculations. Prediction of the C_Q values for aluminum-containing crystalline materials yielded good agreement when compared to the DFT-computed values (RMSE of 0.61 MHz; R^2=0.98). While elemental features can significantly improve the performance of the model, simple geometric features still dominated the predictive accuracy. The model with customized features (geometric and elemental) was proven to have a significantly better performance compared with previously published state-of-the-art models (R^2~0.91).

Looking further into the future, the methodology of combining computational methods (DFT, machine learning) with experimental SSNMR has great potential beyond what has been shown in this dissertation. On the experimental side, computational methods can be used as an illustrative guide for experimental setup and efficient pursuit of spectra (by anticipating lineshapes and frequency regions to be probed); the computed results are also vital for the developments of automatic spectrum decomposition and recognition platforms. On the computational side, the development of machine learning algorithms for the prediction of the full NMR tensor is highly approachable; with the assistance of fast computation methods, rapid structure determination using NMR could be possible and the efficiency of many current structure generation methods such as crystal structure prediction (CSP) can be greatly increased. Also, many of the NMR crystallography methods described in this dissertation are transferable to the world of solution NMR; there is a great possibility that NMR crystallography could be applied to pharmaceutical and bio-chemistry problems such as selection of drug candidates, small molecule confirmation exchange and protein ligand interactions.

Appendix A

Tables of experimental and computed chemical shielding (shift) tensors values, and their descriptive algebraic expressions (δiso, ηcsa, ζ).

Table A1. Experimentally reported ^{29}Si chemical shift tensor parameters.

Experimentally reported*								
Structure Name	**Chemical Formula**	**δ_{iso} (ppm)**	**ζ_{δ} (ppm)**	**η_{CSA}**	**δ_{11} (ppm)**	**δ_{22} (ppm)**	**δ_{33} (ppm)**	**REF**
Forsterite	Mg_2SiO_4	-62	38.0	0.47	-24	-72	-90	121
Monticellite	$CaMgSiO_4$	-66	-28.0	0.57	-44	-60	-94	121
Andalusite	Al_2SiO_5	-80	-36.0	0.17	-59	-65	-116	121
Kyanite	Al_2SiO_5	-83	-37.3	0.89	-48	-81	-120	121
Akermanite	$Ca_2MgSi_2O_7$	-73	72.0	0.69	-1	-84	-134	121
Lawsonite	$CaAl_2Si_2O_7(OH)H_2O$	-81	53.3	0.58	-28	-92	-123	121
Enstatite	$MgSiO_3$	-81	51.3	0.59	-40	-70	-132	121
Diopside	$CaMgSi_2O_6$	-84	-64.0	0.66	-31	-73	-148	121
Wollastonite	$CaSiO_3$	-89	-69.3	0.88	-24	-85	-158	121
Strontium silicate	$SrSiO_3$	-85	-69.3	0.59	-30	-71	-154	121
Barium silicate	$BaSiO_3$	-80	-60.0	0.7	-29	-71	-140	121
Sillimanite	Al_2SiO_5	-87	-42.0	0.81	-49	-83	-129	121
Tremolite	$Ca_2Mg_5Si_8O_{22}(OH)_2$	-91	32.0	0	-59	-107	-107	121
Tremolite	$Ca_2Mg_5Si_8O_{22}(OH)_2$	-88	-49.3	0.55	-50	-77	-137	121
Spodumene	$LiAlSi_2O_6$	-92	-50.0	0.56	-53	-81	-142	121
Lithium disilicate	$Li_2Si_2O_5$	-93	-58.0	0.12	-61	-67	-151	121
Sodium disilicate	$Na_2Si_2O_5$	-95	-66.0	0.24	-54	-70	-161	121
α-quartz	SiO_2	-106.3	3.6	0.58	-102.7	-107.1	-109.2	122
Silica zeolite ITQ4	$Si_{32}O_{64}$	-110.4	-16.1	0.93	-94.9	-109.8	-126.5	123
Silica zeolite ITQ4	$Si_{32}O_{64}$	-112.1	-14.4	0.87	-98.6	-111.2	-126.5	123
Silica zeolite ITQ4	$Si_{32}O_{64}$	-109.2	-13.3	0.92	-96.4	-108.7	-122.5	123
Silica zeolite ITQ4	$Si_{32}O_{64}$	-107.5	12.1	0.54	-95.4	-110.3	-116.9	123
β-Ca_2SiO_4	β-Ca_2SiO_4	-71.3	-16.6	0.83	-56.1	-69.9	-87.9	124
γ-Ca_2SiO_4	γ-Ca_2SiO_4	-73.7	-25.5	0.83	-50.4	-71.5	-99.2	124
$Ca_3SiO_4Cl_2$	$Ca_3SiO_4Cl_2$	-73.6	-11.0	0.78	-63.8	-72.4	-84.6	124
Dicalcium silicate hydrate	α-$Ca_2(SiO_3OH)OH$	-72.7	-26.0	0.3	-55.8	-63.6	-98.7	124
Rankinite	$Ca_3Si_2O_7$	-74.5	55.3	0.69	-19.2	-83.0	-121.2	124
Rankinite	$Ca_3Si_2O_7$	-75.9	40.5	0.65	-35.4	-83.0	-109.3	124

Cuspidine	$Ca_4Si_2O_7F_2$	-79.9	58.3	0.61	-21.6	-91.3	-126.8	124
Wollastonite	β-$Ca_3Si_3O_9$	-87.8	-69.4	0.6	-32.3	-73.9	-157.2	124
Wollastonite	β-$Ca_3Si_3O_9$	-89.0	-59.8	0.62	-40.6	-77.6	-148.8	124
Wollastonite	β-$Ca_3Si_3O_9$	-89.5	-52.1	0.68	-45.7	-81.2	-141.6	124
Pseudowollastonite	α-$Ca_3Si_3O_9$	-83.6	-88.9	0.55	-14.7	-63.6	-172.5	124
Scawtite	$Ca_7(Si_6O_{18})$ CO_3 $(H_2O)_2$	-85.1	-49.1	0.7	-43.4	-77.7	-134.2	124
Scawtite	$Ca_7(Si_6O_{18})$ CO_3 $(H_2O)_2$	-86.5	-61.1	0.66	-35.8	-76.1	-147.6	124
Xonotlite	$Ca_6Si_6O_{17}(OH)_{12}$	-86.4	-38.1	0.65	-55.0	-79.7	-124.5	124
Xonotlite	$Ca_6Si_6O_{17}(OH)_{12}$	-87.2	-39.7	0.58	-56.0	-78.7	-126.9	124
Xonotlite	$Ca_6Si_6O_{17}(OH)_{12}$	-97.6	-33.3	0.02	-80.6	-81.3	-130.9	124
Na2SiO3	Na_2SiO_3	-77.1	-73.5	0.56	-19.7	-60.9	-150.6	124
Kyanite	Al_2SiO_5	-82.4	-19.7	0.95	-63.2	-81.9	-102.1	124
Kyanite	Al_2SiO_5	-83.3	-19.4	0.99	-64.0	-83.2	-102.7	124
Forsterite	Mg_2SiO_4	-63.2	-32.2	0.51	-38.8	-55.3	-95.4	125

* shaded cells are experimentally reported values from literature. Numeric values in unshaded cells were obtained through calculation.

Table A2a. CASTEP-computed ^{29}Si chemical shielding tensor elements, their algebraic expressions, and isotropic chemical shift.

CASTEP-computed							
Structure Name	**σ_{iso} (ppm)**	**δ_{iso} (ppm)***	**ζ_σ (ppm)**	**η_{CSA}**	**σ_{11} (ppm)^**	**σ_{22} (ppm)^**	**σ_{33} (ppm)^**
Forsterite	384.83	-61.27	30.25	0.61	360.48	378.93	415.08
Monticellite	390.52	-66.36	31.97	0.20	371.34	377.73	422.49
Andalusite	404.80	-79.12	-18.36	0.87	386.44	405.99	421.97
Kyanite	408.50	-82.42	8.90	0.94	399.87	408.23	417.40
Akermanite	397.35	-72.46	-55.73	0.96	341.62	398.46	451.97
Lawsonite	406.34	-80.49	-31.77	0.81	374.57	409.36	435.09
Enstatite	404.92	-79.22	54.45	0.69	358.91	396.48	459.37
Diopside	410.47	-84.18	66.05	0.67	355.32	399.57	476.52
Wollastonite	415.27	-88.47	72.49	0.55	359.09	398.96	487.76
Strontium silicate	412.22	-85.75	95.59	0.45	342.92	385.93	507.81
Barium silicate	405.09	-79.38	68.35	0.43	356.22	385.61	473.44
Sillimanite	414.34	-87.64	21.69	0.13	402.09	404.91	436.03
Tremolite	418.96	-91.77	19.35	0.28	406.58	412.00	438.31
Tremolite	413.90	-87.25	57.65	0.75	363.46	406.69	471.55
Spodumene	418.02	-90.93	38.07	0.47	390.04	407.93	456.09
Lithium disilicate	417.49	-90.46	65.23	0.17	379.33	390.42	482.72
Sodium disilicate	420.38	-93.04	69.09	0.09	382.73	388.95	489.47
α-quartz	432.78	-104.12	-2.93	0.78	429.85	433.10	435.39
Silica zeolite ITQ4	440.90	-111.37	14.45	0.82	427.75	439.60	455.35

Silica zeolite ITQ4	443.21	-113.44	14.96	0.67	430.72	440.74	458.17
Silica zeolite ITQ4	439.17	-109.83	12.72	0.87	427.28	438.34	451.89
Silica zeolite ITQ4	437.00	-107.89	-11.03	0.36	425.97	440.53	444.50
β-Ca_2SiO_4	400.13	-74.94	15.00	0.91	385.81	399.46	415.13
γ-Ca_2SiO_4	400.15	-74.96	26.73	0.74	376.90	396.68	426.88
$Ca_3SiO_4Cl_2$	400.98	-75.70	14.05	0.51	390.37	397.54	415.03
Dicalcium silicate hydrate	396.71	-71.89	-51.78	0.14	344.93	418.98	426.22
Rankinite	401.47	-76.14	-53.61	0.51	347.86	414.61	441.95
Rankinite	402.57	-77.12	-41.55	0.34	361.02	416.28	430.41
Cuspidine	405.91	-80.11	-57.27	0.24	348.64	427.67	441.42
Wollastonite	415.27	-88.47	72.49	0.55	359.09	398.96	487.76
Wollastonite	417.40	-90.38	60.45	0.59	369.34	405.01	477.85
Wollastonite	417.81	-90.74	52.47	0.70	373.21	409.94	470.28
Pseudowollastonite	410.97	-84.63	93.97	0.44	343.31	384.66	504.94
Scawtite	413.13	-86.56	44.73	0.60	377.35	404.18	457.86
Scawtite	413.16	-86.59	64.72	0.64	360.09	401.51	477.88
Xonotlite	409.46	-83.28	31.19	0.65	383.73	404.00	440.65
Xonotlite	411.90	-85.46	36.61	0.86	377.86	409.34	448.51
Xonotlite	427.61	-99.50	34.71	0.44	402.62	417.89	462.32
Na_2SiO_3	400.85	-75.59	82.37	0.49	339.48	379.84	483.22
Kyanite	408.48	-82.40	8.95	0.58	401.41	406.60	417.43
Kyanite	409.35	-83.18	8.29	0.97	401.19	409.23	417.64
Forsterite	384.83	-61.27	30.25	0.61	360.48	378.93	415.08

* The CASTEP δ_{iso} values were obtained using the linear regression referencing method, discussed in Chapter 3.
^ These values are obtained from CASTEP output in the .magres file.

Table A2b. CASTEP-computed ^{29}Si chemical shift tensor elements, their algebraic expressions, and isotropic chemical shielding (for reference).

CASTEP-computed							
Structure Name	**σ_{iso} (ppm)**	**δ_{iso} (ppm)***	**ζ_δ (ppm) †**	**η_{CSA} †**	**δ_{11} (ppm)***	**δ_{22} (ppm)***	**δ_{33} (ppm)***
Forsterite	384.83	-61.27	-27.03	0.61	-39.51	-56.00	-88.30
Monticellite	390.52	-66.36	-28.56	0.20	-49.22	-54.93	-94.92
Andalusite	404.80	-79.12	16.41	0.87	-62.71	-80.18	-94.46
Kyanite	408.50	-82.42	-7.95	0.94	-74.71	-82.18	-90.38
Akermanite	397.35	-72.46	49.80	0.96	-22.66	-73.45	-121.27
Lawsonite	406.34	-80.49	28.39	0.81	-52.10	-83.19	-106.18
Enstatite	404.92	-79.22	-48.66	0.69	-38.11	-71.68	-127.88
Diopside	410.47	-84.18	-59.02	0.67	-34.90	-74.44	-143.20

Wollastonite	415.27	-88.47	-64.78	0.55	-38.27	-73.90	-153.25
Strontium silicate	412.22	-85.75	-85.41	0.45	-23.82	-62.25	-171.16
Barium silicate	405.09	-79.38	-61.08	0.43	-35.71	-61.97	-140.45
Sillimanite	414.34	-87.64	-19.38	0.13	-76.69	-79.21	-107.02
Tremolite	418.96	-91.77	-17.29	0.28	-80.71	-85.55	-109.06
Tremolite	413.90	-87.25	-51.51	0.75	-42.18	-80.81	-138.76
Spodumene	418.02	-90.93	-34.02	0.47	-65.93	-81.91	-124.95
Lithium disilicate	417.49	-90.46	-58.29	0.17	-56.36	-66.27	-148.74
Sodium disilicate	420.38	-93.04	-61.73	0.09	-59.40	-64.95	-154.77
α-Quartz	432.78	-104.12	2.62	0.78	-101.50	-104.40	-106.45
silica zeolite ITQ4	440.90	-111.37	-12.91	0.82	-99.62	-110.21	-124.29
silica zeolite ITQ4	443.21	-113.44	-13.37	0.67	-102.28	-111.23	-126.81
silica zeolite ITQ4	439.17	-109.83	-11.37	0.87	-99.20	-109.09	-121.19
silica zeolite ITQ4	437.00	-107.89	9.86	0.36	-98.03	-111.04	-114.59
β-Ca2SiO4	400.13	-74.94	-13.40	0.91	-62.15	-74.34	-88.35
γ-Ca2SiO4	400.15	-74.96	-23.88	0.74	-54.19	-71.86	-98.85
Ca3SiO4Cl2	400.98	-75.70	-12.55	0.51	-66.22	-72.63	-88.26
dicalcium silicate hydrate	396.71	-71.89	46.27	0.14	-25.62	-91.79	-98.26
rankinite	401.47	-76.14	47.91	0.51	-28.24	-87.88	-112.31
rankinite	402.57	-77.12	37.12	0.34	-40.00	-89.37	-102.00
cuspidine	405.91	-80.11	51.17	0.24	-28.93	-99.55	-111.84
wollastonite	415.27	-88.47	-64.78	0.55	-38.27	-73.90	-153.25
wollastonite	417.40	-90.38	-54.02	0.59	-47.43	-79.30	-144.39
wollastonite	417.81	-90.74	-46.88	0.70	-50.89	-83.71	-137.63
pseudowollastonite	410.97	-84.63	-83.97	0.44	-24.17	-61.12	-168.60
scawtite	413.13	-86.56	-39.97	0.60	-54.59	-78.56	-126.53
scawtite	413.16	-86.59	-57.83	0.64	-39.17	-76.18	-144.42
xonotlite	409.46	-83.28	-27.87	0.65	-60.29	-78.40	-111.15
xonotlite	411.90	-85.46	-32.71	0.86	-55.04	-83.17	-118.17
xonotlite	427.61	-99.50	-31.02	0.44	-77.17	-90.81	-130.51
Na2SiO3	400.85	-75.59	-73.61	0.49	-20.75	-56.81	-149.19
kyanite	408.48	-82.40	-7.99	0.58	-76.09	-80.72	-90.40
kyanite	409.35	-83.18	-7.40	0.97	-75.89	-83.07	-90.59
Forsterite	384.83	-61.27	-27.03	0.61	-39.51	-56.00	-88.30

* The CASTEP δ_{iso}, δ_{11}, δ_{22} and δ_{33} values were obtained using the linear regression referencing method, discussed in Chapter 3.

† ζ_δ (ppm) and η_{CSA} values were calculated from δ_{11}, δ_{22} and δ_{33} (converted to δ_{XX}, δ_{YY} and δ_{ZZ}) to match Haeberlen convention expressions (for reference).

Table S3. VASP computed chemical shift tensor parameters: isotropic chemical shielding tensor, algebraic expressions, and isotropic chemical shift.

VASP-computed							
Structure Name	**σ_{iso} (ppm)**	**δ_{iso} (ppm)***	**ζ_{δ} (ppm) †**	**η_{CSA} †**	**δ_{11} (ppm)***	**δ_{22} (ppm)***	**δ_{33} (ppm)***
Forsterite	457.29	-61.35	-22.63	0.64	-42.80	-57.28	-83.98
Monticellite	450.84	-66.94	-25.59	0.21	-51.45	-56.82	-92.53
Andalusite	437.71	-78.30	15.89	0.87	-62.41	-79.33	-93.15
Kyanite	434.00	-81.51	7.03	0.98	-74.47	-81.58	-88.47
Akermanite	442.98	-73.74	44.03	0.92	-29.70	-75.50	-116.01
Lawsonite	436.21	-79.60	23.98	0.77	-55.62	-82.35	-100.82
Enstatite	434.07	-81.45	-43.23	0.66	-45.56	-74.10	-124.68
Diopside	431.47	-83.70	-52.44	0.69	-39.39	-75.57	-136.14
Wollastonite	427.46	-87.17	-58.29	0.56	-41.70	-74.34	-145.46
Strontium silicate	430.32	-84.69	-74.53	0.45	-30.66	-64.20	-159.22
Barium silicate	436.54	-79.31	-54.13	0.43	-40.61	-63.88	-133.44
Sillimanite	428.98	-85.85	-14.87	0.16	-77.23	-79.61	-100.72
Tremolite	423.90	-90.25	-15.18	0.28	-80.53	-84.78	-105.43
Tremolite	428.79	-86.02	-46.61	0.73	-45.70	-79.72	-132.62
Spodumene	424.40	-89.82	-31.23	0.46	-67.02	-81.38	-121.05
Lithium disilicate	423.65	-90.47	-52.10	0.15	-60.51	-68.32	-142.56
Sodium disilicate	419.24	-94.28	-55.69	0.07	-64.49	-68.39	-149.97
α-Quartz	406.23	-105.54	-3.06	0.96	-102.54	-105.48	-108.60
silica zeolite ITQ4	398.88	-111.90	-12.87	0.81	-100.25	-110.68	-124.77
silica zeolite ITQ4	396.46	-114.00	-13.71	0.65	-102.69	-111.60	-127.70
silica zeolite ITQ4	400.93	-110.13	-11.64	0.89	-99.13	-109.49	-121.76

silica zeolite ITQ4	403.25	-108.12	10.06	0.42	-98.06	-111.04	-115.26
β-Ca2SiO4	441.43	-75.08	-12.04	0.86	-63.88	-74.23	-87.12
γ-Ca2SiO4	440.15	-76.19	-21.38	0.76	-57.37	-73.62	-97.57
Ca3SiO4Cl2	440.82	-75.61	-11.75	0.44	-67.14	-72.31	-87.36
dicalcium silicate hydrate	445.05	-71.95	40.76	0.15	-31.19	-89.27	-95.38
rankinite	440.31	-76.05	43.19	0.46	-32.86	-87.71	-107.57
rankinite	438.58	-77.55	33.04	0.30	-44.51	-89.11	-99.02
cuspidine	435.28	-80.40	46.50	0.24	-33.90	-98.07	-109.23
wollastonite	427.33	-87.28	-58.64	0.56	-41.54	-74.38	-145.92
wollastonite	425.30	-89.04	-48.09	0.59	-50.81	-79.18	-137.13
wollastonite	424.86	-89.42	-42.14	0.69	-53.81	-82.89	-131.56
pseudowollastonite	431.17	-83.96	-76.32	0.45	-28.62	-62.97	-160.28
scawtite	428.97	-85.86	-35.74	0.58	-57.62	-78.35	-121.60
scawtite	429.06	-85.79	-52.32	0.63	-43.14	-76.10	-138.11
xonotlite	426.49	-88.01	-27.05	0.67	-65.42	-83.54	-115.06
xonotlite	425.92	-88.50	-28.22	0.62	-65.64	-83.14	-116.72
xonotlite	412.91	-99.76	-29.12	0.13	-83.30	-87.09	-128.88
Na2SiO3	438.61	-77.52	-66.55	0.50	-27.60	-60.88	-144.07
kyanite	434.84	-80.78	-7.39	0.67	-74.61	-79.56	-88.17
kyanite	433.99	-81.52	7.03	0.97	-74.48	-81.62	-88.44
Forsterite	457.30	-61.35	-22.62	0.64	-42.79	-57.27	-83.97

* The VASP δ_{iso}, δ_{11}, δ_{22} and δ_{33} values were obtained using the linear regression referencing method, discussed in Chapter 3.

† ζ_{δ} (ppm) and η_{CSA} values were calculated from δ_{11}, δ_{22} and δ_{33} (converted to δ_{XX}, δ_{YY} and δ_{ZZ}) to match Haeberlen convention expressions (for reference).

Appendix B

Benchmark Set: ^{27}Al Experimental and DFT computed NMR quadrupolar parameters and isotropic chemical shift/shielding.

Table B1. Experimentally reported ^{27}Al NMR quadrupolar parameters and isotropic chemical shifts.

Experimentally reported ^{27}Al NMR parameters					
Chemical Formula	**Structure Name**	**δ_{iso}(ppm)**	**η_Q**	**$\lvert C_Q \rvert$(MHz)**	**Reference**
$AlPO_4$	Berlinite	44.5	0.367	4.088	187
$Al_2Ge_2O_7$	Dialuminium digermanate	36	0.4	8.8	188
$LaAlGe_2O_7$	$LaAlGe_2O_7$	35	0.3	7.2	188
$AlCl_3$	Aluminum chloride		0.1	0.425	189
Al_2O_3	α-Al_2O_3		0	2.38	190
Al_2SiO_5	Andalusite	36	0.7	5.9	191
Al_2SiO_5	Andalusite	12	0.12	15.5	191
$Ca_3Al_2O_6$	Nonacalcium tris(dialuminate)	79.5	0.32	8.69	192
$Ca_3Al_2O_6$	Nonacalcium tris(dialuminate)	78.25	0.54	9.3	192
$CaAlF_5$	Jakobssonite		0.1	10.2	193
$CaO\ (Al_2O_3)_2$	Grossite	75.5	0.88	6.25	194
$CaO\ (Al_2O_3)_2$	Grossite	69.5	0.82	9.55	194
$CaO\ Al_2O_3$	Krotite	81.9	0.2	2.5	194
$CaO\ Al_2O_3$	Krotite	83.8	0.75	2.6	194
$CaO\ Al_2O_3$	Krotite	86.2	0.95	2.6	194
$CaO\ Al_2O_3$	Krotite	82.7	0.53	3.32	194
$CaO\ Al_2O_3$	Krotite	81.6	0.39	3.37	194
$CaO\ Al_2O_3$	Krotite	81.2	0.47	4.3	194
$(CaO)_{12}\ (Al_2O_3)_7$	$(CaO)_{12}\ (Al_2O_3)_7$	85.9	0.4	9.7	194
$(CaO)_{12}\ (Al_2O_3)_7$	$(CaO)_{12}\ (Al_2O_3)_7$	80.2	0.7	3.8	194
Al_2SiO_5	Sillimanite	64.5	0.532	6.77	195
Al_2SiO_5	Sillimanite	4	0.462	8.93	195
$Na_8Al_6Si_6O_{24}C_{l2}$	Sodalite	62.9	0.32	0.94	196
$Na_3KAl_4Si_4O_{16}$	Nepheline	61			197

$Na_3KAl_4Si_4O_{16}$	Nepheline	63.5			197
$CaAl_2(Al_2Si_2)O_{10}(OH)_2$	Margarite	76		4.2	164
$CaAl_2(Al_2Si_2)O_{10}(OH)_2$	Margarite	11		6.3	164
$KAl_2(AlSi_3O_{10})(F,OH)_2$	Muscovite	72		2.1	164
$KAl_2(AlSi_3O_{10})(F,OH)_2$	Muscovite	5		2.2	164
$KAlSiO_4$	Kalsilite	61.7		1.93	172
Al_2SiO_5	Kyanite		0.89	3.7	198
Al_2SiO_5	Kyanite		0.27	10.04	198
Al_2SiO_5	Kyanite		0.38	9.37	198
Al_2SiO_5	Kyanite		0.59	6.53	198
AlF_3	α-aluminum fluoride		0	0.23	168
AlF_3	β-aluminum fluoride		0.8	0.8	168
$Al(OH)_3$	Bayerite	9.1	0.25	1.9	199
$Al(OH)_3$	Bayerite	13.1	0.8	1.4	199
$Al(OH)_3$	Gibbsite	12.5	0.45	4.6	200
$Al(OH)_3$	Gibbsite	10.5	0.75	2.1	200
$NaAlO_2$	β-$NaAlO_2$	80.1	0.5	1.4	201
$AlTlO_2$	$AlTlO_2$	69			201
$KAlO_2$	Potassium aluminate	76	0.7	1.1	201
$BaAl_2O_4$	Barium dialuminate	78	0.4	2.4	201
$Al(OH)_3$	δ-$Al(OH)_3$	9.1	0.72	2.76	200
$Al_2Si_4O_{11}$	Pyrophyllite dehydroxylate	29	0.6	10.5	202
$(Al_2O_3)_9(B_2O_3)_2$	Boroaluminate	71	0.45	9	203
$(Al_2O_3)_9(B_2O_3)_2$	Boroaluminate	52	0.06	6.773	203
$(Al_2O_3)_9(B_2O_3)_2$	Boroaluminate	44	0.7	7.94	203
$(Al_2O_3)_9(B_2O_3)_2$	Boroaluminate	8.6	0.4	5.967	203
$Na_5Al_3F_{14}$	Chiolite	-3		6.5	167
$Na_5Al_3F_{14}$	Chiolite	-1		8.2	167
AlF_3	β-aluminum fluoride	-12.5		3.4	167
Na_3AlF_6	Cryolite	1.4		2	167
K_2NaAlF_6	Elpasolite	0.8		1.4	167
$KAlSi_3O_8$	Microcline		0.21	3.22	178
$AlPO_4$	Berlinite		0.37	4.09	178
$ZnAl_2O_4$	Gahnite		0	3.68	178
$MgAl_2O_4$	Spinel		0	3.68	204
$BeAlSiO_4(OH)$	Euclase		0.7	5.17	205

$Al_2SiO_4F_2$	Topaz		0.38	1.67	206
$LiAl(SiO_3)_2$	Spodumene		0.94	2.95	207
$Be_3Al_2(SiO_3)_6$	Beryl		0	3.09	208
$Ca_3Al_2(SiO_4)_3$	Grossular		0	3.61	204
$Ca_2Al_3Si_3HO_{13}$	Zoisite		0.46	8.05	209
$Ca_2Al_3Si_3HO_{13}$	Zoisite		0.16	18.5	209
$NaAlSi_3O_8$	Albite		0.62	3.29	204
$Mg_2Al_4Si_5O_{18}$	Cordierite		0.38	10.6	210
$Mg_2Al_4Si_5O_{18}$	Cordierite		0.34	5.6	210
$Y_3Al_5O_{12}$	Y, Al-garnet		0	0.63	211
$Y_3Al_5O_{12}$	Y, Al-garnet		0	6.02	211
$BeAl_2O_4$	Chrysoberyl		0.94	2.85	212
$BeAl_2O_4$	Chrysoberyl		0.76	2.85	212
$Al_4Si_4O_{10}(OH)_8$	Kaolinite	4			162
$Mg_3Al_2(SiO_4)_3$	Pyrope	2.9	0.5	1	213
$Ca_3Al_2(SiO_4)_3$	Grossular	-3.35	0.2	3.7	213
$(CaO)_{12}(Al_2O_3)_7$	$(CaO)_{12}(Al_2O_3)_7$	85	0.2	11	165
$(CaO)_{12}(Al_2O_3)_7$	$(CaO)_{12}(Al_2O_3)_7$	79	0.9	3.7	165
Al_2SiO_5	Andalusite		0.7	5.9	214
Al_2SiO_5	Andalusite		0.08	15.6	214
$CaO\ (Al_2O_3)_2$	Calcium dialuminate	75.5	0.88	5.25	165
$CaO\ (Al_2O_3)_2$	Calcium dialuminate	69.5	0.82	9.55	165
$(CaO)(Al_2O_3)$	Krotite	81.9	0.85	2.7	165
$(CaO)_4(Al_2O_3)_3$	Tetracalcium trialuminate	80.3	0.95	2.4	165
Al_2O_3	α-Al_2O_3	16	0	2.38	215
$LiAlO_2$	α-$LiAlO_2$	17	0.05	2.8	216
$LiAlO_2$	γ-$LiAlO_2$	82	0.7	3.2	216
Al_2SiO_5	Sillimanite	63.9	0.51	6.74	217
Al_2SiO_5	Sillimanite	4.7	0.49	8.83	217
Al_2SiO_5	Andalusite	35.5	0.69	5.8	218
Al_2SiO_5	Andalusite	13	0.08	15.3	218
Al_2SiO_5	Kyanite	14.9	0.27	10.1	219
Al_2SiO_5	Kyanite	4	0.85	3.6	219
Al_2SiO_5	Kyanite	7.7	0.59	6.6	219
Al_2SiO_5	Kyanite	11	0.38	9.2	219
$YAlO_3$	Yttrium aluminate	10.7			220

$AlVO_4$	Aluminium vanadate(V)	-8.9	0.3	1.64	221
$AlVO_4$	Aluminium vanadate(V)	27.2	0.42	6.73	221
$AlVO_4$	Aluminium vanadate(V)	-1.1	0.58	5.88	221
$Al(OH)_3$	Gibbsite	8			162
$KAl_3Si_3(HO_6)_2$	Moscovite	4			162
$KAl_3Si_3(HO_6)_2$	Moscovite	72			162
$CaAl_2(Al_2Si_2)O_{10}(OH)_2$	Margarite	74			162
$CaAl_2(Al_2Si_2)O_{10}(OH)_2$	Margarite	2			162
$CaAl_2(Al_2Si_2)O_{10}(OH)_2$	Margarite	-10			162

Table B2a. VASP calculated ^{27}Al NMR quadrupolar parameters, isotropic shielding and corresponding isotropic chemical shift.

VASP computed ^{27}Al NMR parameters						
Chemical Formula	**Structure Name**	**σ_{iso} (ppm)**	**δ_{iso}(ppm)***	**η_Q**	**C_Q(MHz)**	**Reference**
$AlPO_4$	Berlinite	245.30	35.35	0.33	4.05	187
$Al_2Ge_2O_7$	Dialuminium digermanate	245.71	35.82	0.42	-8.31	188
$LaAlGe_2O_7$	$LaAlGe_2O_7$	245.95	36.10	0.16	-6.34	188
$AlCl_3$	Aluminum chloride	234.78	23.22	0.15	0.17	189
Al_2O_3	α-Al_2O_3	229.71	17.38	0.00	2.07	190
Al_2SiO_5	Andalusite	244.94	34.93	0.67	4.54	191
Al_2SiO_5	Andalusite	223.33	10.02	0.17	14.93	191
$Ca_3Al_2O_6$	Nonacalcium tris(dialuminate)	281.89	77.53	0.44	-7.52	192
$Ca_3Al_2O_6$	Nonacalcium tris(dialuminate)	281.08	76.59	0.68	-8.06	192
$CaAlF_5$	Jakobssonite	205.69	-10.31	0.11	9.00	193
$CaO\ (Al_2O_3)_2$	Grossite	281.44	77.01	0.73	6.07	194
$CaO\ (Al_2O_3)_2$	Grossite	277.45	72.41	0.79	-8.52	194
$CaO\ Al_2O_3$	Krotite	286.13	82.41	0.52	-1.59	194
$CaO\ Al_2O_3$	Krotite	287.28	83.74	0.77	2.27	194
$CaO\ Al_2O_3$	Krotite	287.52	84.02	0.79	-2.25	194
$CaO\ Al_2O_3$	Krotite	284.92	81.02	0.51	-3.07	194
$CaO\ Al_2O_3$	Krotite	286.37	82.70	0.38	-3.39	194
$CaO\ Al_2O_3$	Krotite	287.76	84.29	0.40	3.80	194

$(CaO)_{12}(Al_2O_3)_7$	$(CaO)_{12}(Al_2O_3)_7$	280.81	76.28	0.63	-8.06	194
$(CaO)_{12}(Al_2O_3)_7$	$(CaO)_{12}(Al_2O_3)_7$	280.73	76.19	0.37	-4.18	194
Al_2SiO_5	Sillimanite	270.17	64.02	0.53	-6.09	195
Al_2SiO_5	Sillimanite	218.19	4.10	0.62	-8.72	195
$Na_8Al_6Si_6O_{24}C_{l2}$	Sodalite	268.44	62.03	0.00	-0.97	196
$Na_3KAl_4Si_4O_{16}$	Nepheline	268.68	62.30	0.78	-1.45	197
$Na_3KAl_4Si_4O_{16}$	Nepheline	263.59	56.43	0.00	0.81	197
$CaAl_2(Al_2Si_2)O_{10}(OH)_2$	Margarite	283.00	78.80	0.42	-5.57	164
$CaAl_2(Al_2Si_2)O_{10}(OH)_2$	Margarite	220.98	7.31	0.14	6.56	164
$KAl_2(AlSi_3O_{10})(F,OH)_2$	Muscovite	279.62	74.91	0.66	2.86	164
$KAl_2(AlSi_3O_{10})(F,OH)_2$	Muscovite	221.64	8.07	0.92	3.01	164
$KAlSiO_4$	Kalsilite	268.93	62.59	0.76	1.62	172
Al_2SiO_5	Kyanite	217.50	3.30	0.97	3.79	198
Al_2SiO_5	Kyanite	224.89	11.82	0.30	9.98	198
Al_2SiO_5	Kyanite	224.11	10.92	0.32	-9.86	198
Al_2SiO_5	Kyanite	220.52	6.79	0.46	-7.02	198
AlF_3	α-aluminum fluoride	195.36	-22.23	0.00	-0.15	168
AlF_3	β-aluminum fluoride	195.92	-21.57	0.19	-1.31	168
$Al(OH)_3$	Bayerite	229.41	17.03	0.34	-3.51	199
$Al(OH)_3$	Bayerite	229.27	16.87	0.86	1.75	199
$Al(OH)_3$	Gibbsite	228.34	15.79	0.36	-5.46	200
$Al(OH)_3$	Gibbsite	226.07	13.18	0.85	2.20	200
$NaAlO_2$	β-$NaAlO_2$	285.28	81.43	0.61	0.80	201
$AlTlO_2$	$AlTlO_2$	270.52	64.42	0.01	-0.01	201
$KAlO_2$	Potassium aluminate	280.37	75.77	0.85	-1.20	201
$BaAl_2O_4$	Barium dialuminate	283.21	79.05	0.67	2.36	201
$Al(OH)_3$	δ-$Al(OH)_3$	226.17	13.30	0.45	3.36	200
$Al_2Si_4O_{11}$	Pyrophyllite dehydroxylate	241.91	31.44	0.64	-11.40	202
$(Al_2O_3)_9(B_2O_3)_2$	Boroaluminate	278.65	73.79	0.41	8.72	203
$(Al_2O_3)_9(B_2O_3)_2$	Boroaluminate	261.46	53.98	0.04	6.67	203
$(Al_2O_3)_9(B_2O_3)_2$	Boroaluminate	253.22	44.48	0.71	6.98	203
$(Al_2O_3)_9(B_2O_3)_2$	Boroaluminate	221.78	8.24	0.40	-5.72	203
$Na_5Al_3F_{14}$	Chiolite	210.41	-4.87	0.17	6.31	167
$Na_5Al_3F_{14}$	Chiolite	207.96	-7.69	0.00	-5.09	167
AlF_3	β-aluminum fluoride	195.92	-21.57	0.19	-1.31	167
Na_3AlF_6	Cryolite	214.09	-0.63	0.77	-0.89	167
K_2NaAlF_6	Elpasolite	214.54	-0.11	N/A	0.00	167

$KAlSi_3O_8$	Microcline	263.90	56.79	0.44	-3.15	178
$AlPO_4$	Berlinite	244.29	34.19	0.40	4.08	178
$ZnAl_2O_4$	Gahnite	232.65	20.76	0.00	-3.93	178
$MgAl_2O_4$	Spinel	228.73	16.25	0.00	-3.70	204
$BeAlSiO_4(OH)$	Euclase	219.71	5.85	0.72	5.28	205
$Al_2SiO_4F_2$	Topaz	215.42	0.91	0.84	-1.66	206
$LiAl(SiO_3)_2$	Spodumene	216.24	1.84	0.71	-2.33	207
$Be_3Al_2(SiO_3)_6$	Beryl	212.72	-2.21	0.00	2.88	208
$Ca_3Al_2(SiO_4)_3$	Grossular	210.42	-4.86	0.00	2.75	204
$Ca_2Al_3Si_3HO_{13}$	Zoisite	223.21	9.89	0.34	-7.65	209
$Ca_2Al_3Si_3HO_{13}$	Zoisite	222.24	8.77	0.15	-17.41	209
$NaAlSi_3O_8$	Albite	268.33	61.90	0.50	-3.29	204
$Mg_2Al_4Si_5O_{18}$	Cordierite	274.28	68.75	0.43	9.64	210
$Mg_2Al_4Si_5O_{18}$	Cordierite	256.75	48.55	0.42	-5.72	210
$Y_3Al_5O_{12}$	Y, Al-garnet	209.51	-5.91	0.00	0.72	211
$Y_3Al_5O_{12}$	Y, Al-garnet	280.15	75.53	0.00	6.12	211
$BeAl_2O_4$	Chrysoberyl	223.80	10.56	0.92	-2.56	212
$BeAl_2O_4$	Chrysoberyl	232.07	20.10	0.91	-3.37	212
$Al_4Si_4O_{10}(OH)_8$	Kaolinite	221.01	7.35	0.69	-3.52	162
$Mg_3Al_2(SiO_4)_3$	Pyrope	215.74	1.27	0.00	0.71	213
$Ca_3Al_2(SiO_4)_3$	Grossular	210.42	-4.86	0.00	2.75	213
$(CaO)_{12}(Al_2O_3)_7$	$(CaO)_{12}(Al_2O_3)_7$	279.76	75.08	0.17	-10.35	165
$(CaO)_{12}(Al_2O_3)_7$	$(CaO)_{12}(Al_2O_3)_7$	280.52	75.95	0.37	-4.38	165
Al_2SiO_5	Andalusite	244.94	34.93	0.67	4.54	214
Al_2SiO_5	Andalusite	223.33	10.02	0.17	14.93	214
$CaO\ (Al_2O_3)_2$	Calcium dialuminate	281.56	77.15	0.87	5.75	165
$CaO\ (Al_2O_3)_2$	Calcium dialuminate	277.10	72.00	0.82	-8.72	165
$(CaO)(Al_2O_3)$	Krotite	287.35	83.83	0.77	2.27	165
$(CaO)_4(Al_2O_3)_3$	Tetracalcium trialuminate	295.41	93.11	0.00	4.41	165
Al_2O_3	α-Al_2O_3	229.71	17.38	0.00	2.07	215
$LiAlO_2$	α-$LiAlO_2$	232.78	20.91	0.00	2.77	216
$LiAlO_2$	γ-$LiAlO_2$	287.72	84.25	0.72	3.33	216
Al_2SiO_5	Sillimanite	270.17	64.02	0.53	-6.09	217
Al_2SiO_5	Sillimanite	218.19	4.10	0.62	-8.72	217
Al_2SiO_5	Andalusite	244.94	34.93	0.67	4.54	218
Al_2SiO_5	Andalusite	223.33	10.02	0.17	14.93	218
Al_2SiO_5	Kyanite	224.89	11.82	0.30	9.98	219
Al_2SiO_5	Kyanite	217.50	3.30	0.97	3.79	219
Al_2SiO_5	Kyanite	220.52	6.79	0.46	-7.02	219

Al_2SiO_5	Kyanite	224.11	10.92	0.32	-9.86	219
$YAlO_3$	Yttrium aluminate	220.07	6.27	0.93	1.66	220
$AlVO_4$	Aluminium vanadate(V)	199.59	-17.34	0.61	1.02	221
$AlVO_4$	Aluminium vanadate(V)	232.43	20.51	0.70	6.39	221
$AlVO_4$	Aluminium vanadate(V)	205.32	-10.74	0.46	-4.70	221
$Al(OH)_3$	Gibbsite	226.07	13.18	0.85	2.20	162
$KAl_3Si_3(HO_6)_2$	Moscovite	221.82	8.28	0.96	2.99	162
$KAl_3Si_3(HO_6)_2$	Moscovite	279.73	75.04	0.58	2.87	162
$CaAl_2(Al_2Si_2)O_{10}(OH)_2$	Margarite	283.00	78.80	0.42	-5.57	162
$CaAl_2(Al_2Si_2)O_{10}(OH)_2$	Margarite	223.19	9.86	0.11	-7.17	162
$CaAl_2(Al_2Si_2)O_{10}(OH)_2$	Margarite	220.98	7.31	0.14	6.56	162

* The VASP δ_{iso} values were obtained using the linear regression referencing method, more details about this method could be found in Ref. [100].

Table B2b. CASTEP calculated ^{27}Al quadrupolar parameters, isotropic shielding and corresponding isotropic chemical shift.

CASTEP computed ^{27}Al NMR parameters						
Chemical Formula	**Structure Name**	**σ_{iso} (ppm)**	**δ_{iso}(ppm)***	**η_Q**	**C_Q(MHz)**	**Reference**
$AlPO_4$	Berlinite	509.59	45.97	0.44	3.94	187
$Al_2Ge_2O_7$	Dialuminium digermanate	521.63	32.65	0.45	-8.67	188
$LaAlGe_2O_7$	$LaAlGe_2O_7$	516.15	38.71	0.15	-7.09	188
$AlCl_3$	Aluminum chloride	538.54	13.94	0.16	0.27	189
Al_2O_3	α-Al_2O_3	539.94	12.39	0.00	2.22	190
Al_2SiO_5	Andalusite	523.46	30.62	0.64	4.82	191
Al_2SiO_5	Andalusite	547.04	4.54	0.17	15.58	191
$Ca_3Al_2O_6$	Nonacalcium tris(dialuminate)	482.02	76.48	0.48	-7.62	192
$Ca_3Al_2O_6$	Nonacalcium tris(dialuminate)	482.40	76.06	0.80	-8.28	192
$CaAlF_5$	Jakobssonite	558.00	-7.59	0.10	9.34	193
$CaO\ (Al_2O_3)_2$	Grossite	479.39	79.39	0.74	6.53	194
$CaO\ (Al_2O_3)_2$	Grossite	483.80	74.51	0.82	-8.95	194
$CaO\ Al_2O_3$	Krotite	478.06	80.86	0.52	-1.70	194
$CaO\ Al_2O_3$	Krotite	476.87	82.17	0.76	2.40	194
$CaO\ Al_2O_3$	Krotite	476.66	82.41	0.76	-2.41	194

$CaO\ Al_2O_3$	Krotite	479.43	79.34	0.50	-3.28	194
$CaO\ Al_2O_3$	Krotite	477.82	81.12	0.36	-3.67	194
$CaO\ Al_2O_3$	Krotite	476.35	82.75	0.39	3.93	194
$(CaO)_{12}\ (Al_2O_3)_7$	$(CaO)_{12}\ (Al_2O_3)_7$	478.76	80.08	0.65	-8.72	194
$(CaO)_{12}\ (Al_2O_3)_7$	$(CaO)_{12}\ (Al_2O_3)_7$	478.57	80.29	0.35	-5.09	194
Al_2SiO_5	Sillimanite	490.57	67.02	0.58	-6.35	195
Al_2SiO_5	Sillimanite	547.17	4.39	0.48	-8.56	195
$Na_8Al_6Si_6O_{24}C_{l2}$	Sodalite	494.53	62.63	0.00	-0.92	196
$Na_3KAl_4Si_4O_{16}$	Nepheline	495.30	61.78	0.86	-1.51	197
$Na_3KAl_4Si_4O_{16}$	Nepheline	501.36	55.08	0.00	0.76	197
$CaAl_2(Al_2Si_2)O_{10}(OH)_2$	Margarite	478.08	80.84	0.46	-6.07	164
$CaAl_2(Al_2Si_2)O_{10}(OH)_2$	Margarite	542.93	9.08	0.14	6.60	164
$KAl_2(AlSi_3O_{10})(F,OH)_2$	Muscovite	485.01	73.17	0.59	3.02	164
$KAl_2(AlSi_3O_{10})(F,OH)_2$	Muscovite	546.39	5.25	0.86	3.12	164
$KAlSiO_4$	Kalsilite	495.67	61.37	0.73	1.68	172
Al_2SiO_5	Kyanite	548.35	3.09	0.93	4.03	198
Al_2SiO_5	Kyanite	540.84	11.40	0.33	10.30	198
Al_2SiO_5	Kyanite	541.65	10.50	0.33	-9.70	198
Al_2SiO_5	Kyanite	545.57	6.16	0.44	-6.78	198
AlF_3	α-aluminum fluoride	566.78	-17.31	0.00	-0.15	168
AlF_3	β-aluminum fluoride	565.32	-15.69	0.19	-1.34	168
$Al(OH)_3$	Bayerite	539.67	12.69	0.32	-3.80	199
$Al(OH)_3$	Bayerite	539.67	12.69	0.83	1.73	199
$Al(OH)_3$	Gibbsite	540.15	12.16	0.36	-5.68	200
$Al(OH)_3$	Gibbsite	542.51	9.55	0.63	2.34	200
$NaAlO_2$	β-$NaAlO_2$	477.57	81.40	0.65	0.83	201
$AlTlO_2$	$AlTlO_2$	493.89	63.34	N/A	0.00	201
$KAlO_2$	Potassium aluminate	483.43	74.92	0.87	-1.25	201
$BaAl_2O_4$	Barium dialuminate	481.00	77.60	0.67	2.47	201
$Al(OH)_3$	δ-$Al(OH)_3$	539.72	12.63	0.40	3.20	200
$Al_2Si_4O_{11}$	Pyrophyllite dehydroxylate	518.61	35.99	0.59	-11.88	202
$(Al_2O_3)_9(B_2O_3)_2$	Boroaluminate	481.12	77.47	0.42	9.64	203
$(Al_2O_3)_9(B_2O_3)_2$	Boroaluminate	500.35	56.19	0.03	7.22	203
$(Al_2O_3)_9(B_2O_3)_2$	Boroaluminate	510.08	45.43	0.82	6.75	203
$(Al_2O_3)_9(B_2O_3)_2$	Boroaluminate	542.46	9.60	0.31	-5.68	203
$Na_5Al_3F_{14}$	Chiolite	552.98	-2.04	0.17	6.60	167
$Na_5Al_3F_{14}$	Chiolite	555.56	-4.89	0.00	-5.35	167

AlF_3	β-aluminum fluoride	565.32	-15.69	0.19	-1.33	167
Na_3AlF_6	Cryolite	543.75	8.18	0.89	-0.92	167
K_2NaAlF_6	Elpasolite	544.45	7.40	N/A	0.00	167
$KAlSi_3O_8$	Microcline	501.15	55.31	0.48	-3.26	178
$AlPO_4$	Berlinite	520.83	33.53	0.39	4.18	178
$ZnAl_2O_4$	Gahnite	536.51	16.19	0.00	-4.18	178
$MgAl_2O_4$	Spinel	540.11	12.20	0.00	-3.94	204
$BeAlSiO_4(OH)$	Euclase	549.23	2.11	0.86	5.12	205
$Al_2SiO_4F_2$	Topaz	552.92	-1.97	0.98	-1.67	206
$LiAl(SiO_3)_2$	Spodumene	553.08	-2.15	0.72	-2.52	207
$Be_3Al_2(SiO_3)_6$	Beryl	555.54	-4.87	0.00	3.06	208
$Ca_3Al_2(SiO_4)_3$	Grossular	555.80	-5.16	0.00	2.83	204
$Ca_2Al_3Si_3HO_{13}$	Zoisite	545.36	6.39	0.31	-8.18	209
$Ca_2Al_3Si_3HO_{13}$	Zoisite	546.57	5.06	0.15	-18.29	209
$NaAlSi_3O_8$	Albite	492.07	65.36	0.63	-3.38	204
$Mg_2Al_4Si_5O_{18}$	Cordierite	489.83	67.83	0.45	10.22	210
$Mg_2Al_4Si_5O_{18}$	Cordierite	508.08	47.64	0.42	-6.05	210
$Y_3Al_5O_{12}$	Y, Al-garnet	552.62	-1.64	0.00	0.76	211
$Y_3Al_5O_{12}$	Y, Al-garnet	482.18	76.30	0.00	6.65	211
$BeAl_2O_4$	Chrysoberyl	545.04	6.75	0.90	-2.73	212
$BeAl_2O_4$	Chrysoberyl	536.26	16.46	0.88	-3.70	212
$Al_4Si_4O_{10}(OH)_8$	Kaolinite	546.43	5.21	0.83	-3.93	162
$Mg_3Al_2(SiO_4)_3$	Pyrope	553.39	-2.49	0.00	0.70	213
$Ca_3Al_2(SiO_4)_3$	Grossular	558.78	-8.45	0.00	2.81	213
$(CaO)_{12}(Al_2O_3)_7$	$(CaO)_{12}(Al_2O_3)_7$	484.65	73.57	0.20	-11.26	165
$(CaO)_{12}(Al_2O_3)_7$	$(CaO)_{12}(Al_2O_3)_7$	482.27	76.20	0.40	-4.44	165
Al_2SiO_5	Andalusite	523.46	30.62	0.64	4.82	214
Al_2SiO_5	Andalusite	547.04	4.54	0.17	15.58	214
$CaO\ (Al_2O_3)_2$	Calcium dialuminate	480.95	77.66	0.90	6.04	165
$CaO\ (Al_2O_3)_2$	Calcium dialuminate	486.46	71.56	0.83	-9.48	165
$(CaO)(Al_2O_3)$	Krotite	476.87	82.17	0.76	2.40	165
$(CaO)_4(Al_2O_3)_3$	Tetracalcium trialuminate	471.42	88.20	0.00	5.95	165
Al_2O_3	α-Al_2O_3	536.66	16.02	0.00	2.28	215
$LiAlO_2$	α-$LiAlO_2$	532.73	20.37	0.00	3.01	216
$LiAlO_2$	γ-$LiAlO_2$	471.25	88.39	0.70	3.74	216
Al_2SiO_5	Sillimanite	490.57	67.02	0.58	-6.35	217
Al_2SiO_5	Sillimanite	547.17	4.39	0.48	-8.56	217
Al_2SiO_5	Andalusite	523.46	30.62	0.64	4.82	218

Al_2SiO_5	Andalusite	547.04	4.54	0.17	15.58	218
Al_2SiO_5	Kyanite	540.84	11.40	0.33	10.30	219
Al_2SiO_5	Kyanite	548.35	3.09	0.93	4.03	219
Al_2SiO_5	Kyanite	545.57	6.16	0.44	-6.78	219
Al_2SiO_5	Kyanite	541.65	10.50	0.33	-9.70	219
$YAlO_3$	Yttrium aluminate	543.42	8.54	0.99	1.81	220
$AlVO_4$	Aluminium vanadate(V)	566.88	-17.42	0.49	1.50	221
$AlVO_4$	Aluminium vanadate(V)	532.88	20.20	0.69	6.83	221
$AlVO_4$	Aluminium vanadate(V)	560.84	-10.73	0.35	-4.65	221
$Al(OH)_3$	Gibbsite	542.51	9.55	0.63	2.34	162
$KAl_3Si_3(HO_6)_2$	Moscovite	546.39	5.25	0.86	3.12	162
$KAl_3Si_3(HO_6)_2$	Moscovite	485.01	73.17	0.59	3.02	162
$CaAl_2(Al_2Si_2)O_{10}(OH)_2$	Margarite	478.08	80.84	0.46	-6.07	162
$CaAl_2(Al_2Si_2)O_{10}(OH)_2$	Margarite	540.39	11.89	0.13	-7.74	162
$CaAl_2(Al_2Si_2)O_{10}(OH)_2$	Margarite	542.93	9.08	0.14	6.60	162

* The CASTEP δ_{iso} values were obtained using the linear regression referencing method, more details about this method could be found in Ref. [100].

Curriculum vitae

SUN, He (Jason)

Tel.: 1-3146828433 **E-mail:** he.sun@wustl.edu **Github:** github.com/wushanyun64

Linkedin: linkedin.com/in/he-jason-sun-574350136

Education and certifications

08/2016-07/2022 Washington University in St. Louis (WUSTL), Computational Chemistry, Ph.D.

09/2012-06/2016 University of Science and Technology of China (USTC), Physical Chemistry, B.S.

11/2021 Neural Networks and Deep Learning (Coursera Certification)

Research Projects

08/2016-Present Research Assistant/Ph.D. Candidate, Washington University in St. Louis

01/2018-08/2021 Structural Investigation of Silver Vanadium Phosphorus Oxide ($Ag_2VO_2PO_4$) and its Reduction Products

- Performed detailed spectroscopic analysis of a potential cathode material ($Ag_2VO_2PO_4$) and related discharged materials with solid-state NMR and XRD.
- Applied quantum chemistry simulations (CASTEP), obtained an optimized structure model via a comparison between theory and experiment, quantitively confirmed local structural motif distortions of $Ag_2VO_2PO_4$ due to the amorphization during the discharging process.

08/2018-01/2022 Machine learning prediction of Nuclear Magnetic Resonance (NMR) spectrum from simple local geometry and chemistry.

- Ingested and stored relevant research data by leveraging an open-source API and relevant Python libraries.
- Performed data wrangling / ETL and exploratory data analysis (EDA) to build insights for data modeling.
- Created novel engineered features utilizing domain knowledge and building into modeling pipeline, increasing model performance from R^2 = 0.93 to R^2=0.98.
- Leveraging the data pipeline and engineered features, developed and tested machine learning models (random forest regression, XGboost) that outperformed published state-of-the-art (SOTA) models (R^2~0.91) for predicting spectroscopic properties.

- Ensured the quality of training data with rigorous benchmarking study between quantum chemistry calculations and experimental data for over 3000 chemical compounds.

01/2020-03/2022 Reliability analysis of spectrum fitting model with Markov Chain Monte Carlo

- Based on Bayes' theorem and Markov Chain Monte Carlo (MCMC), developed a statistics module using python to access the reliability of model parameters for computational simulation versus experimental spectrum.
- The module is to be used as a submodule in an open-source NMR fitting package Mrsimulator.

09/2017-05/2021 Enabling materials informatics for ^{29}Si solid-state NMR of crystalline materials.

- Leveraged quantum chemistry approaches to construct a data infrastructure for silicon materials and confirm the validity of computational methods. (https://contribs.materialsproject.org/)
- Determined the optimal way of expressing spectroscopic data with statistical error analysis.
- Systematically corrected the machine-generated data via extensive benchmarking against experimental results. Proved a conceptual error in the popular quantum chemistry package VASP and improved the information reliability on the VASP wiki. (https://www.vasp.at/wiki/index.php/LCHIMAG)

List of Publications of He (Jason) Sun

2022 Sun, H., West, M., Dwaraknath, S., Ling, H. *et al.* ^{27}Al NMR benchmarking with DFT and fast prediction of quadrupolar coupling constant from simple local geometry. (pre-submit)

2021 Sun, H; Hammann, B; Brady, A; Singh, G; Housel, L; Takeuchi, E; Takeuchi, K; Marschilok, A; Hayes, S; Szczepura, L. Structural Investigation of Silver Vanadium Phosphorus Oxide ($Ag_2VO_2PO_4$) and its Reduction Products. *Chem Mater* (2021). doi: 10.1021/acs.chemmater.1c00446

2021 Cendejas A J, **Sun H**, Hayes S E, et al. Predicting Plasma Conditions Necessary for Synthesis of γ-Al 2 O 3 Nanocrystals[J]. Nanoscale, 2021. doi: 10.1039/D1NR02488D

2020 Sun, H., Dwaraknath, S., Ling, H. *et al.* Enabling materials informatics for ^{29}Si solid-state NMR of crystalline materials. *npj Comput Mater* **6,** 53 (2020). doi: 10.1038/s41524-020-0328-3

2020 Zahan, M., **Sun, H**., Hayes, S. E., Krautscheid, H., Haase, J., & Bertmer, M. (2020). Influence of Alkali Metal Cations on the Photodimerization of Bromo Cinnamates Studied by Solid-State NMR. *The Journal of Physical Chemistry C.* doi: 10.1021/acs.jpcc.0c09826

2020 Malone, M. W., Espy, M. A., **He, S**., Janicke, M. T., & Williams, R. F. (2020). The 1H T1 dispersion curve of fentanyl citrate to identify NQR parameters. *Solid State Nuclear Magnetic Resonance*, 110, 101697. doi: 10.1016/j.ssnmr.2020.101697

List of Presentations of He (Jason) Sun

04/25/2022 He Sun, ^{27}Al NMR quadrupolar and chemical shielding tensors benchmarking with DFT: prediction of quadrupolar coupling constants (C_Q) from simple local geometry and elemental properties. ENC conference 2022. (Oral)

2021 He Sun, Enabling Materials Informatics for ^{29}Si Solid-state NMR of Crystalline Materials. ENC conference. (Poster)

2014 He Sun, Caulobacter Crescentus Multiple Color Displayer, iGEM Summer Camp 2014, NCTU (Oral)

Reference

1. Flygare, W. H. & Benson, R. C. The molecular zeeman effect in diamagnetic molecules and the determination of molecular magnetic moments (G values), magnetic susceptibilities, and molecular quadrupole moments. *Mol. Phys.* **20**, (1971).

2. Gan, Z. *et al.* NMR spectroscopy up to 35.2 T using a series-connected hybrid magnet. *J. Magn. Reson.* **284**, (2017).

3. Masion, A., Alexandre, A., Ziarelli, F., Viel, S. & Santos, G. M. Dynamic Nuclear Polarization NMR as a new tool to investigate the nature of organic compounds occluded in plant silica particles. *Sci. Rep.* **7**, (2017).

4. *NMR Basic Principles and Progress / NMR Grundlagen und Fortschritte*. *NMR Basic Principles and Progress / NMR Grundlagen und Fortschritte* (1972). doi:10.1007/978-3-642-65312-4

5. *Solid-State NMR Spectroscopy Principles and Applications*. *Solid-State NMR Spectroscopy Principles and Applications* (2007). doi:10.1002/9780470999394

6. Harris, R. K. *et al.* Further conventions for NMR shielding and chemical shifts (IUPAC recommendations 2008): international union of pure and applied chemistry physical and biophysical chemistry division. *Magn. Reson. Chem.* **46**, 582–598 (2008).

7. Waugh, J. S. (John S. *Advances in magnetic resonance.* (Academic Press, 1965).

8. Mason, J. Conventions for the reporting of nuclear magnetic shielding (or shift) tensors suggested by participants in the NATO ARW on NMR Shielding Constants at the University of Maryland, College Park, July 1992. *Solid State Nucl. Magn. Reson.* **2**, 285–288 (1993).

9. Herzfeld, J. & Berger, A. E. Sideband intensities in NMR spectra of sample spinning at the magic angle. *J. Chem. Phys.* **73**, 6021–6030 (1980).

10. Massiot, D. *et al.* Modelling one- and two-dimensional solid-state NMR spectra. *Magn. Reson. Chem.* **40**, 70–76 (2002).

11. Bak, M., Rasmussen, J. T. & Nielsen, N. C. SIMPSON: A general simulation program for solid-state NMR spectroscopy. *J. Magn. Reson.* **213**, 366–400 (2011).

12. Bonhomme, C. *et al.* First-Principles Calculation of NMR Parameters Using the Gauge Including Projector Augmented Wave Method: A Chemist's Point of View. (2012). doi:10.1021/cr300108a

13. Carlomagno, T. NMR in natural products: Understanding conformation, configuration and receptor interactions. *Natural Product Reports* **29**, (2012).

14. Lesage, A. Recent advances in solid-state NMR spectroscopy of spin I = 1/2 nuclei. *Phys. Chem. Chem. Phys.* **11**, (2009).

15. Bryce, D. L. NMR crystallography: Structure and properties of materials from solid-state nuclear magnetic resonance observables. *IUCrJ* **4**, (2017).

16. Mueller, L. J. & Dunn, M. F. NMR crystallography of enzyme active sites: Probing chemically detailed, three-dimensional structure in tryptophan synthase. *Acc. Chem. Res.* **46**, (2013).

17. Harris, R. K. NMR studies of organic polymorphs & solvates. *Analyst* **131**, (2006).

18. Bryce, D. L. New frontiers for solid-state NMR across the periodic table: A snapshot of modern techniques and instrumentation. *Dalt. Trans.* **48**, (2019).

19. Zhao, L., Pinon, A. C., Emsley, L. & Rossini, A. J. DNP-enhanced solid-state NMR spectroscopy of active pharmaceutical ingredients. *Magnetic Resonance in Chemistry* **56**, (2018).

20. Urbanová, M., Czernek, J. & Brus, J. Structural characterization of the new generation of drugs and biomaterials: Application of nmr crystallography. *Chem. List.* **112**, (2018).

21. Radoicic, J., Lu, G. J. & Opella, S. J. NMR structures of membrane proteins in phospholipid bilayers. *Quarterly Reviews of Biophysics* **47**, (2014).

22. Hamilton, J. A. NMR reveals molecular interactions and dynamics of fatty acid binding to albumin. *Biochim. Biophys. Acta - Gen. Subj.* **1830**, (2013).

23. Franks, W. T., Linden, A. H., Kunert, B., van Rossum, B. J. & Oschkinat, H. Solid-state magic-angle spinning NMR of membrane proteins and protein-ligand interactions. *European Journal of Cell Biology* **91**, (2012).

24. Martineau-Corcos, C. NMR Crystallography: A tool for the characterization of microporous hybrid solids. *Current Opinion in Colloid and Interface Science* **33**, (2018).

25. Elena, B., Pintacuda, G., Mifsud, N. & Emsley, L. Molecular structure determination in powders by NMR crystallography from proton spin diffusion. *J. Am. Chem. Soc.* **128**, (2006).

26. Elena, B. & Emsley, L. Powder crystallography by proton solid-state NMR spectroscopy. *J. Am. Chem. Soc.* **127**, (2005).

27. Taulelle, F. NMR crystallography: Crystallochemical formula and space group selection. *Solid State Sci.* **6**, (2004).

28. Yates, J. M. *et al.* A combined first principles computational and solid-state NMR study of a molecular crystal: Flurbiprofen. in *Physical Chemistry Chemical Physics* **7**, (2005).

29. Bonhomme, C. *et al.* ChemInform Abstract: First-Principles Calculation of NMR Parameters Using the Gauge Including Projector Augmented Wave Method: A Chemist′s Point of View. *ChemInform* **44**, (2013).

30. Yates, J. R., Pickard, C. J. & Mauri, F. Calculation of NMR chemical shifts for extended systems using ultrasoft pseudopotentials. *Phys. Rev. B - Condens. Matter Mater. Phys.* **76**, (2007).

31. Gervais, C. *et al.* First principles NMR calculations of phenylphosphinic acid C6H5HPO(OH): Assignments, orientation of tensors by local field experiments and effect of molecular motion. *J. Magn. Reson.* **187**, (2007).

32. Harris, R. K. *et al.* Solid-state NMR and computational studies of 4-methyl-2-nitroacetanilide. *Magn. Reson. Chem.* **44**, (2006).

33. Middleton, D. A. *et al.* Conformational analysis by solid-state NMR and its application to restrained structure determination from powder diffraction data. *Chem. Commun.* (2002). doi:10.1039/b204289d

34. Harris, R. K. Applications of solid-state NMR to pharmaceutical polymorphism and related matters. *J. Pharm. Pharmacol.* **59**, (2010).

35. Xu, Y., Viger-Gravel, J., Korobkov, I. & Bryce, D. L. Mechanochemical Production of Halogen-Bonded Solids Featuring P=O···I-C Motifs and Characterization via X-ray Diffraction, Solid-State Multinuclear Magnetic Resonance, and Density Functional Theory. *J. Phys. Chem. C* **119**, (2015).

36. Baias, M. *et al.* Powder crystallography of pharmaceutical materials by combined crystal structure prediction and solid-state 1H NMR spectroscopy. *Phys. Chem. Chem. Phys.* **15**, (2013).

37. Reilly, A. M. *et al.* Report on the sixth blind test of organic crystal structure prediction methods. *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.* **72**, (2016).

38. Day, G. M. & Cooper, T. G. Crystal packing predictions of the alpha-amino acids: Methods assessment and structural observations. *CrystEngComm* **12**, (2010).

39. Day, G. M., Motherwell, W. D. S. & Jones, W. A strategy for predicting the crystal structures of flexible molecules: The polymorphism of phenobarbital. *Phys. Chem. Chem. Phys.* **9**, (2007).

40. Harper, J. K. & Grant, D. M. Enhancing crystal-structure prediction with NMR tensor data. *Cryst. Growth Des.* **6**, (2006).

41. Meejoo, S. *et al.* Structural aspects of the β-polymorph of (E)-4-formylcinnamic acid: Structure determination directly from powder diffraction data and elucidation of structural disorder from solid-state NMR. *Helv. Chim. Acta* **86**, (2003).

42. McGeorge, G., Harris, R. K., Chippendale, A. M. & Bullock, J. F. Conformational analysis by magic-angle spinning NMR spectroscopy for a series of polymorphs of a disperse azobenzene dyestuff. *J. Chem. Soc. Perkin Trans. 2* **8**, (1996).

43. McGeorge, G. *et al.* Analysis of a solid-state conformational rearrangement using 15N NMR and X-ray crystallography. *J. Phys. Chem. A* **102**, (1998).

44. Harris, R. K. *et al.* Cross-polarization/magic-angle spinning NMR studies of polymorphism: Cortisone acetate. *Spectrochim. Acta Part A Mol. Spectrosc.* **46**, (1990).

45. Harris, R. K., Say, B. J., Yeung, R. R., Fletton, R. A. & Lancaster, R. W. Cross-polarization/magic-angle spinning NMR studies of polymorphism: Androstanolone. *Spectrochim. Acta Part A Mol. Spectrosc.* **45**, (1989).

46. Fletton, R. A. *et al.* A comparative spectroscopic Investigation of two polymorphs of 4′-methyl-2′-nitroacetanilide using solid-state infrared and high-resolution solid-state nuclear magnetic resonance spectroscopy. *J. Chem. Soc. Perkin Trans. 2* (1986). doi:10.1039/P29860001705

47. Harris, R. K. NMR crystallography: The use of chemical shifts. *Solid State Sci.* **6**, (2004).

48. Apperley, D., Harris, R. & Hodgkinson, P. *Solid-State NMR: Basic Principles & Practice*. *Solid-State NMR: Basic Principles & Practice* (2012). doi:10.5643/9781606503522

49. Ashbrook, S. E. & Dawson, D. M. Exploiting periodic first-principles calculations in NMR spectroscopy of disordered solids. *Acc. Chem. Res.* **46**, (2013).

50. Charpentier, T. The PAW/GIPAW approach for computing NMR parameters: A new dimension added to NMR study of solids. *Solid State Nuclear Magnetic Resonance* **40**, (2011).

51. Paruzzo, F. M. *et al.* Chemical shifts in molecular solids by machine learning. *Nat. Commun.* **9**, 4501 (2018).

52. Hu, J., Huang, W., Yang, L. & Pan, F. Structure and performance of the LiFePO4cathode material: From the bulk to the surface. *Nanoscale* **12**, (2020).

53. Yang, Z., Dai, Y., Wang, S. & Yu, J. How to make lithium iron phosphate better: a review exploring classical modification approaches in-depth and proposing future optimization methods. *Journal of Materials Chemistry A* **4**, (2016).

54. Padhi, A. K., Nanjundaswamy, K. S. & Goodenough, J. B. Phospho-olivines as Positive-Electrode Materials for Rechargeable Lithium Batteries. *J. Electrochem. Soc.* **144**, (1997).

55. Zhang, W. J. Structure and performance of LiFePO4 cathode materials: A review. *Journal of Power Sources* **196**, (2011).

56. Chung, S. Y., Bloking, J. T. & Chiang, Y. M. Electronically conductive phospho-olivines

as lithium storage electrodes. *Nat. Mater.* **1**, (2002).

57. Yu, F. *et al.* Mechanism studies of LiFePO4 cathode material: Lithiation/delithiation process, electrochemical modification and synthetic reaction. *RSC Advances* **4**, (2014).

58. Zaghib, K. *et al.* Safe Li-ion polymer batteries for HEV applications. *J. Power Sources* **134**, (2004).

59. Gao, L., Xu, Z., Zhang, S., Xu, J. & Tang, K. Enhanced electrochemical properties of LiFePO4 cathode materials by Co and Zr multi-doping. *Solid State Ionics* **305**, (2017).

60. Jiang, W., Wu, M., Liu, F., Yang, J. & Feng, T. Variation of carbon coatings on the electrochemical performance of LiFePO4 cathodes for lithium ionic batteries. *RSC Adv.* **7**, (2017).

61. Yi, T. F. *et al.* Recent developments in the doping and surface modification of LiFePO 4 as cathode material for power lithium ion battery. *Ionics* **18**, (2012).

62. Stenina, I. A. & Yaroslavtsev, A. B. Nanomaterials for lithium-ion batteries and hydrogen energy. in *Pure and Applied Chemistry* **89**, (2017).

63. Tracey, A. S., Willsky, G. R. & Takeuchi, E. S. *Vanadium: Chemistry, Biochemistry, Pharmacology and Practical Applications*. *Comprehensive Inorganic Chemistry II (Second Edition): From Elements to Applications* (2007).

64. Takeuchi, K. J., Marschilok, A. C., Davis, S. M., Leising, R. A. & Takeuchi, E. S. Silver vanadium oxides and related battery applications. *Coordination Chemistry Reviews* **219**–**221**, (2001).

65. Gleason, N. R., Leising, R. A., Palazzo, M., Takeuchi, E. S. & Takeuchi, K. J. Talk\# 248-Microscopic Study of the First Voltage Plateau in the Discharge of SVO and the Consequences on Electronic Conductivity. in *208th Meeting of the Electrochemical Society. Session D2-Rechargeable Lithium and Lithium-Ion Batteries-Battery/Energy Technology* (2005).

66. Anguchamy, Y. K., Lee, J. W. & Popov, B. N. Electrochemical performance of polypyrrole/silver vanadium oxide composite cathodes in lithium primary batteries. *J. Power Sources* **184**, (2008).

67. Morcrette, M. *et al.* A reversible copper extrusion-insertion electrode for rechargeable Li batteries. *Nat. Mater.* **2**, (2003).

68. Morcrette, M. *et al.* Cu1.1V4O11: A new positive electrode material for rechargeable Li batteries. *Chem. Mater.* **17**, (2005).

69. Kang, H. Y., Wang, S. L., Tsai, P. P. & Lii, K. H. Hydrothermal synthesis, crystal structure and ionic conductivity of Ag2VO2PO4: A new layered phosphate of vanadium(V). *J. Chem. Soc. Dalt. Trans.* (1993). doi:10.1039/DT9930001525

70. Patridge, C. J. *et al.* An X-ray absorption spectroscopy study of the cathodic discharge of Ag 2VO2PO4: Geometric and electronic structure characterization of intermediate phases and mechanistic insights. *J. Phys. Chem. C* **115**, (2011).

71. Marschilok, A. C. *et al.* Electrochemical reduction of silver vanadium phosphorous oxide, Ag2VO2PO4: Silver metal deposition and associated increase in electrical conductivity. *J. Power Sources* **195**, (2010).

72. Takeuchi, E. S. *et al.* Electrochemical reduction of silver vanadium phosphorus oxide, Ag 2VO2PO4: The formation of electrically conductive metallic silver nanoparticles. *Chem. Mater.* **21**, (2009).

73. Zhang, R. *et al.* Electrode Reaction Mechanism of Ag2VO2PO4 Cathode. *Chem. Mater.* **28**, (2016).

74. Marschilok, A. C., Takeuchi, K. J. & Takeuchi, E. S. Preparation and electrochemistry of silver vanadium phosphorous oxide, Ag2 v O2 P O4. *Electrochem. Solid-State Lett.* **12**, (2008).

75. Dupree, R. & Ford, C. J. Magnetic susceptibility of the noble metals around their melting points. *Physical Review B* **8**, (1973).

76. Chamberland, B. L. & Porter, S. K. A study on the preparation and physical property determination of NaVO2. *J. Solid State Chem.* **73**, (1988).

77. Nath, R., Tsirlin, A. A., Rosner, H. & Geibel, C. Magnetic properties of BaCdVO (PO4) 2: A strongly frustrated spin- 1 2 square lattice close to the quantum critical regime. *Phys. Rev. B - Condens. Matter Mater. Phys.* **78**, (2008).

78. Lapina, O. B., Khabibulin, D. F., Shubin, A. A. & Terskikh, V. V. Practical aspects of 51V and 93Nb solid-state NMR spectroscopy and applications to oxide materials. *Progress in Nuclear Magnetic Resonance Spectroscopy* **53**, (2008).

79. Lapina, O. B. *et al.* Solid-state 51V NMR for characterization of vanadium-containing systems. in *Catalysis Today* **78**, (2003).

80. Pooransingh-Margolis, N. *et al.* 51V solid-state magic angle spinning NMR spectroscopy of vanadium chloroperoxidase. *J. Am. Chem. Soc.* **128**, (2006).

81. Gupta, R., Yehl, J., Li, M. & Polenova, T. 51V magic angle spinning NMR spectroscopy and quantum chemical calculations in vanadium bio-inorganic systems: Current perspective. *Canadian Journal of Chemistry* **93**, (2015).

82. Schwendt, P., Tatiersky, J., Krivosudský, L. & Šimuneková, M. Peroxido complexes of vanadium. *Coordination Chemistry Reviews* **318**, (2016).

83. Conte, V., Bortolini, O., Carraro, M. & Moro, S. Models for the active site of vanadium-dependent haloperoxidases: Insight into the solution structure of peroxo vanadium

compounds. *J. Inorg. Biochem.* **80**, (2000).

84. Rehder, D. The (biological) speciation of vanadate(V) as revealed by 51V NMR: A tribute on Lage Pettersson and his work. *J. Inorg. Biochem.* **147**, (2015).

85. Rehder, D., Polenova, T. & Bühl, M. Vanadium-51 NMR. *Annual Reports on NMR Spectroscopy* **62**, (2007).

86. Toby, B. H. & Von Dreele, R. B. GSAS-II: The genesis of a modern open-source all purpose crystallography software package. *J. Appl. Crystallogr.* **46**, (2013).

87. Kunwar, A. C., Turner, G. L. & Oldfield, E. Solid-state spin-echo Fourier transform NMR of 39K and 67Zn salts at high field. *J. Magn. Reson.* **69**, (1986).

88. Segall, M. D. *et al.* First-principles simulation: ideas, illustrations and the CASTEP code. *J. Phys. Condens. Matter* **14**, 2717–2744 (2002).

89. Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, (1965).

90. Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **136**, (1964).

91. Jain, A. *et al.* Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Materials* **1**, (2013).

92. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).

93. Perdew, J. P., Burke, K. & Ernzerhof, M. Perdew, Burke, and Ernzerhof Reply. *Physical Review Letters* **80**, (1998).

94. Hammer, B., Hansen, L. B. & Nørskov, J. K. Improved adsorption energetics within density-functional theory using revised Perdew-Burke-Ernzerhof functionals. *Phys. Rev. B - Condens. Matter Mater. Phys.* **59**, (1999).

95. Pickard, C. J. & Mauri, F. All-electron magnetic response with pseudopotentials: NMR chemical shifts. doi:10.1103/PhysRevB.63.245101

96. Profeta, M., Mauri, F. & Pickard, C. J. Accurate first principles prediction of 17O NMR parameters in SiO2: Assignment of the zeolite ferrierite spectrum. *J. Am. Chem. Soc.* **125**, (2003).

97. De Jong, M. *et al.* Charting the complete elastic properties of inorganic crystalline compounds. *Sci. Data* **2**, (2015).

98. Bock, D. C., Marschilok, A. C., Takeuchi, K. J. & Takeuchi, E. S. A kinetics and equilibrium study of vanadium dissolution from vanadium oxides and phosphates in battery electrolytes: Possible impacts on ICD battery performance. *J. Power Sources* **231**,

(2013).

99. Clark, S. J. *et al.* First principles methods using CASTEP. *Zeitschrift fur Krist.* **220**, (2005).

100. Sun, H. *et al.* Enabling materials informatics for 29Si solid-state NMR of crystalline materials. *npj Comput. Mater.* **6**, (2020).

101. Huang, W. *et al.* 51V magic angle spinning NMR spectroscopy of Keggin anions [PVnW12-nO40](3+n)-: Effect of countercation and vanadium substitution on fine structure constants. *J. Am. Chem. Soc.* **126**, (2004).

102. Huang, W., Todaro, L., Francesconi, L. C. & Polenova, T. 51V magic angle spinning NMR spectroscopy of six-coordinate lindqvist oxoanions: A sensitive probe for the electronic environment in vanadium-containing polyoxometalates. Counterions dictate the 51V fine structure constants in polyoxometalate solids. *J. Am. Chem. Soc.* **125**, (2003).

103. Nakamura, K. *et al.* 7Li and 51V NMR study on Li+ ionic diffusion in lithium intercalated LixV2O5. *Solid State Ionics* **177**, (2006).

104. Ooms, K. *et al.* Identification of mixed valence vanadium in ETS-10 using electron paramagnetic resonance, 51v solid-state nuclear magnetic resonance, and density functional theory studies. *J. Phys. Chem. C* **113**, (2009).

105. Leifer, N. D. *et al.* Nuclear Magnetic Resonance and X-Ray Absorption Spectroscopic Studies of Lithium Insertion in Silver Vanadium Oxide Cathodes. *J. Electrochem. Soc.* **154**, (2007).

106. Simons, W. W. & Laboratories., S. R. The Sadtler guide to carbon-13 NMR spectra. 652 p. (1983).

107. Pouchert, C. J. & Behnke, J. *The Aldrich library of 13C [hoch 13 C] and 1H [hoch 1 H] FT NMR spectra*. (Aldrich Chemical Comp., 1993).

108. SDBSWeb. Available at: https://sdbs.db.aist.go.jp.

109. Rosman, K. J. R. & Taylor, P. D. P. Isotopic compositions of the elements 1997 (Technical Report). *Pure Appl. Chem.* **70**, 217–235 (1998).

110. Sklenak, S. *et al.* Aluminum Siting in Silicon-Rich Zeolite Frameworks: A Combined High-Resolution27Al NMR Spectroscopy and Quantum Mechanics / Molecular Mechanics Study of ZSM-5. *Angew. Chemie Int. Ed.* **46**, 7286–7289 (2007).

111. Cui, J., Olmsted, D. L., Mehta, A. K., Asta, M. & Hayes, S. E. NMR Crystallography: Evaluation of Hydrogen Positions in Hydromagnesite by ^{13}C{^{1}H} REDOR Solid-State NMR and Density Functional Theory Calculation of Chemical Shielding Tensors. *Angew. Chemie Int. Ed.* **58**, 4210–4216 (2019).

112. Salager, E. *et al.* Powder Crystallography by Combined Crystal Structure Prediction and High-Resolution 1 H Solid-State NMR Spectroscopy. *J. AM. CHEM. SOC* **132**, 2564–2566 (2010).

113. Perras, ric A. & Bryce, D. L. Multinuclear Magnetic Resonance Crystallographic Structure Refinement and Cross-Validation Using Experimental and Computed Electric Field Gradients: Application to Na. (2012). doi:10.1021/jp308273h

114. Harris, R. K. *et al.* Computation and NMR crystallography of terbutaline sulfate. *Magn. Reson. Chem.* **48**, S103–S112 (2010).

115. Ashbrook, S. E. & McKay, D. Combining solid-state NMR spectroscopy with first-principles calculations-a guide to NMR crystallography. *Chem. Commun.* **52**, (2016).

116. Martineau, C., Senker, J. & Taulelle, F. NMR Crystallography. *Annu. Reports NMR Spectrosc.* **82**, 1–57 (2014).

117. Brouwer, D. H. NMR Crystallography of Zeolites: Refinement of an NMR-Solved Crystal Structure Using ab Initio Calculations of ^{29} Si Chemical Shift Tensors. *J. Am. Chem. Soc.* **130**, 6306–6307 (2008).

118. Vyalikh, A. *et al.* Fluorine Patterning in Room-Temperature Fluorinated Graphite Determined by Solid-State NMR and DFT. (2013). doi:10.1021/jp4028029

119. Zhou, L. *et al.* Interaction between Histidine and Zn(II) Metal Ions over a Wide pH as Revealed by Solid-State NMR Spectroscopy and DFT Calculations. *J. Phys. Chem. B* **117**, 31 (2013).

120. Finkelstein-Shapiro, D. *et al.* Direct Evidence of Chelated Geometry of Catechol on TiO 2 by a Combined Solid-State NMR and DFT Study. (2016). doi:10.1021/acs.jpcc.6b08041

121. Karen Ann Smith. *High-resolution silicon-29 nuclear magnetic resonance spectroscopic study of rock-forming silicates*. *American Mineralogist* **68**, (1983).

122. Spearing, D. R. & Stebbins, J. F. The 29Si NMR shielding tensor in low qaafiz. *Am. Mineral.* **74**, 956–959 (1989).

123. Brouwer, D. H., Moudrakovski, I. L., Darton, R. J. & Morris, R. E. Comparing quantum-chemical calculation methods for structural investigation of zeolite crystal structures by solid-state NMR spectroscopy. *Magn. Reson. Chem.* **48**, S113–S121 (2010).

124. Hansen Michael Ryan, Jakobsen Hans J. & Skibsted, J. 29Si Chemical Shift Anisotropies in Calcium Silicates from High-Field 29Si MAS NMR Spectroscopy. *Inorg. Chem.* **42**, 2368–2377 (2003).

125. Weiden, N. & Rager, H. The Chemical Shift of the 29Si Nuclear Magnetic Resonance in a Synthetic Single Crystal of Mg2SiO4. *Zeitschrift für Naturforsch. A* **40**, 126–130 (1985).

126. Asakura, T. *et al.* Intermolecular Packing in B. mori Silk Fibroin: Multinuclear NMR Study of the Model Peptide (Ala-Gly)15 Defines a Heterogeneous Antiparallel Antipolar Mode of Assembly in the Silk II Form. *Macromol. (Washington, DC, United States)* **48**, 28–36 (2015).

127. Ukmar, T., Kaucic, V. & Mali, G. Solid-state NMR spectroscopy and first-principles calculations: a powerful combination of tools for the investigation of polymorphism of indomethacin. *Acta Chim. Slov.* **58**, 425–433 (2011).

128. Reader, S. W. *et al.* Cation Disorder in Pyrochlore Ceramics: 89 Y MAS NMR and First-Principles Calculations. doi:10.1021/jp906764e

129. Mitchell, M. R. *et al.* 19 Sn MAS NMR and first-principles calculations for the investigation of disorder in stannate pyrochlores. *This J. is c Own. Soc.* **13**, 488–497 (2011).

130. Arnold, A. A. *et al.* Structure of NaYF4 Upconverting Nanoparticles: A Multinuclear Solid-State NMR and DFT Computational Study. *J. Phys. Chem. C* **117**, 25733–25741 (2013).

131. Martineau, C. *et al.* Structural investigation of α-LaZr2F11 by coupling X-ray powder diffraction, 19F solid state NMR and DFT calculations. *J. Solid State Chem.* **199**, 326–333 (2013).

132. Griffin, J. M. *et al.* 77Se solid-state NMR of inorganic and organoselenium systems: A combined experimental and computational study. *J. Phys. Chem. C* **115**, 10859–10872 (2011).

133. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).

134. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).

135. Kresse, G. & Hafner, J. *Ab. initio molecular dynamics for liquid metals*. **47**,

136. Kresse, G. & Furthmüller, J. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).

137. Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50 (1996).

138. Perdew, J. P., Burke, K. & Ernzerhof, M. *Generalized Gradient Approximation Made Simple*. (1996).

139. Lodewyk, M. W., Siebert, M. R. & Tantillo, D. J. Computational Prediction of ^{1}H and ^{13}C Chemical Shifts: A Useful Tool for Natural Product, Mechanistic, and Synthetic Organic Chemistry. *Chem. Rev.* **112**, 1839–1862 (2012).

140. Hartman, J. D., Kudla, R. A., Day, G. M., Mueller, L. J. & Beran, G. J. O. Benchmark fragment-based ^{1}H, ^{13}C, ^{15}N and ^{17}O chemical shift predictions in molecular crystals. *Phys. Chem. Chem. Phys.* **18**, 21686–21709 (2016).

141. Brouwer, D. H. . & Enright, G. D. Probing Local Structure in Zeolite Frameworks: Ultrahigh-Field NMR Measurements and Accurate First-Principles Calculations of Zeolite 29Si Magnetic Shielding Tensors. (2008). doi:10.1021/JA077430A

142. Young, R. P. *et al.* TensorView: A software tool for displaying NMR tensors. *Magn. Reson. Chem.* **57**, 211–223 (2019).

143. Xu, J., Wang, Q. & Deng, F. Metal Active Sites and Their Catalytic Functions in Zeolites: Insights from Solid-State NMR Spectroscopy. *Acc. Chem. Res.* **52**, (2019).

144. Brouwer, D. H. *et al.* Solid-state 29Si NMR spectra of pure silica zeolites for the International Zeolite Association Database of Zeolite Structures. *Microporous Mesoporous Mater.* **297**, (2020).

145. Castellani, F. *et al.* Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature* **420**, (2002).

146. Otting, G. Protein NMR using paramagnetic ions. *Annual Review of Biophysics* **39**, (2010).

147. Wickramasinghe, N. P. *et al.* Nanomole-scale protein solid-state NMR by breaking intrinsic 1 H T1 boundaries. *Nat. Methods* **6**, (2009).

148. Bhattacharyya, R. *et al.* In situ NMR observation of the formation of metallic lithium microstructures in lithium batteries. *Nat. Mater.* **9**, (2010).

149. Hodgkinson, P. NMR crystallography of molecular organics. *Progress in Nuclear Magnetic Resonance Spectroscopy* **118**–**119**, (2020).

150. Martineau, C. NMR crystallography: Applications to inorganic materials. *Solid State Nuclear Magnetic Resonance* **63**, (2014).

151. Hartman, J. D., Kudla, R. A., Day, G. M., Mueller, L. J. & Beran, G. J. O. Benchmark fragment-based 1H, 13C, 15N and 17O chemical shift predictions in molecular crystals. *Phys. Chem. Chem. Phys.* **18**, (2016).

152. Shenderovich, I. G. Experimentally Established Benchmark Calculations of 31 P NMR Quantities . *Chemistry–Methods* **1**, (2021).

153. Benassi, E. Benchmarking of density functionals for a soft but accurate prediction and assignment of 1H and 13C NMR chemical shifts in organic and biological molecules. *J. Comput. Chem.* **38**, (2017).

154. Hodgkinson, P., Ashbrook, S. E., Morris, A. & Yates, J. R. Collaborative Computational

Project for NMR Crystallography. (2013). Available at: www.ccpnc.ac.uk.

155. Cordova, M. *et al.* Structure determination of an amorphous drug through large-scale NMR predictions. *Nat. Commun.* **12**, (2021).

156. Gerrard, W. *et al.* IMPRESSION-prediction of NMR parameters for 3-dimensional chemical structures using machine learning with near quantum chemical accuracy. *Chem. Sci.* **11**, (2020).

157. Venetos, M. C., Dwaraknath, S. & Persson, K. A. Effective Local Geometry Descriptor for29Si NMR Q4Anisotropy. *J. Phys. Chem. C* **125**, (2021).

158. Chaker, Z., Salanne, M., Delaye, J. M. & Charpentier, T. NMR shifts in aluminosilicate glasses via machine learning. *Phys. Chem. Chem. Phys.* **21**, (2019).

159. Sun, H. *et al.* Structural Investigation of Silver Vanadium Phosphorus Oxide (Ag2VO2PO4) and Its Reduction Products. *Chem. Mater.* **33**, (2021).

160. Kobera, L. *et al.* The Nature of Chemical Bonding in Lewis Adducts as Reflected by 27Al NMR Quadrupolar Coupling Constant: Combined Solid-State NMR and Quantum Chemical Approach. *Inorg. Chem.* **57**, (2018).

161. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).

162. Kinsey, R. A., Kirkpatrick, R. J., Hower, J., Smith, K. A. & Oldfield, E. HIGH RESOLUTION ALUMINUM-27 AND SILICON-29 NUCLEAR MAGNETIC RESONANCE SPECTROSCOPIC STUDY OF LAYER SILICATES, INCLUDING CLAY MINERALS. *Am. Mineral.* **70**, (1985).

163. Woessner, D. E. Characterization of clay minerals by 27 Al nuclear magnetic resonance spectroscopy. *Am. Mineral.* **74**, 203–215 (1989).

164. Lippmaa, E., Samoson, A. & Mägi, M. High-Resolution 27A1 NMR of Aluminosilicates. *J. Am. Chem. Soc.* **108**, (1986).

165. Müller, D., Gessner, W., Samoson, A., Lippmaa, E. & Scheler, G. Solid-state 27Al NMR studies on polycrystalline aluminates of the system CaO-Al2O3. *Polyhedron* **5**, (1986).

166. Depmeier, W. Aluminate sodalites - A family with strained structures and ferroic phase transitions. *Phys. Chem. Miner.* **15**, (1988).

167. Dirken, P. J., Jansen, J. B. H. & Schuiling, R. D. Influence of octahedral polymerization on 23Na and 27Al MAS NMR in alkali fluoroaluminates. *Am. Mineral.* **77**, (1992).

168. Chupas, P. J., Ciraolo, M. F., Hanson, J. C. & Grey, C. P. In situ X-ray diffraction and solid-state NMR study of the fluorination of γ-Al2O3 with HCF2Cl. *J. Am. Chem. Soc.* **123**, (2001).

169. Sadoc, A. *et al.* NMR parameters in column 13 metal fluoride compounds (AlF3, GaF3, InF3 and TlF) from first principle calculations. *Solid State Nucl. Magn. Reson.* **59**–**60**, (2014).

170. Pooransingh, N. *et al.* 51V solid-state magic angle spinning NMR spectroscopy and DFT studies of oxovanadium(V) complexes mimicking the active site of vanadium haloperoxidases. *Inorg. Chem.* **42**, (2003).

171. Schweitzer, A. *et al.* 51V solid-state NMR investigations and DFT studies of model compounds for vanadium haloperoxidases. *Solid State Nucl. Magn. Reson.* **34**, (2008).

172. Hovis, G. L., Spearing, D. R., Stebbins, J. F., Roux, J. & Clare, A. X-ray powder diffraction and 23Na, 27Al, and 29Si MAS-NMR investigation of nepheline-kalsilite crystalline solutions. *Am. Mineral.* **77**, (1992).

173. De, S., Bartók, A. P., Csányi, G. & Ceriotti, M. Comparing molecules and solids across structural and alchemical space. *Phys. Chem. Chem. Phys.* **18**, (2016).

174. Jäger, M. O. J., Morooka, E. V., Federici Canova, F., Himanen, L. & Foster, A. S. Machine learning hydrogen adsorption on nanoclusters through structural descriptors. *npj Comput. Mater.* **4**, (2018).

175. Rupp, M., Tkatchenko, A., Müller, K. R. & Von Lilienfeld, O. A. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.* **108**, (2012).

176. Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.* **134**, (2011).

177. Schmidt, J., Marques, M. R. G., Botti, S. & Marques, M. A. L. Recent advances and applications of machine learning in solid-state materials science. *npj Computational Materials* **5**, (2019).

178. Ghose, subrata & Tsang, T. Structural dependence of quadrupole coupling constant e2qQ/h for 27Al and crystal field parameter D for Fe3+ in aluminosilicates. *Am. Mineral. J. Earth Planet. Mater.* **58**, 748–755 (1973).

179. Baur, W. H. The geometry of polyhedral distortions. Predictive relationships for the phosphate group. *Acta Crystallogr. Sect. B Struct. Crystallogr. Cryst. Chem.* **30**, (1974).

180. Cumby, J. & Attfield, J. P. Ellipsoidal analysis of coordination polyhedra. *Nat. Commun.* **8**, (2017).

181. Padro, D. *et al.* Variations of titanium interactions in solid state NMR-correlations to local structure. *J. Phys. Chem. B* **106**, (2002).

182. Ray, W. D., Neter, J., Wasserman, W. & Kutner, M. H. Applied Linear Statistical Models (3rd Edition). *J. Oper. Res. Soc.* **42**, (1991).

183. Farrar, D. E. & Glauber, R. R. Multicollinearity in Regression Analysis: The Problem Revisited. *Rev. Econ. Stat.* **49**, (1967).

184. Breiman, L. Random forests. *Mach. Learn.* **45**, (2001).

185. Ong, S. P. *et al.* Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **68**, (2013).

186. Himanen, L. *et al.* DScribe: Library of descriptors for machine learning in materials science. *Comput. Phys. Commun.* **247**, (2020).

187. Pettifor, D. G. A chemical scale for crystal-structure maps. *Solid State Commun.* **51**, (1984).

188. Massiot, D., Kahn-Harari, A., Michel, D., Muller, D. & Taulelle, F. Aluminium-27 MAS NMR of Al2Ge2O7 and LaAlGe2O7: Two pentacoordinated aluminium environments. *Magn. Reson. Chem.* **28**, S82–S88 (1990).

189. Ding, S. W. & McDowell, C. A. Nutation Magic-Angle-Spinning Spectra of Half-Integer Quadrupole Spin Systems. *J. Magn. Reson. Ser. A* **112**, (1995).

190. Jakobsen, H. J., Skibsted, J., Bildsøe, H. & Nielsen, N. C. Magic-angle spinning NMR spectra of satellite transitions for quadrupolar nuclei in solids. *J. Magn. Reson.* **85**, (1989).

191. Dec, S. F., Fitzgerald, J. J., Frye, J. S., Shatlock, M. P. & Maciel, G. E. Observation of six-coordinate aluminum in andalusite by solid-state 27Al MAS NMR. *J. Magn. Reson.* **93**, (1991).

192. Skibsted, J., Bildsøe, H. & Jakobsen, H. J. High-speed spinning versus high magnetic field in MAS NMR of quadrupolar nuclei. 27Al MAS NMR of 3CaO·Al2O3. *J. Magn. Reson.* **92**, (1991).

193. Body, M. *et al.* Structural investigations of β-CaAlF5 by coupling powder XRD, NMR, EPR and spectroscopic parameter calculations. *J. Solid State Chem.* **178**, (2005).

194. Skibsted, J., Henderson, E. & Jakobsen, H. J. Characterization of Calcium Aluminate Phases in Cements by 27Al MAS NMR Spectroscopy. *Inorg. Chem.* **32**, (1993).

195. Raymond, M. & Hafner, S. S. Nuclear Quadrupole Coupling Tensors of 27 Al in Sillimanite (Al 2 SiO 5) . *J. Chem. Phys.* **53**, (1970).

196. Hassan, I. & Grundy, H. D. The crystal structures of sodalite-group minerals. *Acta Crystallogr. Sect. B* **40**, (1984).

197. HAHN, T. & BUERGER, M. J. The Detailed Structure of Nepheline, KNa3Al4Si4O16. *Zeitschrift für Krist. - Cryst. Mater.* **106**, (2014).

198. Haener, S. & Raymond, M. The Nuclear Quadrupole Coupling Tensors of Al27 in

Kyanite. *Am. Mineral.* **52**, 1632–1642 (1967).

199. Damodaran, K. *et al.* Triple-quantum magic angle spinning 27Al NMR of aluminum hydroxides. *J. Am. Chem. Soc.* **124**, (2002).

200. Xue, X. & Kanzaki, M. High-pressure δ-Al(OH)3 and δ-AlOOH phases and isostructural hydroxides/oxyhydroxides: New structural insights from high-resolution1H and27Al NMR. *J. Phys. Chem. B* **111**, (2007).

201. Müller, D., Gessner, W., Samoson, A., Lippmaa, E. & Scheler, G. Solid-state aluminium-27 nuclear magnetic resonance chemical shift and quadrupole coupling data for condensed AlO4 tetrahedra. *J. Chem. Soc. Dalt. Trans.* (1986). doi:10.1039/DT9860001277

202. Fitzgerald, J. J., Dec, S. F. & Hamza, A. I. Observation of five-coordinated Al in pyrophyllite dehydroxylate by solid-state 27 Al NMR spectroscopy at 14 T. *Am. Mineral.* **74**, 1405–1408 (1989).

203. Massiot, D. *et al.* Double rotation and magic-angle spinning nuclear magnetic resonance study of 27Al: reexamination of the aluminium borate 9Al2O3 · 2B2O3. *Solid State Nucl. Magn. Reson.* **5**, (1995).

204. BRUN, E. & HAFNER, S. Die elektrische Quadrupolaufspaltung von Al27 in Spinell MgAl2O4 und Korund Al2O3. *Zeitschrift für Krist. - Cryst. Mater.* **117**, (2014).

205. Eades, R. G. AN INVESTIGATION OF THE NUCLEAR RESONANCE ABSORPTION SPECTRUM OF Al 27 IN A SINGLE CRYSTAL OF EUCLASE . *Can. J. Phys.* **33**, (1955).

206. Tsang, T. & Ghose, S. Nuclear Magnetic Resonance of 27Al in Topaz, Al2SiO4(F, OH)2. *J. Chem. Phys.* **56**, (1972).

207. Petch, H. E., Cranna, N. G. & Volkoff, G. M. Second Order Nuclear Quadrupole Effects in Single Crystals: Part II. Experimental Results for Spodumene. *Can. J. Phys.* **31**, 837–858 (1953).

208. Caslton Brown, L. & Williams, D. Quadrapolar splitting of the Al27 and Be9 magnetic resonances in beryl crystals. *J. Chem. Phys.* **24**, (1956).

209. Brinkmann, D., Staehli, J. L. & Ghose, S. Nuclear Magnetic Resonance of 27 Al and 1 H in Zoisite, Ca 2 Al 3 Si 3 O 12 (OH) . *J. Chem. Phys.* **51**, (1969).

210. Tsang, T. & Ghose, S. Nuclear magnetic resonance of1H and27A1 and Al-Si order in low cordierite, Mg2Al4Si5O 18·nH2O. *J. Chem. Phys.* **56**, (1972).

211. Brog, K. C., Jones, W. H. & Verber, C. M. 27Al and 89Y nuclear magnetic resonance in yttrium-aluminum garnet. *Phys. Lett.* **20**, (1966).

212. Hockenberry, J. H., Carlton Brown, L. & Williams, D. Nuclear resonance spectrum of

Al27 in chrysoberyl. *J. Chem. Phys.* **28**, (1958).

213. Kelsey, K. E., Stebbins, J. F., Du, L. S. & Hankins, B. Constraining 17O and 27Al NMR spectra of high-pressure crystals and glasses: New data for jadeite, pyrope, grossular, and mullite. *Am. Mineral.* **92**, (2007).

214. Hafner, S. S., Raymond, M. & Ghose, S. Nuclear quadrupole coupling tensors of 27A1 in andalusite (Al 2SiO5). *J. Chem. Phys.* **52**, (1970).

215. Choi, M., Matsunaga, K., Oba, F. & Tanaka, I. Al NMR chemical shifts in oxide crystals: a first-principles study. *J. Phys. Chem. C* **113**, (2009).

216. Müller, D., Gessner, W. & Scheler, G. Chemical shift and quadrupole coupling of the 27Al NMR spectra of LiAIO2 polymorphs. *Polyhedron* **2**, (1983).

217. Massiot, D. Sensitivity and lineshape improvements of MQ-MAS by rotor-synchronized data acquisition. *Journal of Magnetic Resonance - Series A* **122**, (1996).

218. Rocha, J. Direct observation of highly distorted hexa-coordinated aluminium in andalusite by very fast27Al MAS NMR. *Chem. Commun.* (1998). doi:10.1039/a806939e

219. Smith, M. E., Jaeger, C., Schoenhofer, R. & Steuernagel, S. Structural characterisation of the aluminium sites in kyanite by 27Al magic angle spinning centreband, satellite transition and double rotation NMR. *Chem. Phys. Lett.* **219**, (1994).

220. Florian, P., Gervais, M., Douy, A., Massiot, D. & Coutures, J. P. A multi-nuclear multiple-field nuclear magnetic resonance study of the Y2O3-Al2O3 phase diagram. *J. Phys. Chem. B* **105**, (2001).

221. Nielsen, U. G. *et al.* Aluminum orthovanadate (AlVO4): Synthesis and characterization by 27Al and 51V MAS and MQMAS NMR spectroscopy. *Inorg. Chem.* **41**, (2002).