Theoretical advances in signal analysis of DEER spectroscopy

Sarah R. Sweger

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:
Stefan Stoll, Chair
Anne B. McCoy
Cody W. Schlenker

Program Authorized to Offer Degree:
Chemistry

University of Washington

Abstract

Theoretical advances in signal analysis of DEER spectroscopy

Sarah R. Sweger

Chair of the Supervisory Committee:
Professor Stefan Stoll

Electron paramagnetic resonance (EPR) spectroscopy has been established as an indispensable tool for a wide range of scientific fields. Within the field of EPR, pulse methods, particularly those that measure dipolar couplings, are valuable for resolving protein structures. These applications have driven spectroscopists to address some key challenges in the methodology. This thesis discusses two challenges in the field of pulsed dipolar EPR spectroscopy. The first challenge involves improving the applicability of DEER spectroscopy by expanding the toolkit for data analysis. Due to the ill-conditioned nature of the problem, robust uncertainty quantification is a necessity. The work presented in this thesis provides fully quantified uncertainty of DEER data using Bayesian statistical methods. The second addresses the fundamental behavior of spins when placed in a highly polarized environment and identifies both analytically and experimentally how the generated signal deviates from what has been previously predicted.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

There will never be enough words to fully convey the depth of gratitude I have to the many who made my graduate school experience profoundly worthwhile. Firstly, I am indebted to my advisor Professor Stefan Stoll. I am thankful to have had a mentor who always challenged me to push farther, examine every detail, and leave no stone unturned in my pursuit of understanding. His vast depth of knowledge, constant enthusiasm, and remarkable orientation to detail have been so influential in how I approach all of my work. I am so grateful for his willingness to not only help me grow through feedback but also listen to my feedback and make a space where we can all be continually improving. This was crucial to helping me understand the type of mentor and scientist I hope to be throughout my career.

Referring to the Stoll Patrol members as my 'coworkers' does a disservice to the magnitude of their impact on my time at UW. Getting to come to work everyday and share office space with people who felt like my home away from was truly a privilege. Dr. Donald Mannikko was the most humble and patient groupmate one could have the joy of learning from, especially considering his incredible depth of EPR knowledge. Women have an increasing presence in the physical sciences but it was a uniquely special experience to get to work with so many during my time in Stoll Patrol. I am continually in awe of the tenacity Dr. Lizzy Canarie shows towards all aspects of her life and the seemingly bottomless well of support she gives to those around her. I have never met a person who works harder or is more humble than Shutian Lu, working with her everyday challenged me to constantly be better. Rachelle Stowell quickly became one of the most competent and caring group members I have ever worked with. I cannot wait to hear about all of the brilliant science that these women will do in their careers. I owe Dr. Maxx Tessmer a keg's worth of beer for answering my endless

questions about Python. Thank you to Lukas Zha for being a bright, kind mentee and for helping me become a better mentor. To all the other group members I have overlapped with in the Stoll Patrol - Dr. Eric Evans, Sam Jahn, Dr. Ellen Hayes, Dr. Thomas Edwards, Dr. Jeremy Lehner, Dr. Stephan Pribitzer, Dr. Luis Fábregas-Ibáñez, and Jörg Fischer, thank you for being a constant source of knowledge, support, and joy.

During my first year of graduate school I was fortunate enough to find a cohort of friends, Ryan DiRisio, Jason Sandwisch, Rob Weakly, and Diana Roh, who managed to always bring levity to any situation and made it possible to look back on even the longest evenings in Rab's room fondly. I am so grateful to have gotten to watch as they continually awe everyone around them with their intelligence, drive, and hardwork. It is truly a privilege to support and have been supported by such an amazing group of scientists and people. Diana Roh, specifically, is a person who has irrevocably changed my life for the better — I simply could not have asked to find a better person to be in my corner these last five years and I owe much of my success in graduate school to her never-wavering support.

Thank you to my committee members — Professor Anne McCoy, Professor Cody Schlenker, and Professor Paul Wiggins for the continued feedback and excellent scientific discussions during my time at UW.

Though I was only a part of the Empowering Prevention & Inclusive Communities (EPIC) team for my last year at UW, the team has had a permanent impact on my perspective of the goals of higher education. They have taught me invaluable lessons on how reconfigure our communities to ones the promote the success of all it's members. Further, they have given me a wonderful support network and taught me how to be a better team member.

Lastly to my family, particularly my parents, I would not be in this position if it had not been for their constant encouragement and support. They have stood by my side through the roller coaster ride of my education and I cannot thank them enough for the unconditional love and faith they have given me.

DEDICATION

To my rays of sunshine during the cloudy, Seattle days — Bean, Fig, and Bouillon.

Chapter 1

THEORY AND APPLICATIONS OF DEER SPECTROSCOPY

1.1 Motivation

Proteins are the engines that drive the living world. Conformational change within these systems often dictates their function in essential processes such as signal transduction, transport, and catalysis. While there exists a plethora of tools for probing molecular structures, most are limited to obtaining single static structures rather than sampling a range of conformations. These static structures provide incomplete information about the dynamic protein structure in its native environment in the cell. One method with the potential for filling in this gap in the field utilizes a pulse electron paramagnetic resonance (EPR) technique known as double electron–electron resonance (DEER) spectroscopy. DEER elucidates a full range of distances on the nanometer scale, along with the orientation and conformational flexibility of specific structural features. The technique measures an oscillatory time-domain signal that depends on the magnitude of the magnetic dipole–dipole interactions between spin centers in the range of two to ten nanometers [1, 2], often on spin labeled proteins (Fig. 1.1A). From the time-domain signal a full distribution of distances, $P(r)$, is extracted, providing detailed insight into the conformational landscape of the protein (Fig. 1.1B) [3, 4, 5]. DEER has seen use with studies of organic polymers but is most commonly used for structural studies in large biomolecules, such as proteins and nucleic acids [6, 7, 8, 9, 10, 11, 12, 13, 14].

As with any method, DEER faces certain challenges in data acquisition and data analysis. Among these challenges is the difficulty of attaching stable radicals to proteins, developing labeling techniques for biologically relevant scenarios, improving sensitivity, understanding the spin dynamics in all environments (various magnetic fields, temperatures, etc.), and

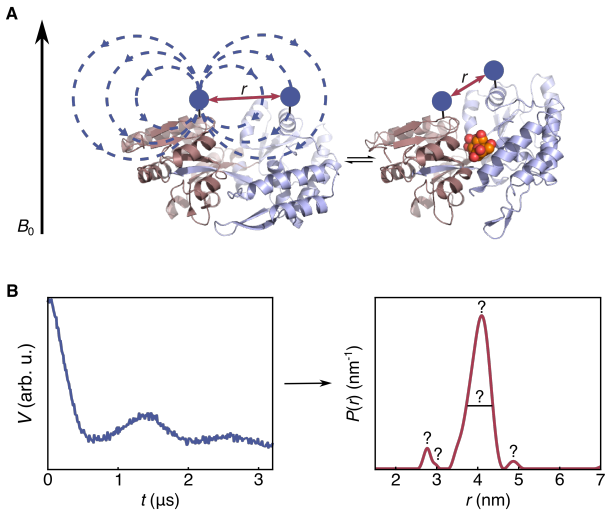

Figure 1.1: A: A spin labeled protein that undergoes a conformational change, shifting the distance, r, between the two spin labels (PDBs: apo: 1OMP, holo: 1ANF). B: Oscillatory time domain signal and extracted distance distribution in solution with distribution features including smaller side peaks, the position of the mean peak, shoulders, and peak widths that can provide structural insight.

signal processing to robustly extract distance distributions [1, 15, 16, 17, 18, 19, 20]. The work presented in this thesis has contributed to two projects addressing the challenges of understanding spin dynamics in different environments and error analysis in signal processing to produce distance distributions with increased confidence.

1.1.1 Thesis Outline

This chapter provides an overview of the theory and application behind electron paramagnetic resonance (EPR). The majority of the focus is on a specific pulse EPR technique known as double electron–electron resonance (DEER), as it is the central method utilized in the following chapters. Chapters 2 and 3 address uncertainty quantification in DEER data analysis through a Bayesian statistical approach. Chapter 2 represents the underlying distance distribution as a mixture of Gaussians while Chapter 3 extends the Bayesian inference approach to analysis that uses a nonparametric histogram to define the distance distribution.

The work in these chapters fully quantifies the error in the analysis procedure and expands the methodology of uncertainty quantification available to DEER spectroscopists. Chapter 4 addresses how signal generation is impacted by conditions of high spin polarization. This work has lead to the discovery, demonstrated analytically and experimentally, of a linear phase dependence of the signal on the magnitude of polarization.

1.2 DEER theory

1.2.1 Time-domain signal

DEER is a two-frequency pump–probe experiment that records the amplitude modulation of an electron spin echo as a function of the pump-pulse position. (Fig. 1.2). The magnetic dipole–dipole coupling between spins leads to a modulation of the echo amplitude whose frequency is dependent on the magnitude of the coupling. The majority of DEER experiments are conducted on samples of dilute and uniformly distributed spin pairs on doubly-labeled molecules or complexes. These samples give rise to an overall signal that is a product of an oscillatory intra-molecular signal and a typically exponentially decaying inter-molecular signal. The latter is often referred to as the background signal.

We base our analysis in the subsequent chapters on the standard model for 3- and 4-pulse DEER [1]. In this model, the noise-free DEER signal is

$$V_{\mathrm{M}}(t) = V_0 \cdot V_{\mathrm{intra}}(t) \cdot V_{\mathrm{inter}}(t) \tag{1.1}$$

where t is the position of the pump pulse relative to the time of the first two pulse echo (Fig. 1.2) and V_0 is the echo amplitude in the absence of the pump pulse. $V_{\mathrm{intra}}(t)$ is the intra-molecular modulation function and is given by

$$V_{\mathrm{intra}}(t) = (1-\lambda) + \lambda \int_0^\infty K(t,r)P(r)\,\mathrm{d}r \tag{1.2}$$

with the modulation depth λ (satisfying $0 < \lambda \leq 1$) and the normalized distribution $P(r)$ of the spin–spin distance r (satisfying $P(r) \geq 0$ and $\int_0^\infty P(r)\mathrm{d}r = 1$). $K(t,r)$ is the dipolar

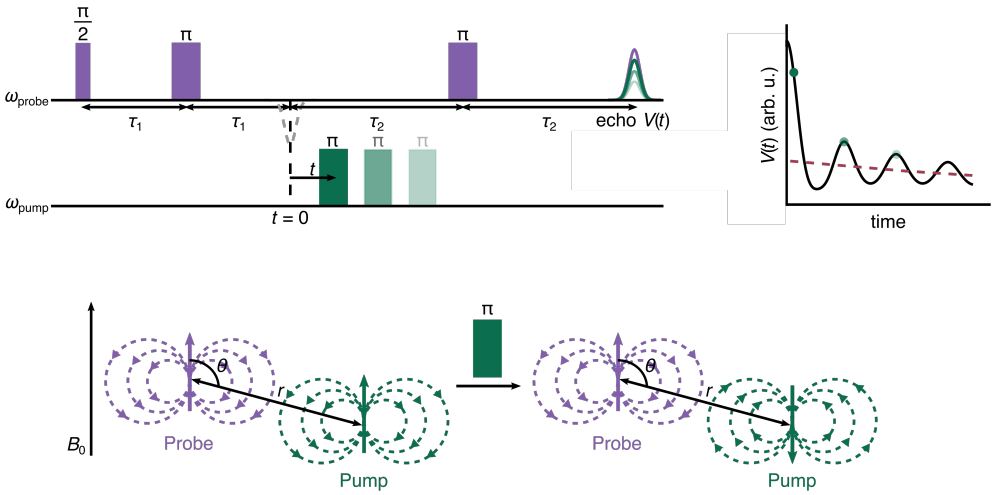

Figure 1.2: DEER pulse sequence. Two spins are shown in the presence of a magnetic field, purple is affected by the probe pulse and green is affected by the pump pulse. When the pump pulse is applied the spins affected at this frequency are inverted, shifting the frequency of the probe spins by the magnitude of the coupling, decreasing the number of spins refocused at the echo. The degree of the attenuation is determined by the positioning of the pump pulse shown in time on both the pulse sequence and deer trace with the respective echoes.

kernel function, given by

$$K(t,r) = \int_0^1 \cos\left((1-3z^2)Dr^{-3}t\right) \mathrm{d}z \tag{1.3}$$

with the dipolar coupling constant $D = (\mu_0/4\pi)(g_\mathrm{e}\mu_\mathrm{B})^2/\hbar$.

$V_\mathrm{inter}(t)$ is the inter-molecular modulation function, also called the background. In this work, we use

$$V_\mathrm{inter}(t) = \exp(-k|t|) \tag{1.4}$$

Here, k is the decay rate constant $k = (8\pi^2/9\sqrt{3})Dc\lambda$, with the spin concentration c (in spins/m^3) and the modulation depth λ.

In practice, the echo amplitude is measured at a set of discrete pump pulse positions t_i up to a maximum t_max, so that the experimental signal consists of a vector $\boldsymbol{V}$ with elements $V_i = V(t_i)$. In addition, the experimental signal is corrupted by measurement noise. This noise is approximately Gaussian, uncorrelated, and of constant variance σ^2 [18]. We can thus

write each measured data point as a random sample from a Gaussian (normal) distribution

$$V_i = V_{\mathrm{M}}(t_i) + \epsilon_i \quad \text{with} \quad \epsilon_i \sim \mathrm{Normal}(0, \sigma^2) \tag{1.5}$$

or, equivalently

$$V_i \sim \mathrm{Normal}(V_{\mathrm{M}}(t_i), \sigma^2) \tag{1.6}$$

where $\sim$ indicates that the quantity on the left is a random sample drawn from the distribution on the right.

While it is possible to use a closed-form expression for the integral in Eq. (1.3), the integral in Eq. (1.2) needs to be approximated numerically. We do this by discretizing $P(r)$ at a set of distances $r = r_j$, giving the n_r-element vector $\boldsymbol{P}$. This gives

$$\boldsymbol{V} = \boldsymbol{V}_{\mathrm{M}} + \boldsymbol{\epsilon} = \boldsymbol{K}\boldsymbol{P} + \boldsymbol{\epsilon} \tag{1.7}$$

with the kernel matrix $\boldsymbol{K}$ with elements $K_{i,j} = V_0 K(t_i, r_j)\Delta r$, where Δr is the increment in the r domain. $\boldsymbol{\epsilon}$ is the vector with elements ϵ_i. The model, dependent upon the time domain, $\boldsymbol{V}$, and distance domain, $\boldsymbol{P}$, can be described using a set of parameters. The same model parameters $\boldsymbol{\theta}$ are utilized for the time domain — modulation depth (λ), echo amplitude (V_0), background decay constant (k), and noise (σ). The parameters for the distance domain are dependent upon the model used for $\boldsymbol{P}$ (Fig. 1.3) and will be specifically detailed in the following sections.

1.2.2 Distance distribution extraction

In the analysis, Eq. (1.1) is fitted with a model that includes a distance distribution. Mathematically, this constitutes an ill-posed inversion problem i.e. small changes in the data can lead to large changes in the fitted distance distribution. Analysis approaches for obtaining a distance distribution range from analytical solutions [21] to deep neural networks [22]. Two methods based on least-squares fitting have seen the widest practical application: Tikhonov regularization and Gaussian mixture models [23, 24, 25, 26, 27, 28, 19]. Tikhonov regularization utilizes a non-parametric distance distribution model and includes a roughness penalty

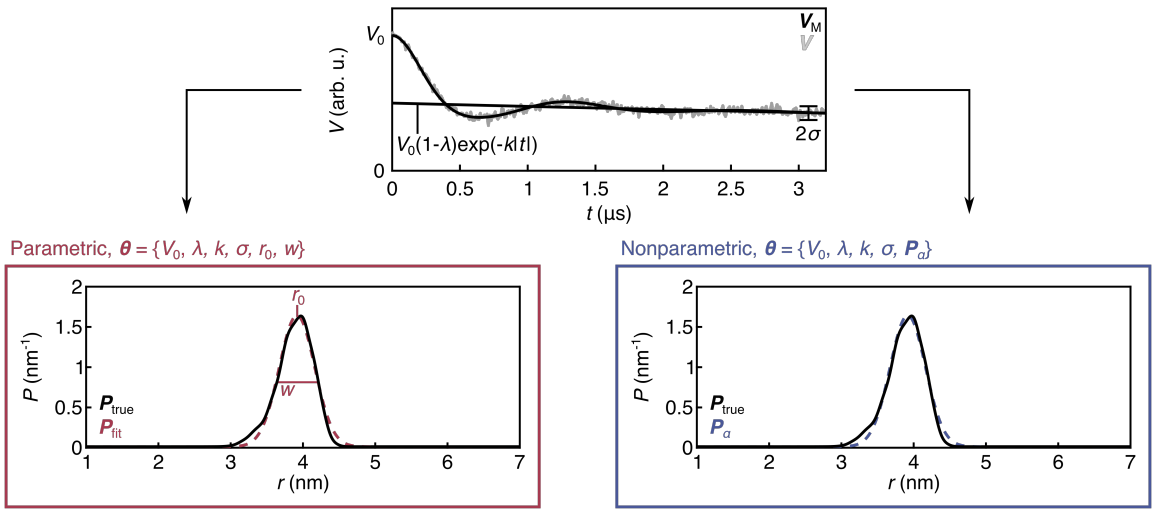

Figure 1.3: The time domain and distance domain models are shown using a relatively simple distance distribution for the protein T4L lysozyme. P_{true} (black) is distribution 3992 from the synthetic T4L test set [19]. A noise-less trace was derived from P_{true} (black) and overlaid with noise (gray). The time–domain parameters k, λ, V_0, and σ are indicated. In an experimental setting, the time domain data would be collected and then fit using a model for the distance distribution. The two most common models utilized for DEER distance distributions and the parameters that define them are shown. The left-hand side shows the parametric model, P_{fit}, is a one-Gaussian fit to P_{true}, with parameters r_0, and w indicated. In cases where a multi-Gauss model is used, a parameter for the peak amplitudes, A, is also included for each Gaussian. The right-hand side shows the nonparametric model, P_α, represents the distance distribution dependent upon a Tikhonov smoothing parameter.

for the distribution into the fitting objective function. Gaussian mixture models are parametric and represent the distribution as a linear combination of a few Gaussian functions. Both Tikhonov regularization and Gaussian models can be fit directly to the raw data in a single fitting step that also fits the background Eq. (1.4) [29].

1.2.3 Gaussian mixture models

In Gaussian mixture models, the spin–spin distance distribution $P(r)$ used in Eq. (1.2) is represented as a linear combination of normalized Gaussian basis functions

$$P(r) = \sum_{i=1}^{m} A_i \,\mathrm{Gauss}(r; r_{0,i}, w_i) \tag{1.8}$$

m is the number of Gaussians, A_i are the amplitudes (with $\sum_i A_i = 1$), $r_{0,i}$ are the centers of the Gaussians, and w_i are the full widths at half maximum. The Gaussian distribution is defined as

$$\mathrm{Gauss}(r; r_0, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(r-r_0)^2}{2\sigma^2}\right) \tag{1.9}$$

where r_0 is the mean of the distribution and σ is the standard deviation. The standard deviation σ relates to the full width at half maximum value w by

$$\sigma = \frac{w}{2\sqrt{2\ln 2}} \tag{1.10}$$

Given that this model for $P(r)$ requires an underlying shape to the distance distribution, it will often be referred to in this work as a 'parametric' model (Fig. 1.3).

1.2.4 Tikhonov regularization

Unlike Gaussian mixture models, nonparametric models do not require the distance distribution to have any underlying shape (Fig. 1.3). The most common method of fitting DEER data using a nonparametric model approach utilizes a form of penalized least-squares fitting known as Tikhonov regularization. This method employs the use of a regularization

parameter to shift minimization in the direction of a particular solution. For DEER data, the objective function is

$$\boldsymbol{P}_\alpha = \min_{P\geq 0}(||\boldsymbol{V} - \boldsymbol{K}\boldsymbol{P}||^2 + \alpha^2||\boldsymbol{L}\boldsymbol{P}||^2) \tag{1.11}$$

where α is the Tikhonov regularization parameter and $\boldsymbol{L}$ is the second-order difference matrix

$$\boldsymbol{L} = \begin{pmatrix} 1 & -2 & 1 & & & 0 \\ & 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ 0 & & & 1 & -2 & 1 \end{pmatrix} \tag{1.12}$$

To obtain physically meaningful results, the minimization problem is constrained to only positive values for $\boldsymbol{P}_\alpha$. The resulting $\boldsymbol{P}_\alpha$ is the result of minimizing the misfit between the model and data ($||\boldsymbol{V} - \boldsymbol{K}\boldsymbol{P}||^2$) with an additional term to penalize for rough or jagged distance distributions (as these are less probable given the flexibility of large biomolecules). Therefore, small α values typically lead to a spiky distance distribution, and probable overfitting of the time-domain data, while larger α values can result in an over-smoothed distance distribution and underfit time-domain data. Historically, a single value for α has been determined by a method of the practitioner's choice [19]. In this thesis, a single value for α is not chosen and rather is used as a tuning or hyperparameter for $\boldsymbol{P}_\alpha$.

1.3 Bayesian inference

The goal of analyzing $\boldsymbol{V}$ given the chosen model is to determine values for the model parameters that are consistent with the data. Now, since $\boldsymbol{V}$ is an incomplete representation of $V(t)$ due to time truncation, time discretization, amplitude discretization, and noise, there will be uncertainty associated with these parameters. It is important to quantify this uncertainty as well. Therefore, the goal of the analysis is to determine the full probability distribution of all parameters, given the data $\boldsymbol{V}$,

$$p(\boldsymbol{\theta}|\boldsymbol{V}) \tag{1.13}$$

This distribution is called the posterior probability distribution, or simply the posterior. It is a probability density defined over the entire N-dimensional parameter space and quantifies how probable any set of parameters is, given the data. Regions in parameter space with high posterior density reveal probable sets of parameters, and the spread of the distribution quantifies the uncertainty. The posterior distribution also reveals the level of correlation between all parameters in the given model.

To calculate the posterior, Bayes' theorem is utilized, which in its full form is shown in Eq. (1.14) [30, 31, 32, 33].

$$p(\boldsymbol{\theta}|\boldsymbol{V}) = \frac{p(\boldsymbol{V}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})}{p(\boldsymbol{V})} \tag{1.14}$$

The denominator, $p(\boldsymbol{V})$, represents the probability of a signal integrated over all possible parameter values.

The first term in the numerator, $p(\boldsymbol{V}|\boldsymbol{\theta})$, is called the likelihood, or the sampling distribution, and defines the probability of a signal given a parameter set. In the case with normally distributed errors, the multi-dimensional Gaussian is defined

$$\begin{aligned} p(\boldsymbol{V}|\boldsymbol{\theta}) &= \text{Normal}(\boldsymbol{V}; \boldsymbol{V}_{\text{M}}(\boldsymbol{\theta}), \sigma^2\mathbf{I}) \\ &\propto \exp\left(-\frac{\|\boldsymbol{V} - \boldsymbol{V}_{\text{M}}(\boldsymbol{\theta})\|^2}{2\sigma^2}\right) \end{aligned} \tag{1.15}$$

with center $\boldsymbol{V}_{\text{M}}$ and isotropic covariance matrix $\sigma^2\mathbf{I}$, where $\mathbf{I}$ is the identity matrix. This distribution quantifies the degree of fit between the data and the model.

The second factor in the numerator in Eq. (1.14), $p(\boldsymbol{\theta})$, is called the prior. It represents information about the parameters prior to taking the observed data $\boldsymbol{V}$ into account. If information about individual parameters θ_i is not correlated, the prior factors into a product

$$p(\boldsymbol{\theta}) = \prod_i p(\theta_i) \tag{1.16}$$

To calculate the posterior from Eq. (1.14), both a likelihood and a full prior distribution must be specified. It is essential to define the prior such that it appropriately captures prior information about the parameters but is diffuse enough to not introduce unwarranted bias

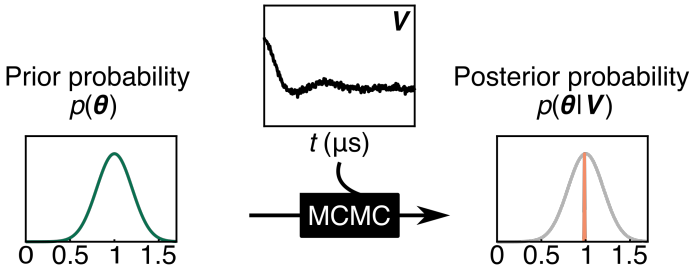

Figure 1.4: Visualization of Bayesian inference. The priors for all parameters (only one shown) and the raw data are input into the 'black box' process of MCMC and the output is the marginalized posterior distribution for each parameter. The marginalized posteriors represent a probability distribution of possible values for each parameter.

into the analysis. A visualization of the general Bayesian process used in this thesis is given in Fig. (1.4).

Chapter 2

BAYESIAN PROBABILISTIC ANALYSIS OF DEER SPECTROSCOPY DATA USING PARAMETRIC DISTANCE DISTRIBUTION MODELS

This chapter introduces a Bayesian probabilistic inference approach to analyze DEER data, using a multi-Gauss mixture model for the distance distribution. The method uses Markov Chain Monte Carlo (MCMC) sampling to determine a posterior probability distribution over model parameter space. This distribution contains all the information available from the data, including a full quantification of the uncertainty about the parameters. The corresponding uncertainty about the distance distribution is captured via an ensemble of posterior predictive distributions. Several synthetic examples illustrate the method. An experimental example shows the importance of model checking and comparison using residual analysis and Bayes factors. Overall, the Bayesian approach allows for more robust inference about protein conformations from DEER spectroscopy[1].

2.1 Motivation

Assessment of uncertainty in the fitted distance distribution is challenging, but is crucial for making sound conclusions about the conformational landscape. In both approaches presented in chapter one, correctly quantifying and visualizing uncertainty is challenging.

For Gaussian mixture models, uncertainty analysis relies primarily on parameter confidence intervals, which are obtained from the covariance matrix or by explicitly exploring the sensitivity of the objective function on the parameter values [26, 27, 28, 20]. The parameter

[1]This chapter is based on the 2020 publication in *The Journal of Physical Chemistry A* by Sweger S. R., Pribitzer S., and Stoll S. titled "Bayesian probabilistic analysis of DEER spectroscopy data using parametric distance distribution models.", volume 124, pages 6193-6202, available at `https://doi.org/10.1021/acs.jpca.0c05026`.

confidence intervals are then propagated to the distance domain to yield error bars on the distribution. This method assumes that the error surface is quadratic and that the parameters are unbounded, neither of which is generally true. Another method to obtain confidence intervals for both approaches is bootstrapping, which generates an ensemble of distributions by analyzing a large number of synthetically generated hypothetical signals based on the fitted model.[29]

Here, we present a Bayesian probabilistic inference approach [30, 32, 31] to analyze DEER data using a multi-Gauss mixture model. The method models and analyses the raw DEER data directly and yields a full posterior probability distribution over all model parameters, providing complete quantitative information about uncertainty and correlations for all parameters, without any limiting assumptions. We also introduce distribution ensembles to more correctly represent uncertainty about the distance distribution, including correlations which are neglected when using visualizations based on error bands.

The chapter is structured as follows. Section 2.2 presents the model used to describe the DEER data with Bayesian inference. Section 2.3 applies this method to a simple distribution while more realistic distributions are evaluated in Section 2.4. Section 2.5 applies the method to experimental data, including model checking and comparison. Section 2.6 provides conclusions regarding the method.

2.2 Model setup

As stated in chapter one, we represent the spin–spin distance distribution $\boldsymbol{P}$ used in Eq. (1.7) as a linear combination of normalized Gaussian basis functions Eq. (1.8). The full set of parameters for this multi-Gauss DEER model includes the distribution parameters $r_{0,i}$, w_i, and A_i, the time-domain parameters k, λ, and V_0, and the noise standard deviation σ (Fig. 1.3). We collect them into a parameter vector

$$\boldsymbol{\theta} = (\{r_{0,i}\}, \{w_i\}, \{A_i\}, k, \lambda, V_0, \sigma) \tag{2.1}$$

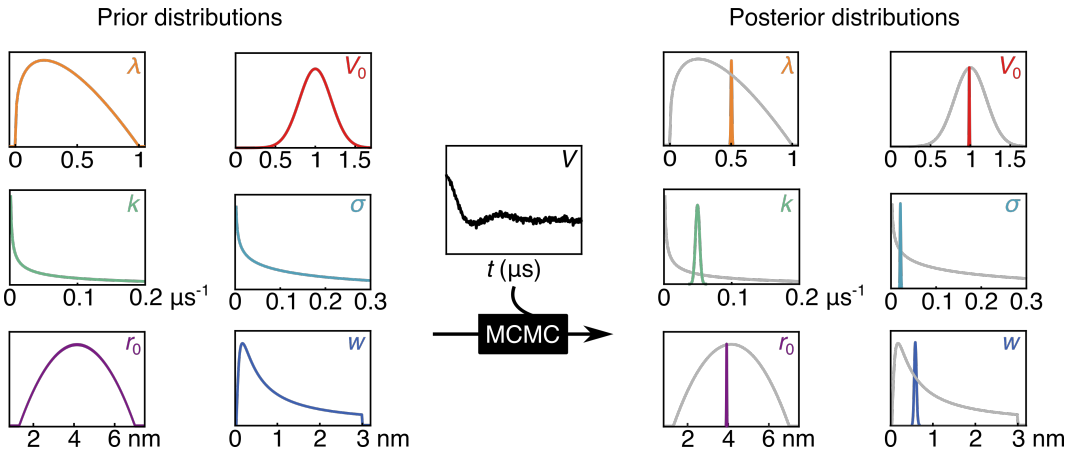

Figure 2.1: Visualization of Bayesian inference. The priors for all parameters (modulation depth (λ), echo amplitude (V_0), background decay constant (k), noise (σ), Gaussian center (r_0), and Gaussian width (w)) and the raw data are input into the 'black box' process of MCMC and the output is the marginalized posterior distribution for each parameter, given the data. The marginalized posteriors represent a probability distribution of possible values for each parameter. Distribution 3992 from the synthetic T4L test set was used as the true distribution from which parameters were derived[19].

Thus, the model has $N = 6$, 10, 13, and 16 parameters for a one-, two-, three-, and four-Gauss distribution, respectively. This number is much smaller than the typical number of experimental data points, which can be several hundred.

2.2.1 Priors

To calculate the posterior from Eq. (1.14), both a likelihood and a full prior distribution must be specified. It is essential to define the prior such that it appropriately captures prior information, about the parameters but is diffuse enough to not introduce unwarranted bias into the analysis. In the following, we describe our choices of priors for all model parameters. For the parametric model, it is necessary to change the notation slightly from what is shown in Eq. (1.14) to incorporate model choice M and any background information I as conditions of the calculated probabilities. The priors are shown on the left-hand side of Fig. 2.1.

Modulation depth λ. The minimal information about λ is that it lies somewhere between 0 and 1, irrespective of the sample and the spectrometer. The associated prior can be represented as a uniform distribution

$$p(\lambda|M, I) = \text{Uniform}(\lambda; 0, 1) \tag{2.2}$$

with lower bound 0 and upper bound 1. If additional information about the spectrum (e.g. nitroxide at Q-band) and the excitation profile of the pump pulse (e.g. a 10 ns π pulse) is available, the modulation depth can be estimated. Then, a more focused probability distribution can be constructed that is centered at this estimated value and has sufficient spread to capture uncertainty in the information. We use a beta distribution

$$p(\lambda|M, I) = \text{Beta}(\lambda; 1.3, 2) \tag{2.3}$$

where the two arguments are the shape parameters α and β, respectively. For our test cases, the choice of prior made no difference on the obtained posterior distributions. This indicates that the data itself contained strong information about λ.

Echo amplitude V_0. Before analysis, we rescale the signal, $\boldsymbol{V} = \boldsymbol{V}/\max(\boldsymbol{V})$. Therefore, for the rescaled signal, we know V_0 has a positive value near 1 as long as the noise is dominating. We capture this information with a truncated Normal distribution bounded above zero with a standard deviation of 0.2

$$p(V_0|M, I) = \text{Bnd}(\text{Normal}(V_0; 1, 0.2^2), 0) \tag{2.4}$$

Background decay rate k. As prior for k, we use the gamma distribution (which ensures k is non-negative)

$$p(k|M, I) = \text{Gamma}(k; 0.5, 2\,\mu\text{s}^{-1}) \tag{2.5}$$

where the two arguments are the shape and the rate parameter, respectively. This distribution has significant density for $k < 0.1\,\mu\text{s}^{-1}$, corresponding to $\lambda c < 0.1\,\text{mM}$. If the spin concentration is known, a tighter prior can be formulated.

Noise level σ. Without taking the data into account, an appropriate prior on the noise standard deviation for the rescaled signal is one which skews heavily towards zero and diminishes towards 1. We choose a gamma distribution

$$p(\sigma|M,I) = \mathrm{Gamma}(\sigma; 0.7, 2) \tag{2.6}$$

Gaussian centers $r_{0,i}$. We choose a prior with significant density around the most common distances (2–6 nm) and diminished probability outside, using a beta distribution

$$p(r_{0,i}|M,I) \propto \mathrm{Beta}\left(\frac{r_{0,i} - r_{\min}}{r_{\max} - r_{\min}}; 2, 2\right) \tag{2.7}$$

where $r_{\min}$ and $r_{\max}$ are 1.3 and 7 nm, respectively, and we disregard the normalization factor $1/(r_{\max} - r_{\min})$.

Gaussian widths w_i. Based upon general knowledge of spin–spin distance distributions in folded proteins, we choose the prior to skew towards values below 1 nm but with substantial probability at larger widths. A bounded inverse gamma distribution captures the desired shape

$$\begin{aligned} &p(w_i|M,I) \\ &= \mathrm{Bnd}(\mathrm{InvGamma}(w_i; 0.1, 0.2\,\mathrm{nm}), 0.05\,\mathrm{nm}, 3\,\mathrm{nm}) \end{aligned} \tag{2.8}$$

The distribution is truncated between 0.05 and 3 nm, since it is very unlikely to have distributions with widths beyond these bounds. If a sample is analyzed where these assumptions are possibly not satisfied, the boundaries can be adjusted.

Gaussian amplitudes A_i. Since we have no prior information on A_i, we use a flat m-dimensional Dirichlet distribution

$$p(A_1, \ldots, A_m|M,I) = \mathrm{Dirichlet}(\mathbf{1}_m) \tag{2.9}$$

where $\mathbf{1}_m$ is the m-dimensional vector of ones. The Dirichlet distribution automatically satisfies $0 \leq A_i \leq 1$ and $\sum_{i=1}^{m} A_i = 1$.

2.2.2 MCMC sampling

Although the posterior is now formulated explicitly via Eqs. (1.14), (1.15), (1.16), and all the individual prior distributions, it is a multidimensional distribution so complicated that the multidimensional integrals necessary to determine its mean and other statistics are impossible to evaluate analytically. Therefore, numerical Markov Chain Monte Carlo (MCMC) sampling methods are used to generate a representation of the posterior in terms of a set of samples in the N-dimensional parameter space. An MCMC sampler generates a correlated chain of samples from the posterior, where each sample depends on the previous one. Once converged, the chain samples represent the posterior. Using this representation, calculating the above statistics and other analysis is straightforward.

In previous work on Bayesian analysis from the Stoll group, a Gibbs sampler [18] was used. This sampler requires specific forms of the priors and is very inefficient in high-dimensional parameter spaces.

In this work, we use the No-U-Turn Sampler (NUTS) [34], a very efficient and self-tuning Hamiltonian Monte Carlo (HMC) sampler [35], as implemented in the Python-based probabilistic programming package PyMC3, version 3.8 [36]. All MCMC simulations were run on standard laptop computers. The Python code that implements the DEER model and the MCMC sampling can be found at `https://github.com/StollLab/dive`.

We use NUTS to generate 5–8 independent MCMC chains. The chains are initialized with different starting points that are randomly sampled from the prior distribution. These points are then propagated for 5,000 steps to tune the sampler. The tuning steps are discarded. The chains are then propagated for 20,000–80,000 steps to generate a large number of posterior samples. The exact sampler settings are noted in the figure captions. The chains are long enough to provide convergence, as assessed via the rank-normalized split $\widehat{R}$ statistic [37, 38, 39]. $\widehat{R}$ is calculated for each parameter separately and compares the between-chains and within-chain variances to determine whether the chains have reached equilibrium for that parameter (corresponding to $\widehat{R} \approx 1$) or not ($\widehat{R} > 1$). In our approach, sampling was

continued until $\widehat{R} < 1.01$ for all parameters.

When using MCMC sampling with mixture models such as Eq. (1.8), one encounters a phenomenon known as label switching. For example, switching the labels of the two Gaussians in a two-Gauss distance distribution changes the location in parameter space ($\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$), but does not affect the distance distribution ($P(\boldsymbol{\theta_1}) = P(\boldsymbol{\theta}_2)$) nor the likelihood or the posterior. This renders the posterior multi-modal, complicating both the sampling and the analysis of the posterior. Different approaches exist to prevent label switching [40, 41]. In this work, we take an approach similar to on-line relabeling [40] and enforce the constraints $r_{0,1} \leq r_{0,2} \leq \cdots \leq r_{0,m}$ after every sample to restrict the parameter space [41]. This technique worked well in most cases that we encountered, and typically provided clean uni-modal marginalized posteriors. Occasionally, due to the imposed constraints, chains get stuck in regions with $r_{0,i} \approx r_{0,j}$, corresponding to the coalescence of two basis functions. Such chains are easily identified via how they degrade $\widehat{R}$ and are removed before further analysis.

2.2.3 Posterior analysis

After convergence, the pooled samples from all chains represent the full N-dimensional posterior $p(\boldsymbol{\theta}|\boldsymbol{V}, M, I)$. Due to its large dimensionality, it is not possible to visualize it directly. Instead we examine each parameter individually using a marginalized posterior, which is obtained by integrating the full posterior over all other parameters. This integral is approximated by generating a histogram of the parameter values from all samples and conducting a kernel density estimation of the histogram, smoothing with a Gaussian with a line width of 1/5 of the standard deviation of the parameter values. This results in a one-dimensional distribution that can easily be plotted and summarized in terms of the mean, mode, and spread. On the right of Fig. 2.1, the marginalized posteriors are shown in color, together with the priors in gray. In this case, they are much narrower than the corresponding priors. The spread of the posterior distribution is both a qualitative and quantitative measure of inferential uncertainty.

However, marginalization discards all information about correlation between parameters. As will be shown, many of the parameters in our model are correlated. Therefore, we also display and examine two-dimensional marginalized posteriors between pairs of parameters (see Sec. 2.3).

Finally, we will additionally visualize the results of the Bayesian inference in terms of a small set of posterior predictive samples of the noise-free signal, $V_M(\boldsymbol{\theta}^{(i)})$, and the distance distribution, $P(\boldsymbol{\theta}^{(i)})$. Here, $\boldsymbol{\theta}^{(i)}$ represent a random sample from the pooled MCMC samples, with i indicating the associated chain sample index.

2.3 Basic illustration

In this and the next section, we illustrate the probabilistic analysis method on several synthetic distributions of increasing complexity. All of the synthetic data are based on distributions taken from the large simulated T4 lysozyme (T4L) test data set published by Edwards and Stoll in 2018 [19, 42]. The distributions in this test data set were generated computationally from an in silico spin-labeled crystal structure of T4L. The indices of the chosen distributions are given in the figure captions. We take the distributions from the test data set as ground truth.

Figures 2.2 and 2.3 show the first example. It uses a distance distribution that resembles a single Gaussian. Two noisy signal traces generated from this distance distribution are shown in Fig. 2.3A. The trace indicated as $V_{\text{good}}(t)$ (blue) has favorable values for the modulation depth, background decay rate, trace length, and noise level ($\lambda = 0.5, k = 0.05$ µs$^{-1}, t_{\max} = 3.2$ µs$, \sigma = 0.02$). Comparatively, the trace $V_{\text{poor}}(t)$ (green) has less ideal values for all of these parameters ($\lambda = 0.2, k = 0.2$ µs$^{-1}, t_{\max} = 1.6$ µs$, \sigma = 0.05$), providing a challenging case with higher uncertainty.

Figure 2.2 summarizes the results of the Bayesian analysis for both cases. The top row shows the marginalized posteriors for each parameter for the poor case (green) while the first column shows the same for the good case (blue). Both cases are shown on all plots, one of them grayed out. The dashed lines for r_0 and w indicate values obtained by directly

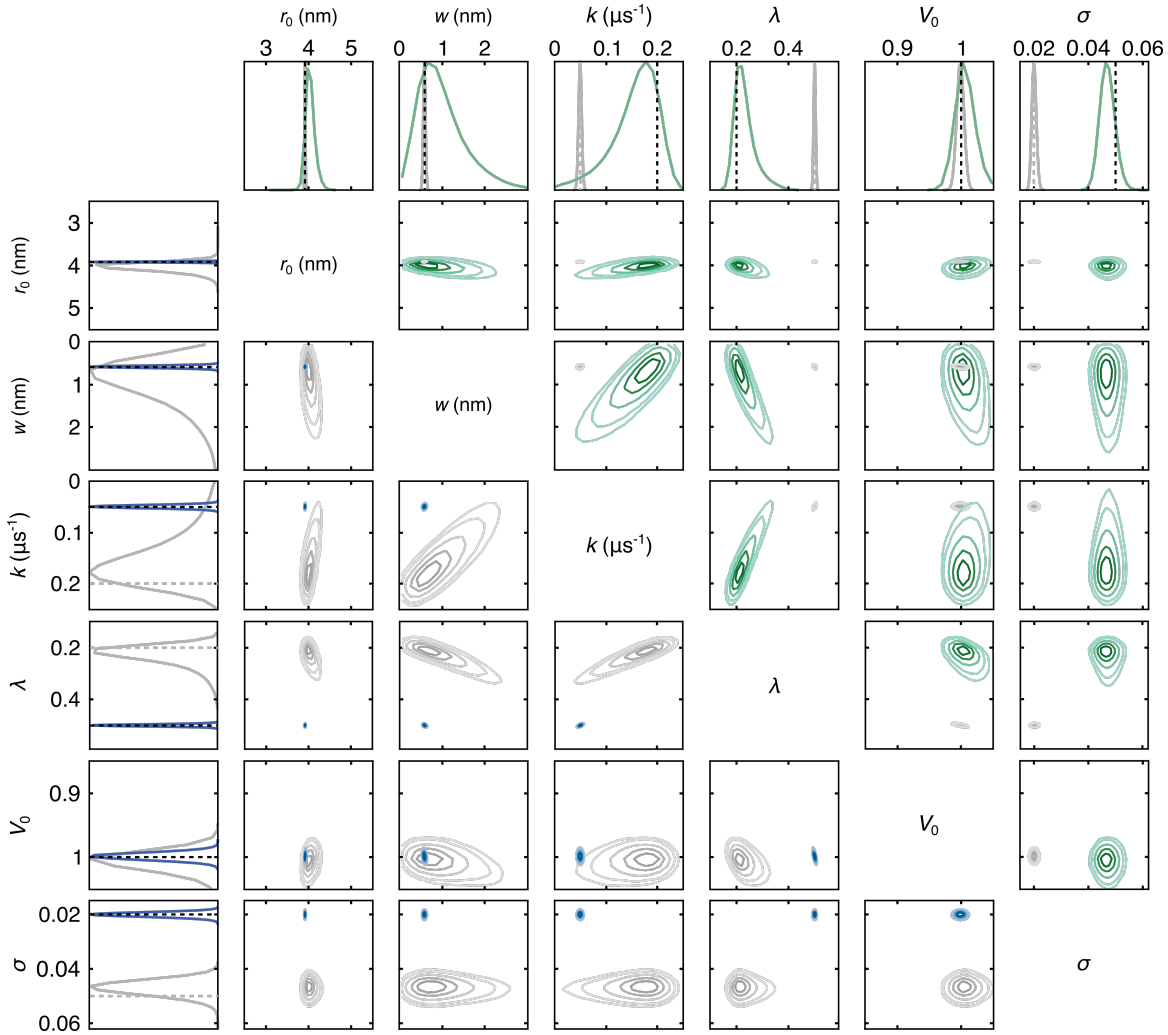

Figure 2.2: Plots of the marginalized parameters posteriors of the two single-Gauss test cases and all pairwise correlations between parameters. The results for the poor test case are indicated in green and those for the good test case are in blue. The dotted lines reference the true parameter values. Both cases are shown in every subplot for comparison. The top row and upper triangle show the poor test case results (green), and the first column and lower triangle show the good test case results (blue). Both test cases are generated from distribution 3992 from the Edwards/Stoll T4L test set. Contour levels are at 0.05, 0.1, 0.25, 0.5, and 0.75 of the maximum of the distribution. The MCMC simulation was run with 5 chains with 20,000 posterior samples each.

fitting a Gaussian to the synthetic distribution, and for the other parameters they indicate the values used in generating the signal. For the longer and less noisy trace, the ground-truth parameters are recovered accurately, and there is very little uncertainty about the inferred parameters, as represented by the narrowness of the distributions. The analysis of the shorter and noisier trace yields posterior modes of each parameter near the expected values. The distributions, however, are much broader, reflecting the detrimental effect of the larger noise level and shorter trace length. The distributions are asymmetric, particularly for w, k, and λ.

The rest of the plot shows the marginalized posteriors of all parameter pairs for both cases, revealing correlations between parameters. Both the upper and lower triangle show the same results, mirroring one another and highlighting one case over the other. Again, the distributions for the good test case are significantly narrower than the ones for the poor case. The angle by which the distribution is skewed from a horizontal or vertical direction indicates the degree of correlation between the two parameters. The results show strong correlations for (k, λ), (λ, w), and (w, k) for the poor trace. k is negatively correlated with w and λ, whereas λ and w are positively correlated.

While the parameter posteriors most directly show the outcome of the Bayesian analysis, they are generally not the main quantities of interest. The most desired quantities are the distance distributions, upon which possible structural conclusions will be based, and the model fit in the time domain.

To show these quantities, we use ensembles of posterior predictions. We draw a small set of random parameter vectors $\boldsymbol{\theta}^{(i)}$ (typically 30–100) from the pooled MCMC chain samples, calculate the associated distance distributions $P(\boldsymbol{\theta}^{(i)})$ and noise-free time-domain signals $V_M(\boldsymbol{\theta}^{(i)})$. This approach is conceptually equivalent to the "spaghetti plots" utilized to visualize predicted hurricane trajectories in weather forecasts.

The calculated time-domain signal ensembles for both cases are shown in Fig. 2.3A against the raw data and show a good match of model and data. The associated ensemble of residuals in Fig. 2.3B show no systematic deviation from zero, indicating that the models

are adequately representing the data. The ensemble of MCMC-based distance distributions is plotted in Fig. 2.3C. For the good test case, there is little scatter among the distributions since the Bayesian analysis recovers the parameters with little uncertainty. For the poor test case, a significant spread of positions and widths is apparent among the members of the ensemble. Consequently, the conclusion about the position or the width of the distribution cannot be precise.

The examination of a posterior-based ensemble of distributions is essential for fully visualizing the information contained in the data and extracted by the Bayesian analysis. Plotting only a single distribution, for instance the one corresponding to the maximum of the posterior (MAP), is misleading, as this discards all information about uncertainty. Also, despite being the point with the highest posterior probability, the MAP is not representative of the posterior in high-dimensional models [35].

In a traditional least-squares fitting approach, a single distribution corresponding to the maximum of the likelihood (Eq. (1.15)) is shown together with error bands based on the curvature around this maximum. These error bands are intended to capture uncertainty, but can be misleading for several reasons. (1) They assume a symmetric Gaussian probability distribution, which is not the case in general. Even in the simple example in Fig. 2.2, many distributions are asymmetric. (2) They do not capture the strong correlations between P at different distances, which is due to the predetermined shape and the normalization of P. (3) They do not capture correlations between r_0 and w. (4) They are shown for a single distance distribution, implying overcertainty about shape and location. Limitation (1) can be overcome by a bootstrapping analysis [29], but the other limitations remain. A distribution ensemble as shown in Fig. 2.3C does not suffer from these drawbacks and therefore constitutes a more complete and prudent way of visualizing the range of distance distributions compatible with the data (given the model).

The lower part of Fig. 2.3 summarizes the results of the parameter inference, in a form more condensed than Fig. 2.2. Panels D and E are the matrices of the pairwise Pearson correlation coefficients for the good case and poor case, respectively. White indicates no

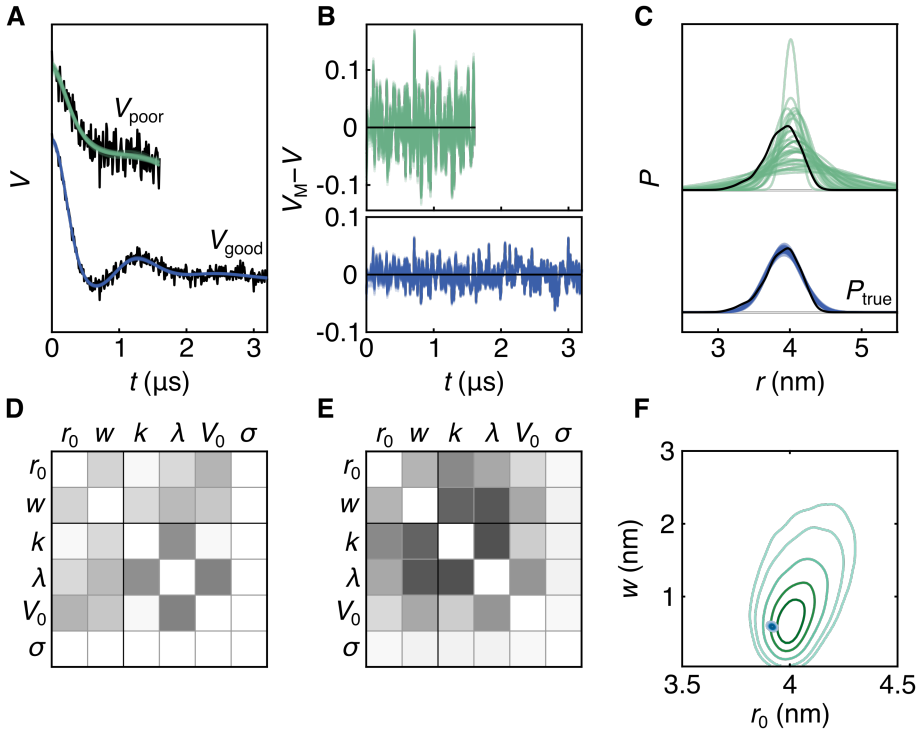

Figure 2.3: Bayesian inference with a single-Gauss distribution model. A: The two data realizations (black) overlaid with the fits predicted from an ensemble of 30 MCMC samples (color). B: Residuals for both traces for all 30 samples (color). C: Ensemble of 30 distributions (color) from MCMC samples plotted with P_{true} (black), distribution 3992 from the synthetic T4L test set [19]. D: Correlation matrix of all parameters for the good trace, white indicating no correlation and black indicating full correlation. E: Correlation matrix of all parameters for the poor trace. F: 2D marginalized posteriors for Gaussian width and center for both cases.

correlation and black indicating full positive or negative correlation. Panel F shows the marginalized 2D posterior for the width and position of both test sets. Among all the correlation plots in Fig. 2.2, this is the most relevant for inference about the distance distribution. It shows a strong difference in the widths for the two cases. For the poor case, the distribution is visibly asymmetric and correlated.

2.4 Multimodal distributions

Most spin–spin distance distributions encountered in DEER spectroscopy of proteins are asymmetric and multimodel and poorly approximated by a single Gaussian. Therefore, we next apply the method to distributions of higher complexity.

First, we analyze a noisy time trace generated from a bimodal distribution from the T4L test set, with two distinct modes, one significantly weaker than the other. The noisy signal is shown in Fig. 2.4A (black), and the underlying distribution is shown in Fig. 2.4C (black). The results of the Bayesian analysis using a two-Gaussian model ($m = 2$ in Eq. 1.8) are also shown in Fig. 2.4. Panels A–C show a subset of posterior samples randomly selected from the MCMC chains. The model works quite well at representing the signal, as evidenced by the good overlap with the data (panel A) and by the absence of any systematic deviation in the ensemble of residuals (panel B). The distribution ensemble is in good agreement with the the true distribution (panel C), and the uncertainty is relatively low given the small scatter of the MCMC distributions. There is increased uncertainty in the region of the minor peak, due to the low amplitude in this region, so that conclusions about location or width of this secondary feature are not precise. Position and width of the major peak are inferred with high confidence, but there is some uncertainty in its amplitude, a consequence of the uncertainty in the amplitude of the minor peak and $A_1 + A_2 = 1$.

Again, a valuable output of this method are the parameter correlations, visually summarized in panel D. All distribution parameters ($r_{0,i}$, w_i, A_i) are heavily correlated with one another while there is little correlation between distribution and time-domain parameters. The strongest correlation among time-domain parameters is between V_0 and λ. Panels E and

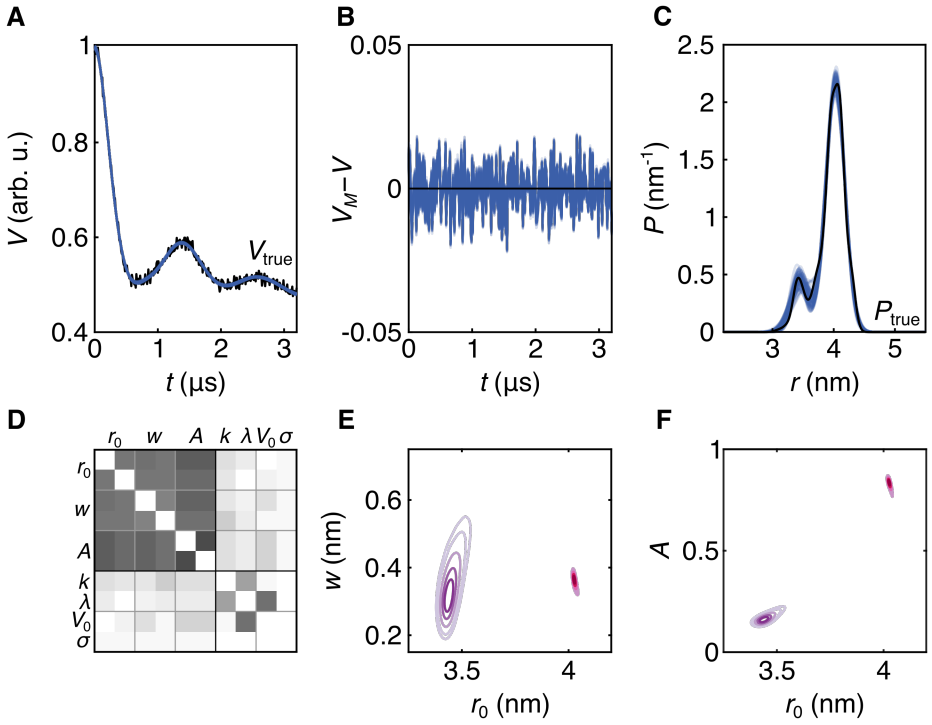

Figure 2.4: Bayesian inference with a two-Gaussian distribution model. A. Time-domain trace 36787 from synthetic data set ($V_{\text{true}}(t)$) (black) and 100 samples from the MCMC chains (blue). B. Residual analysis for the 100 MCMC samples (blue). C. Ensemble of 100 MCMC distributions (blue) and distance distribution 3223 from synthetic data set (black). D. Correlation matrix of all parameters, white indicating no correlation and black indicating full positive or negative correlation. E. 2D marginalized posteriors for widths and positions. The two Gaussian components are distinguished by color. F. 2D marginalized posteriors for amplitudes and positions, with identical coloring as E. The MCMC simulation was run with 5 chains with 30,000 posterior samples each.

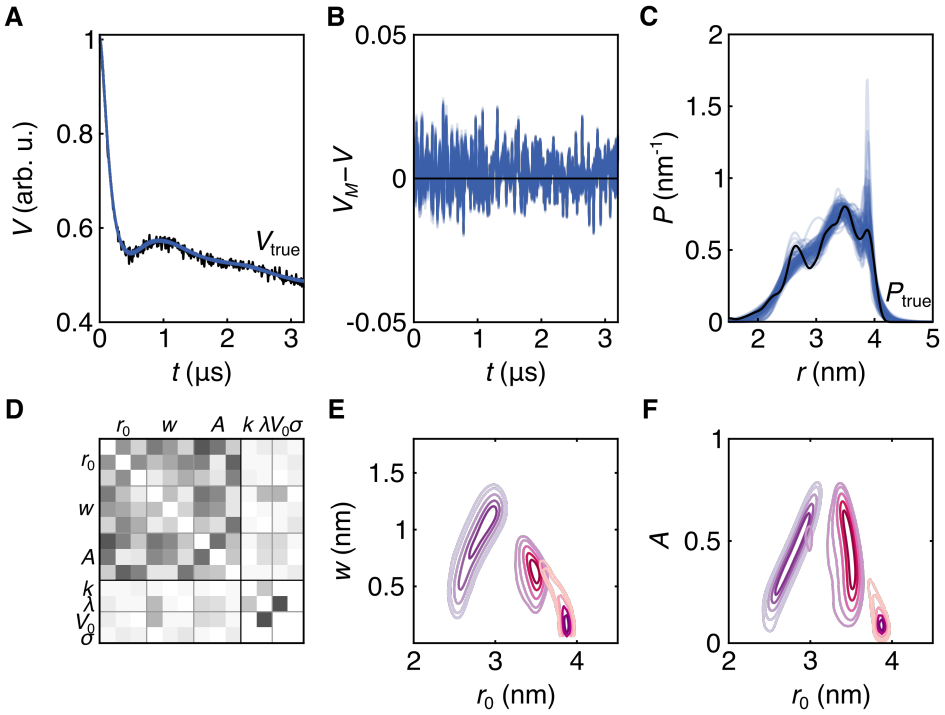

Figure 2.5: Bayesian inference with a three-Gaussian distribution model. A. Time-domain trace 51412 from synthetic data set ($V_{\text{true}}(t)$) (black) and 100 MCMC samples (blue). B. Residual analysis for all MCMC samples (blue). C. MCMC ensemble (blue) and distance distribution 4428 from synthetic data set (black). D. Correlation matrix of all parameters, white indicating no correlation and black indicating full positive or negative correlation. E. 2D marginalized posteriors for widths and positions. Individual components are distinguished by color. F. 2D marginalized posteriors for amplitudes and positions, with identical coloring as E. The MCMC simulation was run with 6 chains, with 40,000 posterior samples each.

F show the obtained 2D marginalized posteriors for the distance distribution parameters. As the position of the short-distance component increases, we see a corresponding increase in its width and in its amplitude. The posteriors for the longer-distance component are very tight, indicating high certainty. There is slight anticorrelation between position and both width and amplitude.

The next synthetic example, shown in Fig. 2.5, is a challenging, broad distribution with several poorly-resolved modes with similar intensities. The Bayesian analysis was conducted using a three-Gaussian distribution model. The ensemble of MCMC samples drawn from

the posterior are shown in Fig. 2.5A, the ensemble of residuals in B, and the distribution ensemble in C. Although the model fits the data well (panels A and B), there is substantial scatter in the ensemble of distance distributions (panel C). This shows that the uncertainty about the shape is significant over the entire distribution. The long-distance edge at 4 nm is fairly well defined, but the location of the short-distance edge around 2 nm is less clear. The distribution ensemble shows that the data are not strong enough (i.e. too noisy or too short) to either exclude or confirm modes at 2.8 and 3.9 nm—although they reveal the possibility of such modes. Also, the location of the mode around 3.4 nm is uncertain. This shows again that it is important to consider a distribution ensemble instead of a single distribution in order to make robust conclusions.

The correlations between distribution parameters (Fig. 2.5D) vary but, as in the previous examples, the correlation between V_0 and λ is substantial. The 2D marginalized posteriors for the distribution parameters (Fig. 2.5E–F) show that there is significant uncertainty about the positions, widths, and amplitudes of the three Gaussian components. The reason for this uncertainty is that in the time-domain signal the noise level is significant relative to the shallow amplitude of the oscillations after the initial drop. The oscillations are shallow because of the large width of the underlying distribution.

2.5 Experimental example

So far, we have utilized synthetic data where we were able to pick ground-truth distributions that can be well approximated by a certain number of Gaussians. In this section, we demonstrate the method with experimental data, where this is not possible.

This brings up an important aspect of DEER data analysis, model checking and comparison [32]. When the ground-truth distribution is not known, is is important to check whether a multi-Gauss model fits the data and to compare the quality of models with different numbers of Gaussians.

Figure 2.6A shows the DEER trace (black) that was collected to determine inter-subunit distances in SthK, a tetrameric bacterial cyclic nucleotide-gated (CNG) ion channel [13].

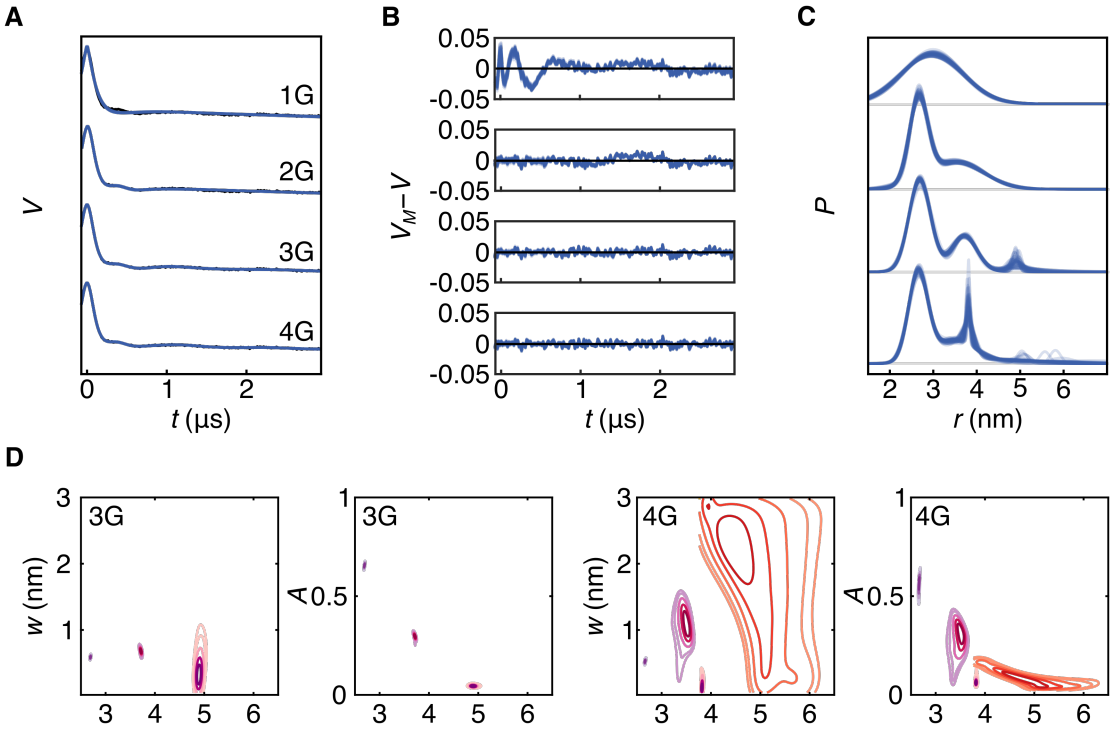

Figure 2.6: Bayesian inference using multi-Gauss models on a DEER trace obtained from the ion channel SthK D261R1 [13]. A. The raw experimental DEER data (black) and 50 MCMC ensembles from the analysis using 1-, 2-, 3-, and 4-Gaussian models (blue). B. Residuals based on the MCMC ensemble. C. Distance distribution ensembles for each of the four models. D. 2D marginalized posteriors for widths and positions for 3G and 4G models. Individual components are distinguished by color. Number of chains and posterior samples per chain for the MCMC simulations: 8 and 20,000 (1G), 5 and 30,000 (2G), 5 and 40,000 (3G), 5 and 80,000 (4G).

Here we analyze the DEER trace using multi-Gauss models with 1 to 4 Gaussians (abbreviated as 1G, 2G, 3G, and 4G). Resulting ensembles of MCMC samples of signals, residuals, and distributions are shown in blue in Fig. 2.6A-C. Based on the systematic deviation of its residuals at early times, it is clear that the 1G model is inappropriate, even though it provides an apparently precise $P(r)$ given how tight the distribution ensemble is. This is consistent with the fact that the ion channel is a homotetramer, and the distribution is expected to be at least bimodal. The 2G model matches the data better, but shows some systematic deviation in the residuals as well. The 3G and 4G models describe the signal similarly well, as their residuals are visually free of systematic deviations.

The distance distribution ensembles in Fig. 2.6C reveal some important details. In the 3G model, two components are quite certain, but the third at about 5 nm has large uncertainty in its width. In the 4G model, three of the components model the distribution below about 4 nm, and the fourth component is at long distances and is very uncertain both in width and position. This is also evident from the 2D marginalized posteriors shown in Fig. 2.6. Whereas all parameter posteriors are relatively tight for the 3G model, the posterior for the long-distance component of the 4G model is very delocalized in position and width. Overall, this suggests that the 4G model is likely overfitting the data.

Within the framework of Bayesian inference, a formal approach for model comparison is available. Two models M_1 and M_2 can be compared via the ratio $p(M_1|\boldsymbol{V}, I)/p(M_2|\boldsymbol{V}, I)$ of their posterior probabilities [43, 32]. This ratio is called the posterior odds and is calculated via

$$\underbrace{\frac{p(M_2|\boldsymbol{V}, I)}{p(M_1|\boldsymbol{V}, I)}}_{\text{posterior odds}} = \underbrace{\frac{p(\boldsymbol{V}|M_2, I)}{p(\boldsymbol{V}|M_1, I)}}_{\text{Bayes factor } B_{2,1}} \cdot \underbrace{\frac{p(M_2|I)}{p(M_1|I)}}_{\text{prior odds}} \tag{2.10}$$

where the first ratio on the right-hand side is known as the Bayes factor, and the second ratio is called the prior odds. The posterior odds summarize how much one model is favored over the other in light of the data and prior information. The Bayes factor represents how likely the data are assuming one model vs. assuming the other. The prior odds quantify the odds for or against one model, prior to taking into account the data. In most applications, this

ratio is set to one (encoding no preference)[32]. Then the Bayes factor equals the posterior odds and can be used to quantify how the data speak for one model over the other. As a rule of thumb, $\log_{10} B_{2,1} > 8$ can be seen as relatively strong indication for M_2 over M_1 [44]. In a comparison of two models that belong to the same family of models, or when one model is a superset of the other (e.g. the 3G model is a superset of the 2G model), the Bayes factor penalises complicated models that might be prone to overfitting [32]. In such cases, if the data are better explained by the simpler model, the Bayes factor in favor of the more complex model is typically small.

The Bayes factors $B_{m,1}$ for all m-Gauss models relative to the 1G model are shown in Fig. 2.7. They show that the 2G model is clearly preferable over 1G, and give preference for 3G and 4G over 2G ($\log_{10} B_{3,2} \approx 25$ and $\log_{10} B_{4,2} \approx 30$). However, there is only a small difference between 3G and 4G, and the Bayes factor $\log_{10} B_{4,3} \approx 5$ indicates that the 4G model does not describe the data much better than the 3G model. In combination with the delocalized posterior, this is a strong indication that the 4G model is overfitting the data.

Neither residuals nor Bayes factors alone are enough for a complete analysis. While the latter compare two models and help with identifying overfitting, they do not contain information on whether the chosen models is appropriate and gives a good fit to the data. Instead, an inadequate model with systematic misfitting can be diagnosed through the residuals. Additionally, even if a model fits the data well and has the largest Bayes factor among a set of models, it might still be physically inappropriate, and considerations outside the Bayesian analysis framework must be used to determine a more appropriate model.

The model comparison outlined here allows assessment of an appropriate number m of Gaussians. This is analogous to selecting an appropriate α regularization parameter in Tikhonov regularization [19]. However, although these selection methods are quantitative, they are not unique and therefore not fully objective.

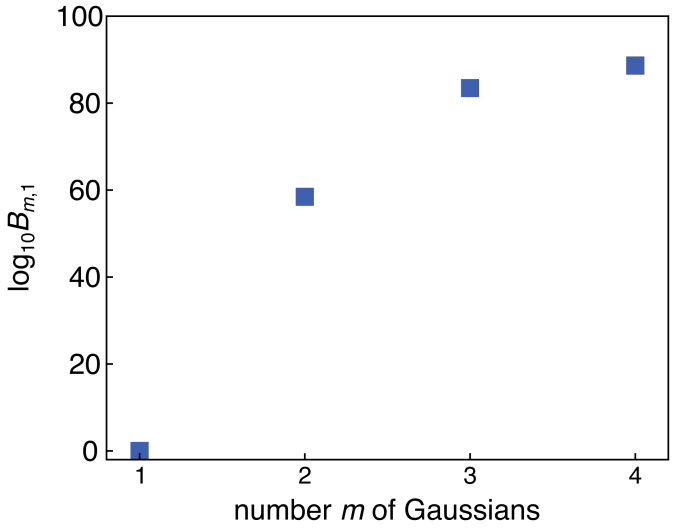

Figure 2.7: Bayes factors comparing the four models from Fig. 2.6 to the 1G model.

2.6 Conclusions

The Bayesian inference approach presented here fully quantifies the uncertainty in model parameters obtained from fitting DEER data using a parametric multi-Gauss distribution model. Its advantage over a least-squares fitting approach is that it provides the posterior distribution that completely quantifies model parameter uncertainty, whereas least-squares fitting only determines a point estimate with confidence intervals. The posterior allows the analysis of spread and asymmetry in the parameter distributions, and of correlations between model parameters.

We showed that a small ensemble of distributions drawn from the MCMC samples is well suited to visually capture both uncertainty and correlation in distance distribution plots. Therefore, it should be preferred over error bands. We also illustrated how to use a combination of residuals, parameter posteriors, and Bayes factors to help with model comparison and to identify under- and overfitting. This can be expanded to include other modern Bayesian methods for model comparison as well [32].

Although we have presented the probabilistic approach using an exponential model for

$V_{\text{inter}}(t)$ and a multi-Gauss model for $P(r)$, the approach is very general. It is applicable to any other parametric models of $V_{\text{inter}}(t)$ and $P(r)$. The background can be extended to include fractal dimensions [45] and to allow for excluded-volume effects [46]. The distance distribution basis functions are not limited to Gaussians; other functions such as 3D Rice functions can be used [47], particularly in cases where there is no clear fit to a sum of Gaussians via residual and Bayes factor analysis. If the noise level is known experimentally (e.g. from analyzing the variance among a series of sequentially acquired traces), then it can be fixed and omitted as a model parameter. In the case the noise level is not constant across the trace, a more sophisticated noise model (and likelihood) can be included.

The probabilistic inference methodology used here for multi-Gauss models is equally applicable to the non-parametric models for $P(r)$ used in Tikhonov regularization. The implementation of which is shown in the next chapter along with a quantitative and systematic comparison between multi-Gauss and Tikhonov models within a Bayesian framework.

Chapter 3

BAYESIAN PROBABILISTIC ANALYSIS OF DEER SPECTROSCOPY DATA USING NONPARAMETRIC DISTANCE DISTRIBUTION MODELS

This chapter expands upon the work presented in the previous chapter by introducing a Bayesian probabilistic inference approach to analyze DEER data assuming a nonparametric distance distribution with a Tikhonov smoothness prior. The method uses Markov chain Monte Carlo (MCMC) sampling with a compositional Gibbs sampler to determine a posterior probability distribution over the entire parameter space, including the distance distribution, given an experimental dataset. This posterior contains all the information available from the data, including a full quantification of the uncertainty about the model. The corresponding uncertainty about the distance distribution is captured via an ensemble of posterior predictive distributions. Several experimental examples are presented to illustrate the method and its relative advantages compared to other methods[1].

3.1 Motivation

As stated in Chapter 2, a robust process for uncertainty analysis with respect to the fitted distance distribution is challenging, but is crucial for making structural conclusions. This chapter aims to extend the uncertainty analysis method presented in the last chapter for a least-squares fitting approach using a nonparametric model for the distance distribution.

For Tikhonov regularization, partial uncertainty analysis is commonly conducted by manually varying some parameters in the analysis (background, modulation depth, noise) and

[1]This chapter is based on the publication in preparation for *The Journal of Physical Chemistry A* by Sweger S. R., Pribitzer S., Zha L. and Stoll S. titled "Probabilistic Inference of Nonparametric Distance Distributions in DEER Spectroscopy".

summarizing the sensitivity of the extracted distance distribution to these parameters into error bands around the fitted distribution [25, 48]. Another partial approach based on Bayesian inference was implemented in prior work from the Stoll group, quantifying the uncertainty in the distribution due to the noise in the signal [18]. Unfortunately, this work required *a priori* background correction and cannot incorporate parameters beyond noise. Another method to obtain confidence intervals for both approaches is bootstrapping, which generates an ensemble of distributions by analyzing a large number of synthetically generated hypothetical signals based on the fitted model [29].

In this chapter, we present a Bayesian inference approach for analyzing a DEER trace in terms of a non-parametric distance distribution with Tikhonov smoothing and an exponential background. The Bayesian inference approach has been implemented previously for use in protein structural analysis for both NMR and EPR [49, 18, 50]. Building on our previous work, the method models and analyses the raw DEER data directly and yields a full posterior probability distribution over all model parameters, providing complete quantitative information about uncertainty and correlations for all parameters, without any limiting assumptions. We also introduce distribution ensembles to more correctly represent uncertainty about the distance distribution, including correlations which are neglected when using visualizations based on error bands.

The chapter is structured as follows. Section 3.2 presents the model, section 3.3 outlines the inference methodology, and section 3.4 shows one example using synthetic data and several examples using experimental data, making comparisons to model selection, distance axis dependence, and other methods for assessing uncertainty. Finally, section 3.5 discusses the relation, advantages, disadvantages of this method in comparison to others.

3.2 Model setup

As stated in chapter one, we extract the spin–spin distance distribution $\boldsymbol{P}$ from Eq. (1.7) using Tikhonov regularization by minimizing the objective function shown in Eq. (1.11). The full set of parameters for the Tikhonov regularization approach using a nonparametric model

for the distance distribution is: the distance distribution vector penalized for roughness $\boldsymbol{P}_\alpha$ using the smoothness hyperparameter δ (related to the Tikhonov smoothness parameter α by $\delta = \alpha^2\tau$), the modulation depth λ, the end point of the background decay $B_{\text{end}} = V_{\text{inter}}(t_{\max})$, the noise precision $\tau = 1/\sigma^2$, and the overall amplitude V_0 [18]. We indicate the parameter set as

$$\boldsymbol{\theta} = \{\boldsymbol{P}_\alpha, \lambda, B_{\text{end}}, \tau, V_0\} \tag{3.1}$$

Based on Eqs. (1.7) & (1.11), the likelihood function in Eq. (1.14) is

$$\begin{aligned} p(\boldsymbol{V}|\boldsymbol{\theta}) &= \text{Normal}\big(\boldsymbol{V}; \boldsymbol{K}\boldsymbol{P}, \tau^{-1}\big) \\ &\propto \exp\left(-\frac{1}{2}\tau\|\boldsymbol{V} - \boldsymbol{K}\boldsymbol{P}\|^2\right) \end{aligned} \tag{3.2}$$

We take the prior as a product of independent distributions over each parameter:

$$p(\boldsymbol{\theta}) = p(\boldsymbol{P}, \delta)p(\lambda)p(V_0)p(\tau)p(k) \tag{3.3}$$

As the prior for $\boldsymbol{P}$, we encode our knowledge that $\boldsymbol{P}$ is element-wise non-negative and that it is expected to be smooth, i.e. P_i and P_j should be similar if the distances r_i and r_j are similar. For this, we write

$$p(\boldsymbol{P}, \delta) = p(\boldsymbol{P}|\delta)p(\delta)f(\boldsymbol{P}) \tag{3.4}$$

We include δ as an additional parameter in $\boldsymbol{\theta}$. The function $f(\boldsymbol{P})$ in Eq. (3.4) is an indicator function that equals one if all elements of $\boldsymbol{P}$ are non-negative, and zero otherwise.

For the smoothness prior, we assume a normal distribution

$$\begin{aligned} p(\boldsymbol{P}|\delta) &= \text{Normal}\left(\boldsymbol{0}, (\delta\boldsymbol{L}^{\mathrm{T}}\boldsymbol{L})^{-1}\right) \\ &\propto \exp\left(-\frac{1}{2}\delta\|\boldsymbol{L}\boldsymbol{P}\|^2\right) \end{aligned} \tag{3.5}$$

This distribution assigns high prior probabilities to smooth distributions (where $\|\boldsymbol{L}\boldsymbol{P}\|^2$ is small) and low prior probabilities to rough distributions (where $\|\boldsymbol{L}\boldsymbol{P}\|^2$ is large). This is motivated physically by the flexibility of the spin labels and the biomacromolecule to which

the pair of spin labels is attached. With $\boldsymbol{L}$ as defined, the endpoints of the distance range are neglected in this smoothness prior. They can be included by extending $\boldsymbol{L}$ with an additional first row with -2 and 1 as the first elements and with an additional last row with 1 and -2 as the last two elements.

For the hyperprior for the regularization parameter δ, we select a gamma distribution[18]

$$p(\delta) = \text{Gamma}(\delta; a_\delta, b_\delta) \tag{3.6}$$

with $a_\delta = 1$ and $b_\delta = 10^{-6}$. This gives a very broad distribution function that decays exponential with increasing δ.

For the priors for the other parameters, the prior distributions are as follows

$$\begin{aligned} p(\lambda) &= \text{Beta}(\lambda; 1.0, 2) && (3.7)\\ p(B_{\text{end}}) &= \text{Beta}(B_{\text{end}}; 1.0, 1.5\,\mu\text{s}) && (3.8)\\ p(V_0) &= \text{Bnd}(\text{Normal}(V_0; 1, 0.2^2), 0) && (3.9)\\ p(\tau) &= \text{Gamma}(\tau; a_\tau, b_\tau) && (3.10) \end{aligned}$$

with $a_\tau = 1$ and $b_\tau = 10^{-4}$. Though the priors are formulated in the model in terms of B_{end} and τ, we will show the posteriors converted to k and σ to remain consistent with the standard practices in DEER data analysis. The same applies for the hyperparameter δ which will be depicted as $\log_{10}(\alpha)$ throughout. The moves from priors to posteriors for each parameter are shown in Fig. 3.1.

With the above expressions, the posterior $p(\boldsymbol{\theta}|\boldsymbol{V})$ is fully defined. It has some structure that is important to recognize. (a) It is a gamma distribution in τ

$$p(\tau|\boldsymbol{V}, \boldsymbol{\theta}_{-\tau}) = \text{Gamma}(\tau; \tilde{a}_\tau, \tilde{b}_\tau) \tag{3.11}$$

where $\boldsymbol{\theta}_{-\tau}$ indicates the set of all parameters except τ. The distribution parameters are $\tilde{a}_\tau = a_\tau + n_t/2$ and $\tilde{b}_\tau = b_\tau + \|\boldsymbol{V} - \boldsymbol{KP}\|^2/2$, where n_t is the number of elements in $\boldsymbol{V}$. (b) It is a gamma distribution in δ

$$p(\delta|\boldsymbol{V}, \boldsymbol{\theta}_{-\delta}) = \text{Gamma}(\delta; \tilde{a}_\delta, \tilde{b}_\delta) \tag{3.12}$$

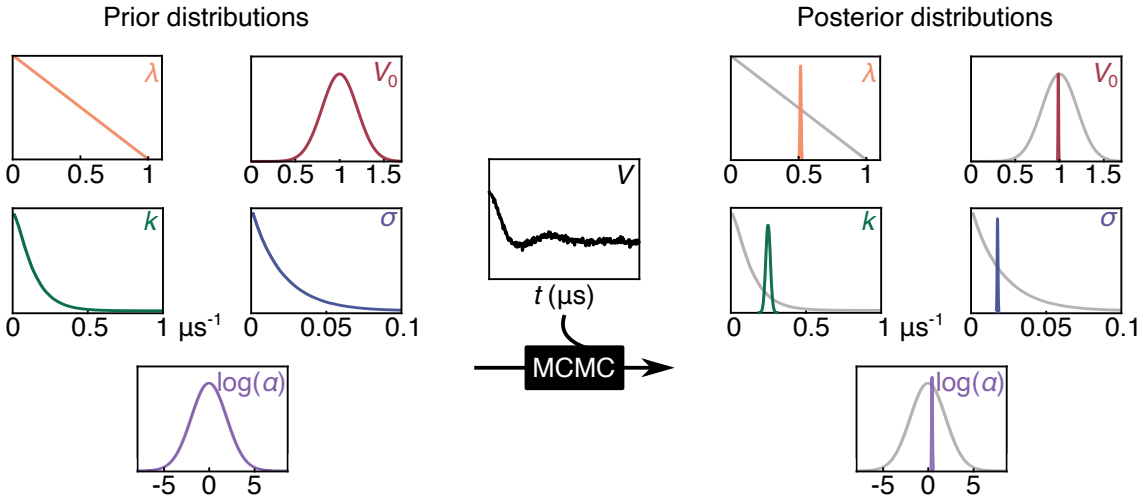

Figure 3.1: Visualization of the Bayes process. The priors for all parameters (modulation depth (λ), echo amplitude (V_0), background decay constant ($k = -1/t_{\max}\log(B_{\text{end}})$), noise ($\sigma = 1/\sqrt{\tau}$), Tikhonov smoothness parameter ($\alpha = \sqrt{\delta/\tau}$) and the raw data are input into the 'black box' process of MCMC and the output is the marginalized posterior distribution for each parameter. The marginalized posteriors represent a probability distribution of possible values for each parameter.

with $\tilde{a}_\delta = a_\delta + n/2$ and $\tilde{b}_\delta = b_\delta + \|\boldsymbol{LP}\|^2/2$ where n is the number of non-zero elements in $\boldsymbol{P}$. (c) It is a truncated multivariate normal distribution in $\boldsymbol{P}$

$$p(\boldsymbol{P}|\boldsymbol{V}, \boldsymbol{\theta}_{-\boldsymbol{P}}) = \text{Normal}\left(\boldsymbol{P}; \bar{\boldsymbol{P}}, \boldsymbol{\Sigma}\right) f(\boldsymbol{P}) \tag{3.13}$$

with center $\bar{\boldsymbol{P}} = \tau\boldsymbol{\Sigma}\boldsymbol{K}^{\mathrm{T}}\boldsymbol{V}$ and covariance matrix $\boldsymbol{\Sigma} = (\tau\boldsymbol{K}^{\mathrm{T}}\boldsymbol{K} + \delta\boldsymbol{L}^{\mathrm{T}}\boldsymbol{L})^{-1}$. We will make use of these structures for the sampling methodology in the next section.

3.3 Inference

The analytical form of the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{V})$ is intractable. In particular, it is not possible to evaluate integrals required to determine the mean, the variance, or marginalized distributions. Therefore, we resort to representing the distribution by a finite set of samples generated numerically, such that the density of samples in various regions in parameter space is proportional to the density (see Figure 3.2). These samples are then used to

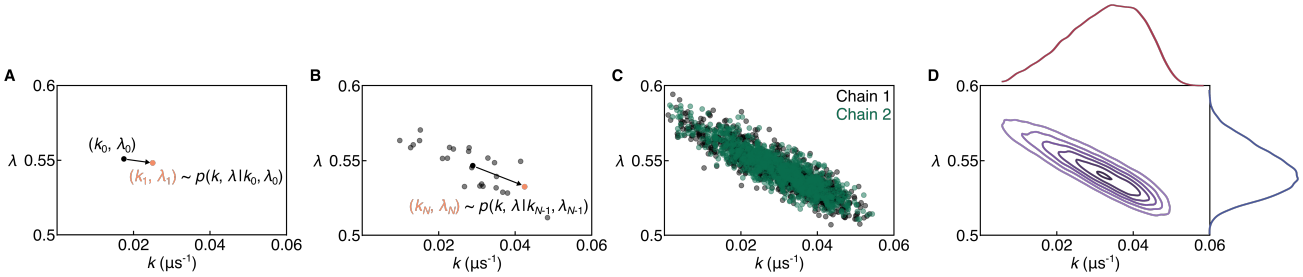

Figure 3.2: Principle of MCMC sampling of the posterior distribution. A. A random starting value is chosen from the joint posterior of k and λ (black). Next, another random sample is chosen from the probability distribution of k and λ given the starting value thus every step is dependent upon the previous step (blue). B. This process continues where each new step is a random draw from a probability distribution for k and λ given the position of the last step. C. Areas of higher probability are sampled more frequently until the chain converges. Multiple chains are run, indicated by color, to verify inter- and intra-chain convergence. D. The results of the sampling are the posterior distributions for each individual parameter or 2D posteriors to visualize correlation between parameters.

(approximately) calculate the aformentioned integrals and to construct visualizations.

To generate samples from the posterior, we use a Gibbs sampling approach that generates a Markov chain of samples [51]. Each sample i, containing $(\boldsymbol{P}_i, \tau_i, \delta_i, V_{0,i}, k_i, \lambda_i)$ is generated in several steps starting from the previous sample $i-1$, containing $(\boldsymbol{P}_{i-1}, \tau_{i-1}, \delta_{i-1}, V_{0,i-1}, k_{i-1}, \lambda_{i-1})$:

(1) Generate the i-th random sample of the precision τ from its full conditional posterior distribution Eq. (3.11) using the values of all other parameters from sample $i-1$.

(2) Generate the i-th random sample of the regularization parameter δ from its posterior distribution Eq. (3.12) using the new value for δ, but the previous values of all other parameters.

(3) Generate the i-th random sample of $\boldsymbol{P}$ from its full conditional posterior distribution Eq. (3.13) with the new values for τ and δ, but with the values for all other parameters from sample $i-1$. To generate a random sample from this distribution, we use a sampling method by Bardsley [52, 53] that relies on a non-negative least-squares solver to enforce the non-negativity constraint imposed by $f(\boldsymbol{P})$. Other algorithms for generating samples from

a truncated multivariate normal distribution are available in the literature.

(4) Generate the i-th random sample of the remaining parameters (V_0, B_{end}, and λ) using the new $\boldsymbol{P}$, τ and δ, but the previous values of V_0, B_{end} and λ. For these parameters, the posterior is not of a form for which independent sampling is possible. Therefore, we need to use Markov chain Monte Carlo (MCMC) sampling. We use the No-U-Turn sampler (NUTS),[34] a type of Hamilton Monte Carlo (HMC) algorithm. HMC methods take the negative logarithm of the distribution to draw from as a potential-energy landscape and the parameters as position variables of fictitious particles. Samples are then generated by simulating particle trajectories on this landscape with classical Hamiltonian dynamics using momenta that are drawn from a multivariate normal distribution[54, 35]. NUTS auto-tunes the step size and the number of steps used in the integration of the Hamiltonian dynamics. A simplified example of the MCMC sampling process is shown in Figure 3.2.

The sampling algorithm is implemented in the Python package PyMC3, version 3.11.4 [36], which uses autodifferentiation for the calculation of the gradient necessary for calculating the Hamiltonian trajectories. For each run, at least 4 chains are used. Convergence is assessed using the rank-normalized split $\widehat{R}$ statistic, which compares intra- and inter-chain variances [37, 38, 55]. Values of $\hat{R}$ very close to 1 indicate that the chains are stationary. All chains in this work were converged to $\hat{R} < 1.05$.

After convergence, the pooled samples from all chains represent the full N-dimensional posterior $p(\boldsymbol{\theta}|\boldsymbol{V})$. Due to its large dimensionality, it is not possible to visualize it directly. Instead we examine each parameter individually using a marginalized posterior, which is obtained by integrating the full posterior over all other parameters. This integral is approximated by generating a histogram of the parameter values from all samples and conducting a kernel density estimation of the histogram by smoothing with a Gaussian with a line width of 1/5 of the standard deviation of the parameter values. This results in a one-dimensional distribution that can easily be plotted and summarized in terms of location (mean, mode), and spread (credible intervals). On the right of Fig. 3.1, the marginalized posteriors are shown in color, together with the priors in gray. In this case, they are much narrower than

the corresponding priors. The spread of the posterior distribution is both a qualitative and quantitative measure of inferential uncertainty.

However, marginalization discards all information about correlation between parameters. It is also possible to display and examine two-dimensional marginalized posteriors between pairs of parameters (Fig. 3.2), which is particularly helpful for assessing issues of convergence often caused by highly correlated parameters.

We additionally visualize the results of the Bayesian inference in terms of a small set of posterior predictive samples of the noise-free signal, $V_M(\boldsymbol{\theta}^{(i)})$, and the distance distribution, $P(\boldsymbol{\theta}^{(i)})$. Here, $\boldsymbol{\theta}^{(i)}$ represent a random sample from the pooled MCMC samples, with i indicating the associated chain sample index.

3.4 Examples

3.4.1 Basic illustrations

In the first example, we illustrate the probabilistic analysis method on a synthetic distribution to allow for validation of the results. The synthetic data is based on distributions taken from the large simulated T4 lysozyme (T4L) test data set published by Edwards and Stoll in 2018 [19, 42]. The distributions in this test data set were generated computationally from an in silico spin-labeled crystal structure of T4L. Distribution 3992 from the test data set is taken as ground truth and two signal traces of differing quality were generated and analyzed with the Bayesian inference method using the nonparametric model implementation. The results are summarized in Figure 3.3. The blue data has favorable values for the modulation depth, background decay rate, trace length, and noise level ($\lambda = 0.5, k = 0.05$ µs$^{-1}, t_{\max} = 3.2$ µs$, \sigma = 0.02$). Comparatively, the data shown in green has less ideal values for all of these parameters ($\lambda = 0.2, k = 0.2$ µs$^{-1}, t_{\max} = 1.6$ µs$, \sigma = 0.05$), providing a challenging case with higher uncertainty.

For both cases, the marginalized posteriors for each parameter are shown (Fig. 3.3A & B, bottom). The black lines indicate values used in generating the signal. There is no line shown

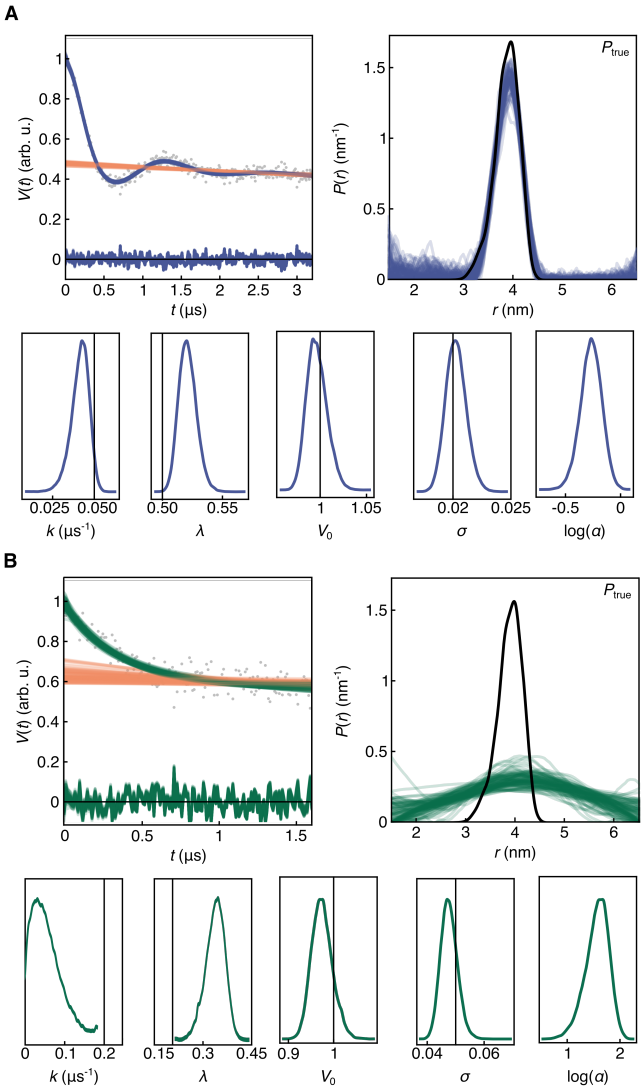

Figure 3.3: Validation of the nonparametric model's performance using synthetic data. Both test cases are generated from distribution 3992 from the Edwards/Stoll test set. The data differ in values for k, λ, σ, and $t_{\max}$. The MCMC simulation of each was run with 4 chains and 20,000 posterior samples per chain. A. Top: The left plot shows the time domain data (gray dots) and plot of posterior predictive signals V (blue) and backgrounds $(1-\lambda)V_{\text{inter}}$ (orange), and associated residuals. The right plot shows the scatter plot of sampled distance distributions plotted with P_{true} (black). All shown fits are random samples from the posterior distribution. Bottom: Plots of the marginalized parameter posteriors for the stronger dataset. The black lines reference the true parameter values. B. The ensemble plots and posteriors for the weaker synthetic data generated from the same starting distance distribution.

for α as this is a parameter that arises from fitting procedure for the distance distribution. For the longer and less noisy trace, the ground-truth parameters V_0 and σ are recovered accurately, and there is very little uncertainty about the inferred parameters, as represented by the narrowness of the distributions. However, for k and λ, the posterior modes do not align with the ground truth parameters and the spread in the case of k is a bit larger. The cause for this will be discussed in more depth in a later section.

The parameter posteriors most directly show the outcome of the Bayesian analysis and are useful for identifying the impact of individual parameters on the overall analysis. However, the fit to the time domain and distance distribution are the primary results of interest as they are the basis of possible structural conclusions. We draw a small set of random parameter vectors $\boldsymbol{\theta}^{(i)}$ (typically 30–100) from the pooled MCMC chain samples, calculate the associated distance distributions $P(\boldsymbol{\theta}^{(i)})$ and noise-free time-domain signals $V_M(\boldsymbol{\theta}^{(i)})$ (Fig. 3.3A & B, top). As with our previous work, this visualization for uncertainty is preferred as it does not put emphasis on any one distribution and more completely encompasses the range of distance distributions compatible with the data. For the strong dataset, there is little scatter among the distributions since the Bayesian analysis recovers the parameters with little uncertainty.

Analysis of the posterior modes for the shorter, noiser dataset (green) show that for V_0 and σ the modes are not correctly identified and the spreads are wider than for the blue dataset. For k and λ, the method is unable to recover the ground truth parameters completely. The difficulty of recovering these parameters, regardless of the dataset quality, is indicative of a problem inherent to the model used to analyze the DEER data – nonidentifiability [56]. The model has difficulty differentiating the background parameters and the long-distance region of the intramolecular signal, skewing the resulting posterior distributions for k and λ. Though the modes are not perfectly recovered for the higher–quality dataset, the data is still strong enough to sample the correct regions of parameter space and the inference results do not show significant spread. The identifiability problems do not hamper the analysis of either the time or distance domain for this test case, though they do cause more uncertainty at the edges of the distance axis. However, the difficulties caused by nonidentifiability become more

significant with shorter and noisier data. For the weaker dataset, the inability to recover the ground truth parameters corresponds to an overall smearing out of the distance distribution, inhibiting any possible structural conclusions.

3.4.2 Model comparison

The rest of the examples presented utilize the DEER data recently published alongside a benchmark test and guidelines paper on the DEER experiment [57]. Four constructs of the Yersinia outer protein O (YopO) from *Yersinia enterocolitica* without its membrane anchor were measured by seven different labs. We have analyzed the data from Lab B. YopO contains an α-helix that is 43 amino acids long, allowing for three site pairs to be chosen that encompass the short- (S585R1/Q603R1), mid- (V599R1/N624R1), and long- (Y588R1/N624R1) range distances accessible by DEER. A fourth site pair was chosen such that one of the spin labels is one a flexible loop giving rise to a very broad distribution (S353R1/Q635R1). The four pairs thus encompass a reasonably comprehensive set of data commonly encountered and analyzed by practitioners.

The results of using the Bayesian inference approach with both the nonparameteric and parametric models are shown in Figure 3.4. For each panel, the darker colored, top ensembles are the results of using the nonparametric model and the lighter colored, bottom ensembles are the parameteric model results with number of Gaussians indicated. For all four samples, convergence is achieved for the nonparametric model and the distance distributions show good agreement to those previously published [57]. Additionally, it appears that the Bayesian implementation does a better job of resolving shoulders and smaller side peaks than other methods. The uncertainty is relatively low given the small scatter of the sampled distributions around the mean features of the distributions. Uncertainty increases at long distances, as is common when using a least-squares fitting approach, but does not inhibit interpretation of the primary features. The parametric results for the short-, mid-range and flexible site pairs (Fig. 3.4A, B, & D) show strong agreement with the nonparametric results. The presence of peaks at longer distances with high uncertainty is indicative of limitations of

the parametric approach and imposing an underlying shape to the distance distribution. For the long-range site pair, the parametric results show similar positions for the main peaks but significantly more uncertainty overall. Neither the short- or long-range parametric results reached convergence according to the $\hat{R} < 1.05$. The multi-Gauss parametric model has more difficulty recovering distributions that have multiple, overlapping peaks of similar width or intensity. Distributions of this nature show larger correlations between distribution parameters, making exploring the parameter space significantly more inefficient and convergence difficult to achieve. This was shown previously when analyzing synthetic data and thus is not a surprising result when analyzing these data [50].

3.4.3 Dependence on distance axis

A non-parametric $\boldsymbol{P}$ is not entirely free of parameters—it depends on the parameters that define the distance axis. For a linear axis, the parameters are the minimum distance $r_{\min}$, the maximum distance, $r_{\max}$, and the resolution Δr. Figure 3.5A shows a series of results for the mid-range dataset with different $r_{\max}$ at a constant resolution Δr. Two effects of $r_{\max}$ can be discerned. First, as the $r_{\max}$ is decreased, the uncertainty in the long-distance region dwindles. Second, it becomes less arduous to obtain converged chains and the resulting posteriors indicate more efficient sampling and certainty around the parameters. When the $r_{\max}$ value is set to 8.5 nm (worth noting that this is past the heuristic value often used to select $r_{\max}$ based on $t_{\max}$ $(108t_{\max})^{1/3}$), the MCMC sampler was unable to converge, evidenced by $\hat{R}$ values far from 1.0. Both effects are a consequence of the fact that the models with long $r_{\max}$ are over-specified - there is not enough information in the data to clearly distinguish between mass in this region of r and the background. The marginal distributions of the background parameter k and the modulation depth λ broaden substantially as $r_{\max}$ is increased. These effects are also noticeable in the spread of background analysis as the $r_{\max}$ increases. Note that these effects are not a peculiarity of the MCMC method, but rather an intrinsic feature of the model given the data. Any analysis method that uses this model will encounter similar difficulties in identifying unique parameter values. Several approaches can

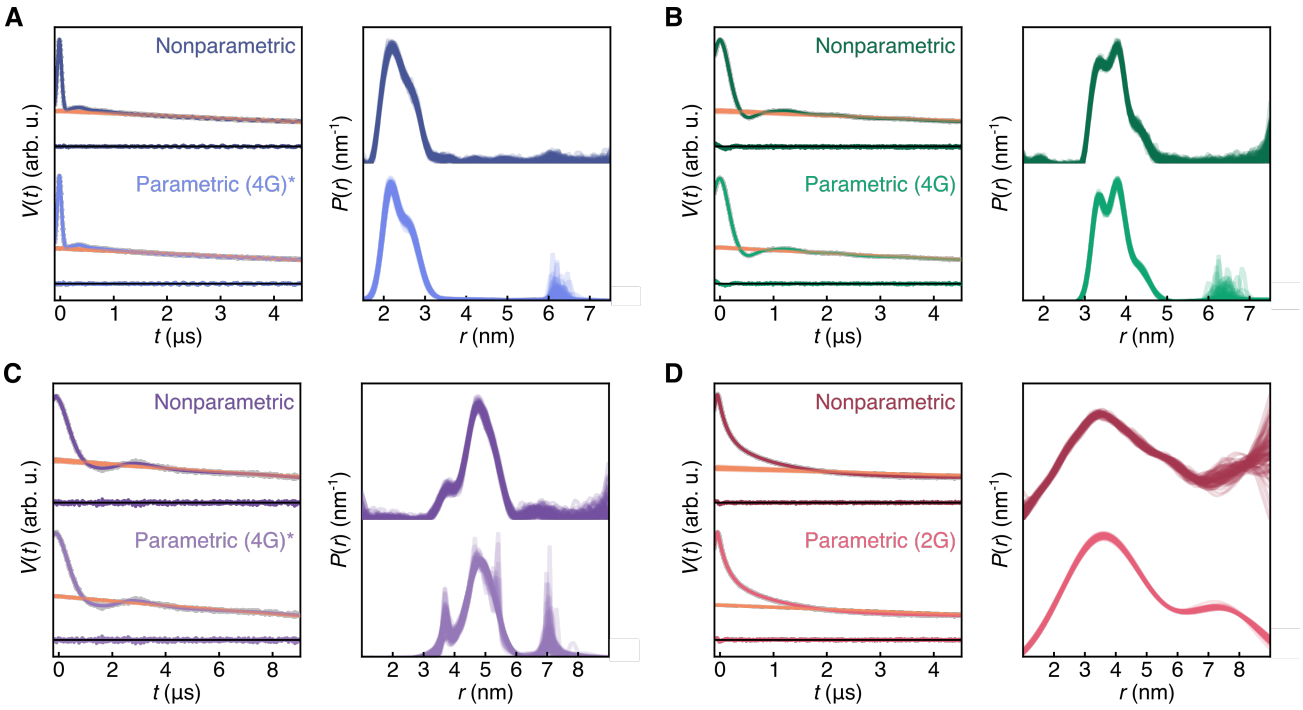

Figure 3.4: Bayesian inference using a multi-Gauss parametric model and a nonparametric model with Tikhonov regularization on DEER data obtained for four constructs of the effector protein YopO[58, 57]. All MCMC data shown were run with 4 chains of 50,000 samples per chain from which 100 samples were randomly selected. The raw experimental data are shown in gray overlaid with the time-domain intra- and intermolecular analysis for the nonparametric model (top, dark) and the parameteric multi-gauss model (bottom, light). Beneath the time-domain data are the residuals from the MCMC ensemble. The distance distributions for each are shown according to the same color scheme. The parametric models indicated with * did not achieve sationarity according to the $\hat{R} < 1.05$. The order is A. S585R1/Q603R1, maximum $\hat{R} = 1.53$ B. V599R1/N624R1, maximum $\hat{R} = 1.00$ C. Y588R1/N624R1, maximum $\hat{R} = 1.85$ D. S353R1/Q635R1, maximum $\hat{R} = 1.00$.

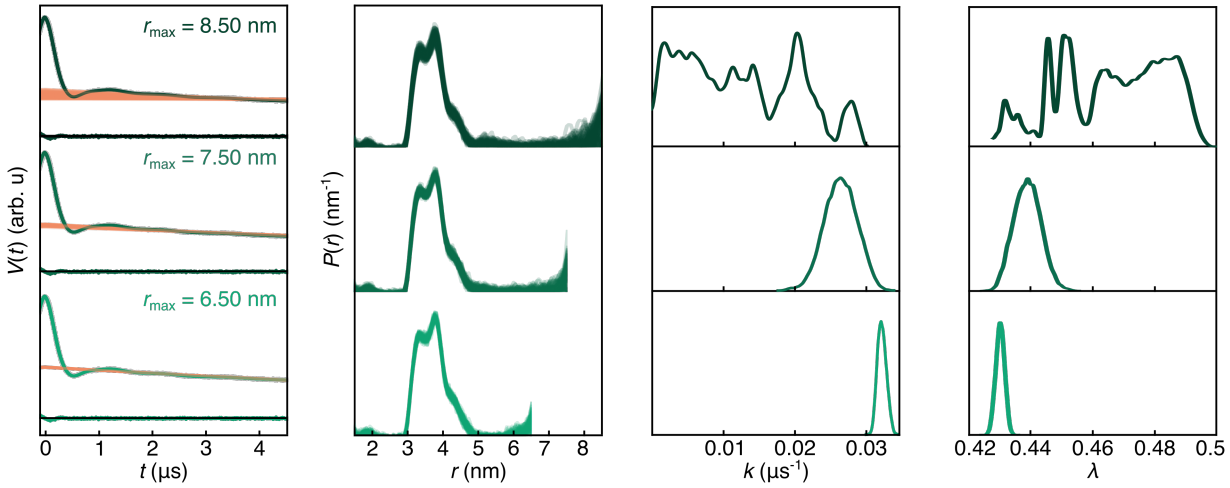

Figure 3.5: Effect of upper distance limit on quality of analysis and uncertainty. Data from the mid-range site pair (V599R1/N624R1) was evaluated for three values of $r_{\max}$ and a constant resolution (Δr). All MCMC data shown were run with 4 chains of 50,000 samples per chain from which 100 samples were randomly selected. The raw experimental data are shown in gray overlaid with the time-domain intra- and intermolecular analysis for the nonparametric model. The saturation of the color decreases with $r_{\max}$. A. Time–domain analysis and resulting distance distributions for $r_{\max}$ values 8.5 nm, 7.5 nm, and 6.5 nm. B. The posterior distributions for the background decay k and modulation depth λ with decreasing $r_{\max}$.

be considered in such situations: (a) restrict $r_{\max}$ to shorter values, (b) include additional information that $P(r)$ is close to zero at long distances in the model definition, or (c) use a less flexible model with stronger assumptions about P, such as a multi-Gauss model.

3.4.4 Comparison with DeerLab

A recently-released, open-source software package, DeerLab, has provided a comprehensive tool for conducting one-step analysis on DEER data using a range of models and approaches to quantifying uncertainty [29]. One approach, known as bootstrapping, allows for full quantification of uncertainty comparable to the analysis done with Bayesian inference. Bootstrapping, as implemented in DeerLab, is a Monte Carlo resampling method in which synthetic data traces are generated by adding varying levels of noise to the fitted signal. These new

traces are then analyzed according to the same procedure as the original experimental data. This results in a number of fitted parameter vectors and distance distributions equal to the number of bootstrap samples taken. The distribution of fitted values compares in nature to the posteriors output from Bayesian inference and can be randomly sampled from to produce spaghetti plots representing the spread of uncertainty around a particular set of data.

In Figure 3.6 we show the same Bayesian analysis using a nonparametric model from above and compare it to the data analyzed with DeerLab bootstrap uncertainty analysis. For each panel, the darker colored, top ensembles are the results of using Bayesian inference and the lighter colored, bottom ensembles are the DeerLab bootstrapping results.For the DeerLab analysis procedure, an initial fit was done using a compactness penalty [56]. The regularization parameter from this fit was frozen for the bootstrap analysis in which 1000 bootstrap samples were taken, i.e. 1000 new signal traces were generated and fit. 100 parameter vectors and distributions were randomly drawn from the 1000 samples and plotted alongside the Bayesian inference ensembles. For all but the fourth, flexible site pair, the ensembles of distance distributions between Bayesian analysis and DeerLab bootstrapping are very similar. There is slightly less jitter in the ensemble for DeerLab due to the nature of the bootstrapping procedure, all the synthetic signals generated are from an initial fit leading to less exploration of the combined parameter space as in the Bayesian approach. For the mid-range site pair (Fig. 3.6B), there is a small peak at about 6 nm in the DeerLab results that is not clearly visible for Bayes. This is likely a result of the compactness penalty added to the DeerLab fit — one approach to mitigating the identifiability problem that has been previously mentioned. This penalty favors distance distributions that are more compact in nature and thus penalizes distributions that would put density in longer distance peaks. Since our model does not currently have this penalty implemented and instead utilizes the priors to bias towards similar distributions, it is not as restrictive and still places some density at long-distance values. A similar occurrence is likely seen in the short-range distance site pair (Fig. 3.6A) at approximately 6 nm and the long-range distance site pair (Fig. 3.6C) at 7 nm where there are more distinct peaks in the DeerLab results and while the same peaks

might be inferred from the Bayesian results, it it too uncertain in those regions to determine conclusively. Interestingly, the results for the flexible site pair with a broad distribution are starkly different between the two analysis tools. Where the Bayesian inference seems to be fairly certain of a broad peak occurring over several nanometers, the DeerLab results appear to be overfitting and predicts a fairly spiky distance distribution. The uncertainty, based upon the spread of the ensemble of distributions, would indicate that none of the individual peaks occurring in the region of 1.5 to 5 nm could be analysed and rather one would assess this as a broad peak similar to the results shown by the Bayesian analysis. Also taking into account what we know about the specific system and, generally about the flexibility of spin labels and biomolecules, a very spiky distribution as determined by DeerLab is physically unrealistic and indicates overfitting.

The visualization of the uncertainty between the two methods shows significant similarity and therefore perform comparably well. There are, however, two key advantages to using Bayesian inference over bootstrapping: (a) run time and computational power, and (b) ease of access to uncertainty for all parameters without restriction. Uncertainty analysis via bootstrapping is an arduous computational process for analyzing DEER data. The standard acceptable procedures for bootstrapping impose that anything below 1000 bootstrapped samples is insufficient for analysis. For the data presented in this paper, that number of samples was only achievable via access to a computing cluster and still required days of computing time. It is possible to parallelize this computation which would dramatically speed up this process. However, the current implementation is under development and the results produced do not indicate reliable sampling. The implementation of the Bayesian inference approach described in this work is not yet parallelized but did not require more than a few hours on a personal computer to complete analysis on even the most cumbersome data presented.

The current DeerLab bootstrapping implementation provides easy access to the sampling results for the model V, distribution P, and time-domain parameters λ, concentration c, and V_0. However, one can not easily access the bootstrapped samples around the regularization

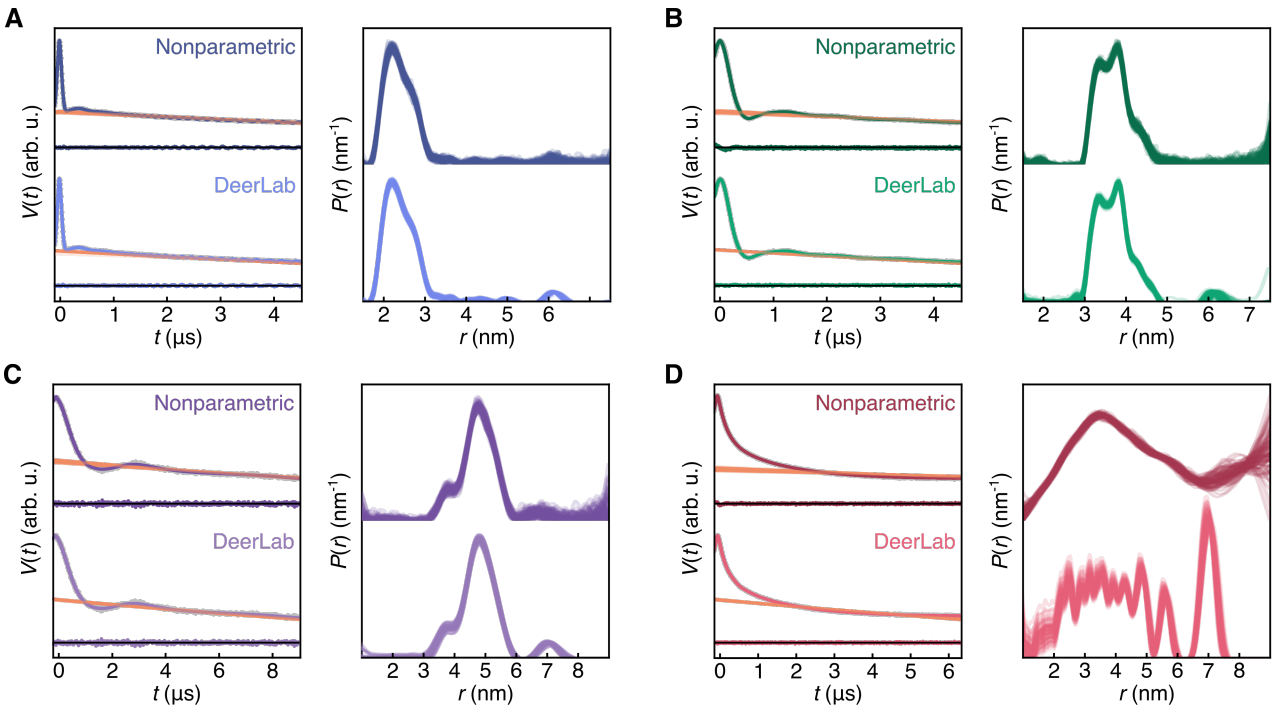

Figure 3.6: Comparison of experimental fits and visualization of uncertainty between Bayesian inference and DeerLab bootstrapping. All MCMC data shown were run with 4 chains of 50,000 samples per chain from which 100 samples were randomly selected. The raw experimental data are shown in gray overlaid with the time-domain intra- and intermolecular fits for the Bayesian analysis (top, dark) and DeerLab bootstrapping (bottom, light). The distance distributions for each are shown according to the same color scheme. The order is A. S585R1/Q603R1 B. V599R1/N624R1 C. Y588R1/N624R1 D. S353R1/Q635R1.

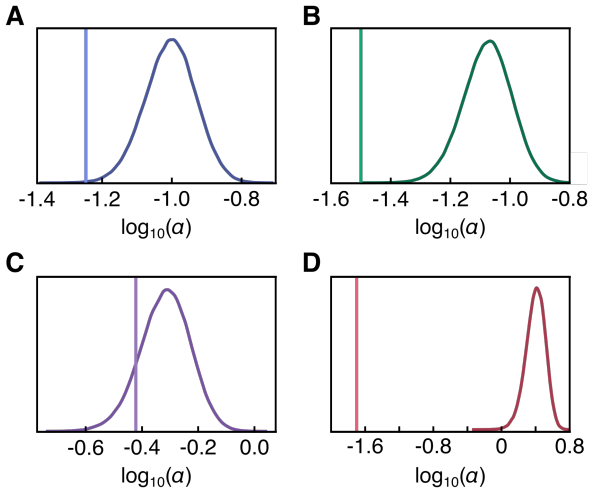

Figure 3.7: Alpha selection uncertainty assessment. The posterior distributions for the Tikhonov smoothness parameter $\log_{10}(\alpha)$ are shown, maintaining the same panel order and coloring as previous figures for the site pairs. Comparatively, the solid lines indicate the single $\log_{10}(\alpha)$ parameter from the DeerLab fits.

parameter, only the mean fit value. This regularization term is a crucial component for assessing whether over- or under-fitting of the data is occurring and uncertainty around it provides insight into the overall shape of the resulting distance distribution. With Bayesian inference, the uncertainty of $\log(\alpha)$, along with all other model parameters, is a direct output of the analysis via the marginalized posterior. These are shown in Figure 3.7 along with lines indicating the regularization parameters determined from the DeerLab fits. The results mirror those seen for the distributions in Figure 3.6, where the Bayesian and DeerLab results look very similar, the regularization parameters are of a similar magnitude. Though it appears DeerLab parameters tend to be smaller, indicative of overfitting the time-domain data while the modes of the marginalized posteriors for the Bayesian method skew a bit larger, indicative of underfitting the time-domain data. This is, again, most noticeable for the flexible site pair where the regularization parameters are significantly different between the two analysis methods.

3.5 Conclusions

The Bayesian inference method outlined in this work presents a rigorous and general inference approach for analyzing DEER data. Given the data and the choice of a particular physical model, it determines the probability distribution of all model parameters. This provides, completely, any information that can be gleaned from the data in the context of the chosen model. As presented, this approach allows for the systematic exploration of the effect of choosing particular values for the lower and upper limits, as well as the resolution of the distance axis. If the data is weak (noisy and truncated), then there will be significant uncertainty about the model parameters, particularly for a nonparameteric model. In this case, it can be advantageous to use a multi-Gaussian model for $\boldsymbol{P}$, which introduces additional constraints (limited number of peaks, symmetric distributions). This will of course bias the analysis and lead to difficulties of its own, depending upon if the underlying shape of the distance distribution is not represented well by the model choice.

Uncertainty quantification (UQ) with Bayesian inference has advantages over other forms of UQ. Compared to confidence intervals based on the Laplace approximation (which assumes the likelihood function is Gaussian around its maximum), it is more complete and accurate and can capture asymmetric uncertainty as well as broad areas in parameter space that lead to equally good model fits. Therefore, it can help identify ambiguity problems with the model. Compared to bootstrapping, which can be used to determine more comprehensive confidence intervals within the maximum-likelihood estimation approach, Bayesian inference does not synthetically generate new datasets. The only dataset that is used in Bayesian inference is the given dataset, saving significantly on computational time.

Although we presented the method using a Tikhonov model for the intra-molecular distance distribution combined with a simple exponential decay for the inter-molecular contribution, this method is very general and can handle more complex models, as well as simpler models such as Gaussian distribution models [50]. Extensions to handle multiple datasets simultaneously, include multiple dipolar pathways in analysis, and extend the models for the

inter-molecular signal are all left for future work. Once implemented, the Bayesian framework will provide an extremely robust tool for widespread DEER data analysis.

Chapter 4

THE EFFECT OF SPIN POLARIZATION IN DEER SPECTROSCOPY

Double electron–electron resonance (DEER) spectroscopy measures the distribution of distances between two electron spins in the nanometer range, often on doubly spin-labeled proteins, via the modulation of a refocused spin echo by the dipolar interaction between the spins. DEER is commonly conducted under conditions where the polarization of the spins is small. Here, we examine the DEER signal under conditions of high spin polarization, thermally obtainable at low temperatures and high magnetic fields, and show that the signal acquires a polarization-dependent out-of-phase component, both for the intra- and the intermolecular contribution. For the latter, this corresponds to a phase shift of the spin echo that is linear in the pump pulse position. We derive a compact analytical form of this phase shift and show experimental measurements using monoradical and biradical nitroxides at several fields and temperatures. The effect highlights a novel aspect of the fundamental spin physics underlying DEER spectroscopy[1].

4.1 Motivation

The behavior of spins depends upon temperature and magnetic field, which together determine the magnitude of thermal polarization. For a spin–1/2, the polarization at thermal equilibrium is

$$\epsilon(B,T) = \frac{N_\beta - N_\alpha}{N_\beta + N_\alpha} = \tanh \frac{g_\mathrm{e}\mu_\mathrm{B}B}{2k_\mathrm{B}T} \tag{4.1}$$

[1]This chapter is based on the 2022 publication in *Magnetic Resonance* by Sweger S. R., Denysenkov V. P., Maibaum L., Prisner T. F., and Stoll, S. titled, "The effect of spin polarization on double electron–electron resonance (DEER) spectroscopy, on volume 3, pages 101-110, available at `https://doi.org/10.5194/mr-3-101-2022`.

where N_β and N_α are the populations of the ground and excited state, respectively, g_e is the g-value for the free electron, μ_B is the Bohr magneton, B is the magnetic field strength, k_B is the Boltzmann constant, and T is the temperature (Fig. 4.1A). DEER spectroscopy using nitroxide radicals is often conducted at 50-70 K and 0.3-1.2 T. Under these experimental conditions, the Zeeman interactions are smaller than the thermal energy, and consequently the thermal spin polarization is small, leading to approximately equal populations of ground and excited state electrons. At higher magnetic fields and lower temperatures, a larger difference in spin populations arises leading to a shift in the total magnetization. The dependence of the thermal polarization on B and T is shown in Fig. 4.1B. Larger-than-thermal polarization can be generated utilizing photo-excited molecular triplet states [59, 60, 61] or by optical pumping in NV centers, which have a triplet ground state [62, 63, 64, 65]. Non-thermal spin polarizations are known to lead to out-of-phase electron spin echo amplitude modulation (OOP-ESEEM) in spin-correlated radical pairs [66, 67].

This chapter considers DEER signals under significant thermal spin polarization, as obtainable at low temperatures and high magnetic fields. Expanding on previous work [16], we show theoretically that the out-of-phase components of both the intra- and inter-molecular signals are polarization-dependent. In particular, we derive a closed-form analytical expression for the inter-molecular signal that reveals a polarization-dependent phase factor. We demonstrate this experimentally with monoradical and biradical nitroxide samples at several fields and temperatures.

The chapter is structured as follows. Section 4.2 steps through the analytical spin dynamics model showing the polarization-dependent echo phase, beginning from a two-spin system and then expanding to a uniform spatial distribution of spins. Section 4.3 describes experimental details, and Section 4.4 presents and discusses the experimental results. Section 4.5 concludes.

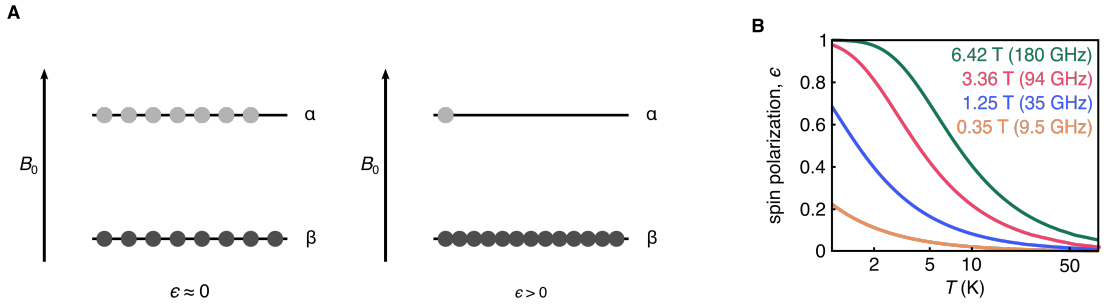

Figure 4.1: A: Polarization in high temperature limit (left). Equivalent populations of spins are aligned with (α) and against (β) the magnetic field. Polarization where the high–temperature approximation cannot be applied (right). Unequal spin populations lead to a shift in the total magnetization. B: Dependence of thermal spin polarization ϵ on magnetic field and temperature.

4.2 Analytical model

In this section, we derive the analytical form of the inter-molecular DEER signal, without the usual assumption of the high-temperature limit. The derivation within the high-temperature limit and the initial analytical expression for the intra-molecular DEER signal of an isolated spin pair beyond the high-temperature limit have previously been reported [68, 16, 69, 70] and is derived in Section A1 of the Appendix. Section A1 derives the 3-pulse DEER signal, the result of which is identical to the 4-pulse DEER model, thus the rest of the discussion will refer to the 4-pulse DEER model to remain consistent with the experimental results.

The analysis is based on the standard model for 4-pulse DEER of a frozen dilute solution of doubly spin-labeled molecules. We define two spectrally non-overlapping and separately addressable subsets of electron spins, A and B. (The sample also contain spins not affected by any pulses due to the broad EPR spectrum, particularly at high fields; we denote them as C spins.) The A-spins are manipulated via pulses at the probe frequency, ω_{A}, while the B-spins are inverted via a pulse at the pump frequency, ω_{B}. Beginning with the noise-free DEER signal shown in Eq. (1.1), this can be rewritten to describe an individual pair of one A-spin located at position $\boldsymbol{r}_{\mathrm{A}}$ and one B-spin located at $\boldsymbol{r}_{\mathrm{B}}$ (not necessarily on the same

molecule), the DEER signal is [16]

$$V_{\mathrm{AB}}(\boldsymbol{r}, t) = (1 - p_{\mathrm{B}}) + p_{\mathrm{B}} \left(\cos(\omega t) + \mathrm{i}\epsilon \sin(\omega t)\right) \tag{4.2}$$

Here, $\boldsymbol{r} = \boldsymbol{r}_{\mathrm{B}} - \boldsymbol{r}_{\mathrm{A}}$ is the inter-spin distance vector, and ω is the dipolar coupling frequency

$$\omega(\boldsymbol{r}) = D \, \frac{1 - 3\cos^2 \theta_{\mathrm{AB}}}{r^3} \qquad r = |\boldsymbol{r}| \qquad \cos\theta_{\mathrm{AB}} = \boldsymbol{r} \cdot \boldsymbol{z}/r \tag{4.3}$$

with the dipolar constant

$$D = \frac{\mu_0}{4\pi} \frac{\mu_{\mathrm{B}}^2 g_{\mathrm{e}}^2}{\hbar} \approx 2\pi \cdot 52.04\,\mathrm{MHz\,nm}^3 \tag{4.4}$$

where μ_0 is the magnetic constant, $\hbar$ is the reduced Planck constant, and $\boldsymbol{z}$ is the unit vector along the direction of the applied magnetic field. For systems with g-values differing from g_{e}, the respective g values can be used instead [71, 72]. In Eq. (4.2), p_{B} is the inversion efficiency of the pump pulse, and ϵ is the polarization of the B-spins, ranging between 0 and 1. Equation (4.2) is derived under the assumption that the dipolar coupling frequency is small compared to the difference in resonance frequencies between the two spins. Equation (4.2) is the standard in-phase DEER signal with an additional polarization-dependent out-of-phase sin term that is proportional to ϵ and therefore significant at low temperatures, high magnetic fields, or in the presence of spin hyperpolarization.

Figure 4.2 uses a vector model in the rotating frame to visualize the dynamics of A-spins through the 4-pulse DEER sequence, illustrating how the polarization affects the echo phase. The vectors indicate the two sub-ensembles of A-spins, one where the neighboring B-spin is in the $m_S = +1/2$ state (blue A-spins) and the other where it is in the $m_S = -1/2$ state (pink A-spins). The lengths of the arrows indicate the respective populations. The first two pulses of the sequence form an A-spin echo at time point 1, with both A-spin sub-ensembles refocused with $-x$ phase. As time passes, the spins precess and dephase due to a distribution of precession frequencies. This distribution is indicated by a set of arrows in varying shades, darker indicating higher frequency. The mean precession frequencies of the two sub-ensembles differ by the dipolar coupling frequency ω. The pump pulse flips the

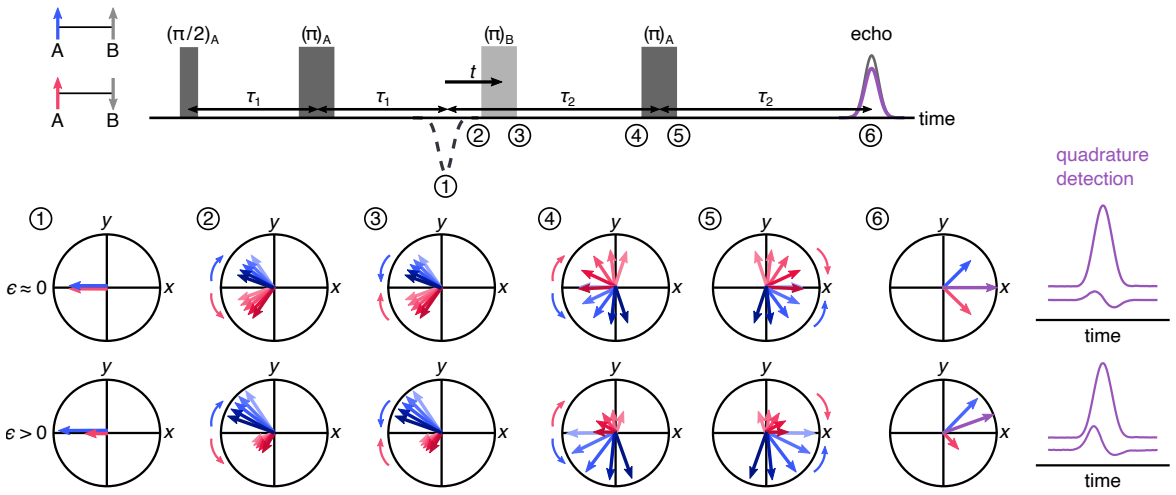

Figure 4.2: Vector model of the spin movement along the 4-pulse DEER sequence. Cases of low and high polarization values are shown where the numbers indicate the vector picture at a given point in time during the pulse sequence. The vectors indicate the two sub-populations of A-spins where those next to up-B-spins are colored blue and those next to down-B-spins are pink. The length of the arrows indicates the population. The overall magnetization is a sum of the two sub-populations, shown in purple. The respective echoes for low and high spin polarization values are shown on the right.

B-spins. This leaves all A-spin phases unaffected, but shifts the frequencies of the two A-spin sub-ensembles in opposite directions by $\pm\omega$, so that the two sub-ensembles now accumulate phase in reversed directions, while continuing to dephase as before the pulse. The final π pulse on the A-spins rotates both sub-ensembles around the y axis, and after evolution for τ_2, they refocus to give two echoes with opposite phases. The measured echo is proportional to the vector sum of these two sub-echoes. In the high-temperature regime ($\epsilon \approx 0$, top), the populations of the two sub-ensembles are equal, and therefore the echo has pure x phase (in-phase), with the quadrature component (y phase) integrating to zero. In the presence of significant polarization ($\epsilon > 0$), the two sub-echoes have significantly different amplitudes, and the overall echo phase is rotated from x, yielding a non-zero quadrature (out-of-phase) component.

The average of V_{AB} over a uniform orientational distribution of A–B spin pairs with fixed

r is

$$V_{\text{intra}}(r,t) = \frac{F_{\text{C}}(z)\cos(\phi) + F_{\text{S}}(z)\sin(\phi)}{z} + \mathrm{i}\epsilon\frac{F_{\text{C}}(z)\sin(\phi) - F_{\text{S}}(z)\cos(\phi)}{z} \tag{4.5}$$

with $\phi = Dt/r^3$ and $z = \sqrt{6\phi/\pi}$. F_{C} and F_{S} are the Fresnel cosine and sine integral functions respectively, which for imaginary arguments (z is imaginary for $t < 0$) have the properties $F_{\text{C}}(a\mathrm{i}) = \mathrm{i}\,F_{\text{C}}(a)$ and $F_{\text{S}}(a\mathrm{i}) = -\mathrm{i}\,F_{\text{S}}(a)$. The in-phase (real) and out-of-phase (imaginary) components of this signal are shown in Fig. 4.3A.

For a uniform, random, spatial distribution of B-spins in a sample, the signal is a product of all individual V_{AB} signals starting from Eq. (4.2), additionally averaged over all B-spin configurations

$$V_{\text{inter}}(t) = \left\langle \prod_{b=1}^{N_{\text{B}}} V_{\text{AB}}(\boldsymbol{r}_b, t) \right\rangle \tag{4.6}$$

Here, N_{B} is the number of B-spins in a configuration, and $\boldsymbol{r}_b$ is the vector from the A-spin to the bth B-spin. To arrive at the product form, the dipolar couplings among B spins are neglected. A somewhat involved derivation (spelled out in Section A.1.2) shows that Eq. (A.30) evaluates to

$$V_{\text{inter}}(t) = \exp(-k|t|) \cdot \exp(\mathrm{i}\alpha\epsilon k t) \tag{4.7}$$

where

$$k = \frac{8\pi^2}{9\sqrt{3}} p_{\text{B}} c_{\text{B}} D \qquad \alpha = \frac{\sqrt{3} + \ln(2-\sqrt{3})}{\pi} \approx 0.13213 \tag{4.8}$$

The first factor in Eq. (4.7) is an exponential decay function, as has been derived and observed before. The decay rate constant k depends on the B-spin concentration c_{B} and on the inversion efficieny p_{B}. The second factor is an additional hitherto unappreciated phase factor with a phase that grows linearly with t and is proportional to the spin polarization ϵ. This phase factor leads to a non-zero out-of-phase signal for $t \neq 0$, as long as the spin polarization ϵ is large enough. The signal is plotted in Fig. 4.3B. Equation (4.7) was also confirmed numerically using Monte Carlo simulations (Fig. A.1).

The extrema in the out-of-phase part of $V_{\text{inter}}(t)$ are located at

$$t = \pm t_{\text{OOP}} = \pm\frac{1}{k} \cdot \frac{\arctan(\alpha\epsilon)}{\alpha\epsilon} \approx \pm\frac{1}{k} \tag{4.9}$$

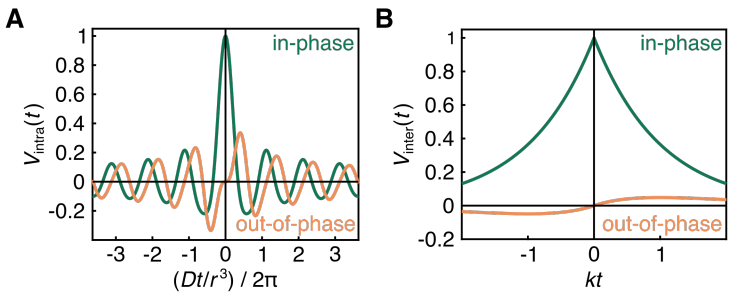

Figure 4.3: A: The theoretical powder-averaged polarized DEER signal for a single A–B spin pair (Eq. (4.5)). B: The theoretical DEER time trace for an A-spin coupled to a uniform three-dimensional distribution of B-spins at full spin polarization (Eq. (4.7), $\epsilon = 1$).

At these points, the magnitude of V_{inter} has decayed to $\mathrm{e}^{-1} \approx 0.368$. These time points depend only on concentration and inversion efficiency, not on polarization. Numerically, these quantities are related via

$$p_{\mathrm{B}} \cdot (c_{\mathrm{B}}/\mathrm{mM}) \cdot (t_{\mathrm{OOP}}/\mu\mathrm{s}) \approx 1 \tag{4.10}$$

At $t = t_{\mathrm{OOP}}$, the out-of-phase amplitude of the inter-molecular signal is

$$\mathrm{Im}(V_{\text{inter}}(t_{\mathrm{OOP}})) \approx \mathrm{e}^{-1} \frac{\alpha\epsilon}{\sqrt{1+(\alpha\epsilon)^2}} \approx \mathrm{e}^{-1}\alpha\epsilon \approx 0.0486\,\epsilon \tag{4.11}$$

At high temperatures and low fields, $\epsilon \approx 0$, and the out-of-phase component vanishes. The slope around $t = 0$ is dictated by the degree of polarization. Even at full polarization, the amplitude of the out-of-phase component is small (Fig. 4.3B), in contrast to that of an isolated A–B pair (Fig. 4.3A), where the out-of-phase signal reaches $\approx 0.335\epsilon$ at $(Dt/r^3)/2\pi \approx 0.403$.

The Fourier transform of Eq. (4.7) is a Lorentzian lineshape

$$\mathrm{FT}(V_{\text{inter}}(t), \omega) = \sqrt{\frac{2}{\pi}} \frac{k}{k^2 + (\omega + \alpha\epsilon k)^2} \tag{4.12}$$

centered at angular frequency $\omega = -\alpha\epsilon k$ and full width at half maximum of $fwhm = 2k$. The frequency ω is proportional to the polarization, the spin concentration, and the inversion efficiency. For $c_{\mathrm{B}} = 1\,\mathrm{mM}$, $p_{\mathrm{B}} = 1$, and $\epsilon = 1$, this gives a frequency of $\omega/2\pi \approx -0.021$ MHz and a line width of $fwhm/2\pi \approx 0.317$ MHz.

4.3 Materials & methods

Q-band DEER. 2,2,6,6-Tetramethylpiperidine-1-oxyl (TEMPO) was obtained from Sigma-Aldrich and used to create an approximately 1 mM solution in 50:50 (w:w) D_2O and d_8-glycerol. For measurements at Q-band, 30-50 μL of sample were syringed into 1.50 mm O.D. 1.1 mm I.D. quartz tubes (Sutter Instrument). Q-band experimental data were measured on a Bruker Elexsys E580 spectrometer equipped with a Bruker D2 dielectric resonator at 33.9 GHz, 1.21 T, at 11 and 40 K. The microwave power was amplified with a 390 W Applied Systems Engineering TWT amplifier. For both temperatures, the measurements were conducted at the field value that gave the maximum echo amplitude at the pump frequency. Optimal pulse lengths were determined using a Rabi nutation experiment. For the 11 K data, a standard 4-pulse DEER sequence was used with rectangular observer pulses at 33.842 GHz, the $\pi/2$ and π pulses being 22 and 44 ns and a sech/tanh pump pulse applied at 33.922 GHz with a 200 ns length and 80 MHz bandwidth. For the 40 K data, rectangular observer pulses with 22 and 44 ns lengths were applied at 33.828 GHz and a sech/tanh pump pulse was applied at 33.908 GHz with a 200 ns length and 80 MHz bandwidth. Experiments at both temperatures utilized a 64-step phase cycle [73]. The values for τ_1 and τ_2 were 4000 and 4200 ns, respectively for the 11 K data and 3000 and 3200 ns for the 40 K data. To keep the sample at thermal equilibrium for each echo, the spectrum was collected with 1 shot per point and a 1 s repetition time at 11 K. At 40 K, the data was collected with 10 shots per point and a 3 ms repetition time.

G-band DEER. 4-Hydroxy-2,2,6,6-tetramethylpiperidine-1-oxyl (TEMPOL) was obtained from Sigma Aldrich Chemie GmbH and used to create an approximately 1.0 mM solution with 45:55 (v/v) D_2O:d_8-glycerol. The homo-biradical nitroxide was synthesized according to the procedure previously published [74], and 0.16 mg were dissolved in 0.61 mL of deuterated toluene, giving a final concentration of 0.25 mM, corresponding 0.5 mM total spin concentration. The solution was placed in a 0.55 mm O.D. 0.4 mm I.D. quartz sample capillary that resulted in 250 nL of sample volume inside the TE011 cylindrical cav-

ity resonator. Samples were frozen after insertion into the the cryostat. G-band experiment were performed on a home-built pulse EPR spectrometer operating at 180 GHz, 6.42 T, and temperatures of 5, 40, and 50 K [75, 76]. For the monoradical experiments, a standard 4-pulse DEER sequence with rectangular pulses was used with the observer $\pi/2$ and π pulses being 36 and 58 ns, respectively, while the pump pulse was 58 ns with a frequency offset of +50 MHz. The values for τ_1 and τ_2 were 4000 and 5000 ns, respectively. Each DEER trace has 80 points and was recorded with 10 shots per point with 500 ms SRT for 5 K and 100 shots per point with 6 ms SRT for 40 K. This repetition time value was chosen as a 4-fold longer time with respect to the longitudinal relaxation time constant T_1 for nitroxide radicals in deuterated glycerol/D_2O solution at 6.4 T and 40 K. For the experiments at 5 K, the repetition time was chosen according to the saturation recovery experiment (Fig. A.5). Each trace was phased individually. For the biradical experiments, a standard 4-pulse DEER sequence with rectangular pulses was used with the observer $\pi/2$ and π pulses being 48 and 90 ns, respectively, while the pump pulse was 80 ns with a frequency offset of +80 MHz. The values for τ_1 and τ_2 were 500 and 5000 ns, respectively. Each DEER trace has 80 points and was recorded with 10 shots per point with repetition time of 6 ms at 50 K, and 500 ms at 5 K.

4.4 Results & discussion

To experimentally observe the polarization-dependent phase of the DEER signal, we performed DEER experiments on a 1 mM TEMPO sample in D_2O:d_8-glycerol at several magnetic fields and temperatures. The results are shown in Fig. 4.4. To fit the model from Eq. (4.7) to these data, we extended the model to

$$V(t) = V_0 \cdot \exp(-k|t - t_0|) \cdot \exp(\mathrm{i}\alpha\epsilon q_\mathrm{B} k(t - t_0)) \tag{4.13}$$

where the fit parameters are the overall signal amplitude V_0, the decay rate constant k, the zero-time shift t_0, and an additional phenomenological fit factor q_B discussed below. ϵ was calculated from temperature and magnetic field using Eq. (4.1) and is given in Fig. 4.4.

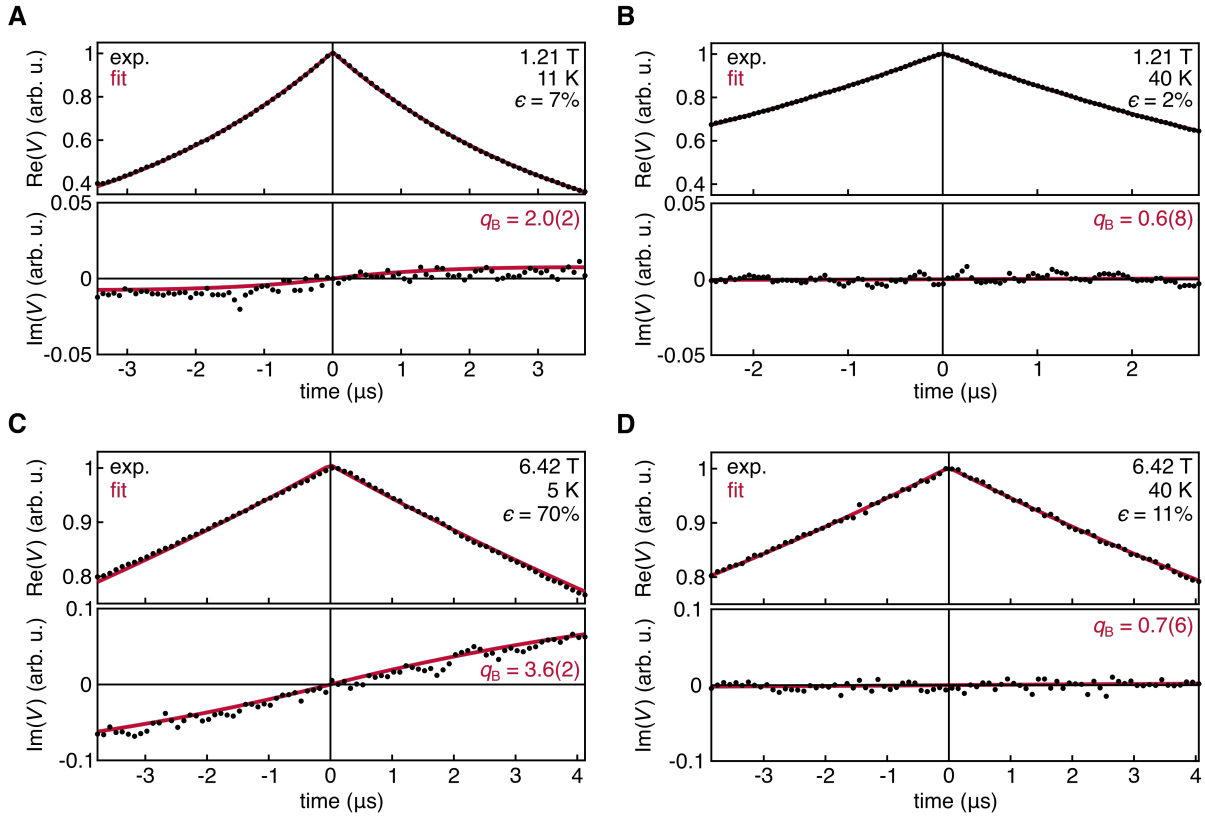

Figure 4.4: Experimental DEER traces for a 1.0 mM solution of TEMPO in 50:50 D_2O:d_8-glycerol, measured at 33.9 GHz at 11 K (A) and 40 K (B). Experimental DEER traces for a 1.0 mM solution of TEMPOL in 45:55 D_2O:d_8-glycerol, measured at 180 GHz and 6.42 T at 5 K (C) and 4d K (D). The data were fit with Eq. (4.13) and are shown with their 95% confidence interval in parentheses. The values of all fit parameters are listed in Table A1.

Figure 4.4A shows the DEER decay measured at 33.9 GHz, 1.21 T, and 11 K where the thermal polarization is about 7%. The out-of-phase signal is small but observable. The shape of the observed out-of-phase signal is reproduced by the fit. However, it grows with t faster than predicted by theory. This is indicated by the fitted value $q_B \approx 2.0(2)$, whereas we expect $q_B = 1$ from theory (Eq. (4.7)). When the temperature is raised to 40 K, Figure 4.4B, thermal polarization is reduced significantly to 2%, and the out-of-phase signal flattens and is not observable as expected.

Figure 4.4C & D show the DEER decays measured at 180 GHz and 6.42 T, at 5 K and

40 K where the thermal polarization is 70% and 11%, respectively. Again, the model of Eq. (4.13) was fit to the data. In the 5 K data, the out-of-phase signal is now clearly visible, and the experiment again confirms the shape predicted by the model, with a discrepancy in the phase rotation rate ($q_{\mathrm{B}} \approx 3.6(2)$). There is also a very slight asymmetry in the in-phase signal that is not captured by the model. At 40 K, the amplitude of the out-of-phase signal is negligible and the in-phase signal is symmetric, consistent with the theoretical prediction.

The experimental results in Fig. 4.4 were obtained with a monoradical solution. To further investigate the origins of the observed deviation of the intermolecular contribution from the theoretical prediction, we measured a frozen solution of the nitroxide biradical shown in Fig. 4.5A at 180 GHz, 6.42 T, at 5 K and 50 K. Under these conditions, the thermal polarization is 70% and 9%, respectively. The results are shown in Fig. 4.5. The intermolecular signal was fit according to Eq. (4.13) and the intramolecular signal was fit using an adaptation of Eq. (4.5) that incorporates an additional phenomenological factor q_{F}, analogous to q_{B} in Eq. (4.13).

$$V_{\mathrm{AB}}(r,t) = \frac{F_{\mathrm{C}}(z)\cos(\phi) + F_{\mathrm{S}}(z)\sin(\phi)}{z} + \mathrm{i}\epsilon q_{\mathrm{F}} \frac{F_{\mathrm{C}}(z)\sin(\phi) - F_{\mathrm{S}}(z)\cos(\phi)}{z} \tag{4.14}$$

This was added to verify experimental temperature and to test whether the observed discrepancy in the intermolecular background signals ($q_{\mathrm{B}} \neq 1$) is also present in the intramolecular signal. The intra-biradical distance distribution was modelled as a Gaussian distribution as defined in Eq. (1.9).

The experimental data are shown in Fig. 4.5. The dipolar oscillations are clearly resolved, and the theory once again matches the overall shape of the experimental data with a discrepancy for the intermolecular signal captured by the fit factor $q_{\mathrm{B}} = 4.2(3)$. However, this discrepancy does not appear for the intramolecular signal, where the fit factor of $q_{\mathrm{F}} = 1.0(4)$ confirms both that the sample temperature is 5 K $\pm$ 0.1 K and that the theoretical expression in Eq. (4.5) is correct. As the temperature is raised to 40 K and the polarization value drops significantly, the out-of-phase component disappears, consistent with theory.

Together, the experiments reveal that the theory predicts the out-of-phase component

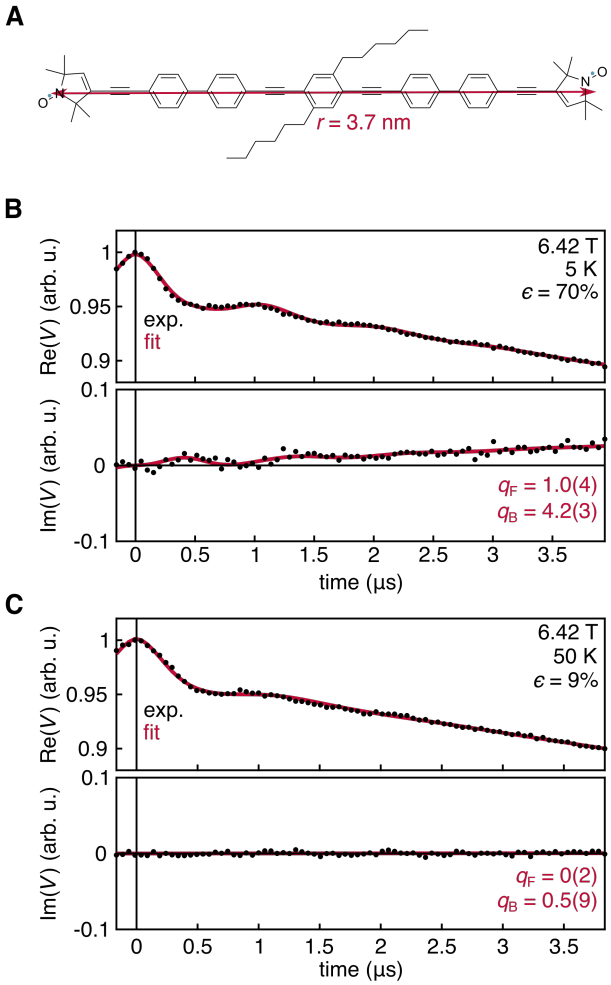

Figure 4.5: A: Nitroxide biradical with an interspin distance of 3.7 nm [77, 78]. Experimental DEER traces for a 0.25 mM solution of the biradical in deuterated toluene, measured at 180 GHz and 6.42 T at 5 K (B) and 50 K (C). The data was fit with Eqs. (4.13), (4.14), and (1.8). Fit parameters are shown with their 95% confidence interval. The values of all fit parameters are listed in Table A2.

of the intramolecular signal quantitatively, but underpredicts it for the intermolecular contribution. Experiments also show a very slight polarization-dependent asymmetry in the intermolecular signal that is not captured by theory.

A range of experimental and theoretical factors were taken into consideration to potentially explain the origin of the observed discrepancy, the results of which are provided in the

SI. Many instrumental factors can be excluded based on control experiments and on the data from the biradical. Pump–probe pulse excitation band overlap for the experimental conditions used (observer frequency 180.000 GHz, pump frequency 180.050 GHz, pulse widths 40–80 ns) is no more than approximately 2.7% (Fig. A.3). This overlap is too small to create significant shifts in the data. Gain imbalance between in-phase and out-of-phase detectors was determined insignificant, as the experiment run with the detector phase rotated by 90 degrees yielded no visible changes in the data (Fig. A.4). Experiments were run with detection both on and off the echo to verify that the observed signal arises from the echo and not background of the resonator (Fig. A.5). The detector phase was shown to drift slowly over the course of the experiments. This was accounted for by running the traces with as short of an acquisition time as possible and phasing each trace individually. However, it appears that this is not the origin of the observed discrepancy as traces with backwards and forward sweeps of t give approximately the same result (Fig. A.6). Temperature drifts were found to be insignificant, as an temperature sensor near the resonator indicated temperature stability within ± 0.1 K during the experiment. The fact that theory fits the out-of-phase component of the intramolecular signal in Fig. 4.5b well excludes many potential instrumental origins of the discrepancy between theory and the intermolecular signal.

Samples were run at sufficient cryoprotectant levels to eliminate aggregation as a source of error, at least for the experiments utilizing a monoradical sample.[79] For all 180 GHz data samples were not frozen rapidly outside the spectrometer, but rather after being placed into resonator. For the biradical data, field sweeps of a sample frozen outside and one frozen in the resonator showed little difference (Fig. A.8), further indicating that aggregation is not likely a major source of error.

Possible saturation effects were tested by running a saturation recovery experiment showing that a 500 ms repetition time is sufficiently slow as to not saturate the A-spins (Fig. A.7). One interesting possibility could be the occurrence of selective saturation. If the saturation is selective with respect to the A-spin sub-populations, then if the less populated A-spin sub-population is saturated more, the echo phase shift would increase.

The possibility of spectral diffusion of the A-spins can be excluded from consideration as the cause of the discrepancy as it would decrease the overall signal, equally in both the in-phase and out-of-phase signals. Spectral diffusion of the B-spins is expected to have no effect on DEER. One possibility is that there is selective spectral diffusion, affecting the two sub-populations of A-spins differently, though this is unlikely given the excitation profiles of the pulses shown in Fig. (A.3).

The discrepancy may be due to flip–flop terms of the A–B or B-–B spin coupling Hamiltonians that are neglected in the current theory. Inclusion of the A–B flip-flop term has been found to be important for the correct description of DEER signals from Gd(III)–Gd(III) spin pairs with short distances [80, 81]. For a spin concentration $c = 0.1$ mM $= 6 \cdot 10^{16}$ cm^{-3} (obtained from 1 mM total concentration and 10% excitation), the modal nearest-neighbor distance between spins is about 13 nm (Fig. A.2), which corresponds to a dipolar coupling frequency of less than 0.05 MHz. The flip-flop term only matters if the dipolar coupling frequency between two adjacent spins is of the same order, or larger, than their resonance frequency difference $\Delta\omega$. Since A and B spins are frequency-separated by at least 50 MHz (see Materials & methods), the A–B flip-flop terms are negligible. In addition, the accurate model fit of the intra-molecular signal in Fig. 4.5B (with $q_{\mathrm{F}} = 1$) confirms that A–B flip-flop terms are irrelevant. Since in the G-band experiments the excitation bandwidth of the pump pulse is about $1.7/t_{\mathrm{p}} \approx 30$ MHz, two adjacent B spins have a typical $\Delta\omega$ on the order of a few MHz. Therefore, the B–B flip–flop terms are likely of minor relevance, although it is not certain that they are negligible.

It is possible that demagnetization effects play a role in the observed discrepancy. In samples with high bulk concentrations of polarized spins, the resonance frequencies of individual spins are shifted due to the presence of a non-negligible dipolar field due to spins that are not excited by either the observer or the pump pulses. This field leads to frequency shifts [82, 83]. Frequency shifts up to 11 MHz have been observed in a crystal of the organic radical 1,3-bisdiphenylene-2-phenylallyl with EPR at 240 GHz and 8.56 T [84]. In our samples, the concentration of these spins (which might be termed C-spins) is about 1 mM, so

the demagnetization field and any associated effects are likely negligible.

Ongoing work that is showing promise into resolving the discrepancy takes into consideration the shape of the sample and how that contributes to anisotropy on a large scale. The DEER signal as derived analytically assumes an infinite isotropic sample. Experimentally, however, DEER samples are closer to cylindrical symmetry. There has been much work done in the field of NMR to understand the mechanisms of and quantify sample susceptibility [82, 85, 86]. As the volume and shapes of the samples differ depending on experimental setup, it is possible that similar effects are present in the experimental data shown in this chapter. Preliminary Monte Carlo simulations taking into account sample shape show q_{B} values different from those that have been derived in the current theory. More extensive calculations as well as systematic experimental data regarding sample volume, shape, and experimental setup will be needed to fully validate this theory.

4.5 Conclusions

Low temperatures and high fields lead to substantial thermal spin polarization. We showed theoretically that this leads to a hitherto unappreciated and unobserved additional phase factor in the DEER signal from a background of a uniform three-dimensional distribution of spins, resulting in a non-zero out-of-phase signal whose amplitude is proportional to the spin polarization. Experimental data, over a range of thermal spin polarizations, i.e. at several temperatures and magnetic fields, for both monoradical and biradical samples confirm the theoretical model describing this polarization-dependent phase factor.

The theory quantitatively matches the experiments for the intra-molecular DEER signal from a biradical, the observed amplitude of the out-of-phase component of the inter-molecular signal contribution is more intense than predicted by current theory. Though the origin of this discrepancy has not yet been definitevely identified, promising work is in progress that indicates sample shape may be a significant contributor.

Although derived for the case of two spins-1/2, the model is also applicable to high-spin systems, as long as only one transition of the pump spin is excited by the pump pulse.

The polarization phase factor is negligible for DEER experiments performed at 50–80 K and 0.35–1.25 T, temperatures and fields typically used for DEER experiments of nitroxide-labeled proteins, since the spin polarization is at most 1.6% under these conditions. However, the results are potentially relevant for measurements at lower temperatures and higher fields (such as for Gd spin labels measured at 10 K and W-band) and for situations with strong non-thermal polarization such as photo-induced excited-state triplets and optically pumped ground-state triplets.

Appendix A

SPIN POLARIZATION

A.1 Theory

A.1.1 Theoretical DEER signal from a single A–B spin pair

Here we derive the 3-pulse DEER signal from a single A–B spin pair, without making the high-temperature approximation. This recapitulates the derivation by Marko et al [16], with adjusted notation. Assuming that the effect of the hyperfine interactions can be folded into the resonance frequencies of the individual spins, the spin Hamiltonian contains terms for the Zeeman and dipolar interactions

$$\hat{H}_{\mathrm{lab}} = \hat{H}_{\mathrm{Z}} + \hat{H}_{\mathrm{D}} \tag{A.1}$$

In the high-field limit, the Zeeman interaction term expands to

$$\hat{H}_{\mathrm{Z}} = \hbar\omega_{\mathrm{A}}\hat{S}_{\mathrm{Az}} + \hbar\omega_{\mathrm{B}}\hat{S}_{\mathrm{Bz}} \tag{A.2}$$

where $\hbar$ is the reduced Planck constant, ω_{A} and ω_{B} are the resonance angular frequencies of the two spins, and $\hat{S}_{Az}$ and $\hat{S}_{Bz}$ are the spin operators. In the high-field limit, the dipolar interaction term is

$$\hat{H}_{\mathrm{D}} = \hbar\omega_{\mathrm{AB}}\hat{S}_{\mathrm{Az}}\hat{S}_{\mathrm{Bz}} \tag{A.3}$$

where ω_{AB} is the dipolar coupling angular frequency

$$\omega_{\mathrm{AB}} = D\,\frac{1 - 3\cos^2\theta_{\mathrm{AB}}}{r_{\mathrm{AB}}^3} \tag{A.4}$$

with the dipolar constant

$$D = \frac{\mu_0}{4\pi}\frac{\mu_{\mathrm{B}}^2 g_{\mathrm{e}}^2}{\hbar} \approx 2\pi \cdot 52.04\,\mathrm{MHz\,nm^3} \tag{A.5}$$

where μ_0 is the magnetic constant, μ_B is the Bohr magneton, g_e is g-value of the free electron, r_AB is the length of the interspin vector, and θ_AB is the orientation of interspin vector with respect to the external magnetic field, B_0.

In order to calculate the DEER signal, it is convenient to transform the Hamiltonian from the laboratory frame to a rotating frame that is rotating with angular frequency ω around the z-axis, where ω is the detection angular frequency

$$\hat{H} = \exp(\mathrm{i}\omega t(\hat{S}_\mathrm{Az} + \hat{S}_\mathrm{Bz}))\hat{H}_\mathrm{lab}\exp(-\mathrm{i}\omega t(\hat{S}_\mathrm{Az} + \hat{S}_\mathrm{Bz})) - \hbar\omega(\hat{S}_\mathrm{Az} + \hat{S}_\mathrm{Bz}) \tag{A.6}$$

This gives

$$\hat{H} = \hbar(\Delta\omega_\mathrm{A}\hat{S}_\mathrm{Az} + \Delta\omega_\mathrm{B}\hat{S}_\mathrm{Bz} + \omega_\mathrm{AB}\hat{S}_\mathrm{Az}\hat{S}_\mathrm{Bz}) \tag{A.7}$$

where $\Delta\omega_i = \omega_i - \omega$ are the angular frequency offsets of the two spins (i = A, B).

Deriving the DEER signal requires propagating the density operator through the course of the DEER experiment, beginning from the equilibrium density

$$\hat{\sigma}_\mathrm{eq} = \frac{\exp(-\hat{H}_\mathrm{Z}/k_\mathrm{B}T)}{Z} \tag{A.8}$$

where k_B is the Boltzmann constant, T is the temperature, and $Z = \mathrm{tr}(\exp(-\hat{H}_\mathrm{Z}/k_\mathrm{B}T))$ is the partition function. In Eq. (A.8), we have neglected the dipolar coupling term $\hat{H}_\mathrm{D}$ in the Hamiltonian, which is much smaller than the two Zeeman terms. Next, we use

$$\exp(-\hbar\omega_i\hat{S}_{iz}/k_\mathrm{B}T) = \hat{\mathbb{1}}\cosh\left(\frac{\hbar\omega_i}{2k_\mathrm{B}T}\right) - 2\hat{S}_{iz}\sinh\left(\frac{\hbar\omega_i}{2k_\mathrm{B}T}\right) = \hat{\mathbb{1}}c_i - 2\hat{S}_{iz}s_i \tag{A.9}$$

where $\hat{\mathbb{1}}$ is the identity operator, and c_i and s_i are abbreviations for the cosh and sinh factors. With this, we get

$$\begin{aligned}\hat{\sigma}_\mathrm{eq} &= \frac{\hat{\mathbb{1}}c_\mathrm{A}c_\mathrm{B} - 2\hat{S}_\mathrm{Bz}c_\mathrm{A}s_\mathrm{B} - 2\hat{S}_\mathrm{Az}c_\mathrm{B}s_\mathrm{A} + 4\hat{S}_\mathrm{Az}\hat{S}_\mathrm{Bz}s_\mathrm{A}s_\mathrm{B}}{4c_\mathrm{A}c_\mathrm{B}} \\ &= \frac{1}{4}\hat{\mathbb{1}} - \frac{1}{2}\epsilon_\mathrm{A}\hat{S}_{\mathrm{A}z} - \frac{1}{2}\epsilon_\mathrm{B}\hat{S}_{\mathrm{B}z} + \frac{1}{2}\cdot\epsilon_\mathrm{A}\epsilon_\mathrm{B}\cdot 2\hat{S}_{\mathrm{A}z}\hat{S}_{\mathrm{B}z}\end{aligned} \tag{A.10}$$

where

$$\epsilon_i = \tanh\left(\frac{\hbar\omega_i}{2k_\mathrm{B}T}\right) \tag{A.11}$$

is the thermal spin polarization of spin i.

The evolution in time of the density matrix is governed by the Liouville–von Neumann equation. In its integrated form, it is

$$\hat{\sigma}(t) = \hat{U}(t)\hat{\sigma}_{\mathrm{eq}}\hat{U}^{\dagger}(t) \tag{A.12}$$

where $\hat{U}$ is the total propagator of the DEER sequence and t is the pump pulse position. In the rotating frame and with ideal pulses, the Hamiltonian is piece-wise constant, and therefore the propagator is a composition of simple exponential operators:

$$\hat{U}(t) = \mathrm{e}^{-\mathrm{i}\hat{H}\tau}\mathrm{e}^{-\mathrm{i}\pi\hat{S}_{\mathrm{A}y}}\mathrm{e}^{-\mathrm{i}\hat{H}(\tau-t)}\mathrm{e}^{-\mathrm{i}\pi\hat{S}_{\mathrm{B}y}}\mathrm{e}^{-\mathrm{i}\hat{H}t}\mathrm{e}^{-\mathrm{i}(\pi/2)\hat{S}_{\mathrm{A}y}} \tag{A.13}$$

For describing the pulses, we assume ideal pulses that are selective for A or B spins, and we neglect fast oscillating terms and internal interactions during the pulses. All pulses in $\hat{U}(t)$ are set to have y phase. A more visual way to represent the propagator is

$$U(t) = \xrightarrow{(\pi/2)\hat{S}_{\mathrm{A}y}} \xrightarrow{\hat{H}t} \xrightarrow{\pi\hat{S}_{\mathrm{B}y}} \xrightarrow{\hat{H}(\tau-t)} \xrightarrow{\pi\hat{S}_{\mathrm{A}y}} \xrightarrow{\hat{H}\tau} \tag{A.14}$$

Before applying the propagator to the starting density based on Eq. (A.12), we simplify the propagator. This is possible due the simplicity of the rotating-frame Hamiltonian $\hat{H}$, which only contains terms with $\hat{S}_{\mathrm{A}z}$ and $\hat{S}_{\mathrm{B}z}$ that all mutually commute.

First, inserting the identity operator $\mathbb{1} = \mathrm{e}^{+\mathrm{i}\pi\hat{S}_{\mathrm{B}y}}\mathrm{e}^{-\mathrm{i}\pi\hat{S}_{\mathrm{B}y}}$ to the right of the $\hat{H}t$ propagator gives

$$U(t) = \mathrm{e}^{-\mathrm{i}\hat{H}\tau}\mathrm{e}^{-\mathrm{i}\pi\hat{S}_{\mathrm{A}y}}\mathrm{e}^{-\mathrm{i}\hat{H}(\tau-t)}\underbrace{\mathrm{e}^{-\mathrm{i}\pi\hat{S}_{\mathrm{B}y}}\mathrm{e}^{-\mathrm{i}\hat{H}t}\mathrm{e}^{+\mathrm{i}\pi\hat{S}_{\mathrm{B}y}}}\mathrm{e}^{-\mathrm{i}\pi\hat{S}_{\mathrm{B}y}}\mathrm{e}^{-\mathrm{i}(\pi/2)\hat{S}_{\mathrm{A}x}} \tag{A.15}$$

Next, the bracketed propagator combination can be written as $\mathrm{e}^{-\mathrm{i}\hat{H}_{\mathrm{B}}t}$, where $\hat{H}_{\mathrm{B}}$ is the transformed Hamiltonian

$$\hat{H}_{\mathrm{B}} = \mathrm{e}^{-\mathrm{i}\pi\hat{S}_{\mathrm{B}y}}\hat{H}\mathrm{e}^{+\mathrm{i}\pi\hat{S}_{\mathrm{B}y}} = \Delta\omega_{\mathrm{A}}\hat{S}_{\mathrm{A}z} - \Delta\omega_{\mathrm{B}}\hat{S}_{\mathrm{B}z} - \omega_{\mathrm{AB}}\hat{S}_{\mathrm{A}z}\hat{S}_{\mathrm{B}z} \tag{A.16}$$

This is possible since, in general, $\hat{U}\mathrm{e}^{-\mathrm{i}\hat{A}}\hat{U}^{\dagger} = \mathrm{e}^{-\mathrm{i}\hat{U}\hat{A}\hat{U}^{\dagger}}$.

Since $\hat{H}$ and $\hat{H}_{\mathrm{B}}$ commute (as both contain only z spin operators), we combine them into a single propagator exponent. The total propagator now reads

$$\hat{U}(t) = \mathrm{e}^{-\mathrm{i}\hat{H}\tau}\mathrm{e}^{-\mathrm{i}\pi\hat{S}_{\mathrm{A}y}}\mathrm{e}^{-\mathrm{i}(\hat{H}(\tau-t)+\hat{H}_{\mathrm{B}}t)}\mathrm{e}^{-\mathrm{i}\pi\hat{S}_{\mathrm{B}y}}\mathrm{e}^{-\mathrm{i}(\pi/2)\hat{S}_{\mathrm{A}y}} \tag{A.17}$$

Inserting the identity operator $\mathbb{1} = \mathrm{e}^{+\mathrm{i}\pi\hat{S}_{\mathrm{A}y}}\mathrm{e}^{-\mathrm{i}\pi\hat{S}_{\mathrm{A}y}}$ to the right of the $\hat{H}/\hat{H}_{\mathrm{B}}$ propagator yields

$$\hat{U}(t) = \mathrm{e}^{-\mathrm{i}\hat{H}\tau}\underbrace{\mathrm{e}^{-\mathrm{i}\pi\hat{S}_{\mathrm{A}y}}\mathrm{e}^{-\mathrm{i}(\hat{H}(\tau-t)+\hat{H}_{\mathrm{B}}t}\mathrm{e}^{+\mathrm{i}\pi\hat{S}_{\mathrm{A}y}}}\underbrace{\mathrm{e}^{-\mathrm{i}\pi\hat{S}_{\mathrm{A}y}}\mathrm{e}^{-\mathrm{i}\pi\hat{S}_{\mathrm{B}y}}\mathrm{e}^{-\mathrm{i}(\pi/2)\hat{S}_{\mathrm{A}y}}} \tag{A.18}$$

The propagator combination of the right bracket simplifies to $\mathrm{e}^{-\mathrm{i}\pi\hat{S}_{\mathrm{B}y}}\mathrm{e}^{-\mathrm{i}(3\pi/2)\hat{S}_{\mathrm{A}y}}$, since $\hat{S}_{\mathrm{A}y}$ and $\hat{S}_{\mathrm{B}y}$ commute. The propagator combination of the left bracket can again be rewritten with the transformed Hamiltonians:

$$\hat{U}(t) = \mathrm{e}^{-\mathrm{i}\hat{H}\tau}\mathrm{e}^{-\mathrm{i}(\hat{H}_{\mathrm{A}}(\tau-t)+\hat{H}_{\mathrm{AB}}t}\mathrm{e}^{-\mathrm{i}\pi\hat{S}_{\mathrm{B}y}}\mathrm{e}^{-\mathrm{i}(3\pi/2)\hat{S}_{\mathrm{A}y}} \tag{A.19}$$

where

$$\hat{H}_{\mathrm{A}} = \mathrm{e}^{-\mathrm{i}\pi\hat{S}_{\mathrm{A}y}}\hat{H}\mathrm{e}^{+\mathrm{i}\pi\hat{S}_{\mathrm{A}y}} = -\Delta\omega_{\mathrm{A}}\hat{S}_{\mathrm{A}z} + \Delta\omega_{\mathrm{B}}\hat{S}_{\mathrm{B}z} - \omega_{\mathrm{AB}}\hat{S}_{\mathrm{A}z}\hat{S}_{\mathrm{B}z} \tag{A.20}$$

$$\hat{H}_{\mathrm{AB}} = \mathrm{e}^{-\mathrm{i}\pi\hat{S}_{\mathrm{A}y}}\hat{H}_{\mathrm{B}}\mathrm{e}^{+\mathrm{i}\pi\hat{S}_{\mathrm{A}y}} = -\Delta\omega_{\mathrm{A}}\hat{S}_{\mathrm{A}z} - \Delta\omega_{\mathrm{B}}\hat{S}_{\mathrm{B}z} + \omega_{\mathrm{AB}}\hat{S}_{\mathrm{A}z}\hat{S}_{\mathrm{B}z} \tag{A.21}$$

After combining the adjacent exponential operators with the Hamiltonians in the exponent, we get

$$\hat{U}(t) = \mathrm{e}^{-\mathrm{i}(\hat{H}\tau+\hat{H}_{\mathrm{A}}(\tau-t)+\hat{H}_{\mathrm{AB}}t)}\mathrm{e}^{-\mathrm{i}\pi\hat{S}_{\mathrm{B}y}}\mathrm{e}^{-\mathrm{i}(3\pi/2)\hat{S}_{\mathrm{A}y}} \tag{A.22}$$

The sum of the Hamiltonian terms is

$$\hat{H}\tau + \hat{H}_{\mathrm{A}}(\tau-t) + \hat{H}_{\mathrm{AB}}t = 2\Delta\omega_{\mathrm{B}}(\tau-t)\hat{S}_{\mathrm{B}z} + \omega_{\mathrm{AB}}t\cdot 2\hat{S}_{\mathrm{A}z}\hat{S}_{\mathrm{B}z} \tag{A.23}$$

Since $\hat{S}_{\mathrm{B}z}$ and $2\hat{S}_{\mathrm{A}z}\hat{S}_{\mathrm{B}z}$ commute, we can separate the propagator into two, one for each term. With this, the final expression for the total propagator is

$$\hat{U}(t) = \mathrm{e}^{-\mathrm{i}\omega_{\mathrm{AB}}t\,2\hat{S}_{\mathrm{A}z}\hat{S}_{\mathrm{B}z}}\mathrm{e}^{-\mathrm{i}2\Delta\omega_{\mathrm{B}}(\tau-t)\hat{S}_{\mathrm{B}z}}\mathrm{e}^{-\mathrm{i}\pi\hat{S}_{\mathrm{B}y}}\mathrm{e}^{-\mathrm{i}(3\pi/2)\hat{S}_{\mathrm{A}y}} \tag{A.24}$$

or, more visually,

$$\hat{U}(t) = \xrightarrow{(3\pi/2)\hat{S}_{\mathrm{A}y}} \xrightarrow{\pi\hat{S}_{\mathrm{B}y}} \xrightarrow{2\Delta\omega_{\mathrm{B}}(\tau-t)\hat{S}_{\mathrm{B}z}} \xrightarrow{\omega_{\mathrm{AB}}t\,2\hat{S}_{\mathrm{A}z}\hat{S}_{\mathrm{B}z}} \tag{A.25}$$

Next, to get the final density $\sigma(t)$, we apply this propagator to the starting density from

Eq. (A.10):

$$\begin{aligned}
\hat{\sigma}_{\mathrm{eq}} =& \frac{\hat{\mathbb{1}}}{4} - \frac{\epsilon_{\mathrm{A}}}{2}\hat{S}_{\mathrm{A}z} - \frac{\epsilon_{\mathrm{B}}}{2}\hat{S}_{\mathrm{B}z} + \frac{\epsilon_{\mathrm{A}}\epsilon_{\mathrm{B}}}{2}\cdot 2\hat{S}_{\mathrm{A}z}\hat{S}_{\mathrm{B}z} \xrightarrow{(3\pi/2)\hat{S}_{\mathrm{A}y}} \\
& \frac{\hat{\mathbb{1}}}{4} + \frac{\epsilon_{\mathrm{A}}}{2}\hat{S}_{\mathrm{A}x} - \frac{\epsilon_{\mathrm{B}}}{2}\hat{S}_{\mathrm{B}z} - \frac{\epsilon_{\mathrm{A}}\epsilon_{\mathrm{B}}}{2}\cdot 2\hat{S}_{\mathrm{A}x}\hat{S}_{\mathrm{B}z} \xrightarrow{\pi\hat{S}_{\mathrm{B}y}} \\
& \frac{\hat{\mathbb{1}}}{4} + \frac{\epsilon_{\mathrm{A}}}{2}\hat{S}_{\mathrm{A}x} + \frac{\epsilon_{\mathrm{B}}}{2}\hat{S}_{\mathrm{B}z} + \frac{\epsilon_{\mathrm{A}}\epsilon_{\mathrm{B}}}{2}\cdot 2\hat{S}_{\mathrm{A}x}\hat{S}_{\mathrm{B}z} \xrightarrow{2\Delta\omega_{\mathrm{B}}(\tau-t)\hat{S}_{\mathrm{B}z}} \\
& \frac{\hat{\mathbb{1}}}{4} + \frac{\epsilon_{\mathrm{A}}}{2}\hat{S}_{\mathrm{A}x} + \frac{\epsilon_{\mathrm{B}}}{2}\hat{S}_{\mathrm{B}z} + \frac{\epsilon_{\mathrm{A}}\epsilon_{\mathrm{B}}}{2}\cdot 2\hat{S}_{\mathrm{A}x}\hat{S}_{\mathrm{B}z} \xrightarrow{\omega_{\mathrm{AB}}t\,2\hat{S}_{\mathrm{A}z}\hat{S}_{\mathrm{B}z}} \\
& \frac{\hat{\mathbb{1}}}{4} + \frac{\epsilon_{\mathrm{A}}}{2}\hat{S}_{\mathrm{A}x}\cos(\omega_{\mathrm{AB}}t) + \frac{\epsilon_{\mathrm{A}}}{2}\cdot 2\hat{S}_{\mathrm{A}y}\hat{S}_{\mathrm{B}z}\sin(\omega_{\mathrm{AB}}t) + \frac{\epsilon_{\mathrm{B}}}{2}\hat{S}_{\mathrm{B}z} \\
& \quad + \frac{\epsilon_{\mathrm{A}}\epsilon_{\mathrm{B}}}{2}\cdot 2\hat{S}_{\mathrm{A}x}\hat{S}_{\mathrm{B}z}\cos(\omega_{\mathrm{AB}}t) + \frac{\epsilon_{\mathrm{A}}\epsilon_{\mathrm{B}}}{2}\cdot \hat{S}_{\mathrm{A}y}\sin(\omega_{\mathrm{AB}}t) = \hat{\sigma}(t)
\end{aligned} \tag{A.26}$$

(Note that the $\hat{S}_{\mathrm{B}z}$ propagator has left the density unchanged.)

The echo signal is

$$V(t) = \mathrm{tr}(\hat{S}_{\mathrm{A}+}\hat{\sigma}(t)) \tag{A.27}$$

The only two terms from the final density in Eq. (A.26) that give a non-zero trace with $\hat{S}_{\mathrm{A}+}$ are those containing $\hat{S}_{\mathrm{A}x}$ and $\hat{S}_{\mathrm{A}y}$, since $\mathrm{tr}(\hat{S}_{\mathrm{A}+}\hat{S}_{\mathrm{A}x}) = 1$ and $\mathrm{tr}(\hat{S}_{\mathrm{A}+}\hat{S}_{\mathrm{A}y}) = \mathrm{i}$. Therefore, only these terms survive:

$$V(t) = \frac{\epsilon_{\mathrm{A}}}{2}\cos(\omega_{\mathrm{AB}}t) + \mathrm{i}\frac{\epsilon_{\mathrm{A}}\epsilon_{\mathrm{B}}}{2}\sin(\omega_{\mathrm{AB}}t) = \frac{\epsilon_{\mathrm{A}}}{2}\big[\cos(\omega_{\mathrm{AB}}t) + \mathrm{i}\epsilon_{\mathrm{B}}\sin(\omega_{\mathrm{AB}}t)\big] \tag{A.28}$$

Dropping the prefactor (which is absorbed into the overall amplitude factor V_0), we get the final result

$$V(t) = \cos(\omega_{\mathrm{AB}}t) + \mathrm{i}\epsilon_{\mathrm{B}}\sin(\omega_{\mathrm{AB}}t) \tag{A.29}$$

In a thermally equilibrated sample, the polarization of pump and probe spins are identical, therefore we can set $\epsilon = \epsilon_{\mathrm{A}} = \epsilon_{\mathrm{B}}$.

A.1.2 Theoretical DEER signal from A-spins in a homogeneous distribution of B-spins

For a sample with an A-spin in a uniform, random, spatial distribution of B-spins, the total signal is a product of all individual V_{AB} signals from Eq. (A.29), additionally averaged over

all B-spin configurations

$$V_{\text{inter}}(t) = \left\langle \prod_{b=1}^{N_{\text{B}}} V_{\text{AB}}(\boldsymbol{r}_b, t) \right\rangle \tag{A.30}$$

Here, N_{B} is the number of B-spins in a configuration, $\boldsymbol{r}_b$ is the vector from the A-spin to the bth B-spin, and the angled brackets indicate an average over B-spin configurations. To arrive at this product form, the dipolar couplings among B spins are neglected.

With the assumption that the positions of the B-spins are uncorrelated, the averaged product can be replaced with the product of the averages.

$$V_{\text{inter}}(t) = \prod_{b=1}^{N_{\text{B}}} \langle V_{\text{AB}}(\boldsymbol{r}_b, t) \rangle \tag{A.31}$$

Additionally, with the assumption that all B-spins are equally distributed, all averages are equal and independent of b, yielding

$$V_{\text{inter}}(t) = \langle V_{\text{AB}}(\boldsymbol{r}, t) \rangle^{N_{\text{B}}} \tag{A.32}$$

Defining a (large) spherical region S_R with a radius R, $V_{\text{inter}}(t)$ is obtained by calculating the signal for all N_{B} B-spins within S_R and then taking the limit $R \to \infty$

$$V_{\text{inter}}(t) = \lim_{R\to\infty} \left(\langle V_{\text{AB}}(\boldsymbol{r}, t) \rangle_R \right)^{N_{\text{B}}} \tag{A.33}$$

with

$$\langle V_{\text{AB}}(\boldsymbol{r}, t) \rangle_R = 1 + p_{\text{B}} \langle \cos(\omega t) - 1 \rangle_R + \mathrm{i}\epsilon p_{\text{B}} \langle \sin(\omega t) \rangle_R \tag{A.34}$$

based on Eq. (A.29). The averages over the oscillatory terms are

$$\langle \cos(\omega t) - 1 \rangle_R = \frac{1}{V_R} \int_{S_R} \mathrm{d}\boldsymbol{r} \left(\cos(\omega(\boldsymbol{r})t) - 1 \right) \equiv \frac{1}{V_R} C_R(t) \tag{A.35}$$

$$\langle \sin(\omega t) \rangle_R = \frac{1}{V_R} \int_{S_R} \mathrm{d}\boldsymbol{r} \, \sin(\omega(\boldsymbol{r})t) \equiv \frac{1}{V_R} S_R(t) \tag{A.36}$$

where $V_R = \frac{4\pi}{3} R^3$ is the volume of the sphere. The integrals $C_R(t)$ and $S_R(t)$ cannot be evaluated analytically, but if we take the limit $R \to \infty$, the integrals give

$$C(t) = \lim_{R\to\infty} C_R(t) = -\frac{8\pi^2}{9\sqrt{3}} \cdot D|t| \tag{A.37}$$

$$S(t) = \lim_{R\to\infty} S_R(t) = \frac{8\pi}{27} \left[3 + \sqrt{3} \ln(2 - \sqrt{3}) \right] \cdot Dt \tag{A.38}$$

The derivations are given separately below in sections A.1.2 and A.1.2 below. Note that $C(t)$ is proportional to $|t|$, whereas $S(t)$ is proportional to t. Using $C(t)$ and $S(t)$ instead of $C_R(t)$ and $S_R(t)$, Eq. (A.34) becomes

$$\langle V_{\mathrm{AB}}(\boldsymbol{r},t)\rangle_R = 1 + p_{\mathrm{B}}\frac{C(t)+\mathrm{i}\epsilon S(t)}{V_R} \tag{A.39}$$

with a small error for finite R that will vanish once we take the limit $R\to\infty$. Given a volume concentration of B-spins of c_{B}, the number of B-spins within S_R is $N_{\mathrm{B}} = c_{\mathrm{B}}V_R$, and the volume is $V_R = N_{\mathrm{B}}/c_{\mathrm{B}}$. Combining this with Eq. (A.39) and inserting into Eq. (A.33) gives

$$V_{\mathrm{inter}}(t) = \lim_{R\to\infty}\left(1 + p_{\mathrm{B}}c_{\mathrm{B}}\frac{C(t)+\mathrm{i}\epsilon S(t)}{N_{\mathrm{B}}}\right)^{N_{\mathrm{B}}} \tag{A.40}$$

The limit $R\to\infty$ corresponds to $V_R\to\infty$ and $N_{\mathrm{B}}\to\infty$, so that we get

$$V_{\mathrm{inter}}(t) = \lim_{N_{\mathrm{B}}\to\infty}\left(1 + p_{\mathrm{B}}c_{\mathrm{B}}\frac{C(t)+\mathrm{i}\epsilon S(t)}{N_{\mathrm{B}}}\right)^{N_{\mathrm{B}}} = \exp\big[p_{\mathrm{B}}c_{\mathrm{B}}(C(t)+\mathrm{i}\epsilon S(t))\big] \tag{A.41}$$

Inserting the expressions for $C(t)$ and $S(t)$ from Eqs. (A.37) and (A.38) yields the final expression

$$V_{\mathrm{inter}}(t) = \exp\left(-k|t| + \mathrm{i}\alpha\epsilon k t\right) = \exp(-k|t|)\cdot\exp(\mathrm{i}\alpha\epsilon k t) \tag{A.42}$$

where

$$k = \frac{8\pi^2}{9\sqrt{3}}p_{\mathrm{B}}c_{\mathrm{B}}D \approx 5.0651\, p_{\mathrm{B}}c_{\mathrm{B}}D \qquad \alpha = \frac{\sqrt{3}+\ln(2-\sqrt{3})}{\pi} \approx 0.13213 \tag{A.43}$$

The cos integral

Here we solve the $C(t)$ integral from Eq. (A.37). Its explicit form is

$$C(t) = \int_0^{2\pi}\mathrm{d}\phi\int_0^{\pi}\mathrm{d}\theta\sin\theta\int_0^{\infty}\mathrm{d}r\, r^2\left[\cos\left(\frac{Dt}{r^3}(1-3\cos^2\theta)\right)-1\right] \tag{A.44}$$

Since the integrand is independent of ϕ, the ϕ integral reduces to a prefactor of 2π. Next, we simplify the notation using the substitution $z=\cos\theta$

$$\int_0^{\pi}\mathrm{d}\theta\,\sin\theta = \int_{-1}^{1}\mathrm{d}\cos\theta = \int_{-1}^{1}\mathrm{d}z \tag{A.45}$$

Since the integrand is symmetric about $z = 0$, we can reduce the z integration interval to $[0, 1]$ and prepend a factor of 2. The integral now is

$$C(t) = 4\pi \int_0^1 \mathrm{d}z \int_0^\infty \mathrm{d}r\, r^2 \left[\cos\left(\frac{Dt}{r^3}(1 - 3z^2) \right) - 1 \right] \tag{A.46}$$

Next, we make the substitution $x = r^{-3}$ with

$$\frac{\mathrm{d}x}{\mathrm{d}r} = -\frac{3}{r^4} \qquad \mathrm{d}r = -\frac{r^4}{3}\mathrm{d}x \qquad r^2\mathrm{d}r = -\frac{r^6}{3}\mathrm{d}x = -\frac{1}{3x^2}\mathrm{d}x \qquad \int_0^\infty r^2 \mathrm{d}r = \frac{1}{3}\int_0^\infty x^{-2}\mathrm{d}x \tag{A.47}$$

Inserting this substitution gives

$$C(t) = \frac{4\pi}{3} \int_0^1 \mathrm{d}z \int_0^\infty \mathrm{d}x\, \frac{\cos(Dt(1 - 3z^2)x) - 1}{x^2} \tag{A.48}$$

The integral of $(\cos(ax) - 1)/x^2$ over x can be solved using integration by parts with $u(x) = \cos(ax) - 1$ and $v'(x) = 1/x^2$. This gives

$$\int \mathrm{d}x\, \frac{\cos(ax) - 1}{x^2} = \frac{1 - \cos(ax)}{x} - a \int \mathrm{d}x\, \frac{\sin(ax)}{x} \tag{A.49}$$

The first term is zero for both $x = 0$ and $x \to \infty$, so that we get

$$\int_0^\infty \mathrm{d}x\, \frac{\cos(ax) - 1}{x^2} = -a \int_0^\infty \mathrm{d}x\, \frac{\sin(ax)}{x} = -a \cdot \frac{\pi}{2}\mathrm{sgn}(a) = -\frac{\pi}{2}|a| \tag{A.50}$$

This gives

$$C(t) = -\frac{2\pi^2}{3}|Dt| \int_0^1 \mathrm{d}z\, |1 - 3z^2| \tag{A.51}$$

Evaluating the remaining z integral gives

$$\int_0^1 \mathrm{d}z\, |1 - 3z^2| = \int_0^{1/\sqrt{3}} \mathrm{d}z\, (1 - 3z^2) + \int_{1/\sqrt{3}}^1 \mathrm{d}z\, (3z^2 - 1) = \frac{2}{3\sqrt{3}} + \frac{2}{3\sqrt{3}} = \frac{4}{3\sqrt{3}} \tag{A.52}$$

so that we finally get

$$C(t) = -\frac{2\pi^2}{3}|Dt| \cdot \frac{4}{3\sqrt{3}} = -\frac{8\pi^2}{9\sqrt{3}} \cdot D|t| \tag{A.53}$$

where we have also used the fact that D is positive.

The sin integral

Next, we solve the integral $S(t)$ from Eq. (A.38). This turned out to be somewhat involved. The explicit form is

$$S(t) = \int_0^{2\pi} \mathrm{d}\phi \int_0^{\pi} \mathrm{d}\theta \sin\theta \int_0^{\infty} \mathrm{d}r\, r^2 \sin\left(\frac{Dt}{r^3}(1 - 3\cos^2\theta)\right) \tag{A.54}$$

Again, the ϕ integral adds a prefactor of 2π. Applying the same $\cos\theta \to z$ and $1/r^3 \to x$ substitutions as for $C(t)$ gives

$$S(t) = \frac{4\pi}{3} \int_0^1 \mathrm{d}z \int_0^{\infty} \mathrm{d}x\, \frac{\sin\left(Dt(1-3z^2)x\right)}{x^2} \tag{A.55}$$

As a result of the $r \to x$ substitution, the integrand now diverges for $x \to 0$. Therefore, we need to be cautious and write the overall integral in terms of a limit:

$$S(t) = \frac{4\pi}{3} \lim_{u\to 0} \int_0^1 \mathrm{d}z \int_u^{\infty} \mathrm{d}x\, \frac{\sin\left(Dt(1-3z^2)x\right)}{x^2} \tag{A.56}$$

(In principle, we could already write the initial integral expression with a limit $r_{\max} \to \infty$.) Depending on t and z, the argument of the sin function can be positive or negative. To solve the x integral, we need a non-negative argument. Utilizing $\sin(\xi) = \mathrm{sgn}(\xi)\sin|\xi|$, we get

$$S(t) = \frac{4\pi}{3}\mathrm{sgn}(Dt) \lim_{u\to 0} \int_0^1 \mathrm{d}z\, \mathrm{sgn}(1-3z^2) \int_u^{\infty} \mathrm{d}x\, \frac{\sin\left(|Dt||1-3z^2|x\right)}{x^2} \tag{A.57}$$

The integral of $\sin(ax)/x^2$ over x can be solved using integration by parts with $u(x) = \sin(ax)$ and $v'(x) = 1/x^2$. For $a \geq 0$, this gives

$$\int \mathrm{d}x\, \frac{\sin(ax)}{x^2} = a\,\mathrm{Ci}(ax) - \frac{\sin(ax)}{x} \tag{A.58}$$

with the cosine integral function $\mathrm{Ci}(\cdot)$, defined only for non-negative arguments. For $x \to \infty$, both terms evaluate to 0, so that the overall integral can now be written as

$$S(t) = \frac{4\pi}{3}\mathrm{sgn}(Dt) \lim_{u\to 0} \int_0^1 \mathrm{d}z\, \mathrm{sgn}(1–3z^2) \left[\frac{\sin(|Dt||1-3z^2|u)}{u} - |Dt||1-3z^2|\mathrm{Ci}(|Dt||1-3z^2|u)\right] \tag{A.59}$$

We can pull $|Dt|$ out of the integral and replace $|Dt|u \to x$. We get

$$S(t) = \frac{4\pi}{3} Dt \lim_{x\to 0} \int_0^1 \mathrm{d}z \, \left[\frac{\sin((1-3z^2)x)}{x} - (1-3z^2)\mathrm{Ci}\left(|1-3z^2|x\right) \right] \tag{A.60}$$

The divergence of the integrand is due to the cosine integral function. To deal with this, we rewrite it using

$$\mathrm{Ci}(\xi) = \gamma + \ln(\xi) + \mathrm{Cin}(\xi) \tag{A.61}$$

where γ is Euler's constant and $\mathrm{Cin}(\cdot)$ is the modified cosine integral function which satisfies $\mathrm{Cin}(0) = 0$. The full integral is now

$$S(t) = \frac{4\pi}{3} Dt \lim_{x\to 0} \int_0^1 \mathrm{d}z \, \left[\frac{\sin((1-3z^2)x)}{x} - (1-3z^2)\left[\gamma + \ln|1-3z^2| + \ln x + \mathrm{Cin}(|1-3z^2|x)\right] \right] \tag{A.62}$$

The only term in the integrand that diverges for $x \to 0$ is $(1–3z^2)\ln x$. However, the z integral over this term is zero, since the z integral over $(1-3z^2)$ is zero. Therefore, we can drop the $\ln x$ term from the integrand. The divergent term along x disappears due to symmetry along z! A similar argument allows us to drop the $(1-3z^2)\gamma$ term as well. All remaining terms in the integrand are finite everywhere and converge for $x \to 0$, so we can swap the x limit and the z integral. (In more technical terms, the integrand converges uniformly over the entire z interval, and we can apply the Lebesgue dominated convergence theorem to interchange the x limit and the z integral.)

$$S(t) = \frac{4\pi}{3} Dt \int_0^1 \mathrm{d}z \, \lim_{x\to 0} \left[\frac{\sin((1-3z^2)x)}{x} - (1-3z^2)\ln|1-3z^2| - (1-3z^2)\mathrm{Cin}(|1-3z^2|x) \right] \tag{A.63}$$

For $x = 0$, the first term becomes $(1-3z^2)$, which integrates to zero. The third term equals zero for $x = 0$ (since $\mathrm{Cin}(0) = 0$). We are left with

$$S(t) = \frac{4\pi}{3} Dt \int_0^1 \mathrm{d}z \, (3z^2-1)\ln|1-3z^2| \tag{A.64}$$

Evaluation of this with the computer algebra software Mathematica yields

$$S(t) = \frac{8\pi}{27} \left[3 + \sqrt{3}\ln(2-\sqrt{3})\right] Dt \tag{A.65}$$

A.2 Numerical calculations

A.2.1 Monte Carlo simulation of intermolecular signal

To verify the analytical expression of the polarized background signal in Eq. (A.42) and to investigate how it depends on neighboring spins, we performed Monte Carlo simulations. We start with Eq. (A.30), and use an average over a finite number N_{conf} of configurations and a product over a finite number N_{B} of B-spins:

$$V_{\mathrm{inter}}(t_j) \approx \frac{1}{N_{\mathrm{conf}}} \sum_{c=1}^{N_{\mathrm{conf}}} \prod_{b=1}^{N_{\mathrm{B}}} V_{\mathrm{AB}}(\boldsymbol{r}_b, t_j) \tag{A.66}$$

with the analytical signal Eq. (A.29) for $V_{\mathrm{AB}}(\boldsymbol{r}_b, t_j)$. t_j is a set of time points over which V_{inter} is evaluated.

The input parameters to this Monte Carlo model are the polarization ϵ, the pump efficiency p_{B}, the B-spin concentration c_{B}, the number N_{B} of B-spins per configuration, and the number N_{conf} of configurations. For each configuration, an A-spin is placed at the origin of a spherical volume of radius $R = (3N_{\mathrm{B}}/4\pi c_{\mathrm{B}})^{1/3}$, and N_{B} B-spins are placed at random uniformly distributed positions $\boldsymbol{r}_b$ within the sphere using

$$\boldsymbol{r}_b = R \frac{u_r^{1/3}}{\sqrt{n_x^2 + n_y^2 + n_z^2}} \begin{pmatrix} n_x \\ n_y \\ n_z \end{pmatrix} \tag{A.67}$$

where u_r is drawn from the standard uniform distribution Uniform$(u_r; 0, 1)$, and n_x, n_y, and n_z are drawn from the standard normal distribution Normal$(n_{x,y,z}; 0, 1)$.

Figure A.1 shows Monte Carlo simulations as a function of the number of excited B-spins in a configuration. The signal predicted analytically is captured perfectly. These simulations provide interesting insight into B-spin contributions to the decay. Remarkably, it is sufficient to consider configurations with only a single *excited* B-spin ($p_{\mathrm{B}} N_{\mathrm{B}} = 1$) in order to accurately recapture the initial 20% drop of the signal from its value at time zero. Both in-phase and out-of-phase signals are close to converged with 5 excited B-spins ($p_{\mathrm{B}} N_{\mathrm{B}} = 5$). With 50 excited

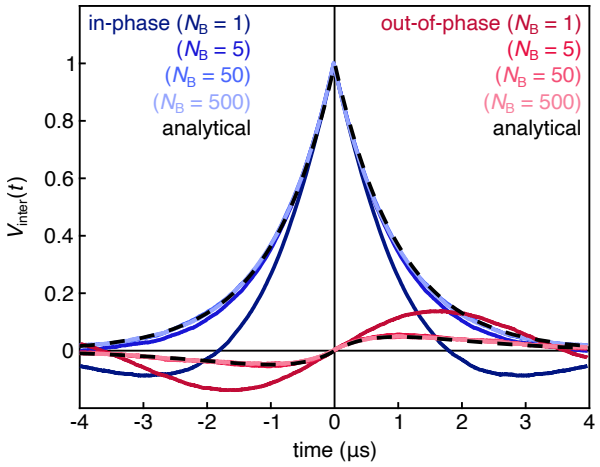

Figure A.1: Monte Carlo model simulating the inter-molecular DEER signal V_{inter} for maximum polarization ($\epsilon = 1$) and $p_{\text{B}} \cdot c_{\text{B}} = 1$ mM, assuming a sample with one A-spin and 75,000 configurations of a varying number N_{B} of uniformly distributed B-spins. Changing $p_{\text{B}} \cdot c_{\text{B}}$ only affects the time scale. The analytical signals are shown as black dashed lines.

B-spins, the entire decay curve is visually indistinguishable from the analytical model. These simulations show that the decay is dominated by just a few *excited* B-spins that are closest to the A-spin.

A.2.2 Nearest-neighbor distance distribution

In a uniform three-dimensional distribution of spins with number concentration c, the distribution of nearest-neighbor distances r_{nn} is given by [87]

$$P(r_{\text{nn}}) = 4\pi c\, r_{\text{nn}}^2 \exp\left(-\frac{4}{3}\pi c\, r_{\text{nn}}^3\right) \tag{A.68}$$

The mode of this distribution is at $(2\pi c)^{-1/3} \approx 0.542\, c^{-1/3}$, and the mean is at $\Gamma(1/3)/(36\pi c)^{1/3} \approx 0.554\, c^{-1/3}$. The nearest-neighbor distance distributions for $c = 1$ mM and 0.1 mM are shown in Fig. A.2.

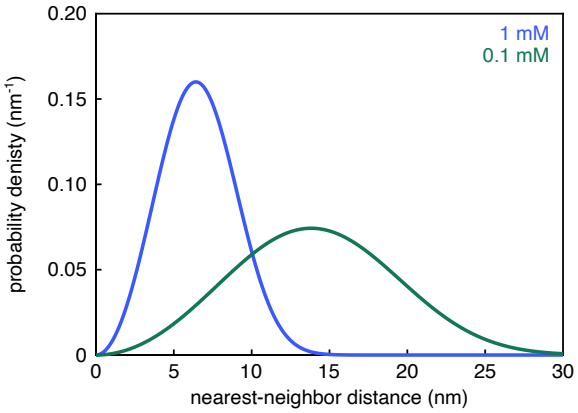

Figure A.2: Nearest-neighbor distributions of spins in uniform three-dimensional distributions with overall concentrations 1 mM and 0.1 mM, based on Eq. (A.68). The modes are at 6.4 and 13.8 nm, respectively.

A.3 Fit parameters

This section contains all of the fit parameters for the Q-band and G-band monoradical data (Table A.1) and the G-band biradical data (Table A.2) at various temperatures. Fitting was done using Eqs. (14)-(16), defined in the main text, with MATLAB's lsqcurvefit. All fit parameters are listed with the their 95% confidence intervals.

Table A.1: The monoradical data were fit according to Eq. (14), where the fit parameters are the decay rate constant k, the overall signal amplitude V_0, the zero-time shift t_0, and an additional phenomenological fit factor q_{B}.

Q-band	11 K		50 K	
Parameter	Fit value	95% confidence values	Fit value	95% confidence values
k	0.2784	0.2773, 0.2794	0.1639	0.1633, 0.1646
V_0	1.0064	1.0048, 1.0081	1.0051	1.0042, 1.0060
t_0	3.4343	3.4308, 3.4378	2.4444	2.4414, 2.4474
q_{B}	2.0587	1.8266, 2.2909	0.5769	-0.2485, 1.4023
G-band	5 K		40 K	
Parameter	Fit value	95% confidence values	Fit value	95% confidence values
k	0.0632	0.0619, 0.0645	0.0566	0.0566, 0.0585
V_0	1.0057	1.0030, 1.0085	1.0020	1.0000, 1.0041
t_0	3.8749	3.8531, 3.8967	3.9496	3.9309, 3.9682
q_{B}	3.5640	3.4294, 3.6985	0.7387	0.1355, 1.3419

A.4 Experimental validation

This section contains the results of the various experiments conducted towards the aim of verifying the observed signal and resolving any discrepancies with the theoretical predictions. Among these experiments are pump–probe pulse excitation band overlap (Fig. A.3), gain imbalance between the in-phase and out-of-phase detectors (Fig. A.4), signals recorded on

Table A.2: The biradical data were fit according to Eqs. (14)-(16), where the fit parameters are the inversion efficiency p_{B}, the decay rate constant k, overall signal amplitude V_0, the zero-time shift t_0, additional phenomenological fit factor for the intramolecular signal q_{F} and intermolecular signal q_{B}, the peak position r_0 and the standard deviation w.

G-band	5 K		50 K	
Parameter	Fit value	95% confidence values	Fit value	95% confidence values
p_{B}	0.0328	0.0294, 0.0362	0.0329	0.0308, 0.0349
k	0.0189	0.0181, 0.0196	0.0192	0.0187, 0.0196
V_0	0.9987	0.9957, 1.0016	1.0016	0.9998, 1.0034
t_0	0.4604	0.4406, 0.480	0.1648	0.1524, 0.1772
q_{F}	1.0341	0.6547, 1.4135	0.0000	-1.9133, 1.9133
q_{B}	4.1766	3.8969, 4.4563	0.5342	-0.4072, 1.4756
r_0	3.7166	3.6603, 3.7728	3.8656	3.7892, 3.9420
w	0.0897	0.0515, 0.1280	0.2198	0.1620, 0.2775

and off the echo to observe resonator background and phase instability (Fig. A.5), and saturation recovery data (Fig. A.7).

If there is a significant amount of pulse overlap, the pump pulse could partially excite A spins (and vice versa). This could lead to additional contributions to the DEER signal from secondary dipolar pathways [56]. The spectrum and pulse excitation profiles for the experimental conditions in the monoradical experiments (6.42 T, observer freuqency 180.000 GHz, pump frequency 180.050 GHz, pulse widths 58 ns) are shown in Fig. A.3. Total excitation by the probe pulse is approx. 3.7% and the pump pulse is approx. 4.3%. The overlap between pulse excitation profiles is approx. 2.7%. This is too small to create significant amplitudes in secondary dipolar pathways. Experiments with varying pump–probe offsets between 40 MHz and 120 MHz showed the same behavior (data not shown).

One possibility to get an overly intense out-of-phase component is that the second channel (out-of-phase with the echo) has a higher gain than the first channel. To test for gain

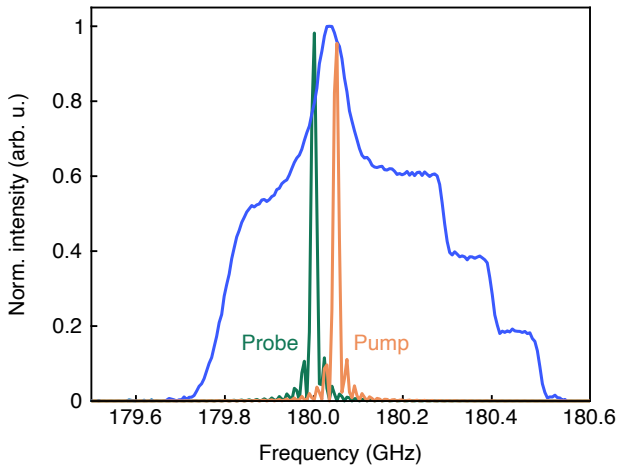

Figure A.3: Frequency-swept spectrum (blue), converted from experimental field sweep, and excitation profiles for the pump (orange) and probe (green) pulses. The field sweep data was obtained for 1.0 mM solution of TEMPOL in 45:55 D_2O:d_8-glycerol at 50 K. Simulation parameters match those used in the monoradical experiments (observer frequency 180.000 GHz, pump frequency 180.050 GHz, pulse widths 36/58 ns, τ 500 ns, and repetition time 10 ms).

imbalance, multiple traces of the monoradical data were recorded (Fig. A.4). The first was run in the same manner as the experiments shown in the main text (blue), the second with the detector phase rotated by 90 degrees (green). Although not completely identical, there is no substantial difference in the relative amplitudes of the out-of-phase signals.

To verify that the observed out-of-phase signal is entirely due to the refocusing spins and not due to instrumental offsets, two traces of the monoradical sample were recorded, the first being run in the same manner as the experiments shown in the main text, the second recorded with the detector window offset in time from the echo (Fig. A.5). The data show that there is no slope in the out-of-phase signal when recording off-echo, confirming that there is no instrumental offset or phase drift as a function of pump pulse position. The data also reveal that the actual out-of-phase signal is much noisier than the instrumental baseline obtained off-echo. This likely arises from some small phase jitter in the instrument.

The detector phase drifts slowly over the (wall clock) course of the experiments. This

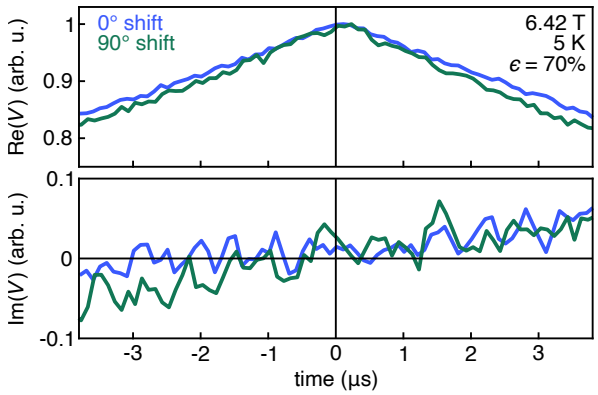

Figure A.4: Experimental DEER traces for a 1.0 mM solution of TEMPOL in 45:55 D_2O:d_8-glycerol, measured at 180 GHz and 6.42 T at 5 K, run with no shift in detector phase (blue) and 90° shift (green). Experimental conditions are identical to those shown in the main text for the monoradical data at 6.42 T and 5 K.

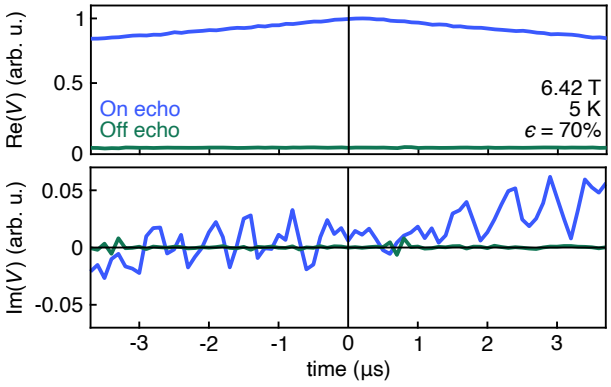

Figure A.5: Experimental DEER traces for a 1.0 mM solution of TEMPOL in 45:55 D_2O:d_8-glycerol, measured at 180 GHz and 6.42 T at 5 K, run on echo (blue) and off echo (green). Experimental conditions are identical to those shown in the main text for the monoradical data at 6.42 T and 5 K.

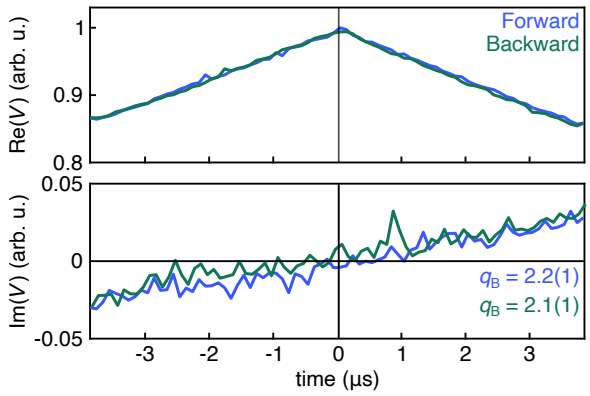

Figure A.6: Experimental DEER traces for a 1.0 mM solution of TEMPOL in 45:55 D_2O:d_8-glycerol, measured at 180 GHz and 6.42 T at 5 K, run in the forward direction (blue) and the reverse direction (green). Experimental conditions are identical to those shown in the main text for the monoradical data at 6.42 T and 5 K. The data were fit with Eq. (14) and are shown with their 95% confidence interval in parentheses.

could be a possible contributor to the observed mismatch between experiment and theory. However, it can be excluded because backwards and forward sweeps of t give approximately the same result (Fig. A.6), indicated by the identical fit factors (within 95% confidence). Regardless, the experiments were run with the shortest feasible acquisition times while still obtaining sufficient signal-to-noise to eliminate as much drift as possible.

To ensure that spin saturation was not occurring during experiments, a saturation recovery experiment was conducted to select the repetition time (Fig. A.7). From the data shown, a 500 ms repetition time was used for all 5 K data shown.

The G-band setup typically utilizes a freezing method where the sample freezes after being placed inside the cold resonator. To ensure that the freezing procedure was not impacting the experimental signal, i.e. causing aggregation upon freezing, two field sweeps (normalized) were obtained at 50 K for 250 μM biradical in d_8 toluene with different freezing procedures. The resulting spectra are shown in Fig. A.8. The black trace shows the data for a sample that was frozen inside the resonator and the red trace shows data for a sample plunge frozen in liquid nitrogen outside the resonator. There is no significant difference.

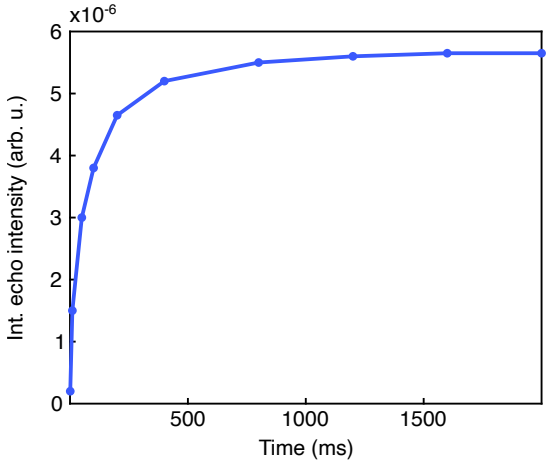

Figure A.7: Saturation recovery experiment of a 1.0 mM solution of TEMPOL in 45:55 D_2O:d_8-glycerol sample, measured at 180 GHz and 6.42 T at 5 K.

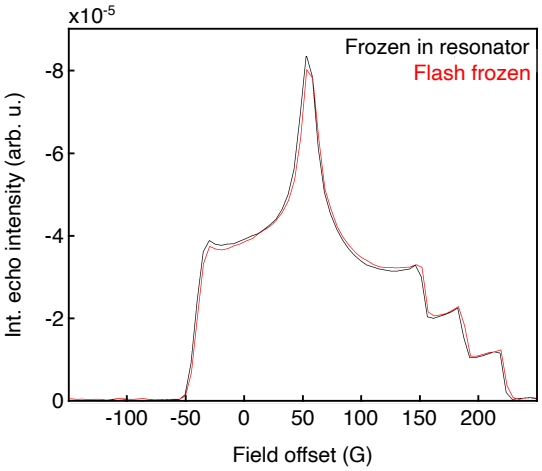

Figure A.8: Field sweeps for a 0.25 mM solution of the biradical in deuterated toluene, measured at 180 GHz and 6.42 T at 50 K, for a sample frozen inside the resonator (black) and plunge frozen in liquid nitrogen (red).

BIBLIOGRAPHY

[1] A. D. Milov, K. M. Salikhov, and M. D. Shchirov. Application of the double resonance method to electron spin echo in a study of the spatial distribution of paramagnetic centers in solids. *Phys. Solid State*, 23:565–569, 1981.

[2] A. D. Milov and Yu. D. Tsvetkov. Double electron–electron resonance in electron spin echo: Conformations of spin-labeled poly-4-vinylpyridine in glassy solutions. *Appl. Magn. Reson.*, 12:495–504, 1997.

[3] M. Pannier, S. Veit, A. Godt, G. Jeschke, and H. W. Spiess. Dead-time free measurement of dipole–dipole interactions between electron spins. *J. Magn. Reson.*, 142:331–340, 2000.

[4] G. Jeschke and Y. Polyhach. Distance measurements on spin-labelled biomacromolecules by pulsed electron paramagnetic resonance. *Phys. Chem. Chem. Phys.*, 9:1895–1910, 2007.

[5] G. Jeschke. DEER distance measurements on proteins. *Annu. Rev. Phys. Chem.*, 63:419–446, 2012.

[6] O. Schiemann and T. F. Prisner. Long-range distance determinations in biomacromolecules by EPR spectroscopy. *Q. Rev. Biophys.*, 40:1–53, 2007.

[7] N. A. Tavenor, K. I. Silva, S. Saxena, and S. W. Horne. Origins of structural flexibility in protein-based supramolecular polymers revealed by DEER spectroscopy. *J. Phys. Chem. B*, 118:9881–9889, 2014.

[8] O. Duss, M. Yulikov, G. Jeschke, and F. H. T. Allain. EPR-aided approach for solution structure determination of large RNAs or protein–RNA complexes. *Nat. Commun.*, 5:1–9, 2014.

[9] A. Manglik, T. H. Kim, C. Masureel, M.and Altenbach, Z. Yang, D. Hilger, M. T. Lerch, T. S. Kobilka, F. S. Thian, W. L. Hubbell, R.S. Prosser, and B.K. Kobilka. Structural insights into the dynamic process of β2-adrenergic receptor signaling. *Cell*, 161:1101–1111, 2015.

[10] R. Verhalen, B.and Dastvan, S. Thangapandian, Y. Peskova, H. A. Koteiche, R. K. Nakamoto, E. Tajkhorshid, and H. S. Mchaourab. Energy transduction and alternating access of the mammalian ABC transporter P-glycoprotein. *Nature*, 543:738–741, 2017.

[11] K. Barth, S. Hank, P. E. Spindler, T. F. Prisner, R. Tampé, and B. Joseph. Conformational coupling and trans-inhibition in the human antigen transporter ortholog TmrAB resolved with dipolar EPR spectroscopy. *J. Am. Chem. Soc.*, 140:4527–4533, 2018.

[12] R. Dastvan, S. Mishra, Y. B. Peskova, R. K. Nakamoto, and H. S. Mchaourab. Mechanism of allosteric modulation of p-glycoprotein by transport substrates and inhibitors. *Science*, 364:689–692, 2019.

[13] E. G. B. Evans, J. L. W. Morgan, F. DiMaio, W. N. Zagotta, and S. Stoll. Allosteric conformational change of a cyclic nucleotide-gated ion channel revealed by DEER spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.*, 117:10839–10847, 2020.

[14] A. Born, J. Soetbeer, F. Breitgoff, M. A. Henen, N. Sgourakis, Y. Polyhach, P. J. Nichols, D. Strotz, G. Jeschke, and B. Vögeli. Reconstruction of coupled intra- and interdomain protein motion from nuclear and electron magnetic resonance. *J. Am. Chem. Soc.*, 143:16055–16067, 2021.

[15] J. Voss, L. Salwiński, H. R. Kaback, and W. L. Hubbell. A method for distance determination in proteins using a designed metal ion binding site and site-directed spin labeling: evaluation with T4 lysozyme. *Proc. Natl. Acad. Sci. U.S.A.*, 92:12295–12299, 1995.

[16] A. Marko, V. Denysenkov, and T. F. Prisner. Out-of-phase PELDOR. *Mol. Phys.*, 111:2834–2844, 2013.

[17] G. Jeschke. Distance measurements in the nanometer range by pulse EPR. *ChemPhysChem*, 3:927–932, 2002.

[18] T. H. Edwards and S. Stoll. A Bayesian approach to quantifying uncertainty from experimental noise in DEER spectroscopy. *J. Magn. Reson.*, 270:87–97, 2016.

[19] T. H. Edwards and S. Stoll. Optimal Tikhonov regularization for DEER spectroscopy. *J. Magn. Reson.*, 288:58–68, 2018.

[20] E. J. Hustedt, F. Marinelli, R. A. Stein, J. D. Faraldo-Gómez, and H. S. Mchaourab. Confidence analysis of DEER data and its structural interpretation with ensemble-biased metadynamics. *Biophys. J.*, 115:1200–1216, 2018.

[21] A. G. Matveeva, V. M. Nekrasov, and A. G. Maryasov. Analytical solution of the PELDOR inverse problem using the integral Mellin transform. *Phys. Chem. Chem. Phys.*, 19:32381–32388, 2018.

[22] S. G. Worswick, J. A. Spencer, G. Jeschke, and I. Kuprov. Deep neural network processing of DEER data. *Sci. Adv.*, 4:eaat5218, 2018.

[23] G. Jeschke, A. Koch, U. Jonas, and A. Godt. Direct conversion of EPR dipolar time evolution data to distance distributions. *J. Magn. Reson.*, 155:75–82, 2002.

[24] Y.-W. Chiang, P. P. Borbat, and J. H. Freed. Maximum entropy: A complement to tikhonov regularization for determination of pair distance distributions by pulsed ESR. *J. Magn. Reson.*, 177:184–196, 2005.

[25] G. Jeschke, V. Chechik, P. Ionita, A. Godt, H. Zimmermann, J. Banham, C. R. Timmel, D. Hilger, and H. Jung. DeerAnalysis2006—a comprehensive software package for analyzing pulsed ELDOR data. *Appl. Magn. Reson.*, 30:473–498, 2006.

[26] K. I. Sen, T. M. Logan, and P. G. Fajer. Protein dynamics and monomer-monomer interactions in AntR activation by electron paramagnetic resonance and double electron-electron resonance. *Biochem.*, 46:11639–11649, 2009.

[27] S. Brandon, A. H. Beth, and E. J. Hustedt. The global analysis of DEER data. *J. Magn. Reson.*, 218:93–104, 2012.

[28] R. A. Stein, A. H. Beth, and E. J. Hustedt. A straightforward approach to the analysis of double electron–electron resonance data. *Methods Enzymol.*, 563:531–567, 2015.

[29] L. Fábregas Ibáñez, G. Jeschke, and S. Stoll. DeerLab: A comprehensive toolbox for analyzing dipolar EPR spectroscopy data. *Magn. Reson.*, 1:209–224, 2020.

[30] E. T. Jaynes. *Probability Theory. The Logic of Science*. Cambridge University Press, St. Louis, 2003.

[31] K. P. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, Cambridge, 2012.

[32] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis*. CRC Press, 2014.

[33] R. McElreath. *Statistical Rethinking. A Bayesian Course With Examples in R and STAN*. CRC Press, London, 2020.

[34] M. D. Hoffman and A. Gelman. The No-U-Turn Sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.*, 15:1593–1623, 2014.

[35] Michael Betancourt. A conceptual introduction to Hamiltonian Monte Carlo. *arXiv e-prints*, page arXiv:1701.02434 [stat.ME], 2017.

[36] J. Salvatier, T. V. Wiecki, and C. Fonnesbeck. Probabilistic programming in Python using PyMC3. *Peer J. Comput. Sci.*, 2:e55, 2016.

[37] A. Gelman and D. B. Rubin. Inference from iterative simulation using multiple sequences. *Stat. Sci.*, 7:457–472, 1992.

[38] S. P. Brooks and A. Gelman. General methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Stat.*, 7:434–455, 1998.

[39] Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of MCMC. *arXiv e-prints*, page arXiv:1903.08008 [stat.CO], 2019.

[40] M. Stephens. Dealing with label switching in mixture models. *J. R. Stat. Soc. B Methodol.*, 62:795–809, 2000.

[41] A. Jasra, C. C. Holmes, and D. A. Stephens. Markov Chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Stat. Sci.*, 20:50–67, 2005.

[42] T. H. Edwards and S. Stoll. Synthetic test data set for DEER spectroscopy based on T4 lysozyme. DOI: 10.6069/H5S75DCG, 2018.

[43] R. Kass and E. Raftery. Bayes Factors. *J. Am. Stat. Assoc.*, 90:773–795, 1995.

[44] E. Thrane and C. Talbot. An introduction to Bayesian inference in gravitational-wave astronomy: Parameter estimation, model selection, and hierarchical models. *Publ. Astron. Soc. Aust.*, 36:e010, 2019.

[45] Y. E. Kutsovsky, A. G. Mariasov, Y. I. Aristov, and V. N. Parmon. Electron spin echo as a tool for investigation of surface structure of finely dispersed fractal solids. *React. Kinet. Catal. Lett.*, 42:19–24, 1990.

[46] D. R. Kattnig, J. Reichenwallner, and D. Hinderberger. Modeling excluded volume effects for the faithful description of the background signal in double electron–electron resonance. *J. Phys. Chem. B*, 117:16542–16557, 2013.

[47] S. Domingo Köhler, M. Spitzbarth, K. Diederichs, T. E. Exner, and M. Drescher. A short note on the analysis of distance measurements by electron paramagnetic resonance. *J. Magn. Reson.*, 208:167–170, 2011.

[48] C. Altenbach. LongDistances. https://sites.google.com/site/altenbach/labview-programs/epr-programs/long-distances. Accessed: 11.06.2022.

[49] T. Ikeya, S. Ikeda, T. Kigawa, Y. Ito, and P. Güntert. Protein NMR structure refinement based on Bayesian inference. In *Journal of Physics: Conference Series*, volume 699, page 012005. IOP Publishing, 2016.

[50] S. R. Sweger, S. Pribitzer, and S. Stoll. Bayesian probabilistic analysis of DEER spectroscopy data using parametric distance distribution models. *J. Phys. Chem. A*, 124:6193–6202, 2020.

[51] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, New York, 2010.

[52] J. M. Bardsley and C. Fox. An MCMC method for uncertainty quantification in nonnegativity constrained inverse problems. *Inverse Probl. Sci. Eng.*, 20:477–498, 2012.

[53] J. M. Bardsley and P. C. Hansen. MCMC algorithms for computational UQ of nonnegativity constrained linear inverse problems. *SIAM J. Sci. Comput.*, 42:A1269–A1288, 2020.

[54] R. M. Neal. *MCMC Using Hamiltonian Dynamics*. CRC Press, Boca Raton, 2011.

[55] A. Vehtari, A. Gelman, D. Simpson, B. Carpenter, and P.-C. Bürkner. Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of MCMC. *Bayesian Anal.*, 1:667–718, 2021.

[56] L. Fábregas-Ibáñez, M. H. Tessmer, G. Jeschke, and S. Stoll. Dipolar pathways in dipolar EPR spectroscopy. *Phys. Chem. Chem. Phys.*, 24:2504–2520, 2022.

[57] O. Schiemann, C. A. Heubach, D. Abdullin, K. Ackermann, M. Azarkh, E. G. Bagryanskaya, M. Drescher, B. Endeward, J. H. Freed, L. Galazzo, et al. Benchmark test and guidelines for DEER/PELDOR experiments on nitroxide-labeled biomolecules. *J. Am. Chem. Soc.*, 143(43):17875–17890, 2021.

[58] M. F. Peter, A. T. Tuukkanen, C. A. Heubach, A. Selsam, F. G. Duthie, D. I. Svergun, O. Schiemann, and G. Hagelueken. Studying conformational changes of the Yersinia type-iii-secretion effector YopO in solution by integrative structural biology. *Structure*, 27(9):1416–1426, 2019.

[59] M. Di Valentin, M. G. Dal Farra, L. Galazzo, M. Albertini, T. Schulte, E. Hofmann, and D. Carbonera. Distance measurements in peridinin-chlorophyll a-protein by light-induced PELDOR spectroscopy. analysis of triplet state localization. *Biochim. Biophys. Acta. Bioenerg*, 1857:1909–1916, 2016.

[60] M. G. Dal Farra, C. Martin, E. Bergantino, Y. E. Kandrashkin, A. van der Est, and M. Di Valentin. Electron spin polarization transfer induced by triplet–radical interactions in the weakly coupled regime. *Phys. Chem. Chem. Phys.*, 22:19982–19991, 2020.

[61] O. A. Krumkacheva, I. O. Timofeev, L. V. Politanskaya, Y.F. Polienko, E.V. Tretyakov, O. Y. Rogozhnikova, D. V. Trukhin, V.M. Tormyshev, A.S. Chubarov, E.G. Bagryanskaya, and M.V. Fedin. Triplet fullerenes as prospective spin labels for nanoscale distance measurements by pulsed dipolar EPR spectroscopy. *Angew. Chem. Int. Ed.*, 58:13271–13275, 2019.

[62] R. Takahashi, S.and Hanson, J. Van Tol, M. S. Sherwin, and D. D. Awschalom. Quenching spin decoherence in diamond through spin bath polarization. *Phys. Rev. Lett.*, 101:047601–1–047601–4, 2008.

[63] V. Stepanov and S. Takahashi. Determination of nitrogen spin concentration in diamond using double electron–electron resonance. *Phys. Rev. B*, 94:024421–1–024421–9, 2016.

[64] A. O. Sushkov, I. Lovchinsky, N. Chisholm, R. L. Walsworth, H. Park, and M. D. Lukin. Magnetic resonance detection of individual proton spins using quantum reporters. *Phys. Rev. Lett.*, 113:197601, 2014.

[65] F. Shi, Q. Zhang, P. Wang, H. Sun, J. Wang, X. Rong, M. Chen, C. Ju, F. Reinhard, H. Chen, J. Wrachtrup, J. Wang, and J. Du. Single-protein spin resonance spectroscopy under ambient conditions. *Science*, 347:1135–1138, 2015.

[66] K. M. Salikhov, Yu. E. Kandrashkin, and A. K. Salikhov. Peculiarities of free induction and primary spin echo signals for spin-correlated radical pairs. *Appl. Magn. Reson.*, 3:199–216, 1992.

[67] J. Tang, M. C. Thurnauer, and J. R. Norris. Electron spin echo envelope modulation due to exchange and dipolar interactions in a spin-correlated radical pair. *Chem. Phys. Lett.*, 219:283–290, 1994.

[68] A. D. Milov, K. M. Salikhov, and Yu. D. Tsvetkov. Effect of spin dipole–dipole interaction on phase relaxation in magnetically dilute solid bodies. *J. Exp. Theor. Phys.*, 36:1229–1232, 1973.

[69] M. G. Dal Farra, S. Ciuti, M. Gobbo, D. Carbonera, and M. Di Valentin. Triplet-state spin labels for highly sensitive pulse dipolar spectroscopy. *Mol. Phys.*, 117:2673–2687, 2019.

[70] A.G. Maryasov, S.A. Dzuba, and K.M. Salikhov. Spin-polarization effects on the phase relaxation induced by dipole–dipole interactions. *J. Magn. Reson.*, 50:432–450, 1982.

[71] D. Abdullin, P. Brehm, N. Fleck, S. Spicher, S. Grimme, and O. Schiemann. Pulsed EPR dipolar spectroscopy on spin pairs with one highly anisotropic spin center: The low-spin Fe^{III} case. *Eur. J. Chem.*, 25:14388–14398, 2019.

[72] A. M. Bowen, M. W. Jones, J. E. Lovett, T. G. Gaule, M. J. McPherson, J. R. Dilworth, C. R. Timmel, and J. R. Harmer. Exploiting orientation-selective DEER: Determining molecular structure in systems containing Cu(II) centres. *Phys. Chem. Chem. Phys.*, 18:5981–5994, 2016.

[73] C. E. Tait and S. Stoll. Coherent pump pulses in double electron–electron resonance spectroscopy. *Phys. Chem. Chem. Phys.*, 18:18470–18485, 2016.

[74] B. E. Bode, D. Margraf, J. Plackmeyer, G. Dürner, T. F. Prisner, and O. Schiemann. Counting the monomers in nanometer-sized oligomers by pulsed electron–electron double resonance. *J. Am. Chem. Soc.*, 129:6736–6745, 2007.

[75] M. Rohrer, O. Brügmann, B. Kinzer, and T. Prisner. High-field/high-frequency EPR spectrometer operating in pulsed and continuous-wave mode at 180 GHz. *Appl. Magn. Reson.*, 21:257–274, 2001.

[76] M. M. Hertel, V. P. Denysenkov, M. Bennati, and T. Prisner. Pulsed 180-GHz EPR/ENDOR/PELDOR spectroscopy. *Magn. Reson. Chem.*, 23:S248–S255, 2005.

[77] P. Schöps, P.E. Spindler, A. Marko, and T.F. Prisner. Broadband spin echoes and broadband SIFTER in EPR. *J. Magn. Reson.*, 250:55–62, 2015.

[78] P. Schöps, J. Plackmeyer, and A. Marko. Separation of intra-and intermolecular contributions to the peldor signal. *J. Magn. Reson.*, 269:70–77, 2016.

[79] E. R. Canarie, S. M. Jahn, and S. Stoll. Quantitative structure-based prediction of electron spin decoherence in organic radicals. *J. Phys. Chem. Lett.*, 11:3396–3400, 2020.

[80] A. Dalaloyan, M. Qi, S. Ruhtstein, S. Vega, A. Godt, A. Feintuch, and D. Goldfarb. Gd(III)–Gd(III) EPR distance measurements - the range of accessible distances and the impact of zero field splitting. *Phys. Chem. Chem. Phys.*, 17:18464–18476, 2015.

[81] N. Manukovsky, A. Feintuch, I. Kuprov, and D. Goldfarb. Time domain simulation of Gd^{3+}–Gd^{3+} distance measurements by EPR. *J. Chem. Phys.*, 147:044201–1–044201–9, 2017.

[82] H. T. Edzes. The nuclear magnetization as the origin of transient changes in the magnetic field in pulsed NMR experiments. *J. Magn. Reson.*, 86:293–303, 1990.

[83] D. J.-Y. Marion, G. Huber, L. Dubois, P. Berthault, and H. Desvaux. ^{1}H and ^{129}Xe NMR absorption line shapes in the presence of highly polarized and concentrated Xenon solutions in high magnetic field. *J. Magn. Reson.*, 187:78–87, 2007.

[84] C. Blake Wilson, Devin T. Edwards, Jessica A. Clayton, Songi Han, and Mark S. Sherwin. Dressed Rabi oscillation in a crystalline organic radical. *Phys. Rev. Lett.*, 124:047201–1–047201–6, 2020.

[85] W. C. Dickinson. The time average magnetic field at the nucleus in nuclear magnetic resonance experiments. *Phys. Rev.*, 81:717–731, 1951.

[86] L. Li. Magnetic susceptibility quantification for arbitrarily shaped objects in inhomogeneous fields. *Magn. Reson. Med*, 46:907–916, 2001.

[87] M. Berberan Santos. On the distribution of the nearest neighbor. *Am. J. Phys.*, 54:1139–1141, 1985.

[88] A. Abragam. *Principles of Nuclear Magnetism*. Oxford University Press, 1961.

[89] K. Keller, M. Qi, C. Gmeiner, I. Ritsch, A. Godt, G. Jeschke, A. Savitsky, and M. Yulikov. Intermolecular background decay in RIDME experiments. *Phys. Chem. Chem. Phys.*, 21:8228–8245, 2019.

[90] R. G. Larsen and D. J. Singel. Double electron–electron resonance spin–echo modulation: Spectroscopic measurement of electron spin pair separations in orientationally disordered solids. *J. Chem. Phys.*, 98:5134–5146, 1993.

[91] G. Celeux, M. Hurn, and C. P. Robert. Computational and inferential difficulties with mixture posterior distributions. *J. Am. Stat. Assoc.*, 95:957–970, 2000.

[92] M. Stephens. *Bayesian Methods for Mixtures of Normal Distributions*. D.Phil. dissertation, Dept. Statistics, Univ. Oxford, 1997.

[93] G. Celeux. Bayesian Inference for Mixture: The Label Switching Problem. In Roger Payne and Peter Green, editors, *COMPSTAT*, pages 227–232. Physica-Verlag HD, Heidelberg, 1998.

[94] A. D. Milov, A. B. Ponomarev, and Yu. D. Tsvetkov. Electron–electron double resonance in electron spin echo: Model biradical systems and the sensitizied photolysis of decalin. *Chem. Phys. Lett.*, 110:67–72, 1984.

[95] J. R. Klauder and P. W. Anderson. Spectral diffusion decay in spin resonance experiments. *Phys. Rev.*, 125:912–932, 1962.

[96] L. Fábregas Ibáñez and G. Jeschke. General regularization framework for DEER spectroscopy. *J. Magn. Reson.*, 300:28–40, 2019.

[97] B. Efron and R. Tibshirani. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci* , pages 54–75, 1986.

[98] H. T. Banks, K. Holm, and D. Robbins. Standard error computations for uncertainty quantification in inverse problems: Asymptotic theory vs. bootstrapping. *Math. Comput. Model.*, 52:1610–1625, 2010.

[99] M. K. Bowman, A. G. Maryasov, N. Kim, and V. J. DeRose. Visualization of distance distribution from pulsed double electron–electron resonance data. *Appl. Magn. Reson.*, 26:23, 2004.

[100] J. Glaenzer, M. F. Peter, and G. Hagelueken. Studying structure and function of membrane proteins with PELDOR/DEER spectroscopy–the crystallographers' perspective. *Methods*, 147:163–175, 2018.