Neural networks in pulsed dipolar spectroscopy: a practical guide

Jake Keeley[1], Tajwar Choudhury[1], Laura Galazzo[2], Enrica Bordignon[2],
Akiva Feintuch[3], Daniella Goldfarb[3], Hannah Russell[4], Michael J. Taylor[4],
Janet E. Lovett[4], Andrea Eggeling[5], Luis Fabregas Ibanez[5], Katharina Keller[5],
Maxim Yulikov[5], Gunnar Jeschke[5], Ilya Kuprov[1,*]

[1]School of Chemistry, University of Southampton, Southampton SO17 1BJ, United Kingdom.
[2]Department of Physical Chemistry, University of Geneva,
Quai Ernest Ansermet 30, CH-1211 Geneva, Switzerland.
[3]Department of Chemical Physics, Weizmann Institute of Science, Rehovot 7610001, Israel.
[4]SUPA School of Physics and Astronomy, and BSRC, University of St Andrews,
North Haugh, St Andrews, KY16 9SS, United Kingdom.
[5]Department of Chemistry and Applied Biosciences, Swiss Federal Institute of
Technology in Zurich, Vladimir Prelog Weg 2, CH-8093 Zürich, Switzerland.

*i.kuprov@soton.ac.uk

Abstract

This is a methodological guide to the use of deep neural networks in the processing of pulsed dipolar spectroscopy (PDS) data encountered in structural biology, organic photovoltaics, photosynthesis research, and other domains featuring long-lived radical pairs and paramagnetic metal ions. PDS uses distance dependence of magnetic dipolar interactions; measuring a single well-defined distance is straightforward, but extracting distance distributions is a hard and mathematically ill-posed problem requiring careful regularisation and background fitting. Neural networks do this exceptionally well, but their "robust black box" reputation hides the complexity of their design and training – particularly when the training dataset is effectively infinite. The objective of this paper is to give insight into training against simulated databases, to discuss network architecture choices, to describe options for handling DEER (double electron-electron resonance) and RIDME (relaxation-induced dipolar modulation enhancement) experiments, and to provide a practical data processing flowchart.

1. Introduction

The recent evolution of statistical analysis and numerical regression into machine learning and neural nets has been a remarkable success. The reason has been known since 1980s: some neural networks are universal approximators [1,2] – but it was not before teraflop scale computing power had arrived that highly visible applications emerged. One of those is deconvolution, defined formally as solving Fredholm equations [3]. Neural nets are apparently able to learn the kernel and produce numerically stable inversions of the convolution operator [4,5]. Applications include image and voice recognition [6-8], deeply subwavelength optical microscopy [9], magnetic resonance imaging [10], and – in our present case – pulsed dipolar spectroscopy (PDS) of electron spin pairs.

The energy of the magnetic dipole interaction between unpaired electrons depends on the inverse cube of the distance. Electron magnetic moments can be determined precisely, and the resulting distance measurement techniques are called pulsed dipolar spectroscopy [11,12]. Double electron-electron resonance (DEER) is the most popular one; it is essentially a spin echo experiment, modified to keep only the dipole term in the effective Hamiltonian, and implemented so as to avoid the dead times of microwave electronics [13,14]. A similar experiment, called relaxation-induced dipolar modulation enhancement (RIDME) uses electron spin relaxation as a replacement for one of the magnetisation inversion steps [15,16]. Both methods are popular in structural biology because they return distributions of distances and therefore offer a window into nanometre-scale conformational mobility [17,18]; RIDME is particularly useful with paramagnetic metal ions [19,20]. When unpaired electrons are not present naturally (*e.g.* in metalloproteins), they are introduced by site-directed spin labelling: most commonly, selected amino acids are mutated into cysteines and stable radicals are attached to the thiol sulphur [21]. The most popular spin label is *S*-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulphonothioate (MTSL). It is highly reactive towards cysteines; it is also small and flexible enough to avoid disturbing protein folds [22].

In rigid and precisely oriented molecules, distances between 15 and 80 Å are easily measured with high accuracy [23], but in soft matter the problem is more difficult. Firstly, a distribution of distances would normally be present. Secondly, biological samples rarely have directional order – experimental data is an average over all orientations. Thirdly, the presence of out-of-pair interactions creates a background signal that depends on placement topology and concentration of external spins [24]. All of this, and the inevitable presence of noise, makes PDS data processing an ill-posed problem in the Hadamard sense [25]: in the absence of regularisation, the solution is neither unique nor a smooth function of the experimental data [26]. Mathematically, the difficulty is somewhere between the inverse Fourier transform [27] and the inverse Laplace transform [28] – not as easy as the former, but not as woefully unstable as the latter. In common with other deconvolution problems, neural networks do well [29] for reasons that are only starting to be explored [30].

A recent application of pulsed dipolar spectroscopy is nanometre-scale distance measurement performed in frozen biological cells [31-33], in particular using Gd(III) spin labels [34-37] that are chemically stable in the cytoplasm. Gd(III) tags produce an easily detectable ESR signal because their $(-1/2) \rightarrow (+1/2)$ electron spin transition has a narrow line at W-band and above [38]. They are also

useful in mixed labelling methods where the partner spin comes from an organic radical, such as nitroxide [39,40] or trityl [41]. However, the complicated quantum dynamics in Gd(III)-Gd(III) spin systems (S=7/2, large zero-field splitting) yields uninterpretable data unless certain safety margins are observed that allow the use of the weak coupling approximation [42,43] – pulse frequencies must be chosen so that the system stays in the weak dipole-dipole coupling regime [44,45]. Within those margins, Gd(III) spin labels are useful and informative [34,36-41], but the general form of Gd(III)-Gd(III) DEER kernel is unknown. A similar situation exists in RIDME spectroscopy where high-frequency overtones are present in high-spin systems, meaning that no simple analytical form exists for the kernel function [46]. This brings us back to neural networks that, from a large enough collection of examples, can learn the kernel (Figure 1) in a way that does not depend on magnetic or structural parameters of any specific physical system – the only requirements are: (a) knowing the mathematical class of distance distributions that can be encountered in PDS measurements; (b) being able to construct a training set that prepares the neural network for dealing with any distribution from that class.
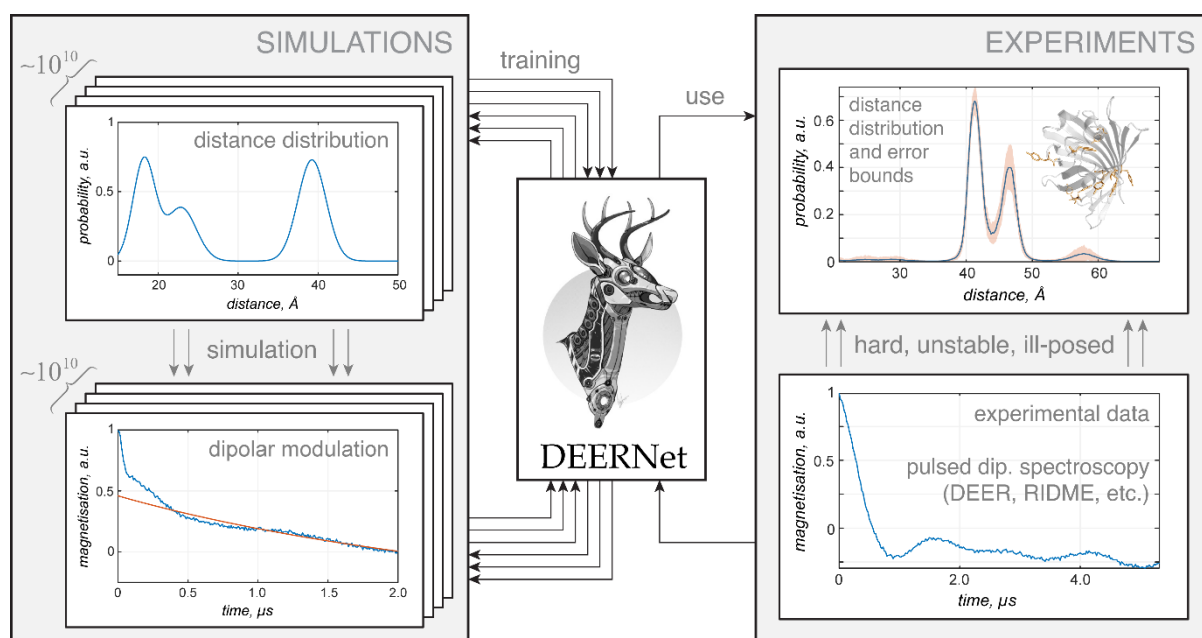

Figure 1. DEERNet training and inference flowchart. The block schematic shows the relationship between the computationally generated training database, the neural network, and its practical application in the context of pulsed dipolar spectroscopy (DEER, RIDME, etc.) distance measurement. To converge the training process, a carefully designed database with many billions of question-answer pairs is necessary; it can only be generated by highly accurate simulations including models of instrumental artefacts and noise. At the point when the network is used, it has no adjustable parameters.

An attractive feature of a trained neural net is the absence of user-adjustable parameters. Existing DEER and RIDME data processing tools feature sophisticated regularisation [47,48] and model fitting [49] algorithms with many unobvious settings, including the choice of background model and uncertainty estimation, each with its own parameters. Inexperienced users picking unsuitable parameter combinations has long been acknowledged as a problem. DEERNet does not expose any adjustable parameters to the user, yet its performance matches existing regularisation-based tools [29,50]. Unusually for machine learning, the important question about *exactly how* the neural network produces

the output signal has been answered [30] – DEERNet appears to be a combination of digital filters and regularised integral transforms. It is not fully transparent, but it is translucent.

The utility of neural networks for solving Fredholm equations [3] that arise in pulsed dipolar spectroscopy is firmly established [5,29,30]; the challenges at this point are logistical: network architecture, training database generation, convergence of the training process, uncertainty estimation, processing sparsely sampled data, and enforcing applicability for the approximations and assumptions made in the underlying models. These logistical aspects are the focus of the present paper.

2. DEERNet architecture

DEERNet is a signal processing network trained to solve the following inhomogeneous Fredholm equations of the second kind for the distance distribution function $p(r)$:

$$\left[(1-\mu)+\mu \int p(r)\gamma(r,t)dr\right]b(t)+n(t)=s(t)$$

$$b_{\mathrm{DEER}}(t)=\exp\left[(-kt)^{N/3}\right] \qquad b_{\mathrm{RIDME}}(t)=\exp\left[-a_1 t - a_2 t^2\right]$$

(1)

where $\gamma(r,t)$ is a known kernel [48,51], $s(t)$ is the experimentally recorded signal [51], $n(t)$ is Gaussian white noise with unknown standard deviation, k is an unknown DEER background decay rate [11], N is an unknown DEER background dimension [24], $a_{1,2}$ are unknown RIDME background decay parameters [52], and μ is an unknown modulation depth [11,51]. The problem is exceedingly difficult [26], but much progress was made over the last twenty years [47,48,51]. Neural networks were found to be surprisingly effective [29] – a recent in-depth inspection revealed digital filters that isolate the integral in Eq (1) followed by spectral filtering regularisation as a means of inverting the convolution operation [30].
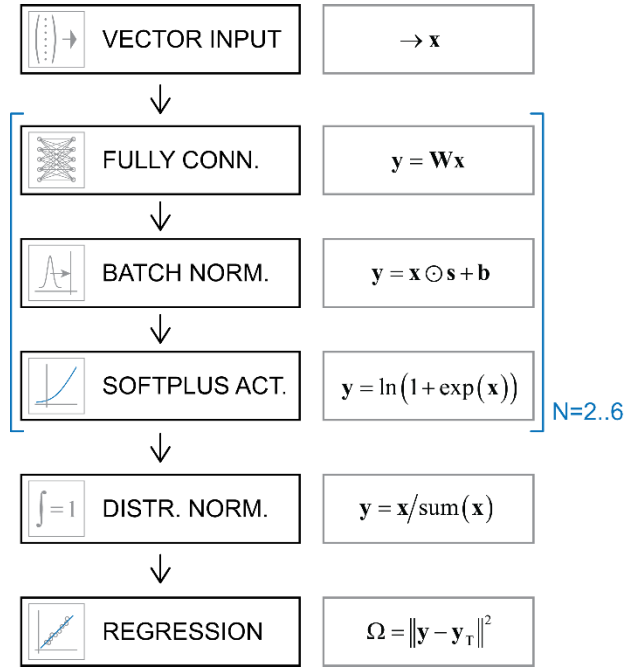
Figure 2. Architecture of the second generation DEERNet. *The blocks on the right show the mathematical operations performed by the corresponding layers. The adjustable parameters are weight matrices **W** in the fully connected layers, as well as scale and shift vectors **s** and **b** in the batch normalisation layers. The strong performance relative to the state of the art [29] is due to the fact that the mathematical operations in question are close to those needed in a Fredholm solver [4,5].*

The architecture of the original DEERNet [29,30] did not respect the physical requirement for asymptotic linearity of the DEER data processing problem – all activation functions were sigmoidal or logsigmoidal. The first generation network also made no use of convergence acceleration tools such as batch normalisation [53], and had no internal provision for normalising the output, which has a physical meaning of probability density. These matters are rectified in the second-generation architecture shown in Figure 2 and implemented in *Spinach 2.6* and later [54]. The new DEERNet is a feed-forward neural network with repeating triads of fully connected (FC), batch normalisation (BN), and softplus activation (SA) layers, terminated by a probability distribution normalisation layer.

Each FC-BN-SA triad applies the following transformation to the input vector $\mathbf{x}$:

$$\mathbf{y} = \ln\left\{\exp\left[(\mathbf{W}\mathbf{x})\odot\mathbf{s}+\mathbf{b}\right]+1\right\} \tag{2}$$

where $\odot$ stands for element-wise multiplication. This equation and the flowchart in Figure 2 warrant an extended explanation. The matrix-vector multiplication $\mathbf{W}\mathbf{x}$ is a linear transformation that – in the absence of instabilities and distortions – the DEER processing problem should have been. The subsequent scaling by $\mathbf{s}$ and shift by $\mathbf{b}$ are also a linear transformation – it is called batch normalisation [53]; it accounts for statistical variations between training data batches, thereby improving the convergence of the training process (Figure 3, top left). The remaining $\mathbf{y} = \ln\left\{\exp\left[...\right]+1\right\}$ transformation, called softplus [55], is the non-linearity required by the universal approximator theorem [1,2] – without it, the network could be rearranged into a single linear operation. Softplus function is non-linear at the origin, but asymptotically linear:

5

$$\ln\left\{\exp[x]+1\right\} = x + O\left(x^{-1}\right), \quad x \gg 1 \tag{3}$$

At the last FC-BN-SA triad in Figure 2, the choice of softplus activation function was dictated by the nature of DEERNet output – probability density, a non-negative quantity without an upper bound. At the intermediate triads, softplus appears empirically (not shown) to be a better choice than the more popular logsigmoidal function; this is likely because the network as a whole approximates an unstable but linear transformation. The number of FC-BN-SA blocks in Figure 2 is set to six because there is no improvement beyond that (Figure 3, bottom row). The final distribution normalisation (DN) layer enforces the physical requirement for all probabilities to sum up to 1; its presence yields a significant performance boost (Figure 3, top right), particularly for narrow distance distributions. It also makes the output independent of the modulation depth parameter μ in Equation (1).
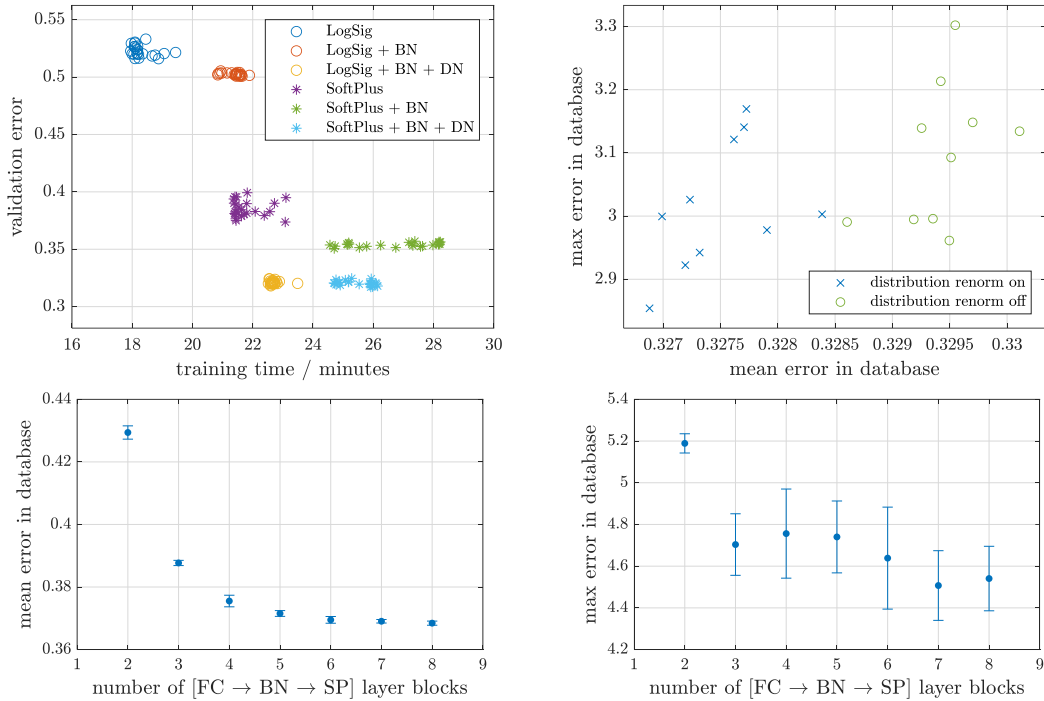

Figure 3. Performance effect of neural network architecture decisions. *Scatter plots and statistics refer to sets of independently trained (different random initial guess, different databases) DEERNets on a batch of 64,000 datasets generated as described in [29] and Section 3 below. The error is defined as root mean square deviation between the network output and the ground truth for the 512-element non-negative output vector, normalised (for numerical accuracy reasons in single-precision arithmetic) to have the mean value of 1. **(Top left)** Validation error and training time for different choices of activation function (logsigmoidal vs. softplus [55]) in the presence or absence of batch normalisation (BN) layers [53], and distribution normalisation (DN) layers. **(Top right)** Effect of the distribution normalisation layer on the mean and maximum (over the test database) regression error by the networks. **(Bottom row)** Mean and maximum error (over the test database) as functions of the number of [fully connected]-[batch normalisation]-[softplus activation] triads in the architecture shown in Figure 2.*

3. Infinite training database

The number of parameters in a typical DEERNet (Figure 2) is in the millions – there is not enough real experimental DEER or RIDME data in existence to train a network of this size. However, accurate sim-

ulations [48,54] are computationally affordable, and decades of experimental work have yielded practical intervals for the parameters of those simulations. Second-generation DEERNets in *Spinach 2.6* and later are trained on simulated signal databases with the parameters listed in Table 1.

DEER background parameters are well understood [11], but the choice of RIDME background parameter ranges in Table 1 merits a discussion, because it is empirical: $b_{\mathrm{RIDME}}(t)$ in Eq (1) only applies rigorously for $T_{1\mathrm{E}}/T_{2\mathrm{E}} > 10$ and no distribution of $T_{1\mathrm{E}}$ times in the sample – elsewhere, it merely has the status of something that fits the data well (here, $T_{1\mathrm{E}}$ refers to the pumped spin and $T_{2\mathrm{E}}$ refers to the detected spin [11]). Neural network training databases must be representative of the experimental data that the network would later see, but the definition of "representative" is subjective. From a recent theoretical treatment [52] and from practical experience [20], it is reasonable to assume that, if a RIDME background function is initially steady or growing, that it would have reached a turning point by $t = t_{\max}/5$ where $t_{\max}$ is the duration of the RIDME trace. We therefore seek such a combination of parameters $a_{1,2}$ as would guarantee that:

$$\begin{cases} \left[\dfrac{\partial}{\partial t} \exp\left(-a_1 t - a_2 t^2\right) \right]_{t=0} \geq 0 \\[2mm] \left[\dfrac{\partial}{\partial t} \exp\left(-a_1 t - a_2 t^2\right) \right]_{t \leq t_{\max}/5} = 0 \end{cases} \quad \Rightarrow \quad \begin{cases} a_1 \leq 0 \\[2mm] a_2 \geq -5a_1/2t_{\max} \end{cases} \tag{4}$$

Practical experience also indicates that, if a RIDME background function is initially decreasing, that it must continue decreasing for the duration of the experiment:

$$\begin{cases} \left[\dfrac{\partial}{\partial t} \exp\left(-a_1 t - a_2 t^2\right) \right]_{t=0} < 0 \\[2mm] \left[\dfrac{\partial}{\partial t} \exp\left(-a_1 t - a_2 t^2\right) \right]_{t=t_{\max}} < 0 \end{cases} \quad \Rightarrow \quad \begin{cases} a_1 > 0 \\[2mm] a_2 > -a_1/2t_{\max} \end{cases} \tag{5}$$

Physically, a_1 is the rate constant associated with the loss of spin correlation; users may be instructed to truncate the data set to a time point $t_{\max}$ when the dipolar modulation is no longer visible, corresponding roughly to $a_1 = 3t_{\max}^{-1}$ – we therefore chose this to be the standard deviation of the normal distribution of the a_1 parameter in the training database (Table 1). A similar extent of endpoint decay with respect to the a_2 parameter is achieved for $a_2 = 3t_{\max}^{-2}$, but its distribution is also bounded by the inequalities in Eqs (4) and (5). Accordingly, a half-normal distribution is used for a_2, from those bounds out into the positive direction (Table 1).

This kind of loose reasoning invites criticism; its justification comes from the nature of the background elimination process in DEERNet, which is known to be a digital filter [30] that simply destroys the background. Accordingly, we need only show the network the general class of signal components which it must learn to destroy without quantification – we need not be rigorous.

Table 1. Parameters used in the generation of training databases. *Where ranges are given, a value is picked randomly with the specified probabilities (discrete quantities) or from the specified distributions (continuous quantities) for each input-output data pair in the database. Variables refer to Equation (1).*

Parameter	Values
Network input and output vector dimension	512
Number of distance peaks in the distribution	1 - 3 (equal probability)
Full width at half-magnitude for distance peaks	from 5% of the distance to 50% of the distance range
Background decay rate k, fraction of $t_{\max}^{-1}$	0.0 - 1.0 (uniform distribution)
Background dimension N	2.0 - 3.5 (uniform distribution)
RIDME background parameter a_1	normal distribution $\langle a_1 \rangle = 0, \quad \sigma[a_1] = 3t_{\max}^{-1}$
RIDME background parameter a_2	half-normal distribution $\langle a_2 \rangle = \max\{-2a_1/5, -a_1/2\}$ $\sigma[a_2] = 3t_{\max}^{-2}$
Modulation depth μ	0.01 - 1.00 (uniform distribution)
Standard deviation of the noise in $n(t)$	$[0.00 - 0.05] \cdot \mu$ (uniform distribution)

Although the maximum number of distance peaks in the training database is set to a sensible value of three, the resulting networks can apparently handle more [29]. Asymmetric distance peak shapes are also reproduced (Supplementary Information, Section S2) even though the training database only contains Gaussians [29,30]. This stability to excursions outside the training range is unusual in machine learning; it may be the consequence of the unique appropriateness of deep networks with fully connected layers specifically for the deconvolution problem – rigorous convergence bounds exist [5] for approximating solutions of integral equations with neural nets, and Fredholm equations were among the earliest applications [4]. Apparently, networks are learning to actually solve Eq (1) as opposed to memorising or interpolating solutions [30].

Since the data source is artificial (Figure 1), the size of the training database is theoretically infinite. This removes the complications associated with overfitting – it is possible to design a database (in this case, a *Matlab* datastore object) that serves previously unseen data every time a batch is requested by the stochastic gradient descent algorithm [56], meaning that each converged network has effectively been trained on an infinite database. This approach has a useful side effect of reducing random-access memory and disk storage requirements: by the time the training process has converged, a typical DEERNet will have seen a few billion question-answer pairs – terabytes of data. Storing that in a

pre-computed form would have been impractical. A minor logistical optimisation is that the calculation of data batches may be done outside the thread that runs the training. Annotated source code and API documentation for all of the above is supplied with *Spinach* (http://spindynamics.org).

4. Background and modulation depth

Distance distribution is not the only information contained in the time-domain DEER data: the background function and the modulation depth are also useful [57-59]. However, because both are destroyed within the neural network [30], their extraction must follow a roundabout path: first the probability density is converted back into the form factor, then the form factor is mixed with an analytical background model, and then the mixture is retrofitted to the time-domain data. The stages are:

1. The normalised distance distribution is converted into the form factor $f(t)$ using the standard DEER kernel [26,47]. This is a stable and therefore unproblematic transformation:

$$f(t) = \int p(r)\gamma(r,t)\,dr \qquad \Leftrightarrow \qquad \mathbf{f} = \mathbf{\Gamma}\mathbf{p} \tag{6}$$

   where $\mathbf{f}$ and $\mathbf{p}$ are a discrete representations of $f(t)$ and $p(r)$ on finite grids, and $\mathbf{\Gamma}$ is the corresponding matrix representation of $\gamma(r,t)$.

2. The parameters of the background model are extracted from a least squares fit of the noiseless DEER signal model in Eq (1) to the experimental data:

$$\{\alpha, \mu, k, N\} = \arg\min \left\| s_{\text{expt}}(t) - s_{\text{theo}}(t) \right\|^2$$
$$s_{\text{theo}}(t) = \alpha \left[(1-\mu) + \mu f(t) \right] b(t) \tag{7}$$

   where α is a multiplier reflecting the overall scaling of the input signal, μ is the modulation depth, and $b(t)$ is a background model. First generation DEERNet [29] had a separate network set for background extraction, but the retrofitting approach advocated here appears to be superior in our practical testing. It also enables processing of batches of signals that share a common background; this feature is implemented in the new version.
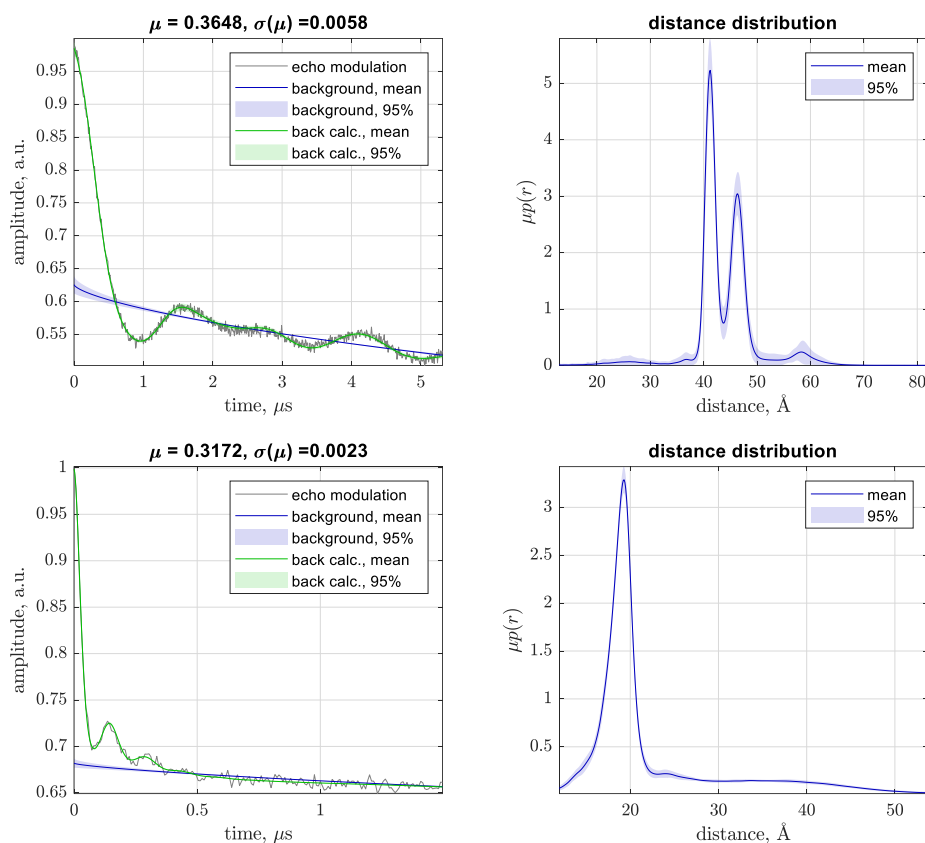
Figure 4. Distance distribution and background extraction performance. *The illustrations match those of the previous generation DEERNet reported in Figures 11 and 15 of [29]; note the improved distance range.* ***(Top row)*** *V96C/I143C site pair in the lumenal loop of a double mutant of light harvesting complex II, using iodocateamido-PROXYL spin labels attached to the indicated cysteines [60]. Residue 96 is located in the lumenal loop, and residue 143 is a structurally rigid "anchor" position in the protein core. In agreement with the results reported in the original paper [60], a bimodal distance distribution is obtained, indicating flexibility in the lumenal loop. The low-confidence peak around 57 Angstrom likely results from protein aggregation.* ***(Bottom row)*** *Pairs of nitroxide radicals tethered to the surface of gold nanoparticles, with the thiol tether attachment points diffusing on the surface of the nanoparticle (sample Au 3 after solvolysis and heating in [61]). The broad pedestal is real, it matches the analytical model reported in [61]; existing regularisation methods cannot process this dataset without either broadening the peak or introducing clumping artefacts into the pedestal.*

5. Uncertainty quantification

Output uncertainty analysis in DEER spectroscopy has been extensively researched [48-51,62,63], but neural nets offer two more methods beyond what is normally considered – the network ensemble method and the linear uncertainty propagation. The former is unique to neural networks and the latter proceeds from the fact that, unlike the formal convolution inversion operation, a neural network is stable (in the sense of having finite derivatives) with respect to its inputs.

5.1 Network ensemble method

In the machine learning community a popular way of estimating the uncertainty in the neural network output is to train an ensemble of nets on different databases from a different random initial guess, and to run descriptive statistics on their outputs [29] – examples (shaded blue intervals) are given in

the right hand panels of Figure 4 and Supplementary Information figures. Training 32 networks appears empirically to be sufficient; it is important to note that we did not try to ruggedise the networks against deliberately crafted adversarial inputs [64].

There are two sources of uncertainty in the distance distribution data: one associated with the distance distribution extraction process itself, and the other associated with the modulation depth obtained in Stage 2. These uncertainties are statistically independent because the probability density is normalised, meaning that the first point of the form factor is always exactly 1. The uncertainties of the outputs are therefore obtained by running statistics over sets of products $\left\{\mu^{(k)}p^{(k)}(r)\right\}$, $\left\{s_{\text{theo}}^{(k)}(t)\right\}$, and $\left\{\left(1-\mu^{(k)}\right)b^{(k)}(t)\right\}$ obtained from the network ensemble. The former yields the shaded confidence intervals in the right panels of Figure 4, and the latter two produce the blue (background) and the green (overall signal) shaded confidence intervals in the left panels.

5.2 Linear uncertainty propagation

A vector-in, vector-out neutral net $\mathbf{y} = \Upsilon(\mathbf{x})$ with continuous activation functions is easy to differentiate with respect to the input vector; this may be done by a finite difference approximation [65] or using the same automatic differentiation methods that compute the gradient for the training process [66]. The derivative of the output vector with respect to the input vector is the Jacobian matrix $\mathbf{J}$:

$$\mathbf{y} = \Upsilon(\mathbf{x}), \qquad J_{nk} = \partial y_n / \partial x_k \tag{8}$$

The appearance of the network Jacobian (Figure 5, left panel), relative to the Jacobian obtained by differentiating Eq (1) analytically (Figure 5, right panel), illustrates of the stability of the neural net with respect to perturbations of its input. The network is apparently only sensitive to variations in the input data when those variations influence significant distance peaks (see the top row of Figure 4 for the distance distribution). The values of the derivatives are moderate throughout.
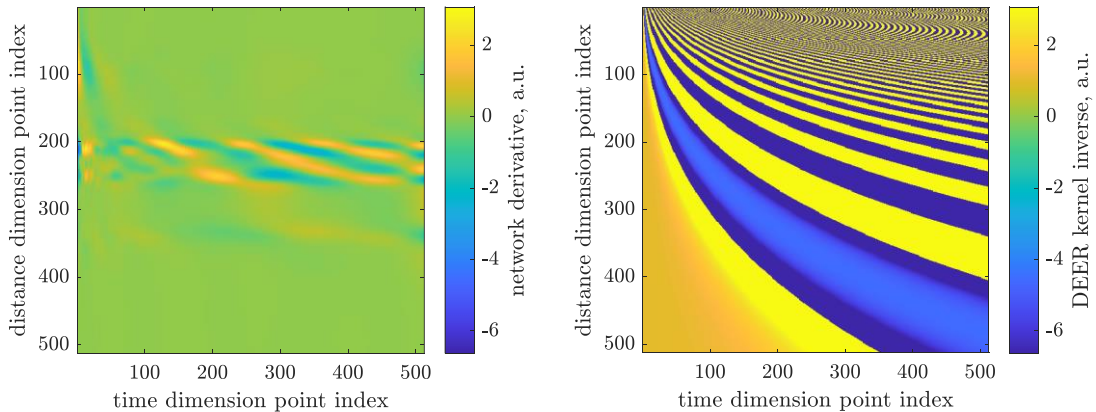

Figure 5. (Left) *A colour map representation of the Jacobian matrix of DEERNet (an average over 32 networks trained on different databases from a different random initial guess) for the input-output pair appearing in the top row of Figure 4. The horizontal dimension corresponds to the input vector and the vertical dimension to the output vector.* **(Right)** *A clipped (to the same range as the left panel) colour map representation Jacobian matrix obtained by computing the point-by-point inverse of the DEER kernel in Eq (9).*

That is emphatically not the case for the Jacobian obtained from Eq (1), where the statistical expectation value of the variational derivative of the distance distribution with respect to the input signal has entire curves filled with infinities because the kernel $\gamma(r,t)$ is zero on those curves:

$$\langle \delta p(r)/\delta s(t) \rangle = \left[\mu \gamma(r,t) b(t) \right]^{-1} \tag{9}$$

This is of course the exact reason why conventional processing of DEER data requires regularisation; it would seem that the neural network has found a way around the problem. At the moment, there is only limited understanding of how it does that [30]; on the bright side, the availability of first derivatives enables linear uncertainty propagation:

$$\sigma_{y_n}^2 \approx \sum_k \left(\frac{\partial y_n}{\partial x_k} \right)^2 \sigma_{x_k}^2 = \sum_k J_{nk}^2 \sigma_{x_k}^2 \tag{10}$$

in which standard deviations σ_{x_k} may be estimated from the fitting residuals in the time domain, or from the difference between the raw signal and a suitably filtered (for example, by the Savitsky-Golay procedure) version. This is illustrated for a RIDME dataset in Figure 6.
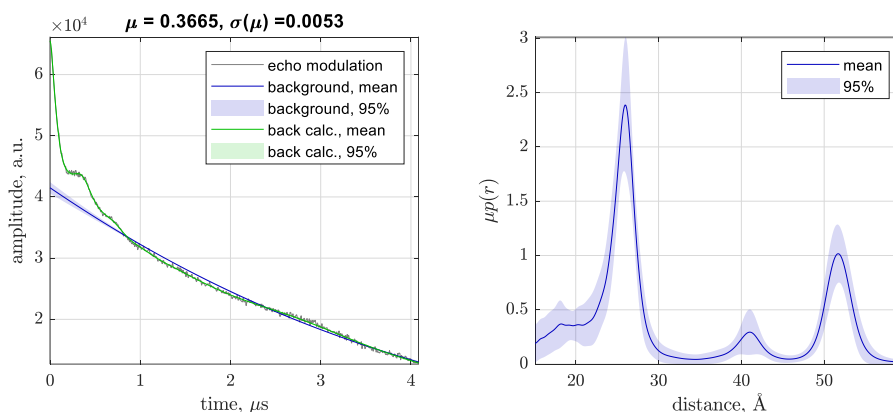

Figure 6. An example of linear uncertainty propagation in the distance distribution. *The sample is a protein homodimer of copper amine oxidase from Arthrobacter globiformis (AGAO) containing a Cu^{2+} ion bound to a surface site on each monomer and one MTSL spin label per monomer. RIDME data was collected with a mixing time of 5 µs. Two dominant Cu^{2+} to MTSL distances are expected: one intra-monomer (26 Å) and one intra-dimer (52 Å). Two-stage regularised processing methods that require background elimination report zero confidence for the biologically certain 52 Å distance (Section S4 in the SI) – single-stage processing with neural networks is clearly superior in this case. Sample preparation procedures and details of data collection may be found in Figure 3a of [67].*

6. Sparsely sampled data

Because DEER and RIDME signals are detected as indirectly incremented dimensions of two-dimensional ESR experiments, there is much scope for instrument time savings using the same sparse sampling methods that are popular in NMR spectroscopy [68-70]. However, an attempt to train a neural net by feeding sparsely sampled data as index-value pairs predictably fails. We also abstain from existing solutions that work well [71], but without – hopefully not for long – an explanation of how.

The key to obtaining interpretable and effectual networks in this case is what we call the "no magic" assumption – based on current evidence [30], we expect digital signal processing networks to invent

efficient implementations of known mathematics. Given a sparsely sampled input, we could expect a fitting or an interpolation procedure to emerge inside; the input should therefore be adjusted to make it easier for the network to do that.

Polynomial interpolation and fitting share the same mathematical structure – a low-order polynomial is assumed to describe the data well in a sufficiently small interval:

$$f(t) = a_0 + a_1 t + a_2 t^2 + ... \tag{11}$$

and its coefficients are obtained by solving, exactly or approximately, the system of equations requiring the polynomial to return values $\{f_n\}$ at locations $\{t_n\}$:

$$\begin{pmatrix} 1 & t_1 & t_1^2 & \cdots \\ 1 & t_2 & t_2^2 & \cdots \\ 1 & t_3 & t_3^2 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ \vdots \end{pmatrix} \tag{12}$$

This system may be overdetermined, in which case it is solved for the coefficients $\{a_k\}$ in the least squares sense. Those coefficients are linear functions of the known values $\{f_n\}$ and rational functions of the time grid locations $\{t_n\}$.
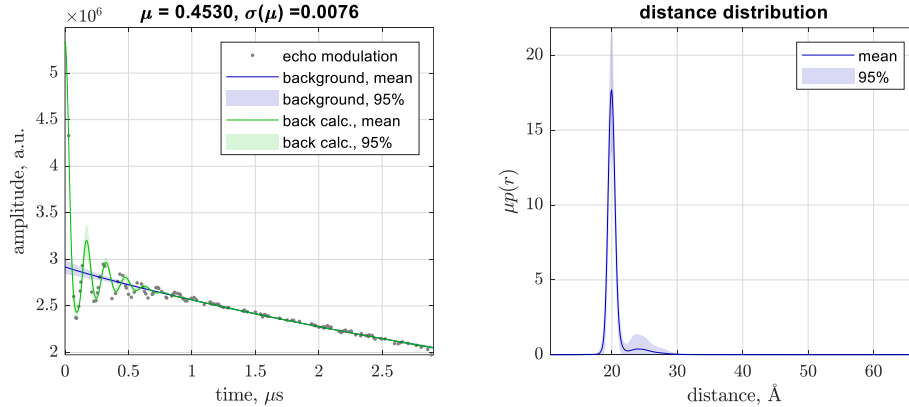

Figure 7. A sparsely sampled DEER dataset and the result of its neural network processing. *Experimental data come from an X-band DEER measurement of model biradical 3 from Ref [26] diluted into o-terphenyl. The dataset contains 512 time grid points, with the probability of a sample being recorded at each point set uniformly to 1/4.*

The internal mechanics of feed-forward neural networks makes it easy for them to learn unary and additive binary functions of the input (sum, inverse, square, *etc.*), but hard to learn multiplicative binary functions – therefore, sets of products like $\{t_n f_n\}$ and $\{t_n^2 f_n\}$ must be supplied as inputs alongside $\{t_n\}$ and $\{f_n\}$. When this is done, network performance approaches the fully sampled case (Figure 7). We found it unnecessary to pick any specific class of sampling schedules [70] at the training stage, but some performance improvements would likely result if sampling schedules are matched between the experimental data and the training database, at the cost of loss of generality.

The only constraint on sparse sampling DEERNets is related to the fixed element count in the input and output vectors of the feed-forward architecture shown in Figure 2: each network ensemble is

trained for a specific number of non-empty samples. 128/512 and 64/512 networks are included with DEERNet in *Spinach 2.6*, as are scripts to train network sets for other time grid sizes and sample counts. Although individual data samples may be missing, the underlying time grid must be uniform to enable the calculation of the distance window in the output.

Sparse sampling raises the delicate question of the minimum and maximum distance that can be reliably expected from the distribution reconstruction process. DEER traces are not sinusoidal, but the frequency spectrum of the oscillatory part of the kernel [29]:

$$\gamma(r,t) = \sqrt{\frac{\pi}{6\omega_D t}} \left[\cos(\omega_D t) \mathrm{FrC}\left(\sqrt{\frac{6\omega_D t}{\pi}}\right) + \sin(\omega_D t) \mathrm{FrS}\left(\sqrt{\frac{6\omega_D t}{\pi}}\right) \right]$$

$$\omega_D = \frac{\mu_0}{4\pi}\frac{\gamma_1\gamma_2\hbar}{r^3}; \quad \mathrm{FrC}(x) = \int_0^x \cos(\pi t^2/2)dt; \quad \mathrm{FrS}(x) = \int_0^x \sin(\pi t^2/2)dt \tag{13}$$

is inside $\left[-2\omega_D, 2\omega_D\right]$, and therefore – by Whittaker's interpolation formula [72] – the minimum distance captured by a uniformly sampled DEER trace with time grid spacing Δt is:

$$\Delta t = \frac{\pi}{2\omega_D^{\max}} \quad \Rightarrow \quad r_{\min}^{\mathrm{DEER}} = \sqrt[3]{\frac{\gamma_1\gamma_2\mu_0\hbar}{2\pi^2}\Delta t} \tag{14}$$

In the uniformly sampled DEERNet, the grid spacing is either that of the original input data, or that of the resampled signal that is supplied to the network, whichever is larger. When individual samples may be missing, no perfect reconstruction conditions are available, but an optimistic choice for Δt is the smallest interval seen between the samples that are present.

The maximum distance is set by the requirement to sample at least half a period of $\omega_D^{\min}$. For a uniformly sampled DEER trace with the last sample at $t_{\max}$:

$$t_{\max} = \frac{\pi}{\omega_D^{\min}} \quad \Rightarrow \quad r_{\max}^{\mathrm{DEER}} = \sqrt[3]{\frac{\gamma_1\gamma_2\mu_0\hbar}{4\pi^2}t_{\max}} \tag{15}$$

Here too, when individual samples may be missing, no perfect reconstruction conditions are available, but an optimistic choice for $t_{\max}$ is the time of the last sample that is recorded. All of this arithmetic is performed internally by DEERNet; it is not visible to the user. Note the sliding multiplicative relationship between $\gamma_{1,2}$, r^3, and t in the $\omega_D t$ product in Eq (13) – variations in experiment duration and magnetogyric ratios are accommodated by distance axis rescaling; this is handled outside the neural nets which internally use scaled and normalised dimensionless units.

The safety margins must be tighter for RIDME data because RIDME background in Eq (1) is not necessarily monotonic and the assumption of single electron T_1 time does not necessarily hold – thus, a full period of the sinusoidal wave is now required for reliably telling dipolar oscillations from the background. Likewise, the short-distance limit of RIDME may contain overtone frequencies in high-spin systems, with the resulting need to sample (and therefore prevent reflections of) frequencies about a factor of two higher than $2\omega_D$. The resulting safety margins are:

$$r_{min}^{RIDME} = \sqrt[3]{\frac{\gamma_1\gamma_2\mu_0\hbar}{\pi^2}\Delta t}, \qquad r_{max}^{RIDME} = \sqrt[3]{\frac{\gamma_1\gamma_2\mu_0\hbar}{8\pi^2}t_{max}} \tag{16}$$

The intervals in Eqs (14)-(16) are used for training database generation and for deciding output distance extents in second-generation DEERNet.

7. Practical flowchart

This section provides practical advice on the use of second-generation DEERNet. The package is distributed with *Spinach 2.6* [54] and later (http://spindynamics.org). It is open-source, but requires *Matlab R2021a* or later, with *Optimisation Toolbox*, *Parallel Computing Toolbox*, *Deep Learning Toolbox*, and *Reinforcement Learning Toolbox* installed.

7.1 Data preprocessing

The responsibility for supplying valid data rests with the user – internal validation only catches glaring errors (array dimension and type mismatch, unphysical values, *etc.*). The input must be:

(a) Sampled, sparsely or fully, on a uniform time grid. This is dictated by the distance edge mathematics in Eqs (14)-(16), which is only unambiguous for uniform time grids. Data array dimension must be the same as that of the time grid array. Data values corresponding to the missing samples must be set to NaN (not-a-number).

(b) Phased as near to zero imaginary part as the signal-to-noise ratio permits. This may be done manually, by picking such φ as to minimise a norm of $\mathrm{Im}\left[e^{i\varphi}s(t)\right]$, or automatically – for example by numerical minimisation of its norm-square:

$$\varphi = \arg\min\left\|\mathrm{Im}\left[e^{i\varphi}s(t)\right]\right\|^2 \tag{17}$$

Data that cannot be phased in this way is corrupted; it must not be used.

(c) Cropped and shifted so as to make the echo modulation maximum correspond to the $t = 0$ point of the time grid. DEERNet assumes time grid units to be seconds.

Further processing may be required for DEER and RIDME data recoded with non-standard pulse sequences. DEERNet assumes that the data came from (or has been transformed to correspond to) the standard 4-pulse DEER [73] or a 5-pulse RIDME [16] experiment. An input filter for the popular BES3T format is provided.

7.2 Input and output data structures

The first argument of the DEERNet function call is a real-valued column vector (or a matrix made of multiple column vectors) containing PDS data; the second argument is a real-valued column vector containing the time grid, and the third argument specifies the experiment ('deer' or 'ridme'):

```
dataset=deernet(input_traces,input_axis,expt);
```

Because feed-forward neural networks have a fixed input dimension and normalisation expectations, appropriate transformations are applied internally; those are automatic and not visible to the user.

The output is a graph with the appearance and content of Figures 4, 6, 7, and a data structure with the fields specified in Table 2. As a matter of policy, DEERNet has no adjustable parameters.

Table 2. Second-generation DEERNet output. *The program returns a Matlab data structure with the indicated fields.*

Field of the output structure	Content
`.input_axis`	time grid (seconds), as received
`.input_traces`	DEER or RIDME data, as received
`.n_networks`	number of neural nets in the ensemble
`.train_params`	neural network training parameters, described in [29]
`.resamp_axis` `.resamp_traces`	time grid and data, resampled to 512 time points to match the input dimension of the neural networks
`.dist_ax`	distance grid (Angstrom) of the output

`.dist_av`
`.dist_lb`
`.dist_ub`
	mean, 95% lower bound, and 95% upper bound on the distance distribution(s); this corresponds to $p(r)$ in Eq (1)

`.backgs_av`
`.backgs_lb`
`.backgs_ub`
	mean, 95% lower bound, and 95% upper bound on the background signals; this corresponds to $(1-\mu)b(r)$ in Eq (1)

`.retros_av`
`.retros_lb`
`.retros_ub`
	mean, 95% lower bound, and 95% upper bound on the back-calculated fit to the experimental data; this corresponds to $s(t)-n(t)$ in Eq (1)

`.mdpths_av` `.mdpths_st`	mean and standard deviation of the modulation depths; this corresponds to μ in Eq (1)

8. Practical applicability considerations

This section deals with the practicalities of running a regularised version of an unstable mathematical operation; these matters are the same for neural networks and conventional DEER/RIDME data processing methods. The absence of user-adjustable parameters in *DEERNet* is a significant advantage, but the principle of *"garbage in, garbage out"* does of course still apply.

8.1 Distances at the long edge

The maximum distance bound in Eqs (15) and (16) ignores the presence of the damping multiplier on the left of the square bracket in Eq (13), of the modulation depth transformation in the square bracket of Eq (7), and of the decaying functions in the baseline term. For this reason, all DEER and RIDME processing methods struggle to tell apart slow dipolar modulations and baseline decays. As a result, distance distributions with significant density at the long edge are unreliable (Figure 8). Because baseline decay is usually significant, the ghost peaks at the long edge of the distance distribution may be reported with high confidence (Figure 8, top right). Those long-distance peaks may sometimes be real (there are examples in the Supplementary Information), but may just as well be artefacts – not a comfortable situation in the data processing context.
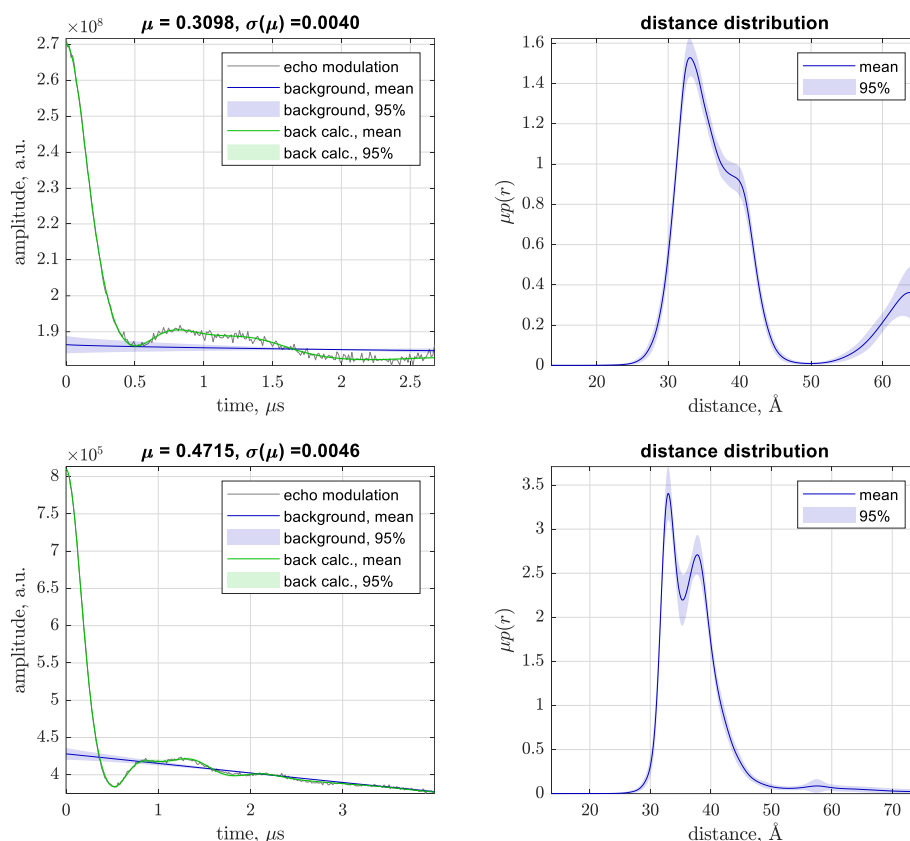
Figure 8. Example of an artefact peak at the long distance edge. *The artefact peak at 65 Å in the top right panel disappears and the true distances become better resolved when a longer DEER trace (bottom row) for the same sample is recorded and processed. Note the incorrect background and therefore incorrect modulation depth in the top panels. Experimental data from Figure S11b of the recent community white paper on PDS data processing [74].*

In such cases, second generation *DEERNet* prints a warning message encouraging the user to record a longer trace, and turns off the output of background signal and modulation depth. This is also done when unreasonable standard deviations are detected in the uncertainty quantification described in Section 5. An appealing feature of *DEERNet* is that background signal retrofitting is optional. This is because backgrounds – with whatever parameters – are eliminated by the notch filter in the first layer of the distance distribution extraction networks [30].

Because RIDME backgrounds are less well understood than DEER backgrounds, two-stage (background elimination followed by deconvolution) Tikhonov-regularised processing methods struggle with longer distances in RIDME data (see the note in the caption of Figure 6). Thus, even in DEERAnalysis2021, the recommended data processing option for such RIDME data is "DEERNet/RIDME". This is further illustrated in Section S4 of the Supplementary Information.

8.2 Low modulation depth

A conceptually difficult scenario for any DEER/RIDME processing method is quantifying distances that do not exist – a sample need not contain any radical pairs. Modulation depth quantification described in Section 5 acts as a safeguard here; it is implemented as two criteria: if the modulation depth returned by any of the networks in the ensemble is negative, or if the ensemble averaged modulation

depth is below 0.5%, second generation *DEERNet* refuses to output the distance distribution and prints a deliberately forceful error message (Figure 9, right panel). The message is essential: open source software technically allows users to turn safeguards off; it is important to discourage that practice.
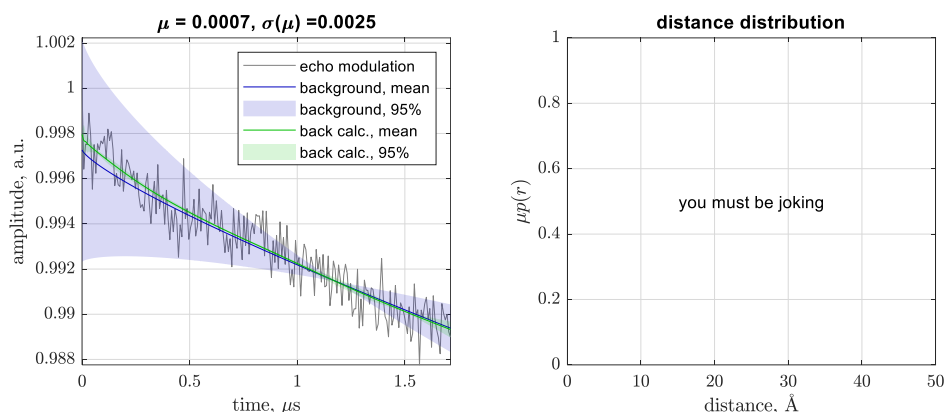

Figure 9. An illustration of the built-in sanity control in the second-generation DEERNet. Distance distribution output is turned off when data is judged by the neural networks to not constrain it sufficiently well. The left panel shows a Q-band DEER measurement of a singly Gd(III) DOTA maleimide labelled protein ligand in the absence of the binding partner and therefore a priori absence of spin echo envelope modulation. The artefact seen in the left panel is due to microwave pulse interference that is not fully suppressed by phase cycling.

8.3 Running echoes and artefacts

The low modulation depth of Gd(III)-Gd(III) data makes it susceptible to microwave pulse interference artefacts that manifest as running echoes that cross the integration window and may survive phase cycling if observer and pump channel are phase-locked [75]. These running echoes are a common sight in RIDME data; they can appear even in Q-band DEER using Gaussian pulses on nitroxides. Another type of distortion arises at ends DEER time traces due to the residual "2+1" pulse train electron spin echoes [76]; those may sometimes be suppressed using Gaussian pulses with non-overlapping excitation profiles, but may also be unavoidable [77].

A surprising observation is that neural networks are – for want of a better term – ignoring these artefacts (Figure 10). This appears to be a beneficial side effect of the digital filter that they evolve in the input layer [30] – running echoes do not pass through that filter. Should the need arise, running echoes may be introduced as procedurally generated artefacts into the training database.
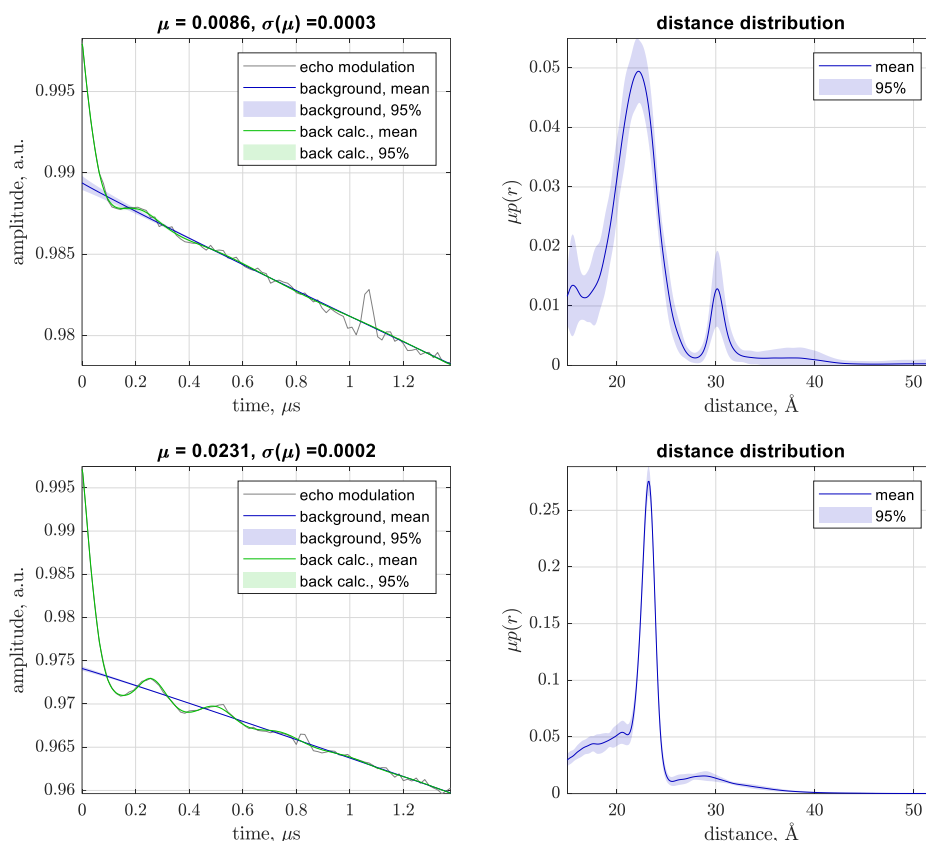
Figure 10*. **An illustration of neural network resilience to running echo artefacts.*** *The time traces are DEER data for a bis-Gd(III)-DOTA model compound, measured with a large frequency separation (top row: 0.363 GHz, bottom row: 1.090 GHz) between the pump pulse and the observe pulse to reduce the effects of non-secular dipolar terms and zero-field splitting [44]. A running echo, ignored by the neural network, is visible at 1.1 μs (top left panel) and 0.8 μs (bottom left panel). The distance peaks at 30 Å and below 20 Å in the right panels are artefacts that depend on the frequency separation between pump and probe pulse; they arise because the kernel in Eq (13) had been derived for spin-1/2 particles, which Gd(III) is not.*

Zero-field splitting makes quantum mechanics of high-spin pairs exceedingly difficult [42,43] – at the moment, there are no computationally affordable ways to process Gd(III) DEER data unless pulse frequencies are positioned so that the system stays in the weak coupling regime where Eq (13) is applicable. That approximation creates artefacts in the distance distribution. At the time of writing, they cannot be eliminated, but can be identified because their position and relative intensity fluctuate as a function of pump and probe pulse frequency separation (Figure 10, right panels).

8.4 Orientation selection

Microwave pulse widths in DEER and RIDME sequences are typically tens of nanoseconds long, corresponding to the excitation bandwidth in the tens of MHz. This is much less than the width of the *g*-tensor powder patterns of common spin labels (~ 500 MHz for nitroxides at W-band); the result is called orientation selection – some orientations of the spin system see the microwave pulse, others do not [78]. Because the kernel in Eq (13) is a full powder average, sending orientationally selected data into DEERNet is equivalent to sending corrupted data: dipolar modulation frequencies and depths strongly depend (Figure 11, left panel) on where the user decides to apply the pulse.
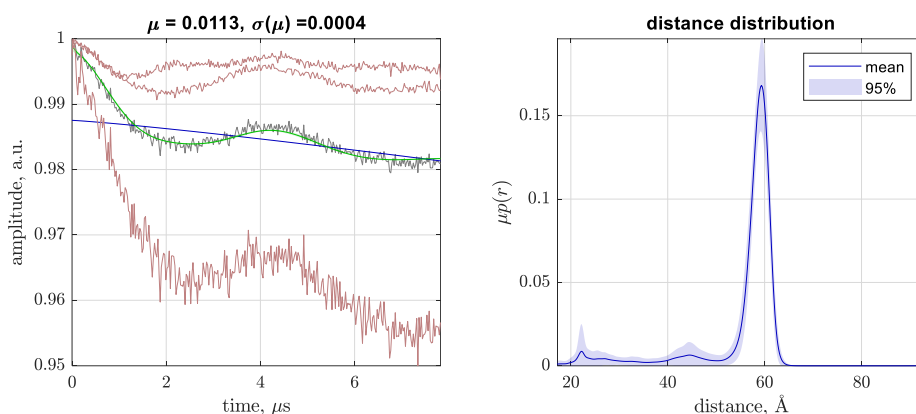
Figure 11. An example of orientation selection in a 4-pulse DEER experiment. *The sample is an ERp29 protein [79] dimer mutant (S114C) spin-labelled with Gd C1 and MTSL tags [80]. The orientation selection effect appears because the width of MTSL powder pattern at W-band is around 500 MHz, but the excitation bandwidth of the pump pulse is only ~10 MHz. The three red traces in the left panel have been recorded at three different pump frequencies inside the MTSL powder pattern; the grey trace with the background signal (blue line) is the average of the three red traces; the distance distribution on the right hand side was computed for that average.*

In situations where the powder pattern width is significantly larger than the pulse excitation width, DEER data must be acquired with pulse position averaging – several data sets with different pump and probe pulse frequencies must be acquired and averaged (Figure 11, grey trace on the left) to make sure that the isotropic powder average kernel in Eq (13) is applicable. It is not possible to train a neural network that would process arbitrary orientationally selected DEER / RIDME data because the one-to-one correspondence between the distance distribution and the modulation is broken.

9. Conclusions and outlook

In the practical testing (Sections S2-S4 in the Supplementary Information) performed by the groups involved in this study, deep neural network processing of DEER and RIDME data matches or exceeds the performance of the existing tools, insofar as "performance" may be defined for experimental da-tasets with uncertain ground truth; the same conclusion was reached for DEER in the recent community white paper [74].The unique advantage of having no user-adjustable parameters, and thus no subjective bias, is pertinent because it removes the element of user discretion and error from the fiddly and subjective process of selecting regularisation parameters.

So far, common use scenarios for neural networks in DEER data processing have been: (a) in difficult cases where user bias in the choice of baseline and regularisation parameters must be eliminated; (b) as a second opinion alongside Tikhonov or model fitting tools; (c) as a source of initial guesses for Tikhonov or model fitting tools; (d) in situations when the users of DEER spectroscopy have no mathematical physics background and no interest in methodology, and only seek a reliable data processing tool. An appealing feature of *DEERNet* is that background signal retrofitting is optional: during distance distribution extraction, neural networks simply eliminate the background with a notch filter [30]; it may be retrofitted, if necessary, as described in Section 4.

The unexpectedly strong performance of neural network based Fredholm solvers merits a special mention – a recurring theme in the user feedback is the resilience of *DEERNet* to dismal experimental realities like low signal-to-noise ratios and incompletely sampled datasets. We would note here that many experimental data processing problems across natural (and unnatural) sciences reduce to solving a Fredholm equation; in this context, neural networks have a bright future.

Acknowledgements

Statement of competing interests

All authors declare that they have no competing interests.

Data and materials availability

All data and program code associated with this manuscript may be obtained by downloading *Spinach 2.7* or later versions from *https://spindynamics.org* web site.

References

[1]   G. Cybenko, Approximation by superpositions of a sigmoidal function, Mathematics of Control, Signals and Systems, 2 (1989) 303-314.

[2]   K. Hornik, Approximation capabilities of multilayer feedforward networks, Neural networks, 4 (1991) 251-257.

[3]   I. Fredholm, Sur une classe d'équations fonctionnelles, Acta mathematica, 27 (1903) 365-390.

[4]   V. Vemuri, G.-S. Jang, Inversion of Fredholm integral equations of the first kind with fully connected neural networks, Journal of the Franklin Institute, 329 (1992) 241-257.

[5]   V. Kůrková, Surrogate modelling of solutions of integral equations by neural networks, in: IFIP International Conference on Artificial Intelligence Applications and Innovations, Springer, 2012, pp. 88-96.

[6]   A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Advances in neural information processing systems, 25 (2012) 1097-1105.

[7]   J. Chorowski, D. Bahdanau, D. Serdyuk, K. Cho, Y. Bengio, Attention-based models for speech recognition, arXiv preprint arXiv:1506.07503, (2015).

[8]   C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1-9.

[9]   T. Pu, J.-Y. Ou, N. Papasimakis, N. Zheludev, Label-free deeply subwavelength optical microscopy, Applied Physics Letters, 116 (2020) 131105.

[10]   A. Pashaei, H. Sajedi, N. Jazayeri, Brain tumor classification via convolutional neural network and extreme learning machines, in: 2018 8th International conference on computer and knowledge engineering (ICCKE), IEEE, 2018, pp. 314-319.

[11]   A. Milov, A. Ponomarev, Y.D. Tsvetkov, Electron-electron double resonance in electron spin echo: model biradical systems and the sensitized photolysis of decalin, Chemical Physics Letters, 110 (1984) 67-72.

[12]   А.Д. Милов, К.М. Салихов, М. Щиров, Применение метода двойного резонанса в электронном спиновом эхо для изучения пространственного распределения парамагнитных центров в твердых телах, Физика Твердого Тела, 23 (1981) 975-982.

[13]   R.E. Martin, M. Pannier, F. Diederich, V. Gramlich, M. Hubrich, H.W. Spiess, Determination of end-to-end distances in a series of TEMPO diradicals of up to 2.8 nm length with a new four-pulse double electron electron resonance experiment, Angewandte Chemie International Edition, 37 (1998) 2833-2837.

[14]   M. Pannier, S. Veit, A. Godt, G. Jeschke, H.W. Spiess, Dead-Time Free Measurement of Dipole–Dipole Interactions between Electron Spins, Journal of Magnetic Resonance, 142 (2000) 331-340.

[15]   L. Kulik, S. Dzuba, I. Grigoryev, Y.D. Tsvetkov, Electron dipole–dipole interaction in ESEEM of nitroxide biradicals, Chemical physics letters, 343 (2001) 315-324.

[16]   S. Milikisyants, F. Scarpelli, M.G. Finiguerra, M. Ubbink, M. Huber, A pulsed EPR method to determine distances between paramagnetic centers with strong spectral anisotropy and radicals: The dead-time free RIDME sequence, Journal of Magnetic Resonance, 201 (2009) 48-56.

[17]   G. Jeschke, DEER Distance Measurements on Proteins, Annual Review of Physical Chemistry, 63 (2012) 419-446.

[18]   G. Jeschke, The contribution of modern EPR to structural biology, Emerging Topics in Life Sciences, 2 (2018) 9-18.

[19]   F.D. Breitgoff, K. Keller, M. Qi, D. Klose, M. Yulikov, A. Godt, G. Jeschke, UWB DEER and RIDME distance measurements in Cu (II)–Cu (II) spin pairs, Journal of Magnetic Resonance, 308 (2019) 106560.

[20]   I. Ritsch, H. Hintz, G. Jeschke, A. Godt, M. Yulikov, Improving the accuracy of Cu (II)–nitroxide RIDME in the presence of orientation correlation in water-soluble Cu (II)–nitroxide rulers, Physical Chemistry Chemical Physics, 21 (2019) 9810-9830.

[21]   J.P. Klare, Site-directed spin labeling EPR spectroscopy in protein research, Biological Chemistry, 394 (2013) 1281-1300.

[22]   L.J. Berliner, J. Grunwald, H.O. Hankovszky, K. Hideg, A novel reversible thiol-specific spin label: Papain active site labeling and inhibition, Analytical Biochemistry, 119 (1982) 450-455.

[23]   S. Santabarbara, I. Kuprov, W.V. Fairclough, S. Purton, P.J. Hore, P. Heathcote, M.C. Evans, Bidirectional electron transfer in photosystem I: determination of two distances between P700+ and A1-in spin-correlated radical pairs, Biochemistry, 44 (2005) 2119-2128.

[24]   A. Milov, Y.D. Tsvetkov, Double electron-electron resonance in electron spin echo: Conformations of spin-labeled poly-4-vinilpyridine in glassy solutions, Applied Magnetic Resonance, 12 (1997) 495-504.

[25]   J. Hadamard, Sur les problèmes aux dérivées partielles et leur signification physique, Princeton university bulletin, (1902) 49-52.

[26]   G. Jeschke, A. Koch, U. Jonas, A. Godt, Direct conversion of EPR dipolar time evolution data to distance distributions, Journal of magnetic resonance, 155 (2002) 72-82.

[27]   J.-B.J. Fourier, Théorie analytique de la chaleur, Firmin Didot, 1822.

[28]   P.S. Laplace, Théorie analytique des probabilités, Courcier, 1820.

[29]   S.G. Worswick, J.A. Spencer, G. Jeschke, I. Kuprov, Deep neural network processing of DEER data, Science advances, 4 (2018) eaat5218.

[30]   J.L. Amey, J. Keeley, T. Choudhury, I. Kuprov, Neural network interpretation using descrambler groups, Proceedings of the National Academy of Sciences, 118 (2021).

[31]   R. Igarashi, T. Sakai, H. Hara, T. Tenno, T. Tanaka, H. Tochio, M. Shirakawa, Distance determination in proteins inside Xenopus laevis oocytes by double electron– electron resonance experiments, Journal of the American Chemical Society, 132 (2010) 8228-8229.

[32]   M. Azarkh, O. Okle, V. Singh, I.T. Seemann, J.S. Hartig, D.R. Dietrich, M. Drescher, Long-Range Distance Determination in a DNA Model System inside Xenopus laevis Oocytes by In-Cell Spin-Label EPR, ChemBioChem, 12 (2011) 1992-1995.

[33]   I. Krstić, R. Hänsel, O. Romainczyk, J.W. Engels, V. Dötsch, T.F. Prisner, Long-range distance measurements on nucleic acids in cells by pulsed EPR spectroscopy, Angewandte Chemie International Edition, 50 (2011) 5070-5074.

[34]   A. Martorana, G. Bellapadrona, A. Feintuch, E. Di Gregorio, S. Aime, D. Goldfarb, Probing protein conformation in cells by EPR distance measurements using Gd3+ spin labeling, Journal of the American Chemical Society, 136 (2014) 13458-13465.

[35]   M. Qi, A. Groß, G. Jeschke, A. Godt, M. Drescher, Gd (III)-PyMTA label is suitable for in-cell EPR, Journal of the American Chemical Society, 136 (2014) 15366-15378.

[36]   F.C. Mascali, H.V. Ching, R.M. Rasia, S. Un, L.C. Tabares, Using Genetically Encodable Self-Assembling GdIII Spin Labels To Make In-Cell Nanometric Distance Measurements, Angewandte Chemie International Edition, 55 (2016) 11041-11043.

[37]   S. Kucher, C. Elsner, M. Safonova, S. Maffini, E. Bordignon, In-Cell Double Electron–Electron Resonance at Nanomolar Protein Concentrations, The Journal of Physical Chemistry Letters, 12 (2021) 3679-3684.

[38]   A. Giannoulis, Y. Ben-Ishay, D. Goldfarb, Characteristics of Gd (III) spin labels for the study of protein conformations, Methods in Enzymology, 651 (2021) 235-290.

[39]   I. Kaminker, H. Yagi, T. Huber, A. Feintuch, G. Otting, D. Goldfarb, Spectroscopic selection of distance measurements in a protein dimer with mixed nitroxide and Gd 3+ spin labels, Physical Chemistry Chemical Physics, 14 (2012) 4355-4358.

[40]   L. Garbuio, E. Bordignon, E.K. Brooks, W.L. Hubbell, G. Jeschke, M. Yulikov, Orthogonal Spin Labeling and Gd (III)–nitroxide distance measurements on bacteriophage T4-lysozyme, The Journal of Physical Chemistry B, 117 (2013) 3145-3153.

[41] A. Giannoulis, Y. Yang, Y.-J. Gong, X. Tan, A. Feintuch, R. Carmieli, T. Bahrenberg, Y. Liu, X.-C. Su, D. Goldfarb, DEER distance measurements on trityl/trityl and Gd (iii)/trityl labelled proteins, Physical Chemistry Chemical Physics, 21 (2019) 10217-10227.

[42] A. Dalaloyan, M. Qi, S. Ruthstein, S. Vega, A. Godt, A. Feintuch, D. Goldfarb, Gd (III)–Gd (III) EPR distance measurements–the range of accessible distances and the impact of zero field splitting, Physical Chemistry Chemical Physics, 17 (2015) 18464-18476.

[43] N. Manukovsky, A. Feintuch, I. Kuprov, D. Goldfarb, Time domain simulation of Gd3+–Gd3+ distance measurements by EPR, The Journal of chemical physics, 147 (2017) 044201.

[44] M.R. Cohen, V. Frydman, P. Milko, M.A. Iron, E.H. Abdelkader, M.D. Lee, J.D. Swarbrick, A. Raitsimring, G. Otting, B. Graham, Overcoming artificial broadening in Gd^{3+}–Gd^{3+} distance distributions arising from dipolar pseudo-secular terms in DEER experiments, Physical Chemistry Chemical Physics, 18 (2016) 12847-12859.

[45] H. El Mkami, R.I. Hunter, P.A. Cruickshank, M.J. Taylor, J.E. Lovett, A. Feintuch, M. Qi, A. Godt, G.M. Smith, High-sensitivity Gd 3+–Gd 3+ EPR distance measurements that eliminate artefacts seen at short distances, Magnetic Resonance, 1 (2020) 301-313.

[46] S. Razzaghi, M. Qi, A.I. Nalepa, A. Godt, G. Jeschke, A. Savitsky, M. Yulikov, RIDME spectroscopy with Gd (III) centers, The journal of physical chemistry letters, 5 (2014) 3970-3975.

[47] G. Jeschke, V. Chechik, P. Ionita, A. Godt, H. Zimmermann, J. Banham, C. Timmel, D. Hilger, H. Jung, DeerAnalysis2006 - a comprehensive software package for analyzing pulsed ELDOR data, Applied Magnetic Resonance, 30 (2006) 473-498.

[48] L. Fábregas Ibáñez, G. Jeschke, S. Stoll, DeerLab: a comprehensive software package for analyzing dipolar electron paramagnetic resonance spectroscopy data, Magnetic Resonance, 1 (2020) 209-224.

[49] R.A. Stein, A.H. Beth, E.J. Hustedt, A Straightforward approach to the analysis of double electron–electron resonance data, Methods in enzymology, 563 (2015) 531-567.

[50] O. Schiemann, C.A. Heubach, D. Abdullin, K. Ackermann, M. Azarkh, E.G. Bagryanskaya, M. Drescher, B. Endeward, J.H. Freed, L. Galazzo, D. Goldfarb, T. Hett, L. Esteban Hofer, L. Fábregas Ibáñez, E.J. Hustedt, S. Kucher, I. Kuprov, J.E. Lovett, A. Meyer, S. Ruthstein, S. Saxena, S. Stoll, C.R. Timmel, M. Di Valentin, H.S. McHaourab, T.F. Prisner, B.E. Bode, E. Bordignon, M. Bennati, G. Jeschke, Benchmark Test and Guidelines for DEER/PELDOR Experiments on Nitroxide-Labeled Biomolecules, Journal of the American Chemical Society, 143 (2021) 17875-17890.

[51] G. Jeschke, Dipolar Spectroscopy - Double-Resonance Methods, EMagRes, 5 (2016) 1459-1476.

[52] K. Keller, M. Qi, C. Gmeiner, I. Ritsch, A. Godt, G. Jeschke, A. Savitsky, M. Yulikov, Intermolecular background decay in RIDME experiments, Physical Chemistry Chemical Physics, 21 (2019) 8228-8245.

[53] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: International conference on machine learning, PMLR, 2015, pp. 448-456.

[54] H. Hogben, M. Krzystyniak, G. Charnock, P. Hore, I. Kuprov, Spinach – a software library for simulation of spin dynamics in large spin systems, Journal of Magnetic Resonance, 208 (2011) 179-194.

[55] V. Nair, G.E. Hinton, Rectified linear units improve restricted boltzmann machines, in: Icml, 2010.

[56] D.P. Kingma, J. Ba, ADAM: a method for stochastic optimization, arXiv preprint arXiv:1412.6980, (2014).

[57] B.E. Bode, D. Margraf, J. Plackmeyer, G. Dürner, T.F. Prisner, O. Schiemann, Counting the Monomers in Nanometer-Sized Oligomers by Pulsed Electron–Electron Double Resonance, Journal of the American Chemical Society, 129 (2007) 6736-6745.

[58] J.L. Wort, K. Ackermann, A. Giannoulis, A.J. Stewart, D.G. Norman, B.E. Bode, Sub-Micromolar Pulse Dipolar EPR Spectroscopy Reveals Increasing CuII-labelling of Double-Histidine Motifs with Lower Temperature, Angewandte Chemie, 131 (2019) 11807-11811.

[59] A. Meyer, J.J. Jassoy, S. Spicher, A. Berndhäuser, O. Schiemann, Performance of PELDOR, RIDME, SIFTER, and DQC in measuring distances in trityl based bi- and triradicals: exchange coupling, pseudosecular coupling and multi-spin effects, Physical Chemistry Chemical Physics, 20 (2018) 13858-13869.

[60] N. Fehr, C. Dietz, Y. Polyhach, T. von Hagens, G. Jeschke, H. Paulsen, Modeling of the N-terminal section and the lumenal loop of trimeric light harvesting complex II (LHCII) by using EPR, Journal of Biological Chemistry, 290 (2015) 26007-26020.

[61] P. Ionita, A. Volkov, G. Jeschke, V. Chechik, Lateral diffusion of thiol ligands on the surface of Au nanoparticles: An electron paramagnetic resonance study, Analytical chemistry, 80 (2008) 95-106.

[62] S.R. Sweger, S. Pribitzer, S. Stoll, Bayesian probabilistic analysis of DEER spectroscopy data using parametric distance distribution models, The Journal of Physical Chemistry A, 124 (2020) 6193-6202.

[63] T.H. Edwards, S. Stoll, A Bayesian approach to quantifying uncertainty from experimental noise in DEER spectroscopy, Journal of magnetic resonance, 270 (2016) 87-97.

[64] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, R. Fergus, Intriguing properties of neural networks, arXiv preprint arXiv:1312.6199, (2013).

[65] B. Fornberg, Generation of finite difference formulas on arbitrarily spaced grids, Mathematics of computation, 51 (1988) 699-706.

[66] H.M. Bücker, G. Corliss, P. Hovland, U. Naumann, B. Norris, Automatic differentiation: applications, theory, and implementations, Springer Science & Business Media, 2006.

[67] H. Russell, R. Stewart, C. Prior, V.S. Oganesyan, T.G. Gaule, J.E. Lovett, DEER and RIDME Measurements of the Nitroxide-Spin Labelled Copper-Bound Amine Oxidase Homodimer from Arthrobacter Globiformis, Applied Magnetic Resonance, 52 (2021) 995-1015.

[68] S.G. Hyberts, K. Takeuchi, G. Wagner, Poisson-gap sampling and forward maximum entropy reconstruction for enhancing the resolution and sensitivity of protein NMR data, Journal of the American Chemical Society, 132 (2010) 2145-2147.

[69] K. Kazimierczuk, V. Orekhov, Non-uniform sampling: post-Fourier era of NMR data collection and processing, Magnetic Resonance in Chemistry, 53 (2015) 921-926.

[70] M. Mobli, M.W. Maciejewski, A.D. Schuyler, A.S. Stern, J.C. Hoch, Sparse sampling methods in multidimensional NMR, Physical Chemistry Chemical Physics, 14 (2012) 10835-10843.

[71] D.F. Hansen, Using deep neural networks to reconstruct non-uniformly sampled NMR spectra, Journal of biomolecular NMR, 73 (2019) 577-585.

[72] E.T. Whittaker, On the functions which are represented by the expansions of the interpolation-theory, Proceedings of the Royal Society of Edinburgh, 35 (1915) 181-194.

[73] M. Pannier, S. Veit, A. Godt, G. Jeschke, H.W. Spiess, Dead-time free measurement of dipole–dipole interactions between electron spins, Journal of magnetic resonance, 213 (2011) 316-325.

[74] O. Schiemann, C.A. Heubach, D. Abdullin, K. Ackermann, M. Azarkh, E.G. Bagryanskaya, M. Drescher, B. Endeward, J.H. Freed, L. Galazzo, Benchmark test and guidelines for DEER/PELDOR experiments on nitroxide-labeled biomolecules, Journal of the American Chemical Society, 143 (2021) 17875-17890.

[75] F.D. Breitgoff, J. Soetbeer, A. Doll, G. Jeschke, Y.O. Polyhach, Artefact suppression in 5-pulse double electron electron resonance for distance distribution measurements, Physical Chemistry Chemical Physics, 19 (2017) 15766-15779.

[76] V.V. Kurshev, A.M. Raitsimring, Y.D. Tsvetkov, Selection of dipolar interaction by the "2+1" pulse train ESE, Journal of Magnetic Resonance, 81 (1989) 441-454.

[77] M. Teucher, E. Bordignon, Improved signal fidelity in 4-pulse DEER with Gaussian pulses, Journal of Magnetic Resonance, 296 (2018) 103-111.

[78] A. Meyer, D. Abdullin, G. Schnakenburg, O. Schiemann, Single and double nitroxide labeled bis(terpyridine)-copper(ii): influence of orientation selectivity and multispin effects on PELDOR and RIDME, Physical Chemistry Chemical Physics, 18 (2016) 9262-9271.

[79] N.N. Barak, P. Neumann, M. Sevvana, M. Schutkowski, K. Naumann, M. Malešević, H. Reichardt, G. Fischer, M.T. Stubbs, D.M. Ferrari, Crystal structure and functional analysis of the protein disulfide isomerase-related protein ERp29, Journal of molecular biology, 385 (2009) 1630-1642.

[80] I. Kaminker, I. Tkach, N. Manukovsky, T. Huber, H. Yagi, G. Otting, M. Bennati, D. Goldfarb, W-band orientation selective DEER measurements on a Gd3+/nitroxide mixed-labeled protein dimer with a dual mode cavity, Journal of Magnetic Resonance, 227 (2013) 66-71.