

ARTICLE OPEN

Enabling materials informatics for ^{29}Si solid-state NMR of crystalline materialsHe Sun¹, Shyam Dwaraknath², Handong Ling³, Xiaohui Qu², Patrick Huck¹, Kristin A. Persson³ and Sophia E. Hayes¹

Nuclear magnetic resonance (NMR) spectroscopy is a powerful tool for obtaining precise information about the local bonding of materials, but difficult to interpret without a well-vetted dataset of reference spectra. The ability to predict NMR parameters and connect them to three-dimensional local environments is critical for understanding more complex, long-range interactions. New computational methods have revealed structural information available from ^{29}Si solid-state NMR by generating computed reference spectra for solids. Such predictions are useful for the identification of new silicon-containing compounds, and serve as a starting point for determination of the local environments present in amorphous structures. In this study, we have used 42 silicon sites as a benchmarking set to compare experimentally reported ^{29}Si solid-state NMR spectra with those computed by CASTEP-NMR and Vienna Ab Initio Simulation Program (VASP). 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 validated, and in some cases corrected, in an effort to catalog these for the local spectroscopy database infrastructure (LSDI), where over 10,000 ^{29}Si NMR tensors for crystalline materials have been computed. Knowledge of specific tensor values can serve as the basis for executing NMR experiments with precision, optimizing conditions to capture the elements accurately. The ability to predict and compare experimental observables from a wide range of structures can aid researchers in their chemical assignments and structure determination, since the computed values enables the extension beyond tables of typical chemical shift (or shielding) ranges.

npj Computational Materials (2020)6:53; <https://doi.org/10.1038/s41524-020-0328-3>

INTRODUCTION

Nuclear magnetic resonance (NMR) spectroscopy has revolutionized organic and biological chemistry fields, owing to its ability to provide precise structural detail through investigation of ^1H and ^{13}C spectra. Assignments of these spectra rely on 50+ years of comprehensive and detailed data, many of which have been cataloged in guides from Sadtler¹ and Aldrich² and subsequently in databases, such as the AIST Spectral Database for Organic Compounds SDBS³.

For inorganic species, there are far fewer resources, and through the local spectroscopy data infrastructure (LSDI), we seek 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 (MP) database. ^{29}Si is attractive, because it is a nuclear spin, I , 1/2 species, found at moderate natural abundance (4.68%)⁴, 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^{5–8}. In this case, other experimental techniques like neutron diffraction and recently NMR have been employed to lend accuracy. This NMR refinement of structures is termed “NMR crystallography”^{7,9–12}. 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^{13–15}. 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. As we have shown in a separate ^{13}C study⁶, determination of the chemical shift tensor values enabled refinement of the H positions in a polycrystalline sample.

¹Department of Chemistry, Washington University, St. Louis, MO 63130, USA. ²Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. ³Department of Materials Science and Engineering, UC Berkeley, Berkeley, CA 94720, USA. ✉email: shyamd@lbl.gov; hayes@wustl.edu

Possessing catalogs 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 cataloging 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} .

NMR definitions

There is a tendency to use the terms “chemical shielding” and “chemical shift” interchangeably, even though these are different but related quantities. We set forth definitions for shielding (σ) first, followed by nomenclature and expressions using chemical shift (δ).

The NMR chemical shielding tensor describes the interaction between an externally applied magnetic field and the shielding effect of electrons, which leads to an induced magnetic field. Nuclear spins sense both the external field and the induced field, expressed by¹⁶

$$\mathbf{B}_{\text{ind}}(\mathbf{r}) = -\vec{\sigma}(\mathbf{r})\mathbf{B}_0 \quad (1)$$

$\vec{\sigma}(\mathbf{r})$ is a second rank tensor, the chemical shielding tensor, and it consists of symmetric and asymmetric contributions. The symmetric contribution can be diagonalized to yield three principal components of the shielding tensor, referred to in Haeberlen notation as σ_{XX} , σ_{YY} , and σ_{ZZ} .

$$\begin{pmatrix} \sigma_{XX} & 0 & 0 \\ 0 & \sigma_{YY} & 0 \\ 0 & 0 & \sigma_{ZZ} \end{pmatrix} \quad (2)$$

Isotropic shielding σ_{iso} is defined as the numerical average of the principal components.

$$\sigma_{\text{iso}} = \frac{\sigma_{XX} + \sigma_{YY} + \sigma_{ZZ}}{3} \quad (3)$$

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” convention¹⁷, Herzfeld Berger¹⁸, Haeberlen¹⁹, and the “Maryland” convention²⁰. 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 Dmfit²¹ and SIMPSON²² programs.

Here, we report many of the comparisons between experiment and computation using the popular Haeberlen¹⁷ convention, in part, because most literature uses this system for reporting the full chemical shielding (or shift) tensor. In Haeberlen, σ_{XX} , σ_{YY} and σ_{ZZ} are defined based on their distance from the isotropic shielding, σ_{iso} :

$$|\sigma_{ZZ} - \sigma_{\text{iso}}| \geq |\sigma_{XX} - \sigma_{\text{iso}}| \geq |\sigma_{YY} - \sigma_{\text{iso}}| \quad (4)$$

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 follows:

$$\zeta_\sigma = \sigma_{ZZ} - \sigma_{\text{iso}} \quad (5)$$

(Please note: 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 .)

While Eq. (5) expresses the algebraic definition for this quantity, the reduced anisotropy can be visualized in terms of the relative location of the most intense point in the static lineshape—to the right or left—of the isotropic shielding, σ_{iso} .

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

$$\eta_{\text{CSA}} = \frac{\sigma_{YY} - \sigma_{XX}}{\sigma_{ZZ} - \sigma_{\text{iso}}} \quad (6)$$

The value ranges from 0 to 1, 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.

It is worthwhile to note, in the “Mehring” convention¹⁷ the diagonalized shielding tensor elements are expressed as σ_{11} , σ_{22} , and σ_{33} . Unlike the Haeberlen labels, these are assigned based on their absolute order in frequency (and are not arranged relative to the isotropic shielding, as seen in Eq. (4) above):

$$\sigma_{33} \geq \sigma_{22} \geq \sigma_{11} \quad (7)$$

This latter point will have important consequences for data accuracy and correlations between experiment and computation, as we show below.

RESULTS AND DISCUSSION

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 three 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.

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 model²³ described in detail in Supplementary Information, Section III²⁰.

Shown in Fig. 1a 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 1b 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 $\sigma_{\text{reference}} = 316.26$ ppm, and the slope of the correlation plot is -1.12 . The resultant value for reference isotropic chemical shielding within VASP is $\sigma_{\text{reference}} = 528.18$ 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 1c 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 Supplementary Information, Section I, Supplementary Tables 1–3.

Beyond isotropic shift, the additional two algebraic expressions (ζ_σ and η_{CSA}) can be directly linked to the individual tensor

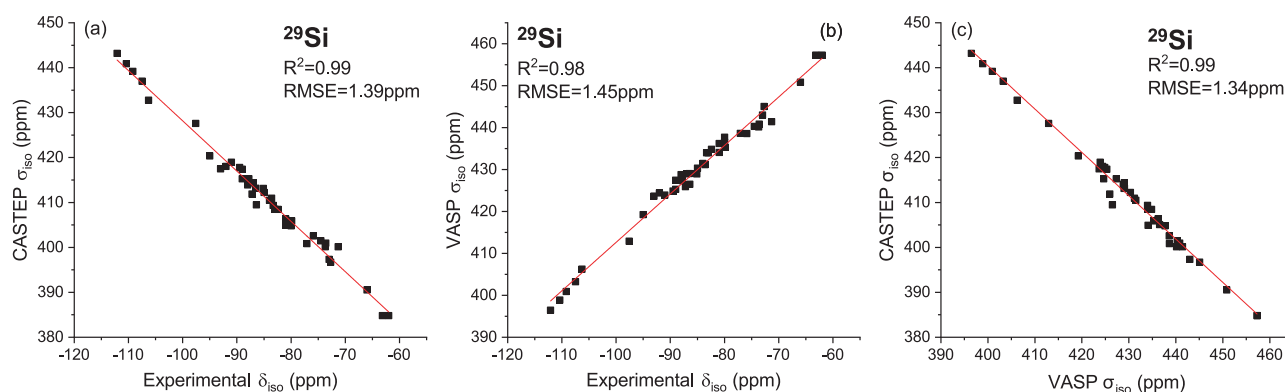


Fig. 1 Comparisons of computed chemical shielding and experimental chemical shifts. **a** CASTEP computed ^{29}Si isotropic chemical shielding values (σ_{iso}) versus experimentally reported ^{29}Si isotropic chemical shifts (δ_{iso}), $R^2 = 0.99$ and $\text{RMSE} = 1.39$ ppm. **b** VASP computed ^{29}Si isotropic chemical shielding values (σ_{iso}) versus experimentally reported ^{29}Si isotropic chemical shifts (δ_{iso}), $R^2 = 0.98$ and $\text{RMSE} = 1.45$ ppm. **c** CASTEP-computed versus VASP-computed ^{29}Si isotropic chemical shielding (σ_{iso}) values with a linear regression, $R^2 = 0.99$ and $\text{RMSE} = 1.34$ ppm. For all three plots, the fit to a linear regression is shown in the figure.

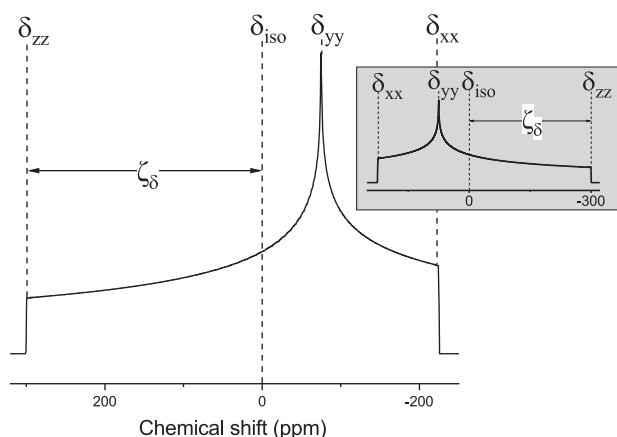


Fig. 2 Simulated static lineshape dominated by chemical shift anisotropy (CSA) interactions. Based on this lineshape, the three principal components of the chemical shift tensor can be identified individually as δ_{xx} , δ_{yy} and δ_{zz} based on the notation from Haeberlen. These values are usually reported as: isotropic 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.

elements that express the shape of the experimental lineshape, whether static NMR or a manifold of spinning sidebands under MAS NMR. Figure 2 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.

Challenges for cataloging full tensors

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 cataloging data. Figure 3 depicts a comparison of 42

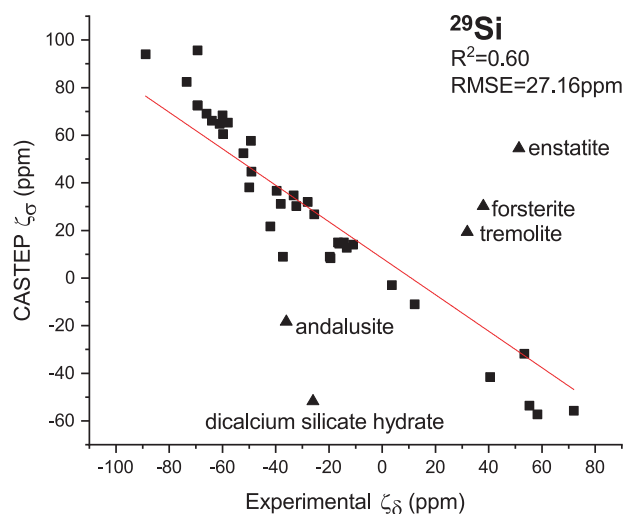


Fig. 3 Comparison of CASTEP computed ^{29}Si reduced shielding anisotropy ζ_δ versus experimentally reported reduced chemical shift anisotropy ζ_δ . The outliers (shown by \blacktriangle symbols) were closely analyzed in terms of experimental and computational error. (See comparison to VASP-computed values in Supplementary Fig. 10).

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. Supplementary Fig. 3 shows excellent agreement between the computed values for reduced anisotropy for CASTEP versus VASP, giving 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 ($\zeta_\delta = \delta_{zz} - \delta_{\text{iso}}$), 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 Fig. 4. 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. 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

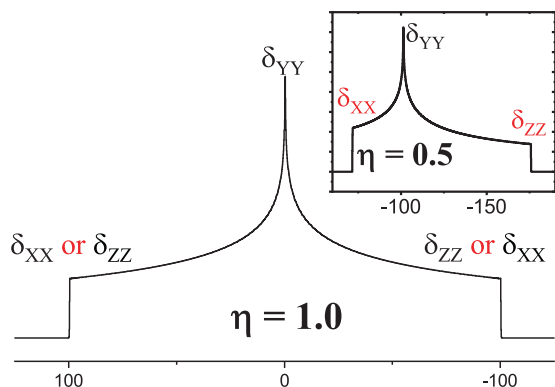


Fig. 4 Scheme of static NMR powder pattern lineshapes for two different values of the asymmetry parameter, η_{CSA} . For values of η_{CSA} less than ~ 0.7 , there is an unambiguous assignment of δ_{ZZ} and δ_{XX} tensor components. However, for η_{CSA} values of ~ 0.7 to 1.0 , the determination between the tensors elements can be mis-assigned.

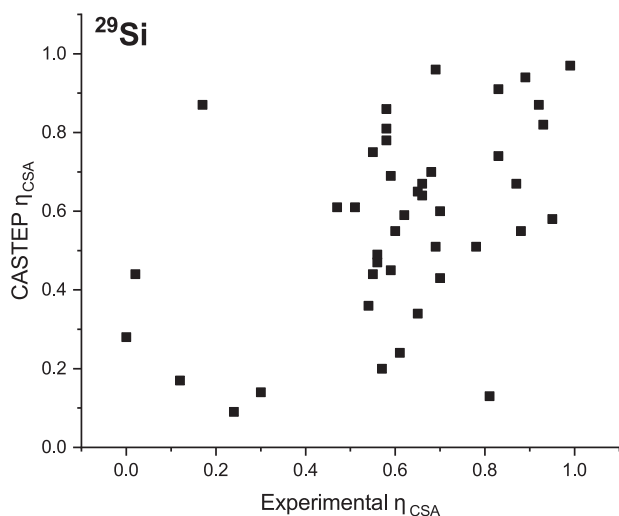


Fig. 5 Comparison of CASTEP computed versus experimentally reported ^{29}Si η_{CSA} asymmetry parameters. The plot compares computed and experimental values for 42 crystal structures. (See comparison to VASP-computed values in Supplementary Fig. 11).

assign that shoulder to one side or another based on sparse data (as illustrated schematically in the Supplementary Information, Section V). 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 imperfections²⁴, the wrong value for δ_{ZZ} may be assigned—importantly, to the incorrect “side” of the lineshape.

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 Fig. 5. (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 Supplementary Fig. 4.)

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 Fig. 6a, 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” convention¹⁷, where σ_{11} , σ_{22} and σ_{33} are the three counterparts, organized in terms of high-frequency to low-frequency for any lineshape, the algebraic solutions for the experimentally reported values become reconciled, as shown in Fig. 6b. (A depiction of the distribution of residuals for each of the principal components after linear rescaling can be found in Section IX of the Supplementary Information.)

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.

It is important to note that the shielding (σ) and the chemical shift (δ) should have a negative correlation with respect to one another (see Eq. (S3)). One finding in creation of this catalog 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 cataloged in the LSDI based on a convention (ref. ¹⁷, IUPAC 2008) (Eq. (7)), namely that the tensor elements are ordered numerically from largest to smallest. A case study is presented in the Supplementary Information, Section VI to illustrate this systematic difference in the shielding tensor elements, that is corrected when producing the individual chemical shift tensor elements. The LSDI catalog will ultimately contain both computed chemical shielding and corrected chemical shift full tensors. Supplementary Figs. 7 and 8 depict this reconciliation between the programs, and Fig. S6 is a matrix of plots CASTEP and VASP principal shielding components showing the mis-correlated components. As a check, we have used the TensorView program²⁵ to render a graphical depiction of the shielding surface ovaloid superimposed onto a Q^3 silicon site in sodium disilicate. In the figure, 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.

Opportunities for applications of the “LSDI” catalog

CASTEP and VASP have particular strengths in the assignment of tensor elements, which will form the basis of the LSDI catalog. The LSDI has already computed over 10,000 unique Si-sites for compounds in the MP using VASP. This continually growing data set, easily accessible application programming interface (API) and collection of software tools is established as a community resource to enable easier in-silico experimentation with solid-state NMR. Having such a catalog 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 three tensor elements. Furthermore, as we depict schematically in Fig. 7, the ability to plan experiments (i.e., to select an ideal MAS spinning frequency, such as shown in Fig. 7c versus that in Fig. 7b) in order

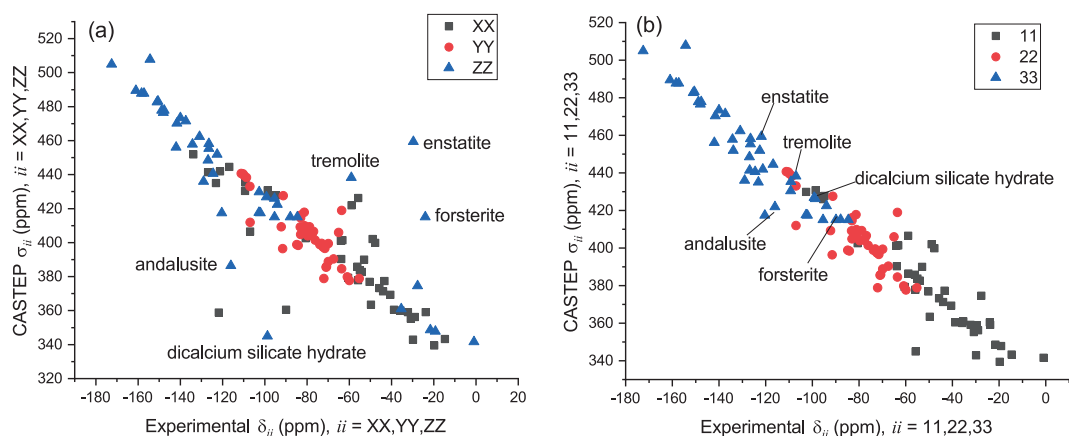


Fig. 6 Correlation plots of CASTEP computed values (σ) versus experimentally reported values (δ). **a** Depicts σ_{XX} , σ_{YY} , and σ_{ZZ} , and **b** depicts computed values expressed instead as σ_{11} , σ_{22} , and σ_{33} , versus the chemical shift (δ) tensor values from experimental reports, extracted algebraically as δ_{11} , δ_{22} , δ_{33} . (See comparison to VASP-computed values in Supplementary Fig. 12).

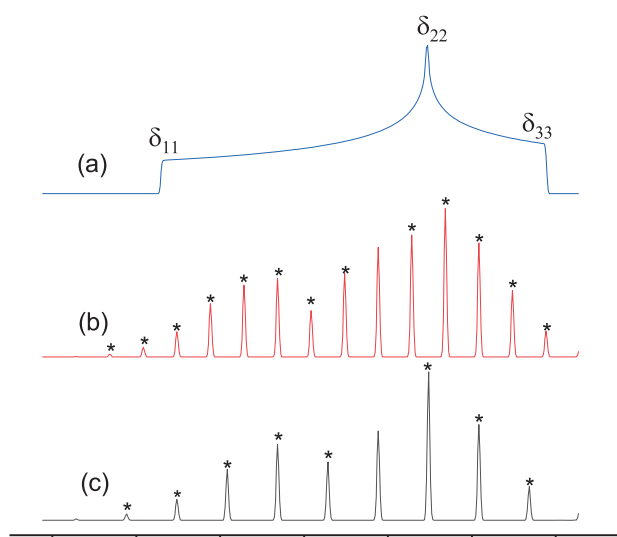


Fig. 7 Comparison of the static CSA lineshape and corresponding NMR MAS spinning sideband manifolds at two different rotational frequencies. **a** Simulated static lineshape with chemical shift principal components labeled. **b** Simulated MAS lineshape with a spinning frequency of 800 Hz. **c** Simulated MAS lineshape with a spinning frequency of 1200 Hz. * Denote spinning side bands.

to accurately map out the tensor values, especially of δ_{22} , is a consequence of possessing such data.

The utility of this catalog 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 four 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 three 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 catalog permits a comparison of isotropic chemical shielding values for >5000 silicate structures.

In Fig. 8, 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

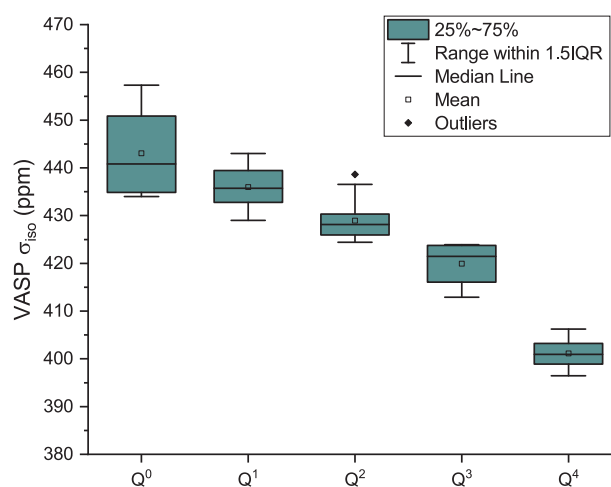


Fig. 8 Distribution of VASP calculated ^{29}Si isotropic shielding parameters. Statistical box plot illustrating the distribution of VASP calculated ^{29}Si isotropic shielding parameters (σ_{iso}) for different Q^n sites from the benchmarking set of 42 silicon sites. The symbols outside the range of 1.5 IQR are outliers. IQR stands for inter-quartile range.

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.

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 LSDI portion of The MP. 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 Haerberlen 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 Supplementary Information). 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.

METHODS

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^{26–30}, with 42 unique silicon sites primarily in minerals such as forsterite, wollastonite, zeolites, and quartz.

Details of DFT computations

CASTEP has been shown to be very effective for calculations of isotropic chemical shifts for nuclei, such as ^1H , ^{13}C , ^{89}Y , and ^{119}Sn ^{9,31–34} as well as diagonalized tensor values for ^{19}F and ^{77}Se ^{35–37} in select systems. DFT calculations using CASTEP were performed within the Perdew–Burke–Erzerhof (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 of stress and chemical shift anisotropy over different cut-off energies and k -points has been performed on α -quartz. The results are depicted in Supplementary Information, Section II. 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 $1/\text{\AA}$.

DFT calculations were also performed using the projector-augmented wave (PAW) method^{38,39} as implemented in the VASP^{40–42} within the PBE GGA formulation of the exchange–correlation functional⁴³. A cut-off for the plane waves of 520 eV is used and a uniform k -point density of $\sim 1000/\text{atom}$ is employed. We note that the computational and convergence parameters were chosen in compliance with the settings used in the MP⁴⁴ to enable direct comparisons with the large set of available MP data.

CASTEP and VASP both use the Gauge Including Projector Augmented Waves (GIPAW) method⁴⁵ 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, Dmfit²¹. 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.

DATA AVAILABILITY

All experimental values and computed NMR tensor values are included in the Supplementary Information as Supplementary Data 1. Each document in the JSON file contains a single site, its corresponding experimental value, the original citing paper, and the computed value. All the geometry optimized structures used for NMR calculation in this paper are included in the Supplementary Information as Supplementary Data 2. In addition, all data for the benchmarking of the computed NMR Tensors including structures, spectra, computed, and experimental tensors are available via the MPContribs platform on the Materials Project at <https://mpcontribs.org/>. The larger database of computed NMR tensors are available via the Materials Project at <https://materialsproject.org/>.

CODE AVAILABILITY

Density functional theory calculations using VASP were computed using the standard parameters of the Materials Project as documented in pymatgen v2020.3.13 using the workflows in atomate v0.9.4. Both pymatgen and atomate are open-source codes obtainable via both the python package index (PyPI) and Github. Density functional theory calculations using CASTEP were performed with the Materials Studio 2017 R2 package. The parameters used for the calculations are the default ones given by the Materials Studio CASTEP manual, unless specified otherwise in the “Methods” section.

Received: 18 December 2019; Accepted: 10 April 2020;

Published online: 12 May 2020

REFERENCES

1. Simons, W. W. (ed.) *The Sadtler Guide to Carbon-13 NMR Spectra*. (Sadtler Research Laboratories, 1983).
2. Pouchert, C. J. & Behnke, J. *The Aldrich Library of ^{13}C and ^1H FT-NMR Spectra* (Aldrich Chemical Co., 1993).
3. Saito, T., Yamaji, T., Hayamizu, K., Yanagisawa, M. & Yamamoto, O. AIST: Spectral Database for Organic Compounds (SDBS). <https://sdb.sdb.aist.go.jp> (2020).
4. Rosman, K. J. R. & Taylor, P. D. P. Isotopic compositions of the elements 1997 (Technical Report). *Pure Appl. Chem.* **70**, 217–235 (1998).
5. Sklenak, S. et al. Aluminum siting in silicon-rich zeolite frameworks: a combined high-resolution ^{27}Al NMR spectroscopy and quantum mechanics/molecular mechanics study of ZSM-5. *Angew. Chem. Int. Ed.* **46**, 7286–7289 (2007).
6. Cui, J., Olmsted, D. L., Mehta, A. K., Asta, M. & Hayes, S. E. NMR crystallography: evaluation of hydrogen positions in hydromagnesite by $^{13}\text{C}\{^1\text{H}\}$ REDOR solid-state NMR and density functional theory calculation of chemical shielding tensors. *Angew. Chem. Int. Ed.* **58**, 4210–4216 (2019).
7. Salager, E. et al. Powder crystallography by combined crystal structure prediction and high-resolution ^1H solid-state NMR spectroscopy. *J. Am. Chem. Soc.* **132**, 2564–2566 (2010).
8. Perras, F. A. & Bryce, D. L. Multinuclear magnetic resonance crystallographic structure refinement and cross-validation using experimental and computed electric field gradients: application to $\text{Na}_2\text{Al}_2\text{B}_2\text{O}_7$. *J. Phys. Chem. C* **116**, 19472–19482 (2012).
9. Harris, R. K. et al. Computation and NMR crystallography of terbutaline sulfate. *Magn. Reson. Chem.* **48**, S103–S112 (2010).
10. Ashbrook, S. E. & McKay, D. Combining solid-state NMR spectroscopy with first-principles calculations—a guide to NMR crystallography. *Chem. Commun.* **52**, 7186–7204 (2016).
11. Martineau, C., Senker, J. & Taulelle, F. NMR crystallography. *Ann. Rep. NMR Spectrosc.* **82**, 1–57 (2014).
12. 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).
13. Vyalikh, A. et al. Fluorine patterning in room-temperature fluorinated graphite determined by solid-state NMR and DFT. *J. Phys. Chem. C* **117**, 7940–7948 (2013).
14. 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**, 8954–8965 (2013).
15. Finkelstein-Shapiro, D. et al. Direct evidence of chelated geometry of catechol on TiO_2 by a combined solid-state NMR and DFT study. *J. Phys. Chem. C* **120**, 23625–23630 (2016).

16. Mills, I., Cvitas, T., Homann, K., Kalley, N. & Kochitsu, K. (eds) *Quantities, Units and Symbols in Physical Chemistry*, 2nd ed. (International Union of Pure and Applied Chemistry, IUPAC, 1993).
17. 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).
18. Herzfeld, J. & Berger, A. E. Sideband intensities in NMR spectra of sample spinning at the magic angle. *J. Chem. Phys.* **73**, 6021–6030 (1980).
19. Haeberlen, U. *High resolution NMR in solids: selective averaging*. (Academic Press, 1976).
20. 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).
21. Massiot, D. et al. Modelling one- and two-dimensional solid-state NMR spectra. *Magn. Reson. Chem.* **40**, 70–76 (2002).
22. 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).
23. Hartman, J. D., Kudla, R. A., Day, G. M., Mueller, L. J. & Beran, G. J. O. Benchmark fragment-based ^1H , ^{13}C , ^{15}N and ^{17}O chemical shift predictions in molecular crystals. *Phys. Chem. Chem. Phys.* **18**, 21686–21709 (2016).
24. Brouwer, D. H. & Enright, G. D. Probing local structure in zeolite frameworks: ultrahigh-field NMR measurements and accurate first-principles calculations of zeolite ^{29}Si magnetic shielding tensors. *J. Am. Chem. Soc.* **130**, 3095–3105 (2008).
25. Young, R. P. et al. TensorView: a software tool for displaying NMR tensors. *Magn. Reson. Chem.* **57**, 211–223 (2019).
26. Smith, K. A., Kirkpatrick, R. J., Oldfield, E. & Henderson, D. M. High-resolution silicon-29 nuclear magnetic resonance spectroscopic study of rock-forming silicates. *Am. Mineral.* **68**, 1206–1215 (1983).
27. Spearing, D. R. & Stebbins, J. F. The ^{29}Si NMR shielding tensor in low quartz. *Am. Mineral.* **74**, 956–959 (1989).
28. 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).
29. Hansen, M. R., Jakobsen, H. J. & Skibsted, J. ^{29}Si chemical shift anisotropies in calcium silicates from high-field ^{29}Si MAS NMR spectroscopy. *Inorg. Chem.* **42**, 2368–2377 (2003).
30. Weiden, N. & Rager, H. The chemical shift of the ^{29}Si nuclear magnetic resonance in a synthetic single crystal of Mg_2SiO_4 . *Z. Naturforsch. A* **40**, 126–130 (1985).
31. Asakura, T. et al. Intermolecular packing in B. mori silk fibroin: multinuclear NMR study of the model peptide $(\text{Ala-Gly})_{15}$ defines a heterogeneous antiparallel antipolar mode of assembly in the Silk II Form. *Macromolecules* **48**, 28–36 (2015).
32. 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).
33. Reader, S. W. et al. Cation disorder in pyrochlore ceramics: ^{89}Y MAS NMR and first-principles calculations. *J. Phys. Chem. C* **113**, 18874–18883 (2009).
34. Mitchell, M. R. et al. ^{119}Sn MAS NMR and first-principles calculations for the investigation of disorder in stannate pyrochlores. *Phys. Chem. Chem. Phys.* **13**, 488–497 (2011).
35. Arnold, A. A. et al. Structure of NaYF_4 upconverting nanoparticles: a multinuclear solid-state NMR and DFT computational study. *J. Phys. Chem. C* **117**, 25733–25741 (2013).
36. Martineau, C. et al. Structural investigation of $\alpha\text{-LaZr}_2\text{F}_{11}$ by coupling X-ray powder diffraction, ^{19}F solid state NMR and DFT calculations. *J. Solid State Chem.* **199**, 326–333 (2013).
37. Griffin, J. M. et al. ^{77}Se solid-state NMR of inorganic and organoselenium systems: a combined experimental and computational study. *J. Phys. Chem. C* **115**, 10859–10872 (2011).
38. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).
39. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).
40. Kresse, G. & Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B* **47**, 558–561 (1993).
41. 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).
42. 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).
43. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
44. Jain, A. et al. Commentary: The Materials Project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
45. Pickard, C. J. & Mauri, F. All-electron magnetic response with pseudopotentials: NMR chemical shifts. *Phys. Rev. B* **63**, 245101 (2001).

ACKNOWLEDGEMENTS

This work is supported by the U.S. National Science Foundation (NSF), Award 1640899. This work 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, Berkeley.

AUTHOR CONTRIBUTIONS

S.E.H. and S.D. led the project examining NMR tensors and conducting DFT calculations from Material Project datasets, and both co-authored the manuscript. H. S. (first author) conducted CASTEP calculations and data curation of solid-state NMR data, and he wrote large portions of the manuscript; H.L. and X.Q. both conducted VASP computations. P.H. contributes in MPContribs platform. K.A.P. is the lead of the Materials Project and contributed time and effort to direct computational efforts.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at <https://doi.org/10.1038/s41524-020-0328-3>.

Correspondence and requests for materials should be addressed to S.D. or S.E.H.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020