



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Healey, AJ;Stacey, A;Johnson, BC;Broadway, DA;Teraji, T;Simpson, DA;Tetienne, J-P;Hollenberg, LCL

Title:

Comparison of different methods of nitrogen-vacancy layer formation in diamond for wide-field quantum microscopy

Date:

2020-10-29

Citation:

Healey, A. J., Stacey, A., Johnson, B. C., Broadway, D. A., Teraji, T., Simpson, D. A., Tetienne, J. -P. & Hollenberg, L. C. L. (2020). Comparison of different methods of nitrogen-vacancy layer formation in diamond for wide-field quantum microscopy. *Physical Review Materials*, 4 (10), <https://doi.org/10.1103/PhysRevMaterials.4.104605>.

Persistent Link:

<https://hdl.handle.net/11343/271001>

Comparison of different methods of nitrogen-vacancy layer formation in diamond for widefield quantum microscopy

A. J. Healey,¹ A. Stacey,^{2,3} B. C. Johnson,^{1,2} D. A. Broadway,^{1,2} T. Teraji,⁴ D. A. Simpson,¹ J.-P. Tetienne,^{1,2,*} and L. C. L. Hollenberg^{1,2,†}

¹*School of Physics, University of Melbourne, Parkville, VIC 3010, Australia*

²*Centre for Quantum Computation and Communication Technology,*

School of Physics, University of Melbourne, Parkville, VIC 3010, Australia

³*School of Science, RMIT University, Melbourne, VIC 3001, Australia*

⁴*National Institute for Materials Science, Tsukuba, Ibaraki 305-0044, Japan*

Thin layers of near-surface nitrogen-vacancy (NV) defects in diamond substrates are the workhorse of NV-based widefield magnetic microscopy, which has applications in physics, geology and biology. Several methods exist to create such NV layers, which generally involve incorporating nitrogen atoms (N) and vacancies (V) into the diamond through growth and/or irradiation. While there have been detailed studies of individual methods, a direct side-by-side experimental comparison of the resulting magnetic sensitivities is still missing. Here we characterise, at room and cryogenic temperatures, ≈ 100 nm thick NV layers fabricated via three different methods: 1) low-energy carbon irradiation of N-rich high-pressure high-temperature (HPHT) diamond, 2) carbon irradiation of δ -doped chemical vapour deposition (CVD) diamond, 3) low-energy N^+ or CN^- implantation into N-free CVD diamond. Despite significant variability within each method, we find that the best HPHT samples yield similar magnetic sensitivities (within a factor 2 on average) to our δ -doped samples, of $< 2 \mu\text{T Hz}^{-1/2}$ for DC magnetic fields and < 100 nT $\text{Hz}^{-1/2}$ for AC fields (for a $400 \text{ nm} \times 400 \text{ nm}$ pixel), while the N^+ and CN^- implanted samples exhibit an inferior sensitivity by a factor 2-5, at both room and low temperature. We also examine the crystal lattice strain caused by the respective methods and discuss the implications this has for widefield NV imaging. The pros and cons of each method, and potential future improvements, are discussed. This study highlights that low-energy irradiation of HPHT diamond, despite its relative simplicity and low cost, is a competitive method to create thin NV layers for widefield magnetic imaging.

I. INTRODUCTION

Widefield imaging based on ensembles of nitrogen-vacancy (NV) defects in diamond [1–7], also known as quantum diamond microscopy, is a promising tool with applications in condensed matter physics [8–10], geology [11, 12] and biology [7, 13]. Common to all these applications is the need for a dense layer of NV defects near the surface of a single crystal diamond substrate. Depending on the exact application, the optimal thickness t of this layer can vary from $t \lesssim 10$ nm to detect mesoscopic fluctuating signals e.g. from paramagnetic ions in solution [14–16], to several μm for imaging of geological samples over mm-scale fields of view [12].

For applications related to magnetic imaging, it is desirable to achieve diffraction limited resolution ($\Delta x \approx 300$ nm) and so the layer thickness should satisfy $t < \Delta x$. This regime is also convenient because it allows the NV layer to be treated as a 2D layer when reconstructing the current density or magnetisation of the sample [8, 17, 18], and it preserves the ability to image the sample via NV photoluminescence (PL) quenching or laser interference effects [19–21]. A thickness of order $t \approx 100$ nm is found to be optimal as it allows one to maximise the number of NVs without impacting the spatial resolution,

while also permitting the addition of a spacing layer (e.g. $t_{\text{sp}} = 100$ nm of Al_2O_3 [17]) grown on the diamond surface, by satisfying $t + t_{\text{sp}} < \Delta x$. The use of such a spacing layer is required to mitigate proximity-induced artefacts arising from strongly magnetic samples [22].

Several approaches exist to create such thin near-surface NV layers [23]. To localise NV centres near the surface one must restrict the creation of vacancies or incorporation of nitrogen (but not necessarily both) to the near-surface region. The N impurities can be introduced during the diamond synthesis, either via the high-pressure high-temperature (HPHT) method or via chemical vapour deposition (CVD) [24], or via ion implantation of an N-containing species [25]. The vacancies are typically introduced by irradiation with electrons, protons or ions [26]. The diamond is then annealed at $800 - 1200^\circ\text{C}$ in a vacuum environment to allow the vacancies to diffuse and form NV centres [27]. The sensitivity of the resulting NV layer depends not only on the number of NVs but also on their “quality”, in particular their charge state (only NV^- is useful for quantum sensing applications, whereas the co-presence of NV^0 produces background PL reducing the sensitivity) and spin properties (coherence times T_2^* and T_2 , optical contrast) [2, 28, 29]. These factors depend on the presence of other defects in the lattice (non converted N impurities or other growth- or irradiation-induced defects) and can vary significantly between different NV creation methods, motivating a comparative study of the resulting magnetic sensitivities.

* jtetienne@unimelb.edu.au

† lloydch@unimelb.edu.au

Assuming shot noise limited measurements, the sensitivity to DC magnetic fields with the pulsed optically detected magnetic resonance (ODMR) protocol is approximately given by [2, 29, 30]

$$\eta_{\text{dc}} \equiv \frac{1}{\gamma_e \mathcal{C} T_2^* \sqrt{\alpha \mathcal{R}}} \quad (1)$$

where $\gamma_e \approx 2.8$ MHz/G is the electron gyromagnetic ratio, $\mathcal{C} \ll 1$ is the spin contrast, T_2^* is the spin dephasing time, \mathcal{R} is the photon count rate under continuous wave (CW) excitation, and $\alpha = \frac{t_R}{t_R + t_D + T_2^*}$ is the readout duty cycle where t_R and t_D are the readout time and dead time per cycle, respectively. As \mathcal{R} will scale with the NV density, itself usually proportional to the density of other unwanted paramagnetic defects such as the substitutional nitrogen defect, it competes with the ensemble's T_2^* and it is not a priori clear which attribute to prioritise. In addition, the sensitivity for given parameters depends on the experimental conditions through t_R and t_D . For widefield imaging, the low laser intensities used require a relatively long t_R (≈ 10 μs), meaning $\alpha \approx 1$ and Equation 1 retains a strong T_2^* dependence. This contrasts with the confocal microscopy case, where often $\alpha \approx t_R/T_2^*$. The sensitivity to AC or fluctuating magnetic fields, η_{ac} , has a similar expression where T_2^* is replaced by T_2 [2, 29, 31], which similarly scales inversely to \mathcal{R} .

In addition to per-pixel measurement sensitivity, the extraction of accurate spatial information using wide-field NV microscopy requires identical ODMR transition frequencies across the field of view. While maintaining background magnetic field homogeneity is usually trivial, inhomogeneous crystal strain due to diamond growth, ion implantation, and surface preparation procedures can also obscure variation of a target field [32, 33].

In this work, we focus on three commonly employed methods to create thin NV layers:

- Method 1: low-energy irradiation of N-rich HPHT diamond;
- Method 2: Irradiation of δ -doped CVD diamond;
- Method 3: low-energy implantation of N-containing species into N-free CVD diamond.

Method 1 aims to create vacancies confined to a distance $\approx t$ from the surface of an HPHT diamond that contains a high density of N throughout the bulk, typically of the order of $[\text{N}] \approx 100$ ppm. This method has been previously demonstrated using ~ 30 keV helium ions (He^+), producing NV layers with $t \sim 100 - 200$ nm [34–36]. Here we employ C^- ions, which is a readily available species in low energy negative ion sources. One advantage over He^+ is the absence of He-related defects in the final NV layer. Method 2 replaces the HPHT substrate with a CVD diamond in which nitrogen was selectively incorporated during the growth so as to form a δ -doped layer of desired N density and thickness, on an otherwise N-free

substrate [24, 37]. The vacancies can then be created by low-energy ion implantation (here we also use C^-) [38–40], although higher energy particles could also be used since the vacancies do not have to be confined in depth, for instance NVs have been formed using 200 keV electrons from a transmission electron microscope [41, 42]. Finally, Method 3 employs low-energy implantation of an N-containing species (here we consider N^+ and CN^-) to simultaneously incorporate both N and V at the desired depth into an otherwise N-free diamond [43–45]. Despite being a very popular method for producing NV sensing layers, this approach creates unnecessary damage to the diamond because of the large number of vacancies created for each N [46, 47], which is especially problematic for dense layers of near-surface NVs [28]. In contrast, in Methods 1 and 2 the irradiation fluence can be adjusted to reach a ratio of created vacancies to the number of N impurities present in the diamond optimal for maximising NV yield, making them better candidates to reach the regime in which T_2^* is limited only by the residual N impurities [29, 48]. In practice, this condition is likely to be given by a combination of the resultant Fermi level position as well as the nitrogen and vacancy concentrations, and thus optimising implant parameters is expected to require a level of fine tuning beyond the scope of this work. For simplicity, we focus on the regime where the number of vacancies created by the implantation process is commensurate with the expected nitrogen concentration.

While these techniques have individually been the subject of previous works [34–47], the large variability in the experimental methods used to assess them make a reliable comparison difficult. In particular, the quantities \mathcal{C} , α and \mathcal{R} in Eq. (1) all depend on the specifics of the optical setup used to measure them, and can vary drastically between confocal and widefield measurements. In wide-field magnetic imaging, the laser intensities are typically much smaller than in confocal measurements, and so the per-NV sensitivity may appear smaller, which is a price to pay to access a large field of view. Thus, the purpose of the present work is to provide a side-by-side comparison of the sensitivities (η_{dc} and η_{ac}) obtained with the three NV creation methods outlined above, while keeping as many experimental parameters fixed as possible. We stress that, given the large number of experimental variables for each method, we do not seek to optimise them. Instead, our aim is to generate a baseline of the performances that can be achieved with the different methods without advanced optimisation, so as to help users to best choose the method that suits their capabilities and application, and to guide future efforts in optimising a particular method.

The paper is organised as follows. In Sec. II, we describe the experimental procedures to form the NV layers studied (II A) and to characterise them (II B). The results of these measurements are presented in Sec. III, followed by a discussion (IV) and conclusion (V).

II. METHODS

A. Samples

The sample creation methods used for this study are detailed in Table I and depicted schematically in Figure 1(a). Method 1 samples (referred to as HPHT samples hereafter) are as-received commercially available (Delaware Diamond Knives) HPHT diamonds polished to a surface roughness <5 nm Ra with specified nitrogen concentration of < 200 ppm. Despite the high initial nitrogen concentration, these samples showed limited NV fluorescence prior to implantation indicating poor N to NV conversion. C^- irradiation with doses between 10^{11} and 10^{13} cm^{-3} , chosen to produce a number of vacancies comparable to the expected nitrogen concentration, resulted in significant PL enhancement (by several orders of magnitude). Samples were held on the implanter stage at a 7° tilt angle and 10° rotation to minimise ion channelling, and the chamber pressure during implantation was $\leq 10^{-6}$ mbar. The implant energy of 100 keV was chosen to confine vacancies produced to a ~ 100 nm layer as shown by Figure 1(b), which is a simulation produced using the Stopping and Range of Ions in Matter (SRIM) software package.

Samples fabricated by Method 2 (hereafter referred to as δ -doped samples) were created via CVD growth over a commercial type Ib HPHT substrate (Element Six). An oxygen etch was performed to remove polishing damage and maximise subsequent overgrowth quality [49]. An electronic grade, N-free layer of a thickness order $10 \mu m$ was then grown to buffer the grown NVs from defects in the substrate and to reduce the PL contribution from the substrate. The δ -doped layer was then grown on top during a short growth designed to create a film between 100 and 200 nm thick and with a level of nitrogen incorporated set by the proportion of nitrogen present in the plasma during growth (varied between samples). The δ -doped layer was grown using a low microwave power of 750 W and under a low pressure of 25 Torr to promote slow, high quality crystal growth [37]. The proportion of methane (natural isotopic proportions) in the gas mixture was $\approx 1\%$. The CVD reactor used to grow the NV layers is multi-purpose and hence not specifically optimised for low density incorporation of growth defects, so the grown samples are not expected to represent benchmark quality. Instead, these samples are intended to represent “standard” and relatively easy to produce δ -doped diamonds. Prior to implantation the doped layer exhibited some fluorescence but it is expected that less than 0.5% of incorporated nitrogen forms NV centres in the as-grown diamond [50]. This PL was increased by over an order of magnitude on average by 100 keV, 10^{11} - 10^{12} cm^{-2} carbon irradiation, indicating an improved NV yield. The implant parameters were chosen to approximately match the vacancy production to the expected N density and depth profile of the doped layer.

Method 3 samples were made from electronic-grade

single-crystal CVD diamonds (purchased from Delaware Diamond Knives) and then overgrown with $2 \mu m$ of ^{12}C -enriched (99.95%) electronic-grade CVD diamond [51], although this step is not normally required for this method. The substrates were then implanted using $^{15}N^+$ (30-100 keV, 4×10^{13} - 1×10^{14} cm^{-2}) or $^{12}C^{14}N^-$ (54 keV, 2×10^{12} cm^{-2}) ions (hereafter referred to as N^+ and CN^- implants), which were chosen as representative nitrogen implant species as they are commonly available at most low energy ion implantation facilities. Figure 1(c) shows this method produces an overlapping region of implanted ions and vacancies but without the ability to tune densities (here the peak vacancy density is two orders of magnitude larger than the peak N density according to SRIM).

Following implantation, all samples were annealed in a vacuum of $\sim 10^{-5}$ Torr using a ramp sequence culminating at $1100^\circ C$ to form NV centres and repair part of the irradiation-induced damage [28]. The samples were then cleaned for 15 minutes in a boiling mixture of sulphuric and nitric acid prior to measurement. This is expected to standardise the surface chemistry across all samples, avoiding the impact of the surface on our comparative measurements. Additional sample production details are given in Appendix C.

B. Measurements

The sensitivity measurements for all samples were carried out on a purpose-built widefield microscope described in detail elsewhere [10, 28]. Microwave delivery is facilitated by a gold resonator deposited on a glass cover slip, onto which diamond chips are mounted NV face-down. A permanent magnet aligned with one family of NV axes provides the bias field required to Zeeman split the NV resonances. NV centres are excited by a 532 nm laser focussed to the back aperture of an oil-immersion microscope objective (Nikon 40x NA = 1.3), where the laser power was measured to be 220 mW. The resulting NV fluorescence was filtered (660-735 nm) and imaged on a scientific complementary metal oxide semiconductor (sCMOS) camera for processing. The measurements were taken over a $50 \mu m \times 50 \mu m$ field of view as this was the region over which the laser intensity was roughly uniform (≈ 3 kW/cm²). This region, typical for widefield imaging experiments, contains 10^6 - 10^7 NVs (depending on the sample) which we average over in our spin measurements. A readout pulse time of $t_R = 5 \mu s$, chosen to reach a compromise between ensemble initialisation and readout contrast, was held constant for all measurements. A dead time of $t_D = 1.5 \mu s$ was introduced following each laser pulse to allow relaxation from the metastable singlet state to the ground state. All room temperature measurements, with the exception of T_2 , were conducted under a weak bias field of 60 G. T_2 measurements were undertaken at a higher field of 475 G to ensure that, for the samples containing a natural abundance of ^{13}C , the

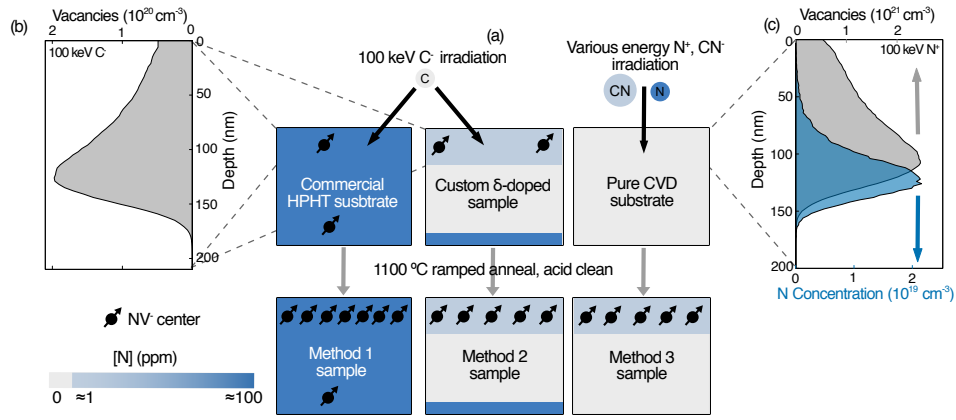


FIG. 1. **Overview of the NV layer formation methods investigated:** a) Schematic illustrating sample production. From left to right: Method 1, carbon irradiation of high-N, low-NV HPHT diamond to create a localised NV-rich layer; carbon irradiation of δ -doped CVD layer to improve NV conversion; implantation of a pure CVD substrate with a chosen nitrogen-containing species to incorporate nitrogen into the crystal at the cost of extra damage by vacancy over-creation. The colour scale used and relative NV densities are schematic only. b) and c) are SRIM simulations for the C^- and N^+ implantation cases respectively, showing depth distributions of vacancies produced (both) and implanted ions (Method 3 only) produced by representative implantation procedures.

Sample name	Substrate type	Surface finish	Implanted species	Energy (keV)	Dose (ions/cm ²)	Projected range (nm)	Estimated NV density (NV/ μm^2)	Estimated NV yield (%)
HPHT-1	HPHT	Polished	$^{12}\text{C}^-$	100	5×10^{11}	142	1.6×10^5	0.90
HPHT-2	HPHT	Polished	$^{12}\text{C}^-$	100	1×10^{12}	142	1.5×10^5	0.85
HPHT-3	HPHT	Polished	$^{12}\text{C}^-$	100	1×10^{13}	142	4.6×10^5	2.60
HPHT-4	HPHT	Polished	$^{12}\text{C}^-$	100	1×10^{13}	142	1.3×10^5	0.73
HPHT-5	HPHT	Polished	$^{12}\text{C}^-$	100	1×10^{13}	142	4.7×10^5	2.66
HPHT-6	HPHT	Polished	$^{12}\text{C}^-$	100	2×10^{11}	142	4.5×10^5	2.54
HPHT-7	HPHT	Polished	$^{12}\text{C}^-$	100	1×10^{13}	142	4.2×10^5	2.37
δ -1	CVD (δ -doped)	As grown	$^{12}\text{C}^-$	100	1×10^{12}	142	2.6×10^4	
δ -2	CVD (δ -doped)	As grown	$^{12}\text{C}^-$	100	1×10^{12}	142	4.8×10^4	
δ -3	CVD (δ -doped)	As grown	$^{12}\text{C}^-$	100	1×10^{11}	142	1.5×10^3	
δ -4	CVD (δ -doped)	As grown	$^{12}\text{C}^-$	100	1×10^{11}	142	2.6×10^3	
δ -5	CVD (δ -doped)	As grown	$^{12}\text{C}^-$	100	1×10^{12}	142	2.9×10^4	
N-1	CVD	As grown	$^{15}\text{N}^+$	30	4×10^{13}	42	1.7×10^4	4.12
N-2	CVD	As grown	$^{15}\text{N}^+$	30	4×10^{13}	42	2.6×10^4	6.43
N-3	CVD	As grown	$^{15}\text{N}^+$	100	4×10^{13}	122	2.6×10^4	4.54
N-4	CVD	As grown	$^{15}\text{N}^+$	100	5×10^{13}	122	2.9×10^4	5.81
N-5	CVD	As grown	$^{15}\text{N}^+$	100	1×10^{14}	122	2.0×10^4	2.00
CN-1	CVD	As grown	$^{12}\text{C}^{14}\text{N}^-$	54	2×10^{13}	42	9.1×10^3	4.55
CN-2	CVD	As grown	$^{12}\text{C}^{14}\text{N}^-$	54	2×10^{13}	42	7.7×10^3	3.87
CN-3	CVD	As grown	$^{12}\text{C}^{14}\text{N}^-$	54	2×10^{13}	42	7.2×10^3	3.61

TABLE I. **List of diamond samples used in this study.** Samples differ in (from left to right in the table) their diamond growth method, whether overgrowth took place after substrate polishing, and the implant species, energy, and dose chosen. NV density estimated by comparing fluorescence to that given by a single NV on a confocal microscope. Yields quoted are calculated by comparing the inferred NV content with the known N fluence for the N implants and taking an estimated N concentration of 100 ppm for HPHT samples. No precise measure of doped nitrogen content is available for δ -doped layers.

resulting revivals were closely spaced enough to allow accurate extraction of the decay envelope. The microwave power was held to give a constant π time of 40 ns for all spin measurements. To remove common mode laser noise

(where applicable), pulse sequences were normalised using a reference measurement taken with respect to the $| -1 \rangle$ basis rather than the usual $| 0 \rangle$ by using a π pulse prior to readout.

To characterise the samples, we measure the quantities present in Equation 1. The photon count rate \mathcal{R} was averaged over the field of view under CW laser excitation over a consistent camera exposure time of 30 ms, and then normalised in time and area. Estimated NV densities listed in Table I were inferred from \mathcal{R} using a scaling factor obtained by comparing the fluorescence collected on a confocal microscope for selected samples to that from a reference single NV under identical conditions. The spin contrast is obtained by coherently driving the NV ensemble, with \mathcal{C} taken as the peak-to-peak amplitude of the Rabi oscillations when fit to a damped sinusoid. T_2^* is measured using the Ramsey pulse sequence and the T_2 is measured first with a Hahn echo sequence and then with Carr-Purcell-Meiboom-Gill (CPMG) sequences with variable numbers of π pulses. We also measure the spin relaxation time T_1 which represents an upper limit to ensemble coherence times and, due to its sensitivity to resonant magnetic fluctuations, forms the basis of many sensing experiments [6, 52]. It is measured simply as the decay of the $|0\rangle$ state under dark evolution. Further details on the measurements and data analysis are given in Appendix B and full results for each sample are given in Table II (Appendix A). In the HPHT samples, different growth sectors were present and so for consistency the measurements were performed on the brightest regions of each sample, corresponding to high-nitrogen growth sectors with an expected nitrogen density of the order of 100 ppm.

These measurements were repeated at lower temperatures (down to 4 K) for representative samples on a similar widefield NV microscope within a close-cycle cryostat [53]. To minimise sample heating, a lower laser power was used for these measurements, which required a longer readout time of $t_R = 10 \mu\text{s}$ to be used to compensate for the reduced rate of ensemble initialisation.

III. RESULTS

A. Basic spin properties

To benchmark sample quality we first consider the basic NV spin properties at room temperature, summarising the results in Figure 2. Focussing solely on PL for now, we plot the photon count rate \mathcal{R} against the implantation doses used for all the samples studied in Figure 2(a), with the different methods separated by colour. In the following, we assume the measured PL is dominantly due to the negative NV charge state, although there may be small contributions from NV^0 as will be discussed below. The HPHT samples are the brightest, which is to be expected given their high initial N concentration (≈ 100 ppm). Method 3 samples exhibit fluorescence within an order of magnitude of each other, mostly in line with the spread in implantation doses as reflected by the fairly constant conversion ratios listed in Table I. These samples are less fluorescent than the HPHT sam-

ples and brightest δ -doped samples despite higher implant doses because the NV production is limited by the nitrogen incorporated during the implant process rather than the vacancy production. δ -doped samples feature the widest spread in brightness due to nitrogen concentration in the CVD gas mixture being varied over several orders of magnitude (see Appendix C), although we do not expect a direct conversion from nitrogen concentration in the reactor gas mixture to that incorporated into the crystal. Indeed, the similar PL of the three brightest δ -doped samples suggests we may be close to saturating the nitrogen concentration under our growth conditions, highlighting a potential detraction of the technique in creating very high density NV ensembles.

In Figure 2(b)-(e) the results of the spin measurements described above are plotted against the \mathcal{R} as this is the parameter from Equation 1 expected to vary most straightforwardly between samples, being proportional to NV density. Figure 2(b) plots T_1 against \mathcal{R} and we see that the T_1 time varies significantly between the different methods and is inversely correlated to PL. T_1 ranges from ~ 3 ms in the lowest density δ -doped samples, which we assume to be a nearly phonon-limited value, to $\lesssim 100 \mu\text{s}$ for the highest PL HPHT samples. We attribute this dramatic reduction in T_1 proportional to the increase in nitrogen defects both to the general increase in magnetic noise and to electron tunnelling to nearby defects that sees the NV^- ionise to NV^0 during dark evolution [54, 55].

Figure 2(c) shows that the Rabi contrast \mathcal{C} also appears to scale inversely with PL, although the variation is milder with most samples lying in the range of 2-4%. As in the T_1 case, the availability of extra tunnelling sites in the higher density samples is likely responsible for the reduced measurement contrast. The presence of NVs that always exist in the NV^0 charge state could also reduce \mathcal{C} , however the trend resembling that present in the T_1 data suggests a dynamical effect. The lowest density δ -doped sample is an exception to this trend, likely because this sample contains too low a nitrogen density to sustain a high NV^- yield in the presence of significant surface-induced band bending, and hence a significant proportion of its fluorescence arising from NV^0 [56].

The inverse correlation between free induction decay time T_2^* and \mathcal{R} (Figure 2(d)) is expected as the increased defect density associated with brighter samples gives a higher level of magnetic noise, which reduces T_2^* . To see how this impacts measurement sensitivity, we plot constant sensitivity (calculated for a $400 \text{ nm} \times 400 \text{ nm}$ “pixel”, representing approximately diffraction-limited spatial resolution) curves (red dashed lines) assuming 3% contrast in Equation 1. The data points roughly follow these curves, indicating that T_2^* and \mathcal{R} offset one another to a large degree, consistent with previous studies [28]. However, there is a large spread in this data and additional trends within and between the methods are worth noting. The largest variation in NV quality relative to PL (vertical spread in figures) is present in the HPHT sam-

ples (with T_2^* ranging from 15 to 120 ns). This variation is due to this method's reliance on a purchased substrate not produced to specification, as well as the variation in defect incorporation in different HPHT growth sectors. The best HPHT samples, however, appear to outperform the nitrogen implanted samples, as do all of the δ -doped samples. This confirms that the additional damage associated with nitrogen implantation produces a source of dephasing that is significant alongside the contribution from the nitrogen spin bath. This is further evidenced by the samples implanted with the larger CN^- ion exhibiting comparable coherence times to the N^+ implants, but with much lower PL. In other words, the extra damage caused per ion (a 54 keV CN^- implant produces 65% and 25% more vacancies than 30 keV and 100 keV N^+ implants respectively) amounts to a similar total paramagnetic defect density for a lower nitrogen concentration.

Figure 2(e) plots Hahn echo T_2 against \mathcal{R} , again with constant sensitivity curves to guide the eye. The trends are similar to the T_2^* case, but with a reduction in spread due to the success of the Hahn echo sequence at removing quasi static magnetic noise contributions.

To better analyse the trends in sensitivity versus fabrication method, we plot the calculated DC and AC sensitivities (using actual Rabi contrast values for each sample) in Figure 3. We note that variance in NV layer thickness of up to a factor of 2 translates to a scaling in sensitivity of up to $\sqrt{2}$. It is also worth noting that the laser power used is far below optical saturation and so sensitivity could be improved by a further 10 times or more for any given sample by focussing the laser and increasing the counts received per unit area by a factor of 100 (at the cost of a decreased field of view). Nevertheless, the trends qualitatively noted earlier are confirmed: the δ -doped samples emerge as the highest quality, with the mean magnetic sensitivity for this method being around a factor of 3 better than that of the others. The HPHT series exhibits the largest amount of spread (almost 100 % of the mean), but the best HPHT samples provide comparable sensitivity to the δ -doped samples. This spread in sensitivity is accompanied by variability in the amount of fluorescence a given implantation dose promotes for HPHT samples, which is indicative of the variability in nitrogen concentration in the purchased crystal. This lack of foreknowledge can be compounded by a failure to correctly tune the vacancy production of the chosen implantation procedure, and hence a failure to optimise the process. Over-implanting will produce unnecessary damage and under-implanting will result in a suboptimal NV:N conversion ratio, and in both cases this will be reflected by a decrease in sensitivity. This suggests that a way to improve the sensitivity of these samples could be to determine the N density prior to implantation and adjust implant parameters accordingly. The existence of different growth sectors within a single diamond also complicates this picture, although our study was restricted to high-nitrogen sectors and PL homogeneity over regions more than sufficient for widefield

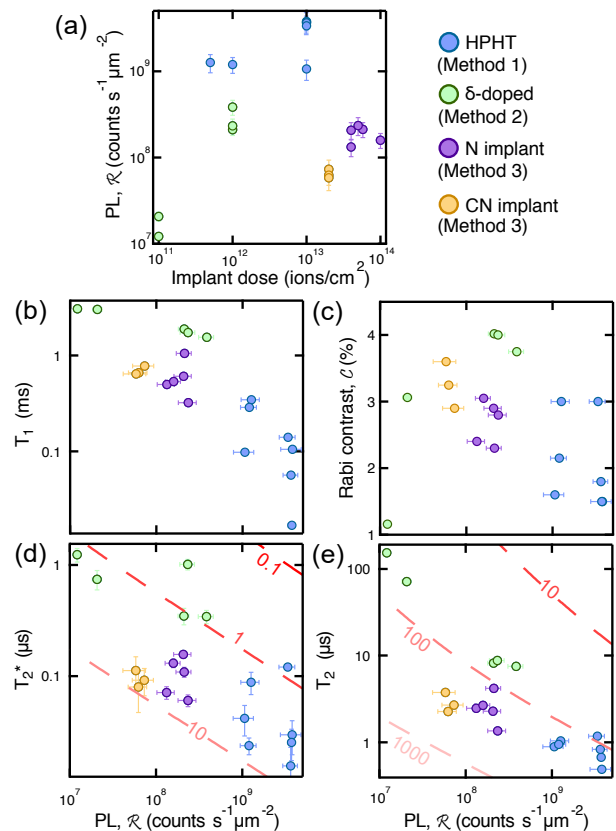


FIG. 2. Sample characterisation at room temperature: Measured variation of **a)** \mathcal{R} with implant dose; and **b)** T_1 , **c)** Rabi contrast, **d)** free induction decay time T_2^* , and **e)** Hahn echo T_2 with \mathcal{R} . All measurements made at room temperature and at 60 G except for the Hahn echo, which was conducted at 475 G to reduce the impact of ^{13}C revivals on the fits. Red lines in **d)** and **e)** denote lines of constant per $400 \text{ nm} \times 400 \text{ nm}$ DC and AC sensitivity (in $\mu\text{T Hz}^{-1/2}$ and $\text{nT Hz}^{-1/2}$ respectively) respectively at a constant Rabi contrast of 3% to guide the eye. The error bars indicate one standard deviation for \mathcal{R} and the standard fitting errors for the spin properties. Where not visible, error bars are smaller than the marker size.

imaging is typical.

The HPHT samples already achieving the best sensitivities are likely those currently best optimised, and so it is reasonable to expect refinement of the process to result in this level of sensitivity being achieved (or improved upon) by a majority of HPHT samples in future. There is also potential to improve the δ -doped samples' sensitivity in this way as the N content is also initially unknown in this case, but this improvement does not exist for the implantation of an N-free substrate as there is no way to alter the ratio of vacancy production to N implanted at a given energy.

These results indicate that the simple and relatively inexpensive Method 1 can offer competitive magnetic sensitivity in widefield NV imaging, with large potential for improvement following further optimisation.

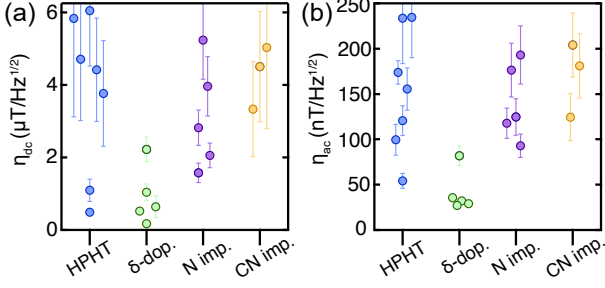


FIG. 3. **Sensitivity vs NV creation method:** Values of **a)** DC and **b)** AC magnetic sensitivities from Equation 1 (using T_2^* and T_2 respectively) for a $400 \text{ nm} \times 400 \text{ nm}$ pixel. Error bars obtained using standard propagation using measurement errors detailed in Figure 2. Sensitivities obtained on a setup with $t_R = 5 \mu\text{s}$, $t_D = 1.5 \mu\text{s}$, and a laser intensity of $3 \text{ kW}/\text{cm}^2$.

B. Extended dynamical decoupling

In sensing AC fields, it is common to use more advanced dynamical decoupling techniques than the Hahn echo pulse sequence to decouple from the spin environment more efficiently, access higher frequencies, and improve spectral selectivity [57, 58]. To benchmark our samples for use in this regime, we test their performance under CPMG sequences of increasing length, characterised by the number of π pulses N .

Figure 4(a) plots the CPMG T_2 as a function of N for representative samples from each method. A clear extension of T_2 is evident in all cases, with saturation above $N \approx 1024$. Fitting the T_2 extension to a power law, $T_2(N) = T_2(1)N^s$, gives values of s between 0.4 and 0.8 across all samples, with the majority of samples exhibiting values below the theoretical $s = 2/3$ for a dilute electron spin bath [59]. In most cases, the maximum attainable T_2 fell short of the $T_2 = T_1/2$ limit previously observed for bulk ensembles [60]. Previous work on shallow ($\leq 20 \text{ nm}$) single NVs saw values of $s < 2/3$ and saturated $T_2 < T_1/2$ [61] and as our layers exist in an intermediate depth regime between this and the bulk case, our results are broadly consistent with these studies. No trends specific to the methods of layer creation were obvious so we leave analysis of decoupling scaling of ensembles like these to future work.

To assess the efficiency of the decoupling across samples, we plot the CPMG-1024 vs Hahn echo T_2 for all samples in Figure 4(b). The enhancement varies between 10 and 100 and is greatest for the CN^- implants and some HPHT samples. The HPHT samples that see the least extension could not be extended to $N = 1024$ due to the finite pulse length (see below), while N implants saw less T_2 extension than the other methods despite the majority supporting this number of pulses. The CN^- implants undergo greater T_2 extension than the N^+ implants, indicating that the decoupling is successful in mitigating the effects of the extra vacancy production.

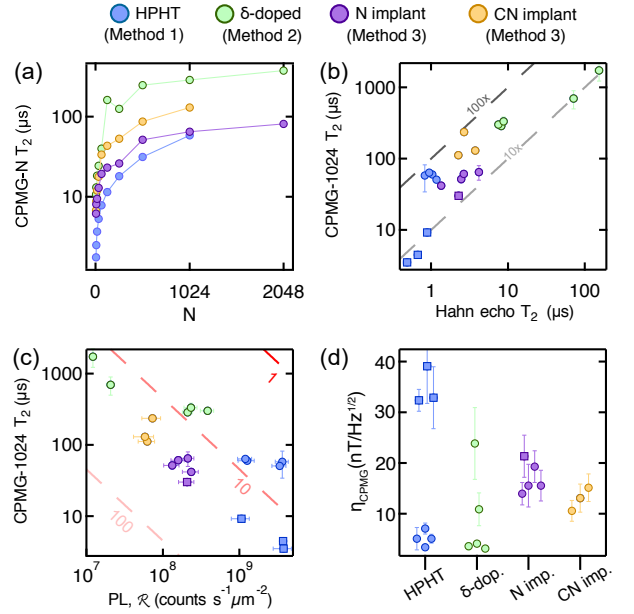


FIG. 4. **Extended decoupling characterisation:** **a)** T_2 scaling of selected representative samples of each sample type as a function of the number of pulses N in the CPMG- N sequence. Data displayed collected from samples δ -1 (green), CN-3 (yellow), N-3 (purple), and HPHT-6 (blue). **b)** Extension of T_2 from $N=1$ (Hahn echo) to $N=1024$, with grey lines showing 10-fold and 100-fold improvement. **c)** Scaling of CPMG-1024 T_2 with \mathcal{R} , with iso-sensitivity curves in $\text{nTHz}^{-1/2}$. **d)** AC sensitivities calculated using Equation 1 (using the CPMG-1024 T_2), grouped by method. Data points marked by squares in panels b)-d) belong to samples that could not be extended beyond $N \approx 256$ and see only modest improvement over Hahn echo values.

Comparing CPMG-1024 T_2 to \mathcal{R} we see (Figure 4(c)), as before, that the data follows constant sensitivity curves but in two apparently separate bands. The HPHT samples that were successfully decoupled to $N = 1024$ display comparable T_2 to the N-implants, which are much less fluorescent, and lie along a similar iso-sensitivity line to the bright δ -doped samples. This is confirmed in Figure 4(d), which shows the best magnetic sensitivities are obtained by δ -doped and HPHT samples.

The lowest density δ -doped samples fall slightly behind in sensitivity compared to their DC or Hahn echo AC sensitivities despite achieving T_2 values close to T_1 . We can understand this by viewing Equation 1 in the long T_2 limit ($T_2 \gg t_R = 5 \mu\text{s}$), in which α tends towards t_R/T_2 rather than 1. This shows that as T_2 is extended the dependence of the equation on \mathcal{R} becomes stronger, and this is part of the reason why the HPHT samples excel in this regime. The overall spread in sensitivity within and between methods (ignoring samples that could not be extended to $N = 1024$) is again reduced with additional decoupling, with the Method 3 samples, although still the least sensitive on average, only obtaining worse sensitivities by a factor ~ 2 .

We note that a key experimental limitation for this section is the microwave infrastructure that imposes a finite pulse duration (here 40 ns for each π pulse). While the idealised instantaneous pulse approximation is valid when only a small number of pulses are applied, as N is increased the microwave duty cycle can become significant and the bulk of dephasing can occur during this driving, limiting the effectiveness of decoupling. This is especially evident for the N-dense HPHT samples and is the reason why the majority of samples reach limiting T_2 values far shorter than T_1 . In principle, faster pulses could see additional coherence extension and the AC sensitivities of the worst-performing HPHT samples fall into line with the others, further incentivising the creation of highly dense NV ensembles, but in practice this is a realistic experimental limitation.

Overall, the conclusions in this sensing regime are much the same as in the previous section, apart from the most optimised HPHT samples becoming even more competitive and the impact of the disadvantages inherent to Method 3 appearing to be reduced.

C. Crystal strain mapping

Separate from magnetic sensitivity, another important quantity for widefield magnetic imaging is the crystal strain homogeneity in the NV layer. In particular, strain inhomogeneity can easily be conflated with the variation of a target field (in DC magnetometry), or constitute variable detuning from resonant driving across the field of view (AC measurements). Additional experiments can be made to normalise out these effects, though this is often impractical as it requires addressing multiple transition frequencies [32]. This could make an experiment less time efficient or be technologically impossible for high fields, hence it is advantageous to have as little strain variation as possible.

Our samples vary in their surface finish (polished vs overgrown) and their growth method (CVD vs HPHT), both factors which will contribute to near-surface strain. The samples also vary in their implant doses, which are in general high and therefore also expected to impact the crystal quality. All samples began as purchased substrates polished by their manufacturer (details unknown) but while the HPHT samples were left as-received prior to implantation, Method 2 and 3 samples all underwent some level of CVD overgrowth designed in part to mitigate polishing-induced strain. CVD growth is known to be susceptible to the incorporation of growth defects such as crystal lattice dislocations that may result in localised strain [33], while the thermodynamically stable HPHT growth process can result in fewer strain-inducing extended growth defects.

To assess the strain homogeneity within our samples, we mapped the strain over the full $50\ \mu\text{m} \times 50\ \mu\text{m}$ field of view using ODMR spectroscopy. ODMR maps focussing on the aligned $|0\rangle \rightarrow |\pm 1\rangle$ transitions under a bias field

of 60 G were obtained for a variety of locations on a range of representative samples. From the two transition frequencies (f_1 and f_2) the zero-field splitting constant $D = \frac{f_1+f_2}{2}$ can be inferred as it is sensitive to strain but not magnetic fields [32]. Assuming temperature and electric fields are homogeneous across fields of view, D maps thus tell us about the strain variations (in units of frequency) within samples. A series of representative strain maps for samples from each method is presented in Figure 5, where the colour scales denote the value of Δf , which is defined as the deviation from $D = 2870$ MHz (though the absolute value is arbitrary due to the possible co-presence of other fields which manifest as a constant shift across the field of view).

In all four cases, strain variations over the scale of a few microns are present over the whole field of view. The amplitude of these variations ranges from a few kHz for the δ -doped sample (close to the noise floor of the measurement) to 10-25 kHz in the others. In addition, stronger isolated features are sometimes visible, for example a 20 kHz spot in the δ -doped sample, and 1 MHz streaks in the HPHT sample. We expect that the δ -doped feature is an example of a CVD growth dislocation, while the HPHT features are polishing marks which are especially prevalent in the absence of any additional surface treatment.

Background variation could originate in the crystal growth, polishing and implantation processes the samples underwent. The directionality of the pattern in Figure 5(b) suggests the substrate's polishing as the source, despite the subsequent $2\ \mu\text{m}$ CVD overgrowth this sample underwent. This highlights the necessity of performing an additional process such as reactive ion etching (RIE), with or without subsequent overgrowth, to completely remove polishing damage [62–64]. Sample CN-1 underwent an RIE process prior to overgrowth and Figure 5(c) shows that this was successful in removing obvious polishing features, although there is still a noticeable strain gradient and variation of up to 25 kHz. This background is unlikely to be due to the CN implantation alone as it is larger than samples subjected to higher doses and so is thought to be damage related to the RIE itself. This shows the importance of optimising the RIE process to best prepare the surface, following an Ar/Cl etch like the one conducted here with an additional oxygen etch [64].

The lowest level of background variation (~ 5 kHz) was found in the δ -doped sample (Figure 5(d)), which benefited from the lowest implantation dose as well as the thickest CVD overgrowth ($> 10\ \mu\text{m}$) on a polished (HPHT) substrate. HPHT substrates were chosen for these growths primarily to inherit their low dislocation densities in the CVD overgrowth [65]. An oxygen etch process designed to remove polishing damage was conducted prior to the CVD overgrowth and appears to have been successful in preventing the appearance of polishing features in the final grown layer [49]. Large dislocation features are sparse and only of the same order as background variation in the other samples, indicating

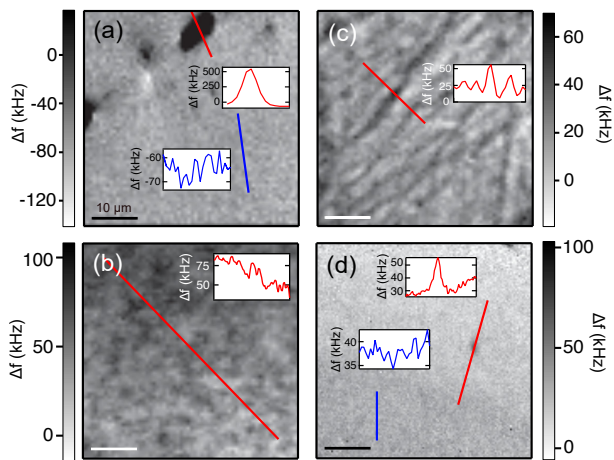


FIG. 5. **Representative strain maps from each method** **a)** Sample HPHT-4: HPHT sample; **b)** Sample N-5: N^+ implant; **c)** Sample CN-1: CN^- implant sample; **d)** Sample δ -2: δ -doped sample. The images show the variation Δf of the zero-field splitting constant D . The insets show linecuts taken along the red and blue solid lines.

that although the CVD process is susceptible to the incorporation of crystal imperfections, this is a relatively small detraction compared to the polishing and implantation processes. This also shows that, with careful selection of growth conditions and underlying substrate, CVD growth is capable of producing mostly strain-free material.

The large polishing marks observed in the HPHT samples underline the importance of undertaking some form of post-polishing process. However they are in general quite sparse, meaning that quick pre-screening of such samples should be adequate to ensure an uncompromised imaging experiment. The background variation in these samples is of similar magnitude to that of Method 3 samples despite the polished surface finish, which is likely an indication of the high strain uniformity of the original crystal.

It is reasonable to expect some degree of variation in the strain homogeneity between and even within samples, however we can still draw informative conclusions from these results. Even away from large features, typical background variation of tens of kHz is significant for high sensitivity magnetic imaging (corresponding to variations on the order of $1 \mu\text{T}$). This means that the specifics of sample surface preparation, which the results of this section show is the single greatest contributor to strain homogeneity even when viewed alongside high ion implantation doses and variation in crystal growth, are especially important for future applications of widefield NV imaging. We emphasise, however, that the typical strain variations over a field of view are still sufficiently small to not limit the measured T_2^* values presented in Section III A, hence T_2^* is likely limited by magnetic noise and the effects of crystal strain will be less noticeable in measurements averaged over a field of view for this

density of NV ensemble. Thus, in this regime surface preparation appears to play a large role in the suitability of a sample to widefield imaging through its influence on strain homogeneity, but only makes a limited contribution to the NV spin properties, while the opposite is true for the NV ensemble creation methods that are the primary focus of this work.

D. Temperature dependence

Some applications of widefield NV imaging require the sample of interest to be cooled to cryogenic temperatures. For example, superconducting phenomena (vortices and transport currents) were recently imaged at 4 K [53]. Thus, it is useful to investigate the temperature dependence of the magnetic sensitivity of the NV layers and so we repeat the suite of measurements from Section III A as a function of temperature from 4-300 K for a series of representative samples.

We first plot T_1 vs temperature in Figure 6(a). At low temperatures, we see significant extension of T_1 for all samples. The largest extension (around two orders of magnitude) is seen for samples with long room temperature T_1 , consistent with them being governed by phonon relaxation. However, even the samples with short room temperature T_1 show a significant extension (of around one order of magnitude for HPHT samples). This suggests that the dominant contribution to T_1 in this case depends on temperature as well, possibly through thermal activation of donor/acceptor defects, although further work is needed to fully understand this behaviour.

In Figure 6(b) we plot Rabi contrast (Sample CN-1) against temperature and see a surprising reduction at 60 G below 100 K. This trend is consistently observed but is most marked for N implants (a factor of 3), while δ -doped samples saw a lesser reduction (factor of 2), though it is unclear from the small sample size whether this difference is significant. This reduction is largely recovered at a higher field of 1970 G. This phenomenon has previously been observed in single NV centres [66] and its origin is currently being investigated. The PL (also plotted in Figure 6(b)) does not vary significantly with temperature and although there is a minor reduction in PL at the higher field due to increased quenching of off-axis NVs [67], this is not enough to explain the differences in measured contrast. No other spin property was observed to change between the two fields apart from a slight extension of T_1 at 1970 G.

Figure 6(c) shows the T_2^* variation with temperature, which is very small for all samples, indicating that temperature-independent magnetic noise dominates this quantity at all temperatures. Hahn echo T_2 (Figure 6(d)) is also broadly constant with temperature, with the exception of a $\approx 25\%$ extension for the δ -doped sample with long room temperature T_2 . Decoupled, saturated T_2 has previously been shown to have a phonon-coupled component for dilute spin baths in bulk and near-surface

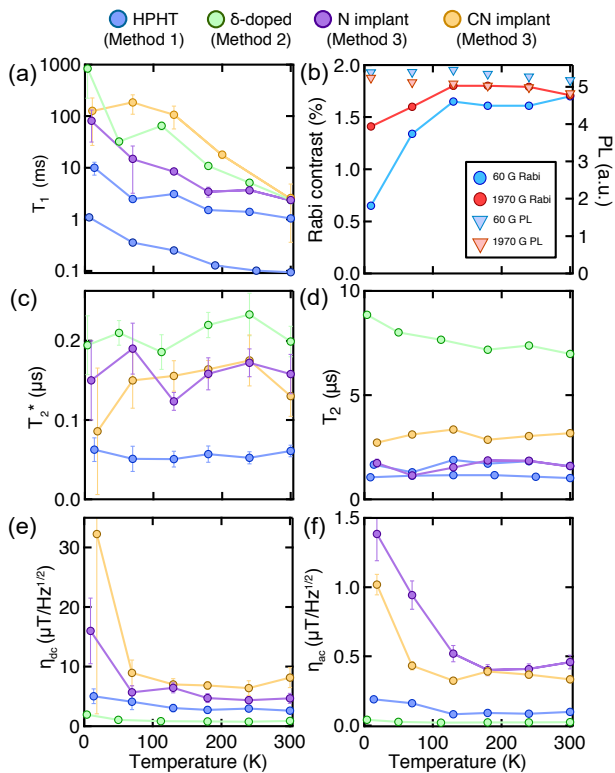


FIG. 6. **Spin property scaling with temperature:** a) T_1 , plotted on a log scale. b): Rabi oscillations taken at 4 K a field of 60 G (blue) and 1970 G (red) in sample CN-1 showing increased measurement contrast at higher fields. Points marked by triangles show PL measurements, showing reduced fluorescence at the higher field but not enough to explain the difference in contrast. c) and d) show T_2^* and T_2 variation with temperature respectively. e) and f) show the DC and AC magnetic sensitivity scaling respectively. Data presented is for samples δ -2 (green), N-2 (purple), CN-1 (yellow), and HPHT-1 (blue) in all comparative plots, with the addition of sample HPHT-3 as the bottom blue set in a) and d).

regimes [60, 61]. Though we only consider Hahn echo AC sensing here for simplicity, the slight extension at low temperatures for the sample with the least dense spin bath suggests a similar phonon component to T_2 which would be more obvious for higher order decoupling sequences.

In Figure 6(e) and (f), the measured quantities are propagated to magnetic sensitivities to DC and AC fields respectively at 60 G. There is little variation in sensitivity with temperature except for the deterioration at low temperature due to reduction in spin contrast \mathcal{C} . The absolute values are higher than in Section III A due to the lower laser power and lower collection efficiency on the cryostat system but the ordering between methods remains unchanged, meaning the conclusions drawn in the previous sections remain valid. This means that although this measurement regime does require additional attention due to non trivial factors such as the contrast loss (resulting in reduced sensitivity below 100 K), the

suitability of samples themselves is similar to the room temperature case and so the findings of this study are equally applicable in the design of low temperature experiments.

IV. DISCUSSION

We now discuss the pros and cons of each method and routes for improvement.

Method 1 is by far the most cost effective and requires only a simple implantation procedure as the HPHT substrates used are commercially available at relatively low cost. Carbon irradiation provides the required depth profile tunability and unnecessary damage to the crystal can be avoided by tuning the process's vacancy creation to the existing N density. Relying on commercial substrates does mean, however, that this method lacks control over the N density. In addition, the exact concentration is not known ahead of time as it is only loosely specified by manufacturers, and can even vary within individual diamonds as they frequently contain different growth sectors that incorporate defects at different rates. This variability between (and within) samples is present in our results as a greater spread in sensitivity within this method than the others, though further optimisation of the procedure is expected to reduce this. The N density in these samples is usually quite high (~ 100 ppm), resulting in shorter coherence times and this has typically been seen as a detraction, resulting in HPHT diamonds seeing little use for widefield NV imaging to date. However, our results show that, following implantation, a bright ensemble providing enough fluorescence to offset the short coherence times is produced. Our data shows that the best (most optimised) HPHT samples were among the highest sensitivity samples studied.

No pre-measurement of nitrogen concentration was undertaken for this work and so there is obvious room for further optimisation in the future. Additional studies into the merit of implanting with lighter ions such as helium could also improve the outlook, producing more even damage to the lattice at the cost of introducing defects related to the implant species.

Like Method 1, in situ δ -doping (Method 2) has the advantage of naturally incorporating nitrogen during growth, allowing the creation of high-yield ensembles following tuned vacancy production at a minimal damage cost. Being a custom process, δ -doping allows control over N density and the depth profile of the doped layer. These benefits are, however, difficult to realise and reproduce due to the complexity of the process and it is in general the most cost intensive of the three methods. CVD growth also carries the detraction of possible incorporation of unwanted point defects such as the NVH complex that limits the NV conversion ratio [68], as well as crystal dislocation defects that may introduce additional strain [33]. The results of Section III C, however, show that material with a high level of strain uniformity

can be produced using growth conditions designed to promote slow growth and careful selection and preparation of the growth substrate. Additionally, it appears that the implant doses useful for creating the high density ensembles that this study focusses on, and in particular the surface preparation of the diamond, play more prominent roles in decreasing strain homogeneity than the method of diamond growth.

Samples grown in the slightly lower density regime (Samples δ -3 and δ -4, with an order of magnitude lower NV density than other Method 2 samples) were observed to be susceptible to uneven nitrogen incorporation. These low density samples exhibited fluorescence striations expected to have arisen as a result of step-bunching growth patterns that see nitrogen defects preferentially incorporate at step-edges. This unevenness is undesirable for widefield imaging, especially in a modality that may seek to image via fluorescence quenching, but would be less relevant in confocal microscopy. This indicates that while in principle δ -doping is highly tunable, the complexity of the process requires careful refinement to realise this. This behaviour appears to be overridden at higher nitrogen densities as the other samples exhibited highly even fluorescence, though the as-grown surface could still contain unwanted features such as growth hillocks that are undesirable for imaging experiments.

The complexity of the CVD growth process also means that there exists a lot of room for improvement within this method, despite it already producing the highest magnetic sensitivities. In addition to the improvements identified for Method 1, we note that further optimisation of (and finer control over) other key growth parameters such as gas composition and growth temperature could see higher quality crystals with fewer unwanted defects. Confining nitrogen to a well-defined layer during growth also carries with it the advantage of being able to explore the merits of introducing vacancies via electron irradiation [41, 42] rather than the implantation of more massive particles. This possible improvement is incompatible with Method 1 as electrons cannot be confined in depth in the same way as carbon ions.

Method 3 has the advantages of giving a predictable N density and depth profile that can be easily tuned. However, this comes at the considerable cost of excess vacancy creation (about two orders of magnitude at the energies considered, as shown in Figure 1(c)) that cannot be tuned. This manifests itself in our results as reduced coherence values relative to nitrogen concentration compared to samples that have N and V concentrations roughly matched, ultimately resulting in inferior magnetic sensitivity. This limitation will be inherent to any procedure that relies on incorporating nitrogen into the crystal via implantation unless new techniques allow better vacancy removal during annealing (discussed below). This method also requires a high purity CVD substrate which are generally more costly than HPHT substrates.

Comparing the two species used to investigate this

method, our results show that implanting with the larger CN^- ion results in more damage to the crystal per ion implanted, compromising both AC and DC sensitivity. Much of the AC sensitivity can be recovered, however, by using longer dynamical decoupling sequences, indicating the effectiveness of this approach in protecting spin coherence against interactions with unwanted paramagnetic defects.

Finally, method-agnostic improvements were beyond the scope of this work but may include dynamic annealing during the implantation stage to better repair crystal damage, and surface treatments to limit surface noise [64, 69]. Optimised RIE may be useful in mitigating strain caused by polishing damage, resulting in a more homogeneous zero field splitting across a field of view to be imaged and possibly improving NV coherence by reducing surface magnetic noise [70]. Fermi level engineering via n-type doping (phosphorus, sulphur) has been shown to promote the NV^- charge state by charging vacancies and other defects which then can be annealed out to improve the NV yield [71, 72]. It is unclear whether this will have a significant effect in the higher nitrogen concentration regime wherein the majority of samples measured in this study belong, so further work is required in this area.

V. CONCLUSION

Our results show that while δ -doping during CVD growth offers the best overall sensitivity across a range of temperatures from cryogenic to room temperature, implantation of nitrogen-rich HPHT substrates is highly competitive, offering superior sensitivities to the much more expensive and currently ubiquitous approach of implanting electronic-grade CVD substrates with nitrogen. Though the HPHT approach has a significant drawback in its lack of reproducibility and precise control over NV density compared to current alternatives, its low cost and untapped potential for further improvement by refinement of the implantation process leaves it as an appealing option for future research.

Even though the CVD growth for the δ -doped samples presented here is not expected to be state-of-the-art, the best sensitivities were achieved by these samples. This indicates that, if the cost requirements are not too strenuous, δ -doping is an achievable method for the creation of high quality NV ensembles of varying density.

The common method of nitrogen implantation resulted in the lowest magnetic sensitivity of the three methods. This is due to the large amounts of damage sustained by the crystal during the high dose implantation procedures required to incorporate the high levels of nitrogen desired for ensemble NV sensing. Unlike the other two methods, implanting nitrogen leaves no scope to tune vacancy production to nitrogen density and this looms as the main limitation of Method 3.

It is clear from our results that the high nitrogen con-

tent of the HPHT samples gives sufficiently high PL following C^- implantation to offset their short coherence times. Additionally, although it offers considerably less flexibility than the other two methods, high PL HPHT samples are particularly advantageous for measurements with very low duty cycles (e.g. T_1 measurements) or very low laser power (e.g. to minimise sample heating [53]). Indeed, in these situations noise sources other than shot noise, such as dark currents in an sCMOS camera, may be dominant if the PL level is too low, deteriorating the sensitivity. The flexibility of δ -doping makes it suitable for the widest range of applications, with the lower density samples excelling in applications that allow long measurement times to reach the shot noise limit. Method 3 is shown to be suboptimal in both absolute sensitivity terms and in practicality from a cost point of view in this work, however we note that further advances in annealing procedures to repair crystal lattice damage sustained during implantation could entirely address the losses in sensitivity.

The NV layer formation methods themselves appear to have relatively little impact on the strain homogeneity of the final crystal compared to the diamond surface preparation. Further work is required to fully determine the contributions different surface preparation processes make to the final crystal strain homogeneity and thus their impact on NV ensemble quality for widefield imaging.

ACKNOWLEDGEMENTS

We acknowledge support from the Australian Research Council (ARC) through grants DE170100129, DE190100336, LP160101515, CE170100012, LE180100037 and DP190101506. We also acknowledge the AFAiR node of the NCRIS Heavy Ion Capability for access to ion-implantation facilities. This work was performed in part at the Melbourne Centre for Nanofabrication (MCN) in the Victorian Node of the Australian National Fabrication Facility (ANFF). A.J.H. and D.A.B. are supported by an Australian Government Research Training Program Scholarship. T.T. acknowledges the support of JSPS KAKENHI (No. 20H02187 and 19H02617), JST CREST (JPMJCR1773) and MEXT Q-LEAP (JPMXS0118068379).

Appendix A: Full room temperature results

In Table II we list the full data plotted in Figures 2, 3, and 4 for comparison with the parameters in Table I.

Appendix B: Data collection

In this appendix we provide greater detail on the measurements undertaken as introduced in the main text.

a. Photoluminescence (PL)

The PL rate, \mathcal{R} , was measured by averaging over an image taken under constant laser illumination over a 30 ms camera exposure and a $50 \mu\text{m} \times 50 \mu\text{m}$ area. The per-pixel average obtained was converted into a per μm^2 per second value as a figure of merit. \mathcal{R} was then converted into an approximate areal NV density by comparing the fluorescence of a representative set of samples on a confocal microscope (recorded by an avalanche photo diode) with that from a single NV (from a sample separate to this study). Multiple samples were compared to confirm an appropriate multiplicative factor to convert the widefield PL measurements to this density. Minor background fluorescence from the HPHT substrate present in the δ -doped samples was subtracted.

b. Rabi contrast

\mathcal{C} was calculated by driving the NV ensemble at a high enough microwave power for the $T_{1\rho}$ decay to be negligible over the first Rabi oscillation and fitting, with \mathcal{C} defined as the peak-to-peak amplitude. This is a measure of the maximum attainable contrast, an upper bound for the actual contrast available for a measurement optimised at T_2 or T_2^* .

As with \mathcal{R} , the impact of background fluorescence artificially decreasing \mathcal{C} was removed as a less fluorescent substrate could be used in future without affecting NV properties.

The lowest density CVD samples also exhibited PL striations arising step bunching growth behaviour. This meant the PL was variable across the field of view used and is also expected to reduce the contrast when averaging over the full region. This effect was not removed as it was observed to be a limitation inherent to growing at low nitrogen concentrations under our growth conditions.

c. Free induction decay time

T_2^* was measured using the Ramsey pulse sequence with a level of detuning of the microwave drive from the hyperfine resonances adjusted to each sample to ensure an appropriate frequency of beating to allow accurate extraction from the resulting fit. The signal was fit to a sum of cosines with frequencies given by the detuning from the respective hyperfine resonances modulated by an exponential envelope with decay constant T_2^* , $e^{-(\tau/T_2^*)} \sum_{i=1}^N \cos(\omega_i \tau + \phi_i)$, where the ω_i are detunings from the individual hyperfine resonances, ϕ_i are phases, and N is the number of hyperfine resonances, equalling 3 and 2 for ^{14}N and ^{15}N respectively. Care was taken to ensure the measurement was not compromised by magnetic field gradients across the field of view so that the measured decay envelope could accurately be attributed to the free induction decay of the ensemble alone.

Sample name	\mathcal{C} (%)	T_2^* (μs)	T_2^{HE} (μs)	T_2^{CPMG} (μs)	T_1 (ms)	η_{dc} ($\mu\text{T Hz}^{-1/2}$)	η_{ac} (nT Hz $^{-1/2}$)	η_{CPMG} (nT Hz $^{-1/2}$)
HPHT-1	3.0 ± 0.1	0.088 ± 0.020	1.04 ± 0.03	60.2 ± 0.7	0.35 ± 0.01	1.10 ± 0.31	99.5 ± 17.1	5.09 ± 0.86
HPHT-2	2.2 ± 0.1	0.024 ± 0.080	0.94 ± 0.01	63.5 ± 0.2	0.29 ± 0.01	5.83 ± 1.32	155.7 ± 23.8	7.10 ± 1.05
HPHT-3	1.5 ± 0.1	0.025 ± 0.008	0.67 ± 0.02	$4.5^* \pm 0.1$	0.02 ± 0.01	4.42 ± 1.43	174.0 ± 12.9	$32.36^* \pm 2.15$
HPHT-4	1.6 ± 0.1	0.042 ± 0.013	0.89 ± 0.01	$9.2^* \pm 0.1$	0.10 ± 0.01	4.71 ± 1.69	233.6 ± 43.5	$32.88^* \pm 6.12$
HPHT-5	1.5 ± 0.1	0.029 ± 0.010	0.49 ± 0.05	$3.5^* \pm 0.1$	0.11 ± 0.01	3.77 ± 1.46	234.7 ± 50.9	$39.07^* \pm 7.35$
HPHT-6	1.8 ± 0.1	0.016 ± 0.007	0.83 ± 0.01	58.2 ± 23.8	0.06 ± 0.01	6.05 ± 2.93	120.5 ± 16.5	5.09 ± 2.19
HPHT-7	3.0 ± 0.1	0.012 ± 0.009	1.17 ± 0.05	50.9 ± 4.0	0.14 ± 0.01	0.49 ± 0.08	54.2 ± 8.3	3.41 ± 0.57
δ -1	4.0 ± 0.1	0.347 ± 0.058	8.19 ± 0.16	285.3 ± 0.2	1.88 ± 0.10	0.52 ± 0.10	32.1 ± 3.3	4.11 ± 0.42
δ -2	3.8 ± 0.1	0.221 ± 0.098	7.49 ± 0.12	301.8 ± 10.0	1.54 ± 0.07	0.64 ± 0.30	27.1 ± 3.8	3.16 ± 0.45
δ -3	1.2 ± 0.1	1.238 ± 0.173	152.53 ± 10.78	1723.5 ± 500.0	3.04 ± 0.09	2.22 ± 0.34	81.7 ± 10.7	23.86 ± 7.06
δ -4	3.1 ± 0.1	0.742 ± 0.147	71.28 ± 6.95	700.0 ± 200.0	3.00 ± 0.10	1.04 ± 0.22	35.5 ± 4.6	10.91 ± 3.25
δ -5	4.0 ± 0.1	1.014 ± 0.111	8.79 ± 0.15	332.1 ± 50.0	1.72 ± 0.04	0.18 ± 0.03	29.1 ± 4.0	3.62 ± 0.74
N-1	2.4 ± 0.1	0.071 ± 0.009	2.46 ± 0.12	51.6 ± 5.6	0.50 ± 0.02	5.23 ± 1.08	176.2 ± 29.5	21.36 ± 4.13
N-2	2.9 ± 0.1	0.157 ± 0.009	2.27 ± 0.05	$30.0^* \pm 0.9$	0.61 ± 0.02	1.56 ± 0.27	124.8 ± 20.3	$19.28^* \pm 3.15$
N-3	2.3 ± 0.1	0.109 ± 0.011	4.20 ± 0.01	64.9 ± 15.0	1.05 ± 0.05	2.78 ± 0.49	92.9 ± 13.0	15.54 ± 4.20
N-4	2.8 ± 0.1	0.060 ± 0.007	1.36 ± 0.01	41.7 ± 4.2	0.32 ± 0.02	3.96 ± 0.82	193.0 ± 32.2	15.53 ± 3.02
N-5	3.1 ± 0.1	0.130 ± 0.012	2.67 ± 0.09	61.0 ± 4.7	0.54 ± 0.03	2.05 ± 0.34	117.9 ± 16.8	13.98 ± 2.21
CN-1	2.9 ± 0.1	0.092 ± 0.025	2.69 ± 0.11	236.8 ± 4.8	0.77 ± 0.05	4.51 ± 1.52	181.1 ± 35.7	10.60 ± 2.06
CN-2	3.3 ± 0.1	0.080 ± 0.033	2.26 ± 0.06	112.1 ± 6.5	0.66 ± 0.04	5.03 ± 1.52	204.1 ± 35.2	15.14 ± 2.73
CN-3	3.6 ± 0.1	0.123 ± 0.038	3.75 ± 0.09	129.7 ± 6.0	0.64 ± 0.04	3.33 ± 1.31	124.2 ± 23.2	13.09 ± 2.77

TABLE II. **Full spin properties:** Summary of room temperature data presented in Section III. All measurements as defined earlier, with T_2^{CPMG} being the CPMG-1024 T_2 value (or highest N CPMG sequence for samples that could not be extended this far, marked with an asterisk).

d. Spin relaxation time

T_1 was measured by initialising the NV ensemble and reading out the state after set periods of dark evolution. The signal was normalised by subtracting the PL obtained by a sequence identical but for the application of a π pulse prior to readout. This is a standard procedure in our measurements but is particularly important in the measurement of T_1 due to the wide range of wait times involved that result in a variable laser duty cycle and thus changing charge state dynamics over the course of a single sweep. The resulting decay was fit to a stretched exponential $e^{-(\tau/T_1)^k}$ with k ranging between 0.5 and 1 at room temperature.

e. Spin coherence time

The sensitivity of the samples to AC fields was measured using a Hahn spin echo sequence at a field of 475 G chosen to ensure that, for the samples containing a natural abundance of ^{13}C , the resulting revivals were closely spaced enough to still allow us to accurately extract T_2 from the decay curve. The envelope of the resulting curves were fit to an exponential decay $e^{-(\tau/T_2)^k}$ and the Hahn echo spin coherence time T_2 extracted. k varied between 0.5 and 1.5.

f. Extended dynamical decoupling

The behaviour of the samples with respect to extended decoupling was assessed using CPMG sequences of increasing length until either T_2 saturation or the total microwave duty cycle exceeded the sample's $T_{1\rho}$ (and hence no further meaningful extension could be achieved). These measurements were carried out at 475 G so as to suppress the impact of any ^{13}C revivals on our ability to fit the resulting curve and reliably extract the decay constant. Fluorescence for all measurements of a sample were normalised to the contrast obtained for the shortest decoupling sequence ($\approx \mathcal{C}$) and fit to an exponential decay of form $e^{-(\tau/T_2)^k}$, modulated by ^{13}C revival structure fit as Lorentzian dips where appropriate. No strong, consistent trend in the variation of k with N was observed across samples.

g. Strain mapping

To obtain maximum fidelity, the ODMR spectra used to obtain crystal strain maps utilised low power microwave driving to achieve T_2^* -limited linewidths. The resulting data was fit to Lorentzian lineshapes (accounting for hyperfine structure where appropriate) and each resonant frequency extracted for each pixel. Pixel size

was set to be comparable to the optical diffraction limit ≈ 300 nm. Images were saved periodically until the level of pixel-to-pixel noise reached a plateau, indicating that photon shot noise was no longer the dominant source of noise and any remaining variation could be attributed to strain inhomogeneity.

h. Low temperature measurements

The cryogenic widefield setup is functionally identical to that used to take room temperature measurements and so measurements taken on it proceeded as described above. The need to use a low laser power (to avoid sample heating) and the lower collection efficiency as well as the loss of measurement contrast at low temperatures increased the measurement uncertainty, particularly for free induction decay measurements on less fluorescent samples (ie CN^- implants). Similarly, there is a large uncertainty in the measurement of long ($\gg 10$ ms) T_1 decays as the readout duty cycle becomes insignificant next to the wait times and so camera dark currents (and their associated noise) begin to account for a large proportion of the total signal. Nevertheless, T_1 extension with decreasing temperature is clear, which is the main observation to make from this section.

Although the base temperature of the cryostat is 4 K, the conditions used to access the full range of temperatures without otherwise altering the experiment meant that the cooling power was not always sufficient to maintain this temperature under the influence of laser heating. The temperature was measured after every sweep of one camera exposure per data point and then averaged for plotting. The temperature throughout an experiment was usually mostly constant as the laser duty cycle does not change much from point to point. The laser heating effect was most significant for the Sample CN-1 T_2^* measurement as this was the closest to a CW laser duty cycle and undertaken under conditions of least cooling power. A minimal number of sweeps were used to maintain a low temperature for this measurement, which contributed to its high uncertainty.

Temperatures above the base temperature were ac-

cessed using a heater placed below the sample and fluctuations within these measurements were minimal. Room temperature measurements conducted on this system occasionally differed slightly from the original set due to differences in ensemble initialisation and microwave driving power and so data is directly comparable within the context of individual sections only.

Appendix C: Further sample details

The samples and their creation methods have been described in the main text, but here we note additional features of some of the samples used. Samples N-1, N-5, HPHT-7 and HPHT-2 had an additional coating on the NV side (80 nm Al and 75 nm of A_2O_3) though this is not expected to affect the measurements presented here. Sample N-4 was imaged face-up as it was mounted as part of a device for separate work, though this is also not thought to impact the following measurements as there is no thick substrate present to hinder fluorescence collection.

HPHT substrates used for the production of Method 2 samples first underwent a 3 hour oxygen etch at 90 Torr / 5200 W (15 sccm O_2 , 478 sccm H_2) before a two-stage N-free CVD overgrowth of 2 hours 85 Torr / 5200 W and 12 hours 85 Torr / 4500 W, each with gas flows of 20 sccm CH_4 and 450 sccm H_2 . The sample temperature was measured to be approximately 850 °C during growth.

The nominal nitrogen flow rates were 0.2, 0.02, 0 (background due to imperfect vacuum only), 10^{-4} , and 2×10^{-3} sccm for samples δ -1,2,3,4,5 respectively. With the H_2 and CH_4 flow rates held to a constant combined value of 502.5 sccm, these values correspond to approximately 400, 40, 0, 0.2, and 4 ppm N_2 respectively.

Sample δ -5 was polished and subjected to an RIE treatment prior to δ -doping. The surface preparation has previously been shown to make no significant difference to the coherence properties of dense ensembles [28], so the effects of these processes are expected to be contained within the bounds of the discussion in Section III C.

-
- [1] Doherty, M. W. *et al.* The nitrogen-vacancy colour centre in diamond. *Physics Reports* **528**, 1–45 (2013).
 - [2] Rondin, L. *et al.* Magnetometry with nitrogen-vacancy defects in diamond. *Rep. Prog. Phys.* **77**, 56503 (2014).
 - [3] Steinert, S. *et al.* High sensitivity magnetic imaging using an array of spins in diamond. *Rev. Sci. Instrum.* **81**, 043705 (2010).
 - [4] Pham, L. M. *et al.* Magnetic field imaging with nitrogen-vacancy ensembles. *New J. Phys.* **13**, 045021 (2011).
 - [5] Chipaux, M. *et al.* Magnetic imaging with an ensemble of NV centers in diamond. *Eur. Phys. J. D* **69**, 166 (2015).
 - [6] Simpson, D. A. *et al.* Magneto-optical imaging of thin magnetic films using spins in diamond. *Sci. Rep.* **6**, 22797 (2016).
 - [7] Glenn, D. R. *et al.* Single-cell magnetic imaging using a quantum diamond microscope. *Nat. Methods* **12**, 736 (2015).
 - [8] Tetienne, J.-P. *et al.* Quantum imaging of current flow in graphene. *Sci. Adv.* **3**, e1602429 (2017).
 - [9] Casola, F., Van Der Sar, T. & Yacoby, A. Probing condensed matter physics with magnetometry based on nitrogen-vacancy centres in diamond. *Nature Reviews Materials* **3**, 17088 (2018).
 - [10] Broadway, D. A. *et al.* Spatial mapping of band bending

- in semiconductor devices using in situ quantum sensors. *Nature Electronics* **1**, 502–507 (2018).
- [11] Fu, R. R. *et al.* Solar nebula magnetic fields recorded in the Semarkona meteorite. *Science* **346**, 1089–1092 (2014).
- [12] Glenn, D. R. *et al.* Micrometer-scale magnetic imaging of geological samples using a quantum diamond microscope. *Geochemistry, Geophysics, Geosystems* **18**, 3254–3267 (2017).
- [13] Le Sage, D. *et al.* Optical magnetic imaging of living cells. *Nature* **496**, 486–9 (2013).
- [14] Steinert, S. *et al.* Magnetic spin imaging under ambient conditions with sub-cellular resolution. *Nat. Commun.* **4**, 1607 (2013).
- [15] DeVience, S. J. *et al.* Nanoscale NMR spectroscopy and imaging of multiple nuclear species. *Nat. Nanotechnol.* **10**, 129 (2015).
- [16] Simpson, D. A. *et al.* Electron paramagnetic resonance microscopy using spins in diamond under ambient conditions. *Nat. Commun.* **8**, 458 (2017).
- [17] Broadway, D. A. *et al.* Imaging domain reversal in an ultrathin van der waals ferromagnet. *Advanced Materials* 2003314 (2020).
- [18] Ku, M. J. *et al.* Imaging viscous flow of the dirac fluid in graphene using a quantum spin magnetometer. *Nature* **583**, 537–541 (2020).
- [19] Tisler, J. *et al.* Single defect center scanning near-field optical microscopy on graphene. *Nano Letters* **13**, 3152–3156 (2013).
- [20] Tetienne, J.-P. *et al.* Apparent delocalization of the current density in metallic wires observed with diamond nitrogen-vacancy magnetometry. *Phys. Rev. B* **99**, 014436 (2019).
- [21] Lillie, S. E. *et al.* Imaging graphene field-effect transistors on diamond using nitrogen-vacancy microscopy. *Phys. Rev. Applied* **12**, 024018 (2019).
- [22] Tetienne, J.-P. *et al.* Proximity-induced artefacts in magnetic imaging with nitrogen-vacancy ensembles in diamond. *Sensors* **18**, 1290 (2018).
- [23] Smith, J. M., Meynell, S. A., Jayich, A. C. B. & Meijer, J. Colour centre generation in diamond for quantum technologies. *Nanophotonics* **8**, 1889–1906 (2019).
- [24] Achard, J., Jacques, V. & Tallaire, A. Chemical vapour deposition diamond single crystals with nitrogen-vacancy centres: a review of material synthesis and technology for quantum sensing applications. *J. Phys. D: Appl. Phys.* **53**, 313001 (2020).
- [25] Pezzagna, S., Rogalla, D., Wildanger, D., Meijer, J. & Zaitsev, A. Creation and nature of optical centres in diamond for single-photon emission – overview and critical remarks. *New J. Phys.* **13** (2011).
- [26] Acosta, V. M. *et al.* Diamonds with a high density of nitrogen-vacancy centers for magnetometry applications. *Physical Review B* **80**, 115202 (2009).
- [27] Deák, P., Aradi, B., Kaviani, M., Frauenheim, T. & Gali, A. Formation of NV centers in diamond: A theoretical study based on calculated transitions and migration of nitrogen and vacancy related defects. *Physical Review B - Condensed Matter and Materials Physics* **89**, 075203 (2014).
- [28] Tetienne, J.-P. *et al.* Spin properties of dense near-surface ensembles of nitrogen-vacancy centers in diamond. *Physical Review B* **97**, 085402 (2018).
- [29] Barry, J. F. *et al.* Sensitivity Optimization for NV-Diamond Magnetometry. *Reviews of Modern Physics* **92**, 15004 (2020).
- [30] Dréau, a. *et al.* Avoiding power broadening in optically detected magnetic resonance of single NV defects for enhanced dc magnetic field sensitivity. *Physical Review B - Condensed Matter and Materials Physics* **84**, 195204 (2011).
- [31] Taylor, J. M. *et al.* High-sensitivity diamond magnetometer with nanoscale resolution. *Nat. Phys.* **4**, 810–816 (2008).
- [32] Broadway, D. A. *et al.* Microscopic Imaging of the Stress Tensor in Diamond Using in Situ Quantum Sensors. *Nano Lett.* **19**, 4543–4550 (2019).
- [33] Kehayias, P. *et al.* Imaging crystal stress in diamond using ensembles of nitrogen-vacancy centers. *Phys. Rev. B - Condens. Matter Mater. Phys.* **100**, 174103 (2019).
- [34] Huang, Z. *et al.* Diamond nitrogen-vacancy centers created by scanning focused helium ion beam and annealing. *Applied Physics Letters* **103**, 081906 (2013).
- [35] McCloskey, D. *et al.* Helium ion microscope generated nitrogen-vacancy centres in type Ib diamond. *Applied Physics Letters* **104**, 031109 (2014).
- [36] Fescenko, I. *et al.* Diamond Magnetic Microscopy of Malarial Hemozoin Nanocrystals. *Phys. Rev. Appl.* **11**, 034029 (2019).
- [37] Ohno, K. *et al.* Engineering shallow spins in diamond with nitrogen delta-doping. *Applied Physics Letters* **101**, 082413 (2012).
- [38] Ohno, K. *et al.* Three-dimensional localization of spins in diamond using ¹²C implantation. *Applied Physics Letters* **105**, 052406 (2014).
- [39] Fávoro de Oliveira, F. *et al.* Toward Optimized Surface δ -Profiles of Nitrogen-Vacancy Centers Activated by Helium Irradiation in Diamond. *Nano Letters* **16**, 2228–2233 (2016).
- [40] Kleinsasser, E. E. *et al.* High density NV sensing surface created via He⁺ ion implantation of (¹²)¹³C diamond. *Appl. Phys. Lett.* **108**, 202401 (2016).
- [41] Kim, E., Acosta, V. M., Bauch, E., Budker, D. & Hemmer, P. R. Electron spin resonance shift and linewidth broadening of nitrogen-vacancy centers in diamond as a function of electron irradiation dose. *Applied Physics Letters* **101**, 082410 (2012).
- [42] McLellan, C. A. *et al.* Patterned Formation of Highly Coherent Nitrogen-Vacancy Centers Using a Focused Electron Irradiation Technique. *Nano Letters* **16**, 2450–2454 (2016).
- [43] Pezzagna, S., Naydenov, B., Jelezko, F., Wrachtrup, J. & Meijer, J. Creation efficiency of nitrogen-vacancy centres in diamond. *New Journal of Physics* **12**, 065017 (2010).
- [44] Toyli, D. M., Weis, C. D., Fuchs, G. D., Schenkel, T. & Awschalom, D. D. Chip-scale nanofabrication of single spins and spin arrays in diamond. *Nano Letters* **10**, 3168–3172 (2010).
- [45] Spinicelli, P. *et al.* Engineered arrays of nitrogen-vacancy color centers in diamond based on implantation of CN⁻ Molecules through nanoapertures. *New Journal of Physics* **13**, 025014 (2011).
- [46] Antonov, D. *et al.* Statistical investigations on nitrogen-vacancy center creation. *Applied Physics Letters* **104**, 012105 (2014).
- [47] de Oliveira, F. F. *et al.* Tailoring spin defects in diamond. *Nat. Comm.* **8**, 15409 (2017).
- [48] Bauch, E. *et al.* Decoherence of dipolar spin ensembles

- in diamond. *pre-print* arXiv:1904.08763 (2019).
- [49] Tallaire, A. *et al.* Oxygen plasma pre-treatments for high quality homoepitaxial CVD diamond deposition. *Phys. Status Solidi* **201**, 2419–2424 (2004).
- [50] Edmonds, A. M. *et al.* Production of oriented nitrogen-vacancy color centers in synthetic diamond. *Phys. Rev. B - Condens. Matter Mater. Phys.* **86**, 035201 (2012).
- [51] Teraji, T. High-quality and high-purity homoepitaxial diamond (100) film growth under high oxygen concentration condition. *J. Appl. Phys.* **118**, 115304 (2015).
- [52] McCoey, J. M. *et al.* Quantum Magnetic Imaging of Iron Biomineralization in Teeth of the Chiton *Acanthopleura hirtosa*. *Small Methods* **4**, 1900754 (2020).
- [53] Lillie, S. E. *et al.* Laser Modulation of Superconductivity in a Cryogenic Wide- field Nitrogen-Vacancy Microscope. *Nano Lett.* **20**, 1855–1861 (2020).
- [54] Bluvstein, D., Zhang, Z. & Jayich, A. C. B. Identifying and Mitigating Charge Instabilities in Shallow Diamond Nitrogen-Vacancy Centers. *Phys. Rev. Lett.* **122**, 76101 (2019).
- [55] Manson, N. B., Hedges, M., Barson, M. S. J., Ahlefeldt, R. & Doherty, M. W. NV⁻N⁺ pair centre in 1b diamond. *New J. Phys.* **20**, 113037 (2018).
- [56] Stacey, A. *et al.* Evidence for Primal sp² Defects at the Diamond Surface : Candidates for Electron Trapping and Noise Sources. *Adv. Mater. Interfaces* **6**, 1801449 (2019).
- [57] Naydenov, B. *et al.* Dynamical Decoupling of a single-electron spin at room temperature. *Phys. Rev. B* **83**, 081201 (2011).
- [58] de Lange, G., Riste, D., Dobrovitski, V. V. & Hanson, R. Single-Spin Magnetometry with Multipulse Sensing Sequences. *Phys. Rev. Lett.* **106**, 080802 (2011).
- [59] de Lange, G., Wang, Z. H., Risté, D., Dobrovitski, V. V. & Hanson, R. Universal dynamical decoupling of a single solid-state spin from a spin bath. *Science* **330**, 60–63 (2010).
- [60] Bar-Gill, N., Pham, L. M., Jarmola, A., Budker, D. & Walsworth, R. L. Solid-state electronic spin coherence time approaching one second. *Nat. Commun.* **4**, 1743 (2013).
- [61] Romach, Y. *et al.* Spectroscopy of surface-induced noise using shallow spins in diamond. *Phys. Rev. Lett.* **114**, 017601 (2015).
- [62] Friel, I. *et al.* Diamond & Related Materials Control of surface and bulk crystalline quality in single crystal diamond grown by chemical vapour deposition. *Diam. Relat. Mater.* **18**, 808–815 (2009).
- [63] Appel, P. *et al.* Fabrication of all diamond scanning probes for nanoscale magnetometry. *Rev. Sci. Instrum.* **87**, 063703 (2016).
- [64] Sangtawesin, S. *et al.* Origins of Diamond Surface Noise Probed by Correlating Single-Spin Measurements with Surface Spectroscopy. *Phys. Rev. X* **9**, 031052 (2019).
- [65] Martineau, P. M. *et al.* High crystalline quality single crystal chemical vapour deposition diamond. *J. Phys. Condens. Matter* **21**, 364205 (2009).
- [66] Maletinsky, P. *et al.* *Private communication* (2019).
- [67] Tetienne, J. P. *et al.* Magnetic-field-dependent photodynamics of single NV defects in diamond: An application to qualitative all-optical magnetic imaging. *New J. Phys.* **14**, 103033 (2012).
- [68] Stacey, A. *et al.* Depletion of nitrogen-vacancy color centers in diamond via hydrogen passivation. *Appl. Phys. Lett.* **100**, 071902 (2012).
- [69] Kucsko, G. *et al.* Critical Thermalization of a Disordered Dipolar Spin System in Diamond. *Phys. Rev. Lett.* **121**, 23601 (2018).
- [70] de Oliveira, F. F. *et al.* Effect of Low-Damage Inductively Coupled Plasma on Shallow NV Centers in Diamond. *Appl. Phys. Lett.* **107**, 073107 (2015).
- [71] Lühmann, T., John, R., Wunderlich, R., Meijer, J. & Pezzagna, S. Coulomb-driven single defect engineering for scalable qubits and spin sensors in diamond. *Nat. Commun.* **10**, 4956 (2019).
- [72] Herbschleb, E. D. *et al.* Ultra-long coherence times amongst room-temperature solid-state spins. *Nat. Commun.* **10**, 3766 (2019).