Statistically Modeling Optical Linewidths of Nitrogen Vacancy Centers in Post-Implanted Nanostructures

M. Kasperczyk, J. A. Zuber, A. Barfuss, J. Kölbl, V. Yurgens,
S. Flågan, T. Jakubczyk, B. Shields, R. J. Warburton, P. Maletinsky*

Department of Physics, University of Basel, Klingelbergstrasse 82, 4056 Basel, Switzerland
(Dated: May 11, 2020)

We investigate the effects of a novel approach to diamond nanofabrication and nitrogen vacancy (NV) center formation on the optical linewidth of the NV zero-phonon line (ZPL). In this post-implantation method, nitrogen is implanted *after* all fabrication processes have been completed. We examine three post-implanted samples, one implanted with ^{14}N and two with ^{15}N isotopes. We perform photoluminescence excitation (PLE) spectroscopy to assess optical linewidths and optically detected magnetic resonance (ODMR) measurements to isotopically classify the NV centers. From this, we find that NV centers formed from nitrogen naturally occuring in the diamond lattice are characterized by a linewidth distribution peaked at an optical linewidth nearly two orders of magnitude smaller than the distribution characterizing most of the NV centers formed from implanted nitrogen. Surprisingly, we also observe a number of ^{15}NV centers with narrow (< 500 MHz) linewidths, implying that implanted nitrogen can yield NV centers with narrow optical linewidths. We further use a Bayesian approach to statistically model the linewidth distributions, to accurately quantify the uncertainty of fit parameters in our model, and to predict future linewidths within a particular sample. Our model is designed to aid comparisons between samples and research groups, in order to determine the best methods of achieving narrow NV linewidths in structured samples.

I. INTRODUCTION

Excellent spectral properties and low spectral noise are a necessity for most quantum communications and entanglement protocols. Whether the goal is to entangle atoms in different cities [1], to relay quantum information across vast distances through a communications channel [2], to couple a qubit to a photonic cavity [3, 4], or to study the interference between two qubits [5], some of the biggest successes of quantum technology rely on quantum sources that are spectrally stable [6]. The nitrogen vacancy (NV) center in diamond has been particularly successful in a variety of quantum information experiments, as the NV spin can be coupled to its optical degree of freedom [3, 5–8]. Yet a key challenge remains in creating NV centers with good spectral properties in nano-structured samples, even though many groups have studied diverse methods of creating NVs. These methods include implantation and annealing [9–11], laser writing of NV centers [12, 13], and high-energy electron irradiation [14], with many studies focusing explicitly on the linewidth properties of the NV centers [15–19]. The transform limited optical linewidth of the NV center is ≈ 13 MHz, which sets the ultimate limit to how narrow the lines can be [11]. Depending on the application, broader optical linewidths can be tolerated: a 100 MHz linewidth is acceptable for a decent microcavity [4], and two-photon interference has been shown using an NV center with an inhomogeneous linewidth as broad as 480 MHz [5].

Here we study the distribution of optical linewidths of NV centers formed with implanted and native nitrogen in diamond nanostructures. We implant one of our samples with ^{15}N, which has a natural abundance of only 0.37%

[20], so that we can distinguish between implanted and native nitrogen by measuring the nitrogen isotope of the NV center. In line with the results of S. B. van Dam *et al.*, we find that implanted nitrogen yields NV centers with generally broader linewidths than native nitrogen does [21]. We also find evidence that implanted nitrogen can yield NV centers with narrow linewidths. Additionally, we demonstrate the novel approach of post-implantation, in which all nano-structuring and fabrication procedures are completed *before* implanting the sample with nitrogen. We do this to reduce the effects of fabrication on the NV center properties, as it is unclear to what degree the fabrication procedures themselves influence the optical linewidth [4, 14]. In studying post-implanted samples, we find a significant proportion of narrow linewidth NV centers, even in structured areas as thin as 1.57 µm. Finally, as determining what influences the NV center coherence properties remains an open question that is actively being explored, we develop a rigourous statistical model to help unify approaches within the community and to more easily compare results across research groups. We discuss our model in depth and show how we can use it to compare different data sets. To aid other researchers in implementing our model, we include a demo Matlab script, available as Supplementary Online Material.

II. OVERVIEW OF SAMPLES

A. Fabrication Processes

In the experimental part of this work, we first study two samples in detail (the third sample is discussed in Sec. V D). Both are made from electronic grade (N < 5 ppb, B < 1 ppb) diamond acquired from Element Six. Our fabrication procedure is summarized by P. Appel *et*

* patrick.maletinsky@unibas.ch

al. [22]. In both samples, we fabricated a membrane of a nonuniform thickness spanning 2.5–5 μm, as well as cantilevers with variable dimension: lengths from 35–70 μm, widths of approximately 4.5 μm, and thickness of roughly 2.5–4 μm. An optical microscope image of Sample B is shown in Fig. 1(a), showing the cantilevers, membrane, and bulk parts of the sample.

B. Implantation Parameters

After all fabrication of the membrane and cantilevers was finished, Sample A was sent to the Helmholtz-Zentrum Dresden-Rossendorf to be implanted with nitrogen, and Sample B was sent to CuttingEdge Ions. Both samples were implanted with 12 keV nitrogen ions at an angle of 7° relative to the sample mount and at a fluence of 10^{11} ions/cm^2. Whereas Sample A was implanted with ^{14}N, Sample B was implanted with ^{15}N, so that the NVs could be isotopically classified. After the samples were implanted, they were annealed with a procedure outlined in P. Appel *et al.* [22], consisting of 4 hours at 400 C°, 10 hours at 800 C°, and 2 hours at 1200 C°. Finally, the samples were cleaned with a tri-acid clean [22, 23].

III. MEASUREMENT METHODOLOGY

We begin by taking a confocal fluorescence map in the target area of the sample (see Fig. 1(b)). We then characterize each potential NV center by taking a photoluminescence spectrum under green (532 nm) laser excitation. Once the zero-phonon line (ZPL) has been identified on the spectrometer, we perform a photoluminescence excitation (PLE) measurement on the NV center by sweeping the wavelength of a red (637 nm) laser across the transition while recording the fluorescence counts on an avalanche photodiode (APD), yielding a measurement of the excited state transition linewidth. We did not use the same intensity of red laser power for each NV center, as the broader linewidths were often too weak to measure at low laser power. Red laser powers for NV centers with narrow linewidths ranged from 10−200 nW, whereas broad linewidths were typically measured with 2 μW of excitation power. Optical linewidths are extracted from the FWHM of a Gaussian fit to the PLE data. Because we include a repump pulse in every iteration of the pulse sequence, our linewidths are broadened by spectral diffusion, making a Gaussian fit suitable [10, 21]. The pulse sequence we use is shown above Fig. 1(c), and representative measurements for ^{14}NV and ^{15}NV centers are shown in Figs. 1(c)&(e). In the case in which both E_x and E_y lines were visible, we used only the narrower linewidth in the dataset, as the goal of this analysis is to analyze the narrowest linewidth measurable on each individual NV center.

After recording the optical linewidth, we use pulsed optical detection of magnetic resonance (ODMR) to measure the hyperfine structure of the NV center ground
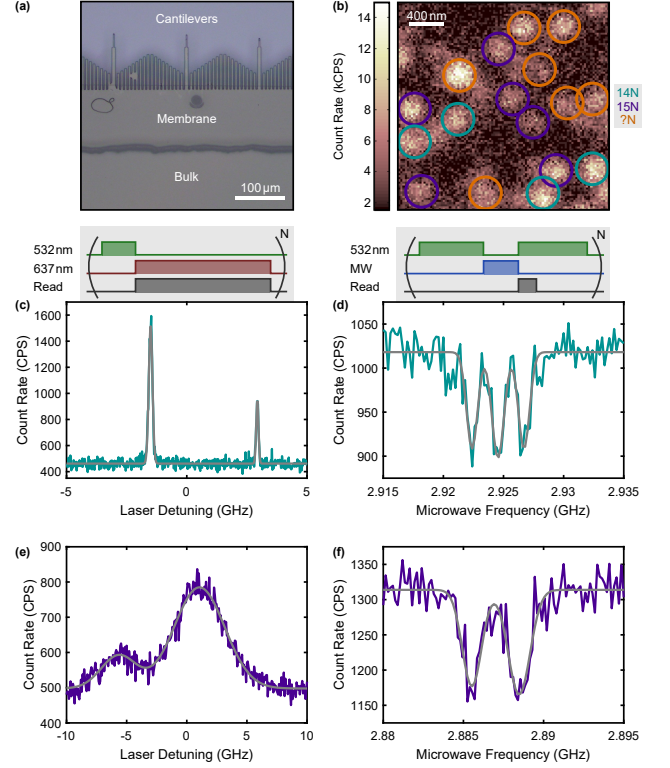

Figure 1. (a) Optical microscope image of Sample B, showing the bulk area (bottom), membrane (center), and cantilevers (top). (b) Fluorescence map in the bulk area of Sample B (50 μm thickness). Each NV center we measured is indicated by a circle around it, with the color of the circle indicating the isotope. Some NV centers indicated did not exhibit measurable PLE. (c) Representative PLE measurement on an ^{14}NV, with the measurement pulse sequence shown above. The two lines have linewidths of 107 MHz and 186 MHz and are split by 4.4 GHz. (d) Representative pulsed ODMR measurement on an ^{14}NV center, with the pulse sequence shown above. Pulsed ODMR reveals the three peaks split by 2.2 MHz characteristic of ^{14}N. (e) Representative PLE measurement on an ^{15}NV center. The two lines have linewidths of 3.4 GHz and 5.1 GHz and are split by 6.8 GHz. (f) Representative pulsed ODMR measurement on an ^{15}NV center. Pulsed ODMR reveals the two peaks split by 3.1 MHz characteristic of ^{15}N. In (c)–(f), the gray lines indicate fits to the data.

state, thereby identifying the isotope of the NV nitrogen. The pulse sequence is shown above Fig. 1(d), and typical hyperfine-resolved ODMR measurements are shown in Figs. 1(d)&(f) for ^{14}N and ^{15}N, respectively. The locations and widths of the ODMR dips are extracted from Gaussian fits to the data. We attempted to measure the ZPL wavelengths, optical linewidths, and hyperfine-resolved ODMR of a total of 159 NV centers in Sample A and 104 NV centers in Sample B. We note, however, that some NV centers did not exhibit any PLE, and others failed to show hyperfine-resolved ODMR. In total, we successfully measured PLE on 78 NV centers in Sample A and 61 NV centers in Sample B. Similarly, we were able to isotopically classify 47 NV centers on Sample B (iso-
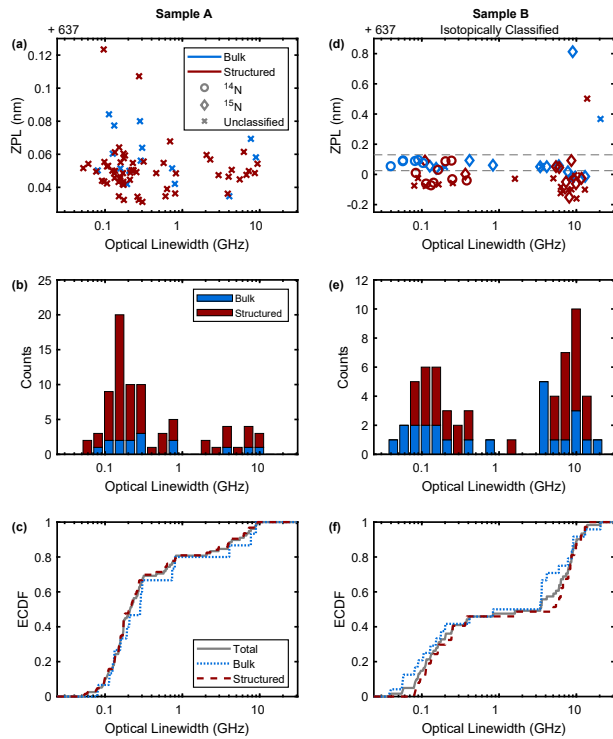
Figure 2. The data for Sample A are shown in the left column, and for Sample B in the right column. (a) A scatter plot showing the ZPL wavelength (in air) for each linewidth measured in Sample A. The marker color indicates which area of the sample the datapoint was taken on. (b) Stacked histograms of optical linewidths in Sample A with the data labelled by sample location. There is evidence of two distinct populations of NV centers, and narrow NV centers can occur in the structured parts of the sample as well as the unstructured. (c) ECDFs of the linewidths in Sample A. The plot shows that the median measured linewidth was $\approx$ 200 MHz. (d) A similar scatter plot as in (a) for Sample B. The marker color again indicates where the datapoint was taken, and the marker shape indicates which isotope the hyperfine structure indicated. The horizontal dashed lines demarcate the limits of the ZPL axis of (a), showing that Sample B showed a much larger variation in ZPL wavelength, suggesting more variability in the local strain environment. (e) Stacked histogram of optical linewidths in Sample B, with data labelled by sample location. Two populations are again evident, and they are not related to location in the sample. (f) ECDF for Sample B. The median linewidth is $\approx$ 3.5 GHz. The plateau indicates a clear separation between the two populations.

topic classifcation was not performed on Sample A, as it was implanted with ^{14}N).

IV. RESULTS AND DISCUSSION

A. Influence of NV Center Location on Linewidth

We summarize the data for Sample A in Fig. 2(a), which plots the measured optical linewidths against the measured ZPL wavelength. The data points are color-coded to indicate which part of the sample they were taken on. The bulk part of the sample is approximately 50 μm thick, whereas the membrane and cantilever dimensions are discussed earlier. According to Wilcoxon ranked sum tests [24], the linewidths of NV centers found in the membrane likely follow the same statistics as those in the cantilevers. We therefore combine the cantilever and membrane measurements into a single category: structured. The ZPL wavelengths for Sample A are tightly clustered (spanning a spectral range of only 0.2 nm), and the sample exhibits no clear relationship between ZPL wavelength and optical linewidth. Binning the linewidths and color-coding them according to the sample location (see Fig. 2(b)) reveals that there are two distinct populations of NV centers: those with narrow linewidths, and those with broad linewidths. Both types of linewidths can be found anywhere on the sample. Fig. 2(c) shows the empirical cumulative distribution functions (ECDFs) for the structured, bulk, and total datasets for Sample A, showing that there is no apparent difference between the three distributions. A Wilcoxon ranked sum test (p-value 0.551) reveals that there is no statistically significant evidence that the structured and bulk linewidth distributions are different. Fig. 2(c) also shows that the median measured linewidth was approximately 200 MHz.

In Figs. 2(d)–(f), we show similar plots for Sample B. In Fig. 2(d) we see that although the ZPL wavelengths are far more scattered in Sample B (spanning a range of 2 nm) than in Sample A, there is still no clear relationship between ZPL wavelength and optical linewidth, indicating that local strain does not play a strong role in determining the linewidth. Plotting the data in a histogram labelled by location of the sample in Fig. 2(e) shows a similar situation as in Fig. 2(b): there are two distinct populations of NV centers, independent of the location on the sample. Likewise, Fig. 2(f) leads to similar conclusions as Fig. 2(c). Again, a Wilcoxon rank sum test (p-value 0.334) indicates that there is no clear evidence for a difference between the linewidth distributions in the bulk and structured areas in Sample B. As we see similar results in two different samples, and across different regions on those samples, we turn to isotopic classification to better understand these two populations.

B. Influence of NV Center Isotope on Linewidth

In Fig. 3(a) we bin the data and color-code the bins by isotope classification. We find that although many NV centers could not be clearly classified, a clear pattern emerges: NV centers formed with native ^{14}N exhibit narrow ($<$ 1 GHz) linewidths, whereas most of the ^{15}NV centers showed broad ($>$ 1 GHz) linewidths, in agreement with the results of S. B. van Dam [21]. Indeed, the median ^{14}NV linewidth in Sample B was roughly 100 MHz. We fit log-normal sampling distributions to the ^{14}NV and ^{15}NV data (dashed lines). In Fig. 3(b) we
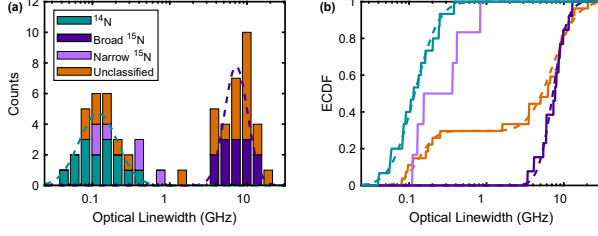
Figure 3. (a) Histogram of optical linewidths for Sample B, with data points labelled by color to indicate isotope. The histogram clearly shows that, with a few exceptions, the ^{14}NV centers and the ^{15}NV centers are completely separated into two distinct populations. The dashed lines indicate fits of log-normal sampling distributions $P(\{x_i\}|\mu, \sigma^2)$ to the ^{14}NV and (broad) ^{15}NV datasets. (b) Isotopically classified ECDFs. The teal and purple dashed lines are the log-normal fits to the ^{14}NV and ^{15}NV datasets, respectively, showing excellent agreement between the ECDFs and the fits. The dashed orange line is obtained by fitting a sum of two log-normal distributions to the unclassified NV centers. The narrow ^{15}NV data are included without a fit in this plot.

plot the ECDFs for the isotopically classified datasets, as well as the cumulative distribution functions (CDFs) for the log-normal fits. The CDFs show exceptional agreement with the ECDFs (see Sec. V C for model diagnostics). The fit curve for the unclassified data set comprises a weighted sum of two log-normal distributions in which the weights are also fit parameters. This fit suggests that all the unclassified data can be attributed to one of the two distributions.

Finally, we note that six of the ^{15}NV center linewidths were well separated from those of the other ^{15}NV centers. Based on an analysis of Q-Q plots (see the discussion of Q-Q plots in Sec. V C), we exclude these NV centers from the fits in Fig. 3(a), as they clearly do not belong to the same population; in Fig. 3(b) we include the ECDF of these data points but do not fit them. Due to the low natural abundance of ^{15}N, it is highly unlikely that the narrow ^{15}NV centers are due to naturally occurring ^{15}N. To wit: in a sample size of 61 PLE lines, there is a mere $1.2 \times 10^{-5}\%$ chance of observing 6 or more naturally occurring ^{15}NV centers, i.e. $P(m \geq 6|n = 61, p = 0.0037) \approx 1.2 \times 10^{-7}$, calculated from the CDF of the binomial distribution with 61 trials and a success rate of 0.37%. Previous studies have reported that implanted nitrogen can lead to crystal damage that degrades the optical properties of NV centers, and that this damage can be at least partially repaired through annealing [21], but it is unclear whether the annealing is the reason we were able to observe narrow linewidths from NV centers formed by implanted nitrogen.

V. STATISTICAL MODEL

A. Building the Model

We now develop a model to describe the two distinct populations we see, as it could be useful to determine how different the populations are. A model could help to decide how we should classify future or unclassified data points, and to predict how narrow future linewidths in the same sample will be. Additionally, having a model will allow us to more quantitatively determine which fabrication procedures yield NV centers with better optical linewidths and quantify how certain we are a new procedure is better. Using a Bayesian approach, we model the likelihood of a particular linewidth x_i with a log-normal likelihood:

$$P(x_i|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma}\frac{1}{x_i}e^{-(\ln(x_i) - \mu)^2/2\sigma^2}, \qquad (1)$$

which is parameterized by a median μ and a standard deviation σ. This is an appropriate distribution for any purely positive quantity that has contributions from multiple independent noise sources [25] (here, e.g., electric field noise, temperature, and strain fluctuations can all influence the optical linewidth [15, 16]).

Using uninformative priors for the parameters, (uniform distribution for μ and the Jeffreys prior for σ [25]), we find their posterior distributions [24, 25]. See Appendix A for details. Broadly speaking, the posterior distributions describe our best guess for the parameters, as well as our confidence in those guesses, given the data we have and the model we use. For ease of notation, we define two constants that depend on the data:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N}(\ln(x_i)),$$

$$\overline{X^2} = \frac{1}{N}\sum_{i=1}^{N}(\ln(x_i))^2,$$

where x_i is the ith linewidth in the dataset (or data subset, if focusing on a particular isotope, for example) and N is the total number of linewidths in the dataset (or subset).

For μ, we find that the posterior distribution $P(\mu|\{x_i\})$ (where $\{x_i\}$ is the dataset of linewidths being analyzed) is given by a location-scale t-distribution:

$$P(\mu|\{x_i\}) =$$
$$\frac{\Gamma\left(\frac{\nu_\mu+1}{2}\right)}{\Gamma\left(\frac{\nu_\mu}{2}\right)\sqrt{\pi\nu_\mu\sigma_\mu^2}}\left(1 + \frac{1}{\nu_\mu}\left(\frac{\mu - \mu_\mu}{\sigma_\mu}\right)^2\right)^{-\frac{\nu_\mu+1}{2}}, \quad (2)$$

where $\mu_\mu = \overline{X}$, $\nu_\mu = N-1$, and $\sigma_\mu = \sqrt{\frac{1}{N-1}\left(\overline{X^2} - \overline{X}^2\right)}$.

For the variance σ^2, the posterior distribution $P(\sigma^2|\{x_i\})$ is an inverse gamma distribution:

$$P(\sigma^2|\{x_i\}) = \frac{\beta_\sigma{}^{\alpha_\sigma}}{\Gamma(\alpha_\sigma)}\left(\frac{1}{\sigma^2}\right)^{(\alpha_\sigma+1)} e^{-\beta_\sigma/\sigma^2}, \qquad (3)$$

where $\alpha_\sigma = \frac{N-1}{2}$ and $\beta_\sigma = \frac{N}{2}\left(\overline{X^2} - \overline{X}^2\right)$.

We next consider what distribution of future linewidths $\tilde{x}$ we expect to measure, given the data we have observed so far. Working in terms of the natural logarithm of the linewidth $\tilde{X} \equiv \ln(\tilde{x})$, we also calculate the posterior predictive distribution $P(\tilde{X}|\{x_i\})$, which describes how likely the next linewidth is to be narrow. We find that $P(\tilde{X}|\{x_i\})$ is a location-scale t-distribution:

$$P(\tilde{X}|\{x_i\}) =$$

$$\frac{\Gamma\left(\frac{\tilde{\nu}+1}{2}\right)}{\Gamma\left(\frac{\tilde{\nu}}{2}\right)\sqrt{\pi\tilde{\nu}\tilde{\sigma}^2}}\left(1+\frac{1}{\tilde{\nu}}\left(\frac{\tilde{X}-\tilde{\mu}}{\tilde{\sigma}}\right)^2\right)^{-\frac{\tilde{\nu}+1}{2}}, \qquad (4)$$

where $\tilde{\mu} = \overline{X}$, $\tilde{\nu} = N - 1$, and $\tilde{\sigma} = \sqrt{\frac{N+1}{N-1}\left(\overline{X^2} - \overline{X}^2\right)}$. Note that this is a location-scale t-distribution for the natural logarithm of the linewidth $\tilde{X}$, not for the linewidth $\tilde{x}$ itself. The posterior predictive distribution for the linewidth $\tilde{x}$ is given by

$$P(\tilde{x}|\{x_i\}) =$$

$$\frac{\Gamma\left(\frac{\tilde{\nu}+1}{2}\right)}{\Gamma\left(\frac{\tilde{\nu}}{2}\right)\sqrt{\pi\tilde{\nu}\tilde{\sigma}^2}}\frac{1}{\tilde{x}}\left(1+\frac{1}{\tilde{\nu}}\left(\frac{\ln(\tilde{x})-\tilde{\mu}}{\tilde{\sigma}}\right)^2\right)^{-\frac{\tilde{\nu}+1}{2}}, \qquad (5)$$

which is not quite a t-distribution. For a more detailed discussion of the derivations of these distributions and their interrelations, see Appendix A.

B. Inferences from the Model

Because these distributions are of a common form, it is straightforward to find their most likely values and their credible intervals. For example, the maximum a posteriori (MAP) estimate (i.e. the most likely value) for the t-distribution $P(\mu|\{x_i\})$ is given by

$$\mu^{\mathrm{MAP}} = \mu_\mu = \overline{X},$$

and the 95% credible interval is given by

$$\left[\mu_\mu - \sigma_\mu t_{(0.975,\nu_\mu)}, \mu_\mu + \sigma_\mu t_{(0.975,\nu_\mu)}\right],$$

where $t_{(f,\nu)}$ is the t-statistic at the fth percentile and with ν degrees of freedom [24, 25]. Note, however, that the parameter μ in the log-normal distribution has units
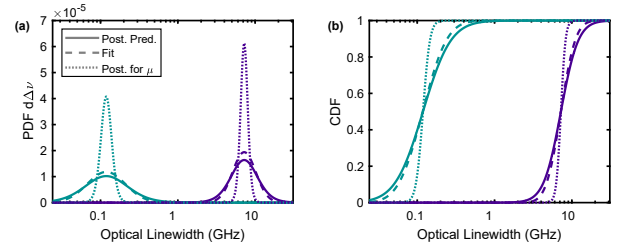

Figure 4. (a) The dashed lines reproduce the log-normal sampling distributions from Fig. 3(a). The solid lines indicate the posterior predictive distributions $P(\tilde{x}|\{x_i\})$. They are slightly broader than the sampling distributions $P(x|\mu,\sigma)$ because they account for the uncertainty in our estimates of μ and σ. The dotted lines are the posterior distributions for the median $P(\mu|\{x_i\})$, showing that the median ^{14}NV and ^{15}NV linewidths are well separated. (b) CDFs of the corresponding distributions in (a).

of ln(MHz) (if the dataset is in MHz); the MAP estimate and credible interval (CI) in terms of MHz are then given by

$$e^{\mu_\mu} \text{ and}$$

$$\left[e^{\mu_\mu-\sigma_\mu t_{(0.975,\nu_\mu)}}, e^{\mu_\mu+\sigma_\mu t_{(0.975,\nu_\mu)}}\right],$$

respectively. Similarly, the MAP estimate and CI for $\tilde{x}$ are given by

$$e^{\tilde{\mu}} \text{ and}$$

$$\left[e^{\tilde{\mu}-\tilde{\sigma}t_{(0.975,\tilde{\nu})}}, e^{\tilde{\mu}+\tilde{\sigma}t_{(0.975,\tilde{\nu})}}\right],$$

respectively. The MAP estimate of σ^2 is given by $\beta_\sigma/(\alpha_\sigma+1)$ [24]. Unfortunately, there is no closed-form solution for the 95% CI of the inverse gamma distribution, but it can be easily estimated through simulated draws, which we describe below [24].

We graphically represent our results in Fig. 4. The dashed lines in Fig. 4(a) are the log-normal fits from Fig. 3(a). As in Fig. 3, the color of the line indicates the isotope. The solid lines are the posterior predictive distributions $P(\tilde{x}|\{x_i\})$, and the dotted lines are the posterior distributions $P(\mu|\{x_i\})$. The posterior predictive distributions $P(\tilde{x}|\{x_i\})$ resemble the sampling distributions $P(\{x_i\}|\mu,\sigma)$ but are slightly broader, as they account for the uncertainty in our estimates of μ and σ. The posterior for μ_{N14} given by $P(\mu_{\mathrm{N14}}|\{x_i\}_{N14})$ is fairly narrow, indicating that only a narrow range of values of μ_{N14} is consistent with the ^{14}NV data. Similar conclusions hold for the ^{15}NV data.

Finally, we simulate draws from the distributions, which allows us to compare the ^{14}NV and ^{15}NV results and give approximate answers to questions such as what is the probability that the next ^{14}NV linewidth is narrower than the next ^{15}NV linewidth $P(\tilde{x}_{\mathrm{N14}} < \tilde{x}_{\mathrm{N15}}|\{x_i\})$ or how likely is the next ^{14}NV linewidth to be below $100\,\mathrm{MHz}$
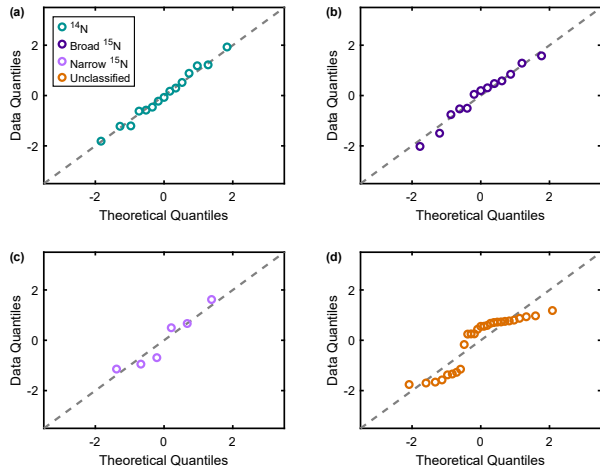
6

Figure 5. (a) Q-Q plot for the ^{14}NV data, showing that a log-normal model fits the data well. (b) Q-Q plot for the ^{15}NV data, again showing the model choice was appropriate. (c) The Q-Q plot for the narrow ^{15}NV data shows that a log-normal model might be appropriate. (d) The Q-Q plot for the unclassified data shows that a single log-normal model is a poor fit to the data, as expected.

$P\left(\tilde{x}_{N14} < 100\,\mathrm{MHz}|\{x_i\}\right)$. For example, using our data and 10^8 simulated draws from each of the posterior distributions, we find that $P\left(\mu_{N14} < \mu_{N15}|\{x_i\}\right) \approx 1$. Similarly, we estimate that we have a roughly 40% chance of finding sub-100 MHz ^{14}NV centers: $P\left(\tilde{x}_{N14} < 100\,\mathrm{MHz}|\{x_i\}\right) \approx 0.398$. For details of the simulated draws, see Appendix B.

C. Model Checking

To check how appropriate our model is for our data, we look at the quantile-quantile (or Q-Q) plots for the different data subsets. By comparing the data quantiles to the expected quantiles from the model, Q-Q plots show whether the spread in the data can be explained by the model and are therefore a useful diagnostic for determining whether a model is appropriate for the data. They can also be useful for identifying outliers in the dataset. The quantile for the ith optical linewidth in the dataset is calculated according to the formula

$$Q_{\mathrm{Data}}^{(i)} = \frac{\ln(x_i) - \overline{X}}{\sqrt{\overline{X^2} - \overline{X}^2}}$$

and therefore summarizes how many standard deviations the data point is from the mean of the dataset [26]. Using a log-normal model to calculate the theoretical quantiles, we plot the Q-Q plots for the ^{14}NV data, the broad ^{15}NV data, the narrow ^{15}NV data, and the unclassified data in Fig. 5. Both the ^{14}NV data in Fig. 5(a) and the broad ^{15}NV data in Fig. 5(b) closely follow the diagonal dashed

line, indicating the quantiles of the measured data match the quantiles we would expect from a log-normal distribution in both cases. Due to the dearth of data points, it is difficult to say how appropriate a log-normal model is for the narrow ^{15}NV data in Fig. 5(c), but our data do show that a log-normal model is promising. From Fig. 5(d), it is clear that a single log-normal model is inappropriate for the unclassified data, as expected.

D. Example with Sample C

As an application of our statistical model, we now examine a third structured sample, Sample C, which was post-implanted by InnovIon with 52 keV ^{15}N ions at an angle of $7°$ and a fluence of 5×10^9 ions/cm^2. In Sample C, we compare two structured parts of the sample: one area that is 1.57 μm thick, and one that is 0.87 μm thick. First, we note that we were able to observe two narrow ($< 250\,\mathrm{MHz}$) linewidths in the 1.57 μm-thick area (see Fig. 6(a)). To our knowledge, these are the narrowest NV ZPL lines reported in such thin structures obtained by standard etching techniques. A recent report, however, suggests that ultra-slow etching can significantly improve surface quality and lead to a further reduction of charge noise, which is at the origin of the inhomogenous broadening [27]. We note that the distributions of the data from the two sample areas strongly overlap (see Fig. 6(a)). In Fig. 6(b), we show log-normal fits to the data and the posterior distributions for μ. We find that the two data subsets have similar MAP estimates for the medians: $\mu_{0.87\,\mu m}^{\mathrm{MAP}} \approx \ln(2.27\,\mathrm{GHz})$ and $\mu_{0.87\,\mu m}^{\mathrm{MAP}} \approx \ln(1.47\,\mathrm{GHz})$. Although the data and fits overlap and the estimates for μ are similar for the two data subsets, the posterior distributions for the medians μ barely overlap. Using the data from Sample C and simulated draws from the posterior distributions for μ, we find that $P\left(\mu_{1.57\,\mu m} < \mu_{0.87\,\mu m}\right) \approx 0.996$, strongly suggesting that the two areas have different median linewidths. For the purposes of estimating the two medians, we exclude the two narrowest linewidths in the 1.57 μm area of the sample and the broadest linewidth in the 0.87 μm area of the sample, as Q-Q plots (not shown) reveal these data points to be outliers. Although our data suggest that the thinner part of the sample has a larger median linewidth, it is unclear whether this change is due the thickness itself or due to confounding variables. We hope our statistical model will aid in determining which variables influence the spectral properties of NV centers in other nano-structured samples.

VI. CONCLUSION

We have shown that NV centers in post-implanted samples exhibit narrow linewidths, even in structured samples as thin as 1.57 μm, and that the narrow lines are primarily due to NV centers being formed from nitrogen native to the diamond. Even so, we observe a few
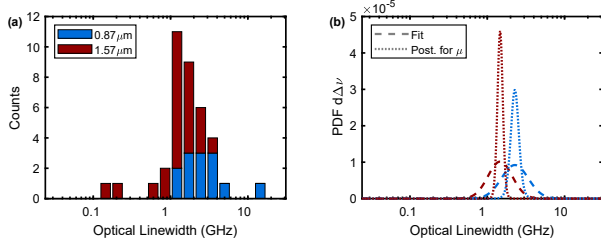
Figure 6. (a) Stacked histograms of data taken in Sample C. The data are color coded to indicate which area of the sample they were taken on. (b) Fits (dashed lines) and posterior distributions for μ (dotted lines) for the two sample areas. The plots are color coded as in (a). Although the fits strongly overlap, the posterior distributions for μ do not, indicating that the thinner part of the sample is characterized by a higher median linewidth.

narrow linewidths that can be attributed to implanted nitrogen. Furthermore, we develop a statistical model to aid in summarizing our results and to enable easy comparison of results between research groups. Indeed, we employ our model to show that in one of our samples, the sample thickness is linked to changes in the linewidth distribution. Our results suggest that post-implantation is a useful method for reducing or avoiding fabrication-induced damage. To further investigate the benefits of post-implantation, we propose testing different fabrication steps on a post-implanted structured sample, to study if and how various common fabrication techniques degrade the NV properties. If post-implantation does consistently improve NV coherence properties, it is worthwhile to study the effects of carbon implantation, as implanted nitrogen rarely leads to narrow linewidths. Finally, the model itself can be developed further, by implementing a hierarchical model (to allow, e.g., isotope abundance to vary across the sample, or to allow the parameters μ and σ to vary with sample location) and by including a model for data sampling and missing data.

Appendix A: Derivation of Posterior Probabilities

We find the joint posterior distribution for the parameters μ and σ by using Bayes law:

$$P\left(\mu, \sigma | \{x_i\}\right) = \frac{P\left(\{x_i\} | \mu, \sigma\right) P\left(\mu, \sigma\right)}{P\left(\{x_i\}\right)},$$

where $P\left(\mu, \sigma | \{x_i\}\right)$ is the joint posterior for μ and σ, $P\left(\{x_i\} | \mu, \sigma\right)$ is the sampling distribution or likelihood, $P\left(\mu, \sigma\right)$ is the joint prior distribution for μ and σ, and $P\left(\{x_i\}\right)$ acts as a normalizing constant. The most important component is the sampling distribution $P\left(\{x_i\} | \mu, \sigma\right)$, as this acts as our model for the data. As mentioned in the main text, we use a log-normal model for the dataset, such that

$$
\begin{aligned}
P\left(\{x_i\} | \mu, \sigma\right) &= \prod_{i=1}^{N} P\left(x_i | \mu, \sigma\right) \\
&= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{x_i} e^{-\left(\ln(x_i) - \mu\right)^2 / 2\sigma^2}.
\end{aligned}
$$

For the prior distribution $P\left(\mu, \sigma\right)$ we choose uninformative priors. Typical uninformative priors for μ and σ are the uniform and Jeffreys priors, respectively [25], but as they are improper distributions (i.e. not normalizable), it is common to start with proper (i.e. normalizable) distributions and take a limit at the end of the calculation to turn the proper priors into the desired uninformative priors [24]. As we can rewrite $P\left(\mu, \sigma\right) = P\left(\mu\right) P\left(\sigma\right)$ (assuming our prior states of knowledge for μ and σ are independent), we have to choose two priors. For $P\left(\mu\right)$, we choose a uniform prior:

$$P\left(\mu\right) = \frac{1}{2\mu_{UB}} I_{-\mu_{UB} \leq \mu \leq \mu_{UB}},$$

where I is the indicator function and causes $P\left(\mu\right)$ to be non-zero only within the bounds set by μ_{UB}. In the limit $\mu_{UB} \rightarrow \infty$, this goes to a uniform distribution that allows all real values of μ. We take this limit after finding $P\left(\mu, \sigma | \{x_i\}\right)$. For $P\left(\sigma\right)$ we choose the Jeffreys prior, which is a uniform distribution on a logarithmic scale and is commonly used for scale parameters such as the standard deviation:

$$P\left(\sigma\right) = \frac{1}{\ln(\sigma_{UB}) - \ln(\sigma_{LB})} \frac{1}{\sigma} I_{\sigma_{LB} \leq \sigma \leq \sigma_{UB}},$$

where σ_{LB} and σ_{UB} are the lower and upper bounds on σ. We next calculate $P\left(\mu, \sigma | \{x_i\}\right)$ by combining our expressions for $P\left(\{x_i\} | \mu\sigma\right)$, $P\left(\mu\right)$, and $P\left(\sigma\right)$, and we also use the fact that $P\left(\{x_i\}\right) = \int_{\sigma_{LB}}^{\sigma_{UB}} \int_{-\mu_{UB}}^{\mu_{UB}} P\left(\{x_i\} | \mu, \sigma\right) P\left(\mu\right) P\left(\sigma\right) \, \mathrm{d}\mu \, \mathrm{d}\sigma$. After taking the limits $\mu_{UB} \rightarrow \infty$, $\sigma_{LB} \rightarrow 0$, and $\sigma_{UB} \rightarrow \infty$, we find

$$P\left(\mu,\sigma|\{x_i\}\right) = \frac{2^{1-\frac{N}{2}} N^{\frac{N}{2}} \left(\overline{X^2} - \overline{X}^2\right)^{\frac{N-1}{2}}}{\sqrt{\pi}\Gamma\left(\frac{N-1}{2}\right)} \left(\frac{1}{\sigma}\right)^{N+1} e^{-N\left(\overline{X^2} - 2\overline{X}\mu + \mu^2\right)/2\sigma^2} I_{-\infty \leq \mu \leq \infty} I_{0 \leq \sigma \leq \infty} \,. \tag{A1}$$

From here on, we leave out the indicator functions $I_{\infty \leq \mu \leq \infty}$ and $I_{0 \leq \sigma \leq \infty}$ for ease of notation, but they are always implicitly there. Note that this derivation relies on the assumption that $N \geq 2$, i.e. the dataset or data subset has at least two data points in it.

Now we can calculate the marginal posteriors for μ and σ, which summarize how much our data determine those parameters. The marginal posterior for μ is defined as follows:

$$P\left(\mu|\{x_i\}\right) = \int_0^\infty P\left(\mu,\sigma|\{x_i\}\right) \, d\sigma \,.$$

Using our expression for $P\left(\mu,\sigma|\{x_i\}\right)$ in Eq. A1, we find

$$P(\mu|\{x_i\}) = \frac{\Gamma\left(\frac{\nu_\mu+1}{2}\right)}{\Gamma\left(\frac{\nu_\mu}{2}\right)\sqrt{\pi\nu_\mu\sigma_\mu{}^2}} \left(1 + \frac{1}{\nu_\mu}\left(\frac{\mu-\mu_\mu}{\sigma_\mu}\right)^2\right)^{-\frac{\nu_\mu+1}{2}},$$

which is Eq. 2 of the main text.

Similarly, $P\left(\sigma|\{x_i\}\right)$ is defined by

$$P\left(\sigma|\{x_i\}\right) = \int_{-\infty}^\infty P\left(\mu,\sigma|\{x_i\}\right) \, d\mu \,,$$

which in our case yields

$$P\left(\sigma|\{x_i\}\right) = \frac{2\beta_\sigma^{\alpha_\sigma}}{\Gamma\left(\alpha_\sigma\right)} \left(\frac{1}{\sigma}\right)^{2\alpha_\sigma+1} e^{-\beta_\sigma/\sigma^2} \,,$$

where α_σ and β_σ are defined in the main text. Note that this is a distribution for the standard deviation σ, not the variance σ^2. To find the distribution for σ^2, we perform a change of variables and find

$$P(\sigma^2|\{x_i\}) = \frac{\beta_\sigma{}^{\alpha_\sigma}}{\Gamma\left(\alpha_\sigma\right)} \left(\frac{1}{\sigma^2}\right)^{(\alpha_\sigma+1)} e^{-\beta_\sigma/\sigma^2} \,,$$

which is the Inverse Gamma distribution of Eq. 3 in the main text.

Finally, we derive the posterior predictive distribution, which summarizes what the next data point could be, based on the data taken so far. The posterior predictive distribution is defined as follows:

$$P\left(\tilde{x}|\{x_i\}\right) = \int_{-\infty}^\infty \int_0^\infty P\left(\tilde{x}|\mu,\sigma\right) P\left(\mu,\sigma|\{x_i\}\right) \, d\sigma d\mu \,,$$

which leads to Eq. 5 of the main text:

$$P(\tilde{x}|\{x_i\}) = $$
$$\frac{\Gamma\left(\frac{\tilde{\nu}+1}{2}\right)}{\Gamma\left(\frac{\tilde{\nu}}{2}\right)\sqrt{\pi\tilde{\nu}\tilde{\sigma}^2}} \frac{1}{\tilde{x}} \left(1 + \frac{1}{\tilde{\nu}}\left(\frac{\ln(\tilde{x})-\tilde{\mu}}{\tilde{\sigma}}\right)^2\right)^{-\frac{\tilde{\nu}+1}{2}},$$

where $\tilde{\nu}$, $\tilde{\mu}$, and $\tilde{\sigma}$ are defined above. Making a change of variables using $\tilde{X} = \ln\left(\tilde{x}\right)$ leads to Eq. 4 of the main text:

$$P(\tilde{X}|\{x_i\}) = $$
$$\frac{\Gamma\left(\frac{\tilde{\nu}+1}{2}\right)}{\Gamma\left(\frac{\tilde{\nu}}{2}\right)\sqrt{\pi\tilde{\nu}\tilde{\sigma}^2}} \left(1 + \frac{1}{\tilde{\nu}}\left(\frac{\tilde{X}-\tilde{\mu}}{\tilde{\sigma}}\right)^2\right)^{-\frac{\tilde{\nu}+1}{2}},$$

which is a location-scale t-distribution.

Appendix B: Simulated Draws

Simulating draws is a common technique in Bayesian statistics to estimate credible intervals and answer probabilistic questions [24]. To simulate draws from the posterior and posterior predictive distributions, we use Matlab's makedist function in the Statistics and Machine Learning Toolbox to define location-scale t-distributions and inverse gamma distributions with parameters determined by the data, as described in Section V.1. Matlab's random function then allows us to sample from the distributions we defined based on our data. Once samples have been drawn, it is simple to estimate the probabilities we describe above. For example, if we label our samples from $P\left(\mu_{\text{N}14}|\{x_i\}\right)$ and $P\left(\mu_{\text{N}15}|\{x_i\}\right)$ as $\{\hat{\mu}_{\text{N}14}\}$ and $\{\hat{\mu}_{\text{N}15}\}$, respectively, then we can estimate $P\left(\mu_{\text{N}14} < \mu_{\text{N}15}|\{x_i\}\right)$:

$$P\left(\mu_{\text{N}14} < \mu_{\text{N}15}|\{x_i\}\right) \approx \text{mean}\left(\hat{\mu}_{\text{N}14} < \hat{\mu}_{\text{N}15}\right),$$

where $\text{mean}\left(\hat{x}\right)$ is the Matlab command for taking the mean of a vector [24].

[1] J. Hofmann, M. Krug, N. Ortegel, L. Grard, M. Weber, W. Rosenfeld, and H. Weinfurter, Science **337**, 72 (2012).

[2] J. Yin, Y. Cao, Y.-H. Li, S.-K. Liao, L. Zhang, J.-G. Ren, W.-Q. Cai, W.-Y. Liu, B. Li, H. Dai, G.-B. Li, Q.-M. Lu, Y.-H. Gong, Y. Xu, S.-L. Li, F.-Z. Li, Y.-Y. Yin, Z.-Q. Jiang, M. Li, J.-J. Jia, G. Ren, D. He, Y.-L. Zhou, X.-X. Zhang, N. Wang, X. Chang, Z.-C. Zhu, N.-L. Liu, Y.-A. Chen, C.-Y. Lu, R. Shu, C.-Z. Peng, J.-Y. Wang, and J.-W. Pan, Science **356**, 1140 (2017).

[3] A. Faraon, C. Santori, Z. Huang, V. M. Acosta, and R. G. Beausoleil, Physical Review Letters **109**, 033604 (2012).

[4] D. Riedel, I. Sllner, B. J. Shields, S. Starosielec, P. Appel, E. Neu, P. Maletinsky, and R. J. Warburton, Physical Review X **7**, 031040 (2017).

[5] H. Bernien, L. Childress, L. Robledo, M. Markham, D. Twitchen, and R. Hanson, Physical Review Letters **108**, 043604 (2012).

[6] A. Batalov, C. Zierl, T. Gaebel, P. Neumann, I.-Y. Chan, G. Balasubramanian, P. R. Hemmer, F. Jelezko, and J. Wrachtrup, Physical Review Letters **100**, 077401 (2008).

[7] S. L. Mouradian, T. Schrder, C. B. Poitras, L. Li, J. Goldstein, E. H. Chen, M. Walsh, J. Cardenas, M. L. Markham, D. J. Twitchen, M. Lipson, and D. Englund, Physical Review X **5**, 031009 (2015).

[8] A. Sipahigil, M. L. Goldman, E. Togan, Y. Chu, M. Markham, D. J. Twitchen, A. S. Zibrov, A. Kubanek, and M. D. Lukin, Physical Review Letters **108**, 143601 (2012).

[9] J. O. Orwa, C. Santori, K. M. C. Fu, B. Gibson, D. Simpson, I. Aharonovich, A. Stacey, A. Cimmino, P. Balog, M. Markham, D. Twitchen, A. D. Greentree, R. G. Beausoleil, and S. Prawer, Journal of Applied Physics **109**, 083530 (2011).

[10] K.-M. C. Fu, C. Santori, P. E. Barclay, and R. G. Beausoleil, Applied Physics Letters **96**, 121907 (2010).

[11] Y. Chu, N. de Leon, B. Shields, B. Hausmann, R. Evans, E. Togan, M. J. Burek, M. Markham, A. Stacey, A. Zibrov, A. Yacoby, D. Twitchen, M. Loncar, H. Park, P. Maletinsky, and M. Lukin, Nano Letters **14**, 1982 (2014).

[12] Y.-C. Chen, P. S. Salter, S. Knauer, L. Weng, A. C. Frangeskou, C. J. Stephen, S. N. Ishmael, P. R. Dolan, S. Johnson, B. L. Green, G. W. Morley, M. E. Newton, J. G. Rarity, M. J. Booth, and J. M. Smith, Nature Photonics **11**, 77 (2017).

[13] J. P. Hadden, V. Bharadwaj, B. Sotillo, S. Rampini, R. Osellame, J. D. Witmer, H. Jayakumar, T. T. Fernandez, A. Chiappini, C. Armellini, M. Ferrari, R. Ramponi, P. E. Barclay, and S. M. Eaton, Optics Letters **43**, 3586 (2018).

[14] M. Ruf, M. IJspeert, S. van Dam, N. de Jong, H. van den Berg, G. Evers, and R. Hanson, Nano Letters **19**, 3987 (2019).

[15] P. Tamarat, T. Gaebel, J. R. Rabeau, M. Khan, A. D. Greentree, H. Wilson, L. C. L. Hollenberg, S. Prawer, P. Hemmer, F. Jelezko, and J. Wrachtrup, Physical Review Letters **97**, 083002 (2006).

[16] K.-M. C. Fu, C. Santori, P. E. Barclay, L. J. Rogers, N. B. Manson, and R. G. Beausoleil, Physical Review Letters **103**, 256404 (2009).

[17] L. Robledo, H. Bernien, I. van Weperen, and R. Hanson, Physical Review Letters **105**, 177403 (2010).

[18] C. Santori, P. E. Barclay, K.-M. C. Fu, R. G. Beausoleil, S. Spillane, and M. Fisch, Nanotechnology **21**, 274008 (2010).

[19] P. Siyushev, H. Pinto, M. Vrs, A. Gali, F. Jelezko, and J. Wrachtrup, Physical Review Letters **110**, 167402 (2013).

[20] J. R. Rabeau, P. Reichart, G. Tamanyan, D. N. Jamieson, S. Prawer, F. Jelezko, T. Gaebel, I. Popa, M. Domhan, and J. Wrachtrup, Applied Physics Letters **88**, 023113 (2006).

[21] S. B. van Dam, M. Walsh, M. J. Degen, E. Bersin, S. L. Mouradian, A. Galiullin, M. Ruf, M. IJspeert, T. H. Taminiau, R. Hanson, and D. R. Englund, Physical Review B **99**, 161203 (2019).

[22] P. Appel, E. Neu, M. Ganzhorn, A. Barfuss, M. Batzer, M. Gratz, A. Tschpe, and P. Maletinsky, Review of Scientific Instruments **87**, 063703 (2016).

[23] K. J. Brown, E. Chartier, E. M. Sweet, D. A. Hopper, and L. C. Bassett, Journal of Chemical Health and Safety **26**, 40 (2019).

[24] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, 3rd ed., Texts in Statistical Science (CRC Press, 2014).

[25] D. S. Sivia, *Data Analyis: A Baysian Tutorial*, 2nd ed. (Oxford University Press, New York, 2006).

[26] H. C. Thode, *Testing for Normality* (Marcel Dekker, Inc., NY, 2002).

[27] I. Lekavicius, T. Oo, and H. Wang, Journal of Applied Physics **126**, 214301 (2019).