Supplementary Methods 1: Measurement of NV Ensemble Magnetometer Sensitivity using Coherently Averaged Synchronized Readout (CASR)

To measure the sensitivity of the NV ensemble magnetometer using the CASR protocol, we performed a calibration in two steps: (i) We first determined the amplitude of an oscillating magnetic field, applied with a nearby coil antenna, using known physical properties of the NV sensor. (ii) We then used that calibration to apply a weak oscillating field of known amplitude, detected it using the CASR protocol, and observed the signal-to-noise ratio (*SNR*) as a function of averaging time.

We started by varying the magnitude of an applied oscillating magnetic field, described by

$$b(t) = b_{ac}\sin(2\pi f_{coil} + \phi)$$

Here, b_{ac} is the field amplitude we wish to calibrate; f_{coil} = 3.742 MHz is the drive frequency, which is within a few kHz of the central magnetometry frequency f_0; and ϕ is a fixed, but random phase with respect to the first NV magnetometry sub-sequence. We varied the AC current supplied to the antenna, and recorded the resulting CASR time series signal for each value of the current (Extended Data Figure 1a). Because the current was not directly measured, we parameterize it here with the control voltage V_c used to set the output of the current supply. (Thus, b_{ac} is linearly proportional to V_c.)

The value of b_{ac} can be calculated directly from the CASR time series data, using the NV magnetic moment and the timing properties of the magnetometry sub-sequences. When the applied oscillating magnetic field is exactly in phase with one of the magnetometry sub-sequences, a minimum appears in the CASR fluorescence time series data, $S(t)$. We define this time as t_{min}. (The value of t_{min} is determined solely by ϕ. In Extended Data Figure 1a, $t_{min} \approx 0.53$ ms.) The CASR signal minimum $S(t_{min})$ decreases with increasing b_{ac} until reaching some absolute lowest value $S_{(\pi/2)}$, corresponding to a $\pi/2$ phase accumulation by the NVs during a single magnetometry sub-sequence, then starts to increase. The magnetic signal amplitude that produces $\pi/2$ phase accumulation is given by

$$b_{ac,(\pi/2)} = \frac{\hbar \pi^2 f_0}{2g\mu_B N}$$

where g = 2 is the Landé g-factor, μ_B the Bohr magneton, and N the number of π-pulses in the magnetometry subsequence at central frequency f_0. For XY8-6 sub-sequences (with N = 48) at f_0 = 3.742 MHz, the oscillating field amplitude required to produce a $\pi/2$ phase accumulation on the NV is $b_{ac,(\pi/2)}$ = 2.19 µT. We plot the amplitude of the CASR signal at time t_{min} against the current-supply control voltage V_c (Extended Data Figure 1b). The lowest CASR signal $S(t_{min})$ = $S_{(\pi/2)}$ occurs for V_c = 0.22 V, giving a magnetic signal calibration of **b_{ac} / V_c = 9.9 µT/V**.

We applied a 1.01 mV control voltage to generate an oscillating magnetic field of 10.0 nT and measured the signal from the ensemble magnetometer for a CASR measurement duration T = 0.96 s. The observed signal-to-noise ratio in the Fourier amplitude spectrum (Extended Data Figure 1c) was *SNR* = 320 ± 38. (Uncertainty is standard deviation of the signal peak height over 500 measurements.) This yields a magnetic field sensitivity of **η_B = 32 ± 4 pT·Hz$^{-\frac{1}{2}}$**. By recording the RMS spectral noise amplitude in the window 5.2 kHz < $(f - f_0)$ < 10.4 kHz as a function of averaging time, we confirmed that the sensitivity improves with the square root of time (Extended Data Figure 1d) for $t \leq 4.5 \times 10^4$ s, approximately equal to the longest measurements reported in the main text. We repeated this measurement using a CASR

measurement duration $T = 0.05$ s and found the same sensitivity, demonstrating that CASR magnetic field sensitivity is independent of spectral resolution. (This is in contrast to NV-NMR detection using correlation spectroscopy techniques, for which magnetic field sensitivity varies as the inverse square root of spectral resolution.[5]) We note that the observed sensitivity depends on details of fluorescence collection and delivery of the NV microwave drive in a given experiment. All experiments described in the main text were performed with sensitivity in the range $\eta_B \approx 25 - 75$ pT·Hz$^{-\frac{1}{2}}$.

Supplementary Note 2: Calculated NMR Magnetic Field Amplitude

We estimated the magnetic field produced by Larmor-precessing protons at a test point in the NV ensemble sensor by integrating over the average magnetic moment of dipoles for each point in the sample volume. Projecting the field from each sample dipole element onto the NV axis, we obtain

$$\vec{B} \cdot \hat{\mu}_{NV} = \frac{\mu_0 \, \rho_{th}}{4\pi} \int \frac{1}{r^3} \, [3(\hat{r} \cdot \vec{m}(t))(\hat{r} \cdot \hat{\mu}_{NV}) - (\vec{m}(t) \cdot \hat{\mu}_{NV}) \,] dV$$

Here ρ_{th} is the equilibrium density of thermally-polarized protons in the sample, $\hat{r}$ is a unit vector along the direction connecting the test point and the infinitesimal sample volume, $\hat{\mu}_{NV}$ is a unit vector parallel to the NV axis, and $\vec{m}(t)$ is the time-varying magnetic moment vector of the precessing proton spin. We can divide the proton magnetic moment into a direction $\hat{m}$ and magnitude γ_p, and pull the latter out of the integral such that all dimensions are collected in a prefactor:

$$\vec{B} \cdot \hat{\mu}_{NV} = C \int \frac{1}{r^3} \, [3(\hat{r} \cdot \hat{m}(t))(\hat{r} \cdot \hat{\mu}_{NV}) - (\hat{m}(t) \cdot \hat{\mu}_{NV}) \,] dV$$

The prefactor that determines the approximate scale of the field produced by the sample is

$$C = \frac{\mu_0 \, \gamma_p}{4\pi} \rho_{Tot} \left[1 - \exp\left(\frac{-2\gamma_p \, B_0}{k_B T} \right) \right] \approx \frac{\mu_0 \, \gamma_p}{4\pi} \rho_{Tot} \left[\frac{2\gamma_p \, B_0}{k_B T} \right]$$

where ρ_{Tot} is the full density of protons in the sample (taken here to be pure water), k_B is the Boltzmann constant, T is the temperature, and B_0 is the magnitude of the static magnetic bias field. Numerically, we obtain

$$C = \frac{\mu_0 \, \gamma_p{}^2 \rho_{Tot} B_0}{2\pi \, k_B T} = \frac{[4\pi \times 10^{-7} \frac{T \cdot m}{A}][1.41 \times 10^{-26} A \cdot m^2]^2 [6.6 \times 10^{28} m^{-3}][0.0882 \, T]}{2\pi \left[1.38 \times 10^{-23} \frac{T \cdot A \cdot m^2}{K} \right] [300 \, K]}$$

$$= 56 \text{ pT}$$

Thus, we expect the NMR magnetic field generated by the protons at the NVs to be on the order of ~50 pT, assuming that the volume integral $G = \int \frac{dV}{r^3} [3(\hat{r} \cdot \hat{m}(t))(\hat{r} \cdot \hat{\mu}_{NV}) - (\hat{m}(t) \cdot \hat{\mu}_{NV})]$ evaluates to something of order ~1 for our sample geometry.

Next we evaluated the dimensionless integral. The z-axis is taken to be perpendicular to the diamond surface. We define $\hat{\mu}_{NV} = (\sqrt{\frac{2}{3}}, 0, \frac{1}{\sqrt{3}})$, as well as two perpendicular unit vectors, $\hat{q}_1 = (\frac{1}{\sqrt{3}}, 0, -\sqrt{\frac{2}{3}})$ and $\hat{q}_2 = (\, 0, 1, 0)$, that describe the plane in which the nuclear spins precess when the bias field $\vec{B}_0$ is aligned

parallel to the NV axis. Thus, the proton magnetization vector is $\hat{m}(t) = \hat{q}_1 \cos(\omega t) + \hat{q}_2 \sin(\omega t)$. By symmetry, the integral G vanishes for the component of $\hat{m}$ along $\hat{q}_2$, and the component along $\hat{q}_1$ determines the amplitude of the oscillating magnetic field at the sensor. Because G is dimensionless, it is scale-independent, in the sense that for a given shape of the sample volume V, the value of G depends only on the ratio $V^{1/3}/d_{NV}$, where d_{NV} is the depth of the NV test point below the diamond surface. Thus, for a sufficiently large and homogeneous sample volume ($V \gg t^3$, for t the NV layer thickness), every NV center in the sensor feels approximately the same magnetic field from the precessing sample spins.

We evaluated G as a function of $V^{1/3}/d_{NV}$ for hemispherical and cube-shaped volumes. Empirically, we found that the asymptotic value (as $V^{1/3}/d_{NV} \to \infty$) of G is between 1 and 4. (The hemispherical volume calculation is plotted in the inset of Fig. 1a.) It tends to be largest for sample volumes of high aspect ratio (long in z, narrow in x and y), which may be of practical importance for sample-holder design in future NV NMR detectors. (For a hemispherical volume, the asymptotic value of G is ~1.4, giving a predicted magnetic signal amplitude of **81 pT**. The measured value reported in Fig. 2a of the main text was 95 ± 8 pT.) For all sample volume shapes we calculated, G reaches half its asymptotic value for $V^{1/3}/d_{NV}$ in the range of $2 - 3$.

Finally, we compare the magnetic noise due to statistically fluctuating sample spin polarization to the mean magnetic signal we just calculated. Following the calculations in Ref. 12 of the main text, we obtain the variance of the magnetic field fluctuations at a test point of depth d_{NV} below the diamond surface due to a semi-infinite proton sample volume:

$$\Delta B^2 = \frac{\rho_{Tot}\, \mu_0{}^2\, \gamma_p{}^2}{96\, \pi\, d_{NV}{}^3}$$

We want to find the condition on d_{NV} to ensure $\vec{B} \cdot \hat{\mu}_{NV} \geq \sqrt{\Delta B^2}$. Substituting from the expressions above, we find this is equivalent to

$$d_{NV} \geq \left(\frac{k_B\, T}{\gamma_p B_0\, G} \right)^{2/3} \left(\frac{1}{\rho_{Tot}} \right)^{1/3}$$

Numerically, for $B_0 = 88$ mT and $G \sim 1.5$ this evaluates to $d_{NV} > 3$ μm. Thus, the shallowest NV centers in our sensor (in the top 3 μm of the sensing volume) are primarily sensitive to proton spin fluctuations, while the deeper NVs give a signal that is mostly dependent on the thermal proton polarization.

Supplementary Note 3: NMR Measurement Volume

To estimate the effective measurement volume of the NV ensemble sensor, we first determined the volume of NV centers comprising the sensor by imaging their fluorescence on a CMOS camera (Extended Data Figure 3a). Because of geometric constraints imposed by the electromagnet, the fluorescence was detected in a plane with normal 45° to the diamond surface; we therefore numerically stretched the fluorescence image by a factor of $2^{1/2}$ in the x direction to recover the image of the laser spot in the plane of the diamond surface. Gaussian fits to the resulting image then yielded a spot size of 27 μm FWHM in x and 20 μm FWHM in y (Extended Data Figures 3b,c). Thus, for numerical estimation of the measurement

volume, we approximated the sensor as a uniform elliptical cylinder of NV centers, with semi-axes of 14 μm in x and 10 μm in y, and a height of 13 μm in z due to the NV layer thickness.

The total fluorescence signal $S(t)$ produced by the sensor is linearly proportional to the sum of the time-varying magnetic field projections along the quantization axes of each NV center in the sensor. We write this as follows:

$$S(t) \propto \int \int \frac{1}{\|\vec{r}_{NV} - \vec{r}_p\|^5} \left[3\left((\vec{r}_{NV} - \vec{r}_p) \cdot \widehat{m}(t)\right)\left((\vec{r}_{NV} - \vec{r}_p) \cdot \hat{\mu}_{NV}\right) \right. \\ \left. - \|\vec{r}_{NV} - \vec{r}_p\|^2 (\widehat{m}(t) \cdot \hat{\mu}_{NV}) \right] dV_{NV} \, dV_p$$

This expression is identical to the first line of the calculation of Supplementary Note 2, except that we have discarded the dimensional constants and integrated over all NV centers in the sensor volume V_{NV}. The measurement volume is denoted V_p.

We evaluated the six-dimensional integral using Monte-Carlo (MC) methods. A sample of 10^3 NV centers were randomly positioned in the model sensor volume, and for each NV center, 10^5 sample protons were randomly positioned in a $(200 \, \mu m)^3$ box above the sensor (Extended Data Figure 3d). (An artificial gap of 1 μm was placed between the diamond surface and the top of the NV layer, to ensure numerical stability given finite MC sample density. This undercounts strong interactions between NV centers and protons near the diamond surface, and therefore causes us to slightly overestimate the measurement volume corresponding to half the total NMR signal.) The dipole-dipole interaction was calculated for each NV-proton pair using the proton magnetic dipole direction $\widehat{m} = \hat{q}_1 = (\frac{1}{\sqrt{3}}, 0, -\sqrt{\frac{2}{3}})$, and recorded at the position of the proton. The interactions were then summed over increasingly large sample volumes V_p, normalized by the asymptotic value $S(V_p \to \infty)$ and plotted (Extended Data Figure 3e). (The asymptotic value was estimated using V = 1 nL.) The full MC integration was repeated ten times, yielding numerical error bars shown in the figure. This procedure was carried out for cube-shaped, hemispherical, and half-ellipsoidal (semi-axis ratio 1:2:1 in x,y and z) sample volumes. In each case, the sample volume at which the integrated signal reached half of the asymptotic value was less than or equal to 10 pL. We therefore estimate the effective measurement volume to be **10 pL**.

Supplementary Note 4: Classical NV Back-Action Calculation

We consider a sensing volume of NV centers at the diamond surface of approximately $(10 \, \mu m)^3$, with polarized density $[NV]_{pol} \approx 0.8 \times 10^{17} cm^{-3}$. (This is equal to one fourth of the estimated total NV density in our sensor, since only NV centers aligned parallel to B_0 contribute to sensor back-action.) During the magnetometry sequence, NV centers are in an equal superposition of $|m_s=0\rangle$ and $|m_s=-1\rangle$ states, and thus have an average magnetization of -0.5 μ_B along the NV axis. The NVs are aligned with the external bias field, so the effective NV magnetization direction is parallel to the proton spin quantization axis. The NV magnetization will produce a field that is strongest near the diamond surface and falls off with a length scale on the order of ~1 – 10 μm. This spatial inhomogeneity may broaden the effective NMR spectral width of the protons being sensed. Furthermore, for synchronized readout measurements, each NV optical readout repolarizes the NV centers to $|m_s = 0\rangle$, temporarily turning off the NV-induced field for

several μs and resulting in a (spatially inhomogeneous) phase jump in the proton spin precession at each readout. We therefore made a numerical estimate for the strength of the magnetic field gradient produced by the NVs in the proton volume when they are in the superposition state.

The magnetic field at position $\vec{r}$ in the proton sample volume, produced by NV spins of density ρ_{NV}, is given by

$$\vec{B}(\vec{r}) = \frac{\mu_0 \rho_{NV}}{4\,\pi} \int \frac{1}{r^3} [3(\hat{r} \cdot \vec{m}_{NV})\hat{r} - \vec{m}_{NV}]dV$$

where $\vec{m}_{NV} = \frac{-1}{2}g\,\mu_B \hat{u}_{NV}$ is the NV magnetic moment along the NV axis $\hat{u}_{NV}$, for g = 2, μ_B the Bohr magneton, and μ_0 is the permeability of free space. (We assume the NV centers are in an equal superposition of $|m_s = 0\rangle$ and $|m_s = -1\rangle$.) If we consider the projection of this field along the proton quantization axis, which is parallel to $\hat{u}_{NV}$, we obtain:

$$\vec{B} \cdot \hat{u}_{NV} = \left[\frac{-\mu_0\,\mu_B}{4\,\pi}\rho_{NV}\right]\int \frac{1}{r^3}[3(\hat{r} \cdot \hat{u}_{NV})^2 - 1]dV$$

The integral contributes a dimensionless geometric factor that depends only on the probe position $\vec{r}$ relative to the NV volume. The dimensional pre-factor sets the scale of magnetic field produced in the proton volume:

$$\left[\frac{\mu_0\,\mu_B}{4\,\pi}\rho_{NV}\right] = \frac{1}{4\pi}[(4\pi \times 10^{-7}\text{ T m/A})(9.3 \times 10^{-24}\text{ J/T})(0.8 \times 10^{23}\text{ m}^{-3})] = \ 74\text{ nT}$$

Thus, the expected scale for the magnetic field inhomogeneity due to back-action is on the order of ~**74 nT**, or approximately 0.8 ppm of our bias magnetic field $B_0 \sim$ 88 mT.

We next computed the geometric integral as a function of space in the proton volume. We take $\hat{u}_{NV} = \sqrt{2/3}\,\hat{x} + \sqrt{1/3}\,\hat{z}$, with the z-axis perpendicular to the diamond surface. We approximate the NV magnetization as two dimensional Gaussian in x and y (with FWHM widths 27 μm and 20 μm respectively) to represent the laser-intensity dependent NV polarization, and a step function in z (such that it is nonzero only between z = -13 μm and 0 μm) to represent the finite extent of the NV layer below the diamond surface. Under these assumptions, we calculated x-y maps of the geometric integral factor in several planes above the diamond surface, i.e., inside the proton volume (Extended Data Figure 4).

Even at a distance of only ~2 μm above the diamond surface, the maximum range of the geometric factor is approximately between ±1, corresponding to a ~±74 nT shift in the B_0 field seen by the protons. This might lead to noticeable broadening for high-resolution spectroscopy on the < 1 ppm level, but only from the small fraction of signal protons that are within ~2 μm of the diamond. Over the full detection volume (roughly a hemisphere of radius 20 – 30 μm above the sensor), the mean value of geometric factor is < 0.1, and the expected broadening is on the order of 10 – 100 ppb. Thus, we conclude that while NV back-action might result in minor line-broadening for a small fraction of the signal, it is unlikely to be the dominant factor for linewidths > 1 ppm. (For B_0 = 88 mT, 1 ppm of the proton Larmor frequency is about 4 Hz).

Supplementary Methods 5: Incoherently Averaged Synchronized Readout with a Nanoscale Sensor

While the present work has focused on using CASR to interrogate a ~10 μm-scale NV ensemble detector, the general synchronized readout (SR) protocol is applicable to sensors of any size, and does not require coherent averaging. In particular, synchronized readout may be applied to signals with random phase, providing spectral resolution proportional to the inverse correlation time of the signal τ_c^{-1}, by incoherent averaging using periodogram techniques [S1].

To illustrate the application of the technique to a nanoscale sensor and random-phase signal, we performed SR measurements of a single NV center, imaged in a confocal microscope. (See below for technical details of the sensor.) A coil antenna was excited continuously (without gating or triggering of the sources) to produce a magnetic signal consisting of three closely-spaced frequencies around 3.7325 MHz; and the SR protocol was performed at a central frequency f_0 = 3.7313 MHz. Due to finite optical collection efficiency, each SR fluorescence readout detected a mean of only ~0.03 photons. We therefore repeated the SR protocol 100 times, and constructed a Bartlett periodogram from the full data set. In the resulting power spectrum (Extended Data Figure 5a), the three signal peaks were clearly distinguishable, and the spectral resolution was 5.2 mHz (FWHM). The averages in this experiment were *incoherent*, in that the phase of the detected signal relative to the start of the first magnetometry sub-sequence was randomized every average. Such an incoherent averaging procedure could, in principle, be used for noise spectroscopy of a statistically-polarized nuclear spin ensemble, with the synchronized readout measurement duration matched to the noise correlation time $n\ \tau_{SR} \approx \tau_c$.

Unfortunately, for nano-NMR applications using single-NV synchronized readout, the requirement of weak sample-sensor coupling, combined with imperfect spin state readout of single-NV experiments [S2], presents a significant technical challenge. In particular, weak coupling implies small NV phase accumulation at each magnetometry pulse sequence. However, the inherently statistical nano-NMR signal will be averaged incoherently over many repetitions of the SR protocol, and is necessarily added in quadrature with large single-NV readout noise (due to Poissonian fluctuations in the number of detected photons) in the SR power spectrum, making the signal difficult to detect. Thus, techniques such as correlation spectroscopy that are optimized for strong NV interactions with randomly-fluctuating sample spins may prove superior to SR at the nanoscale, particularly in view of recent demonstrations of diamond-intrinsic ^{13}C detection with linewidths approaching ~1 Hz [S3, S4]. Irrespective of the detection mechanism, the challenge of short signal correlation times from nanoscale liquid-state NMR samples due to molecular diffusion is not yet solved, limiting the advantages of a spectrally-selective nanoscale sensor. Until translational diffusion can be reliably restricted at these length scales (e.g., by gel media [S5] or nanofabricated encapsulation chambers [S6]) without increased dipolar broadening, SR techniques will likely be of greatest utility in larger (e.g., micron-scale) volumes.

Single-NV magnetic sensor: The single-NV sensor was based on a low NV density diamond chip. The diamond was a 4 mm x 4 mm x 0.5 mm high-purity CVD diamond chip, with 99.99% ^{12}C isotopic purity near the surface, which contained preferentially oriented NV centers with nitrogen concentration [^{14}N] $\approx$ 1×10^{15} cm^{-3} and NV concentration [NV] $\approx 3 \times 10^{12}$ cm^{-3}. The approximate coherence times for the single NV center used in our experiments were $T_1 \approx 1$ ms, $T_2 \approx 500$ μs, and $T_2^* \approx 50$ μs.

Single-NV optical measurements were performed with a home-built scanning laser microscopy system. Confocal scanning of the diamond chip was done by a three-axis motorized stage (Micos GmbH). A 400 mW, 532 nm diode-pumped solid state laser (Changchun New Industries) provided excitation light, and an acousto-optic modulator (Isomet Corporation) operated at 80 MHz was used to time-gate the laser. An oil-immersion objective (100x, 1.3 NA, Nikon CFI Plan Fluor) focused the green laser pulses onto a single NV center. NV red fluorescence was collected by a silicon avalanche photodetector (Perkin Elmer SPCM-ARQH-12) through a 75 μm sized pinhole. NV spin initialization and readout pulses were 3 μs and 0.5 μs, respectively. Microwave pulses were applied to the NV center through nanofabricated 20 μm gapped waveguide, and the pulse sequence was generated using a GHz range signal source (Agilent E4428C), which was I/Q modulated by an AWG (Tektronix 5014c) for the microwave phase control.

The antenna-generated magnetic signals were measured via a SR protocol using CPMG-32 magnetometry sub-sequences; the cycle period was τ_{SR} = 75.04 μs; and the total experiment duration was $T = n\,\tau_{SR}$ = 112.5 s, for $n = 1.5 \times 10^6$ the number of readouts. Apart from using incoherent averaging techniques, the structure of the pulse sequence was identical to the CASR technique described in the main text and methods. We tuned the antenna-generated signal strength such that magnetic field amplitude at the NV sensor was ~3 μT, corresponding to maximum fluorescence contrast for a single CPMG-32 sequence.

Supplementary Methods 6: Spatial Gradients in B_0

We characterized spatial inhomogeneity in the magnetic bias field B_0 by translating the excitation laser spot across the surface of the diamond chip, and performing pulsed NV ensemble ESR measurements at each position. The beam spot was moved by adjusting the position of a lens immediately upstream of the diamond chip, using micrometer-controlled stages. The ESR measurements were performed on both NV electronic spin transitions to account for temperature drifts, and were carried out while the mean B_0 field was held approximately constant using the secondary NV-diamond magnetometer. We defined the (horizontal) translation direction with finite projection on the cylindrical axis of the electromagnet poles as u; and the (vertical) translation direction perpendicular to the cylindrical axis of the poles as v.

The resulting magnetic field profiles (Extended Data Figure 6) showed magnetic gradients of approximately 5 μT/mm in u and 10 μT/mm in v. The displayed error bars are the standard deviation of three measurements recorded at each position during repeated position sweeps, where each sweep required ~20 minutes to complete. The spread of magnetic field values recorded at each point was likely due to slow fluctuations in the mean value of B_0, caused by imperfect stabilization of the electromagnet current. (These fluctuations are routinely corrected during actual NMR measurements by periodically performing ESR measurements at the fixed NV-diamond sensor position and using the result to adjust the setpoint of the fast feedback loop on the secondary diamond magnetometer. The gradient measurement was incompatible with these slow ESR measurements, thus preventing setpoint corrections for the fast feedback loop and resulting in slow drifts over time.) The measured gradients were reproducible over repeated scans across the same range of positions, but changed significantly whenever the sample holder was disassembled and reconstructed, e.g., for cleaning. Nevertheless, the approximate scale of gradients shown in Extended Data Figure 6 (i.e. ~5 − 20 μT/mm) was observed repeatedly over many weeks of operation. We attribute the gradients to local, differential magnetic susceptibility effects produced by

various sensor components including the diamond itself, the microwave delivery wire, and a glass plate used to hold the wire in place.

While we had no means to directly measure the magnetic field gradient in the direction perpendicular to the diamond (which we denote z), it is possible to estimate the contribution to dB/dz arising from susceptibility mismatch between diamond and the sample.[19] We define a cylindrical coordinate system, with radius r and angle ϕ in the plane of the interface, and positive z denoting distance from the interface into the sample. The interface is taken to be circular with maximum radius $R = 0.5$ mm, equal to half the diameter of the top surface of the diamond chip. The magnetic field contribution due to the susceptibility mismatch is given as a function of distance from the diamond surface by $\mathbf{B}(r = 0, z) = -\mu_0 \nabla U$, for μ_0 the permeability of free space, and U a magnetic scalar potential. The scalar potential near the interface may be calculated as follows:

$$U(\vec{r}) = \frac{H_0}{4\pi} \iint \frac{(\hat{n} \cdot \hat{h}_0)(\chi_2 - \chi_1)}{\sqrt{|\vec{r} - \vec{r}'|^2}} \, dS'$$

Here, $\overrightarrow{H_0} = H_0 \hat{h}_0$ is the externally applied magnetic field, $\hat{n} = \hat{z}$ is a unit vector normal to the interface, and the integration is carried out over the plane of the interface. Evaluating in the cylindrical coordinate system gives

$$U(r = 0, z) = \frac{H_0}{4\pi} (\hat{n} \cdot \hat{h}_0)(\chi_2 - \chi_1) \, 2\pi \int_0^R \frac{r' \, dr'}{\sqrt{r'^2 + z^2}}$$
$$= \frac{H_0}{2} \left(\frac{1}{\sqrt{3}}\right)(\chi_2 - \chi_1)\left(\sqrt{R^2 + z^2} - z\right)$$

Here, we have assumed that the external field is aligned parallel to one of the [111] directions in the diamond, corresponding to $\hat{n} \cdot \hat{h}_0 = 1/\sqrt{3}$. Now we can evaluate the z-component of the magnetic field due to the susceptibility mismatch:

$$B_z(z) = -\mu_0 \frac{\partial U}{\partial z}$$
$$= \frac{-\mu_0 H_0}{2} \left(\frac{1}{\sqrt{3}}\right)(\chi_2 - \chi_1)\left(\frac{z}{\sqrt{R^2 + z^2}} - 1\right)$$
$$\approx \frac{B_0}{2\sqrt{3}}(\chi_2 - \chi_1)\left[1 - \frac{z}{R} + \frac{1}{2}\left(\frac{z}{R}\right)^3\right]$$

The last line is obtained by series expansion in z/R. We can now evaluate at the nearest ($z = 0$) and furthest ($z = \kappa \, d_{NV}^{(max)}$) points of the sensing volume, using susceptibilities for water $\chi_2 = -9.0 \times 10^{-6}$, and diamond $\chi_1 = -2.2 \times 10^{-5}$. Then $B_z(z = 31 \text{ μm}) - B_z(0) = (-2.3 \times 10^{-7}) B_0$. This corresponds to a line width of approximately 0.9 Hz (or ~0.2 ppm of the proton Larmor frequency), and is therefore unlikely to be the dominant source of spectral broadening in the CASR signal.

An in-plane gradient of 10 μT/mm (Extended Data Figure 6b) across the ~20 μm diameter of the proton NMR measurement volume would result in a spectral width of ~8.5 Hz for NMR detection (or ~2.3 ppm). This value is consistent with our observed FNP linewidths of 8 − 12 Hz, and also consistent with the observation of a reduction in the linewidth (to ~2.8 Hz) when spin echo measurements were performed on a water sample (Figure 2c). We therefore conclude that spatial B_0 inhomogeneity due to local

susceptibilities are the dominant form of line broadening in the present experiment. In the future, these effects may be reduced and/or made more repeatable by using a fixed co-planar waveguide geometry for microwave delivery, and microfluidics with appropriate susceptibility-matching design to contain the NMR sample.

Supplementary Data 7: Modified Pulse Sequences to Investigate Sensor-Induced Line Broadening

We experimentally investigated two effects that could potentially lead to broadening of the NMR spectral width due to interaction with the NV sensor:

(i)   The microwave field used to drive the NV spins could interact off-resonantly with the proton spins, resulting in an AC-Zeeman shift of the proton energy levels whenever the microwave field is applied. The approximate magnitude of this effect would be $\Delta f \sim \frac{f_\Omega{}^2}{\Delta} \left(\frac{\gamma_p}{\gamma_{NV}}\right)^2 \sim 1$ Hz, for our maximum NV Rabi frequency $f_\Omega \approx 15$ MHz, detuning $\Delta \approx 400$ MHz, proton gyromagnetic ratio $\gamma_p =$ 42.58 MHz/T, and NV gyromagnetic ratio $\gamma_{NV} =$ 28.02 GHz/T. To test for this effect, we decreased f_Ω by up to a factor of 3, with the duration τ of each pulse in the NV magnetometry pulse sequence increased proportionally to compensate. Because the expected NMR signal phase shift for each pulse is $\Delta\phi = \tau \, \Delta f$, any line broadening should be linearly proportional to f_Ω.

(ii)   During the magnetometry pulse sequence, the NV centers are in a superposition of $|m_s = 0\rangle$ and $|m_s = -1\rangle$ states, resulting in a net magnetization along the NV quantization axis and thus a magnetic field gradient outside the diamond, which could produce inhomogeneous broadening of the proton spectra. Numerical calculations (Supplementary Note 4) suggest this effect should only be relevant for a small fraction of the NMR sample, and should result in broadening of those spins on the order of $\sim$1 $-$ 2 Hz. To verify this estimate, we used NV magnetometry sequences (i.e., XY8-2, XY8-4, XY8-6) of varying duration τ_{seq}, while keeping the synchronized readout repetition period τ_{SR} = 24.06 μs fixed. If the gradient produced by the NV magnetization contributed significantly to the proton linewidth, the amount of broadening should be proportional to the fraction of time τ_{seq}/τ_{SR} that the NV centers were in the superposition state.

We carried out these experiments on a sample of pure water, which should have NMR FNP linewidth < 1 Hz in the absence of field gradients or other perturbations. Results are summarized in the following tables:

NV Rabi Frequency f_Ω	Measured FNP Linewidth Γ
16.6 MHz	10 ± 1 Hz
8.3 MHz	8 ± 1 Hz
5.6 MHz	7 ± 1 Hz

Magnetometry Sequence	τ_{seq}/τ_{SR}	Measured FNP Linewidth Γ
XY8-6	0.53	9 ± 1 Hz
XY8-4	0.36	8 ± 1 Hz
XY8-2	0.18	10 ± 1 Hz

Because no significant variation in the measured water NMR FNP linewidths was observed in either experiment, we conclude that interaction with the NV ensemble sensor is not the primary source of proton spin dephasing in our system.

Supplementary Note 8: Calculated Ethyl Formate NMR Spectrum

In the bias magnetic field used for our experiments (B_0 = 88 mT), the NMR spectrum of the ethyl formate molecule contains signatures of multiple chemical shifts and J-couplings of comparable magnitude ($\Delta f \approx$ 1 – 25 Hz). This makes it more complex than the spectra of the other molecules measured in this work (trimethyl phosphate and xylene), necessitating comparison with numerical calculations (see Fig. 3). Ethyl formate contains three groups of protons, which we label with indices corresponding to the number of protons in each group. Thus, the isolated proton has chemical shift $\Delta f_{CS,1}$, the H_2 group has chemical shift $\Delta f_{CS,2}$, and the H_3 has chemical shift $\Delta f_{CS,3}$. Similarly, the J-coupling between the isolated proton H and the protons in H_2 is denoted $\Delta f_{J1,2}$ and that between protons in the H_2 and H_3 groups is $\Delta f_{J2,3}$.

We first obtained high-field NMR spectra (proton reference frequency f_{Ref} = 500 MHz, corresponding to $B_0 \approx 11.7$ T) using a sample volume of ~1 mL to determine the molecular parameters for our calculation (Extended Data Figure 7b). In the high-field system, the chemical shift groups were clearly resolved, and each was split into multiple peaks due to the J-couplings. We fit sums of Lorentzian lineshapes to each group of peaks to extract the molecular parameter values. The line centers for H_3, H_2, and H were at offset frequencies of 1.012 kHz, 2.474 kHz, and 4.398 kHz, respectively, corresponding to chemical shifts of 2.0 ppm, 4.9 ppm, and 8.8 ppm. At low field (B_0 = 88 mT), the chemical shift parameters are therefore $\Delta f_{CS,3}$ = 7.5 Hz, $\Delta f_{CS,2}$ = 18.4 Hz, and $\Delta f_{CS,1}$ = 33.0 Hz. From the line splittings of the H_3 and H_2 groups, we obtained $\Delta f_{J2,3}$ = 8.5 Hz. The expected triplet splitting of the peak corresponding to the isolated H was unresolved, although the observed peak was broadened; we therefore used the largest J-coupling parameter consistent with the observed linewidth, $\Delta f_{J1,2}$ = 3.0 Hz.

Using the molecular parameters obtained from high-field NMR, we calculated the expected low-field spectrum with the SpinDynamica software package [S7]. The Hamiltonian describing the system was as follows:

$$\hat{H}/h = \Delta f_{CS,1}\,\hat{I}_{z1} + \Delta f_{CS,2}(\hat{I}_{z2} + \hat{I}_{z3}) + \Delta f_{CS,2}\,(\hat{I}_{z4} + \hat{I}_{z5} + \hat{I}_{z6}) + \Delta f_{J1,2}\hat{\boldsymbol{I}}_1 \cdot (\hat{\boldsymbol{I}}_2 + \hat{\boldsymbol{I}}_3)$$
$$+ \Delta f_{J2,3}(\hat{\boldsymbol{I}}_2 + \hat{\boldsymbol{I}}_3) \cdot (\hat{\boldsymbol{I}}_4 + \hat{\boldsymbol{I}}_5 + \hat{\boldsymbol{I}}_6)$$

Here, h is Planck's constant, and the individual nuclear spin operators are I_1 for the isolated proton, $I_2 - I_3$ for the protons in the H_2 group, and $I_4 - I_6$ for the protons in the H_3 group. The free nuclear precession (FNP) spectrum was calculated without secular approximations, then numerically broadened to match the experimental linewidth. The calculated FNP spectrum shows good qualitative agreement with our measurement (Fig. 3c and Extended Data Figure 7a). A second measurement was performed (Extended Data Figure 7a) under the same conditions as the original experiment, to verify the repeatability of the observed spectral features.

Supplementary Note 9: Estimated Sensitivity of a Bulk NV Ensemble Sensor

We expect that the CASR technique may also be applicable to larger NV ensemble sensors, providing increased sensitivity for chemical analysis of concentration-limited samples at the mm^3 scale. We therefore estimated the expected sensitivity achievable for a bulk NV ensemble sensor.

The sensor described in the present work has a volume $V_{NV} \approx 6 \times 10^3$ μm^3 defined by the overlap of the excitation laser and NV layer; and operates at a bias magnetic field B_0 = 88 mT. The detected photocurrent due to NV fluorescence at each readout is $I_{det} \approx 30$ µA. With this sensor, we are empirically able to detect an NMR signal (FNP) from a semi-infinite sample of pure glycerol with a signal to noise ratio (SNR) = 3 in $T_{avg} \approx 10$ s of signal averaging:

$$T_{avg} = \left[\frac{\eta_B \ (SNR)}{B_{sig}} \right]^2 \frac{T_{rep}}{T_2^*} \ = \ 11.3 \text{ s}$$

Here, B_{sig} = 95 pT is the measured signal from a semi-infinite sample of glycerol, η_B = 50 pT·Hz$^{-1/2}$ is the typical sensitivity of the ensemble NV sensor; T_2^* = 11 ms is the observed proton spin dephasing time for glycerol in our system; and T_{rep} = 50 ms $\approx 2 \ T_1$ is the experiment repetition time, which is approximately twice the spin relaxation lifetime of glycerol. The concentration of protons in pure glycerol is 110 M. Therefore, in 10 minutes of averaging, the present sensor can detect a sample concentration of C_{samp} = 15.1 M with SNR = 3.

We note that this is equivalent to a limit of detection of **3.7 nmol·Hz$^{-1/2}$** with SNR = 3, or a proton sensitivity of 2.5×10^{14} spins with SNR = 1 in one second of integration. The sensitivity scales with the inverse square of the nuclear magnetic moment, yielding projected sensitivities of, e.g., 3.9×10^{15} Hz$^{-1/2}$ for ^{13}C, and 1.5×10^{15} Hz$^{-1/2}$ for ^{31}P nuclei.

For applications in chemical analysis, it is desirable to scale the sample region up to ~1 µL to accommodate analyte volumes that can be practically handled in the lab. We therefore consider a sensor consisting of a (200 µm)^3 NV ensemble volume (with similar NV density and coherence times to the present sensor), operating at B_0 = 3 T. By employing a light-trapping diamond waveguide geometry [S8], it should be feasible to obtain an increase in detected photocurrent proportional to the sensor volume (i.e., by a factor of ~10^3) to $I_{det} \approx 30$ mA, with only modest increase in the optical excitation intensity (to I_{exc} = 1 W). The increased photocurrent would provide a ~30-fold decrease in photon shot noise, yielding a projected magnetic field sensitivity of $\eta_B \approx 2$ pT·Hz$^{-\frac{1}{2}}$. (While even larger fluorescence signals may be achievable, photon shot noise will at some point cease to be the dominant factor limiting the sensitivity of the instrument. The projected figure of 2 pT·Hz$^{-\frac{1}{2}}$ is approximately consistent with the best reported sensitivities for bulk NV ensemble magnetometers to date [S9, S10], but should not be considered a hard lower bound.)

The combination of ~30 times better magnetic field sensitivity with a ~36-fold increase in NMR signal due to higher sample polarization at B_0 = 3 T would enable detection of ~13 mM proton concentrations with SNR $\approx$ 3 in 10 minutes of averaging. This is comparable to the concentration sensitivity demonstrated with state of the art microcoils operating at much higher fields of B_0 = 7 − 12 T [S11]. The NV detector sensitivity might be further improved with advances in diamond engineering allowing preferential NV orientation [S12] and/or improved N-NV conversion efficiency [S13]. Operation of an NV sensor at several times lower B_0 compared to microcoils could be advantageous, potentially mitigating the challenges

associated with susceptibility mismatches encountered in many microcoil experiments. Furthermore, the diamond platform should be amenable to parallel operation, using an array of chips with independent (cross-talk free) optical readouts for each.

Supplementary Note 10: Linear Magnetometry Mitigates Broadening Due to Sample Diffusion

An NV center at depth d_{NV} below the diamond surface is primarily sensitive to magnetic fields produced by nuclear spins within a radius of $\sim \kappa\, d_{NV}$ of the corresponding point on the diamond surface, for $\kappa \approx 2$ (Fig. 1a). The time-varying magnetic field at the position of the NV center, oscillating at the nuclear Larmor frequency, has two components: (i) a coherent component of constant amplitude and phase, whose amplitude is proportional to the mean number, N, of thermally-polarized spins in the sensing volume ($\sim[3 \times 10^{-7}]\, N$ at bias field B_0 = 88 mT and room temperature); and (ii) a noise component with random phase and amplitude, with amplitude standard deviation proportional to $N^{1/2}$. The correlation time of the noise component is set by the rate at which nuclear spins diffuse in and out of the sensing volume, $\tau_c \approx (\kappa\, d_{NV})^2 / (6\, D_T)$, for D_T the translational diffusion coefficient of the sample molecules. The noise component is dominant for NV centers with d_{NV} < 3 µm, while the coherent component is dominant for deeper NV centers (Supplementary Note 2).

The AC magnetometry pulse sequence applied to the NV centers at each iteration of the CASR protocol effectively mixes the magnetic signal frequency down from the nuclear Larmor frequency to near DC. Thus, when the nuclear Larmor frequency matches the center frequency of the AC magnetometry pulse sequence, $f_{Larmor} = f_0$, the coherent component of the signal is constant in time over successive NV readouts, while the noise component is slowly time-varying on timescale τ_c. In our CASR measurements, the final $\pi/2$-pulse of each NV magnetometry sequence is chosen to be 90° out of phase from the initial $\pi/2$-pulse of the magnetometry sequence, such that the final NV spin population is *linearly* dependent on the amplitude of the total oscillating magnetic field signal. This is a key difference from previous NV-detected NMR protocols, where magnetometry pulse sequences yielded a *quadratic* dependence of the final NV spin population on the magnetic field signal in order to rectify the zero-mean noise component and sense its variance. Because we do not carry out this rectification step, the mean amplitude of the noise component in our fluorescence signal is zero, and its standard deviation increases only like the square root of the number of NV fluorescence readouts. On the other hand, the amplitude of the coherent component in the fluorescence signal increases linearly with the number of NV readouts. Thus, over many synchronized readout cycles, the measurement is sensitive primarily to the coherent component of the signal (with time-independent amplitude proportional to the Boltzmann polarization), and insensitive to the noise component (with phase and amplitude varying on a timescale set by diffusion into and out of the sensing volume). Therefore, diffusion is not an important consideration in determining the FNP signal linewidth in the CASR measurements presented in this work.

[S1] Alessio, S. M. Digital signal processing and spectral analysis for scientists: concepts and applications, Springer, Switzerland (2016).

[S2] Doherty, M. W. *et. al.* The nitrogen-vacancy colour center in diamond, *Phys. Rep.* **528**, 1 – 46 (2013).

[S3] Pfender, M. *et al.* Nonvolatile nuclear spin memory enables sensor-unlimited nanoscale spectroscopy of small spin clusters, *Nat. Commun*. **8**, 834 (2017).

[S4] Maurer, P. C., *et al.*, Room-temperature quantum bit memory exceeding one second, *Science* **336**, 1283 – 1286 (2012).

[S5] Dickson, R. M., Norris, D. J., Tzeng, Y.-L. & Moerner, W. E. Three-dimensional imaging of single molecules solvated in pores of poly(acrylamide) gels, *Science* **274**, 966 – 968 (1996).

[S6] Shon, M. J. & Cohen, A. E. Mass Action at the Single-Molecule Level, *J. Am. Chem Soc.* **134**, 14618 – 14623 (2012).

[S7] Levitt, M. H., Rantaharju, J., & Brinkman, A. SpinDynamica code for Wolfram Mathematica. www.SpinDynamica.soton.ac.uk (2017).

[S8] Clevenson, H. Broadband magnetometry and temperature sensing with a light-trapping diamond waveguide, *Nat. Phys.* **11**, 393 – 397 (2015).

[S9] Barry, J. F. *et. al.* Optical magnetic imaging of single-neuron action potentials using quantum defects in diamond, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 14133 – 14138 (2016).

[S10] Wolf, T. et. al. Subpicotesla diamond magnetometry, *Phys. Rev. X* **5**, 041001 (2015).

[S11] Badilita, V., *et. al.* Microscale nuclear magnetic resonance: a tool for soft matter research, *Soft Matter* **8**, 10583 – 10597 (2012).

[S12] Ozawa, H., Tahara, K., Ishiwata, H., Hatano, M., & Iwasaki, T. Formation of perfectly aligned nitrogen-vacancy ensembles in chemical-vapor-deposition-grown diamond (111), *Appl. Phys. Express* **10**, 045501 (2017).

[S13] Kucsko, G., *et. al.* Critical thermalization of a disordered dipolar spin system in diamond. Preprint at https://arxiv.org/abs/1609.08216 (2016).