Supplementary Materials for
Multi-scale reconstruction of single-ion damage tracks in diamond via nitrogen-vacancy centers

Daniel G. Ang*†, Jiashen Tang*†, Maximilian Shen, Michael Titze, Gavishta Liyanage, Jordan Chapman, Chinmay Bharathulwar, Andrew T. Gilpin, Tanguy Terlier, Edward S. Bielejec, Vsevolod Ivanov, Ronald L. Walsworth*

*Corresponding author. Email: dga@umd.edu (D.G.A.); jstang@umd.edu (J.T.); walsworth@umd.edu (R.L.W.)

†These authors contributed equally to this work.

This PDF file includes:

Supplementary Text (Sections A to G)

Figs. S1 to S13

Tables S1 and S2

A NV photoluminescence (PL) analysis details

A.1 Data analysis procedure for confocal scans

The number of NV centers formed at each implantation site is extracted from the confocal PL data in several stages (Fig. S1). First, 100 μm × 100 μm implantation squares on the sample containing a regular grid of sites at 6 or 10 μm pitch are imaged in a coarse, wide-field two-dimensional scan at 500 nm pixel pitch, focused at ~600 nm below the surface (Fig. S1a). The following procedure is then used to locate candidate implantation sites for finer scans. PL count rates are converted to a number of NV centers using a single-NV reference rate $I_1 = 25$ kcts/s, calibrated on isolated NV centers in a low-background region of the same diamond sample under identical excitation and collection conditions. Sub-regions within implantation squares containing high NV backgrounds from residual subsurface polishing damage are manually excluded, and an intensity window of 5–100 I_1 (125–2500 kcts/s) is applied to suppress the residual background ($\lesssim$ 2 NV per confocal

spot; Fig. S1b).[1]

Next, a rectangular grid of points spaced by the nominal implantation pitch (6 or 10 μm) is overlaid on the image, and an elliptical acceptance region with semi-axes equal to $\pm 3\sigma$ of the measured beam profile is placed at each grid node. Within each ellipse, the brightest pixel $\mathbf{r}_i^{\mathrm{peak}}$ is taken to mark the implantation site. The grid origin is then shifted by an offset (δ_x, δ_y) chosen to minimize the total distance between the grid nodes and their enclosed brightest pixels,

$$D(\delta_x, \delta_y) = \sum_i \left\| \mathbf{r}_i^{\mathrm{grid}}(\delta_x, \delta_y) - \mathbf{r}_i^{\mathrm{peak}} \right\|, \tag{S1}$$

where $\mathbf{r}_i^{\mathrm{grid}}$ is the position of the i-th grid node, which shifts with the applied offset (δ_x, δ_y). Grid nodes whose brightest enclosed pixel exceeds $25\, I_1$ are excluded from the sum as anomalously bright artifacts (Fig. S1c). This procedure is similar to that of Ref. (*1*).

From the procedure above, each located site is matched to a high-resolution three-dimensional confocal scan (~200 nm pixel pitch) acquired at that position, from which the NV number is obtained as $N_{\mathrm{NV}} = (I_{\mathrm{max}} - I_{\mathrm{bg}})/I_1$, where I_{max} is the brightest pixel of the in-focus slice and I_{bg} the median of the scan's border pixels (Fig. S1d). Sites with one to six resolved PL spots are retained, and per-site outliers are removed if the NV count deviates from the population median by more than three times the rescaled median absolute deviation ($1.4826 \times$ MAD, which estimates the standard deviation for normally distributed data). Across the 800 keV region the procedure yields $N = 222$ analyzed sites with a mean NV yield $\langle N_{\mathrm{NV}} \rangle = 17.5 \pm 0.38$ (standard error of the mean, $5.7/\sqrt{N}$) and a site-to-site standard deviation of 5.7 (main text Fig. 2d).

The reported mean NV count is robust to the analysis choices. Tests were performed varying the grid offset, elliptical acceptance area, outlier rejection brightness threshold, and the per-site intensity estimator (brightest pixel versus the mean of the ten brightest pixels), as well as restricting the data set to sites with a single instead of multiple resolved spots. None of these data analysis variations resulted in significant changes to $\langle N_{\mathrm{NV}} \rangle$.

[1] See section G for results of additional diamond surface preparation experiments to suppress such backgrounds for future implantations.

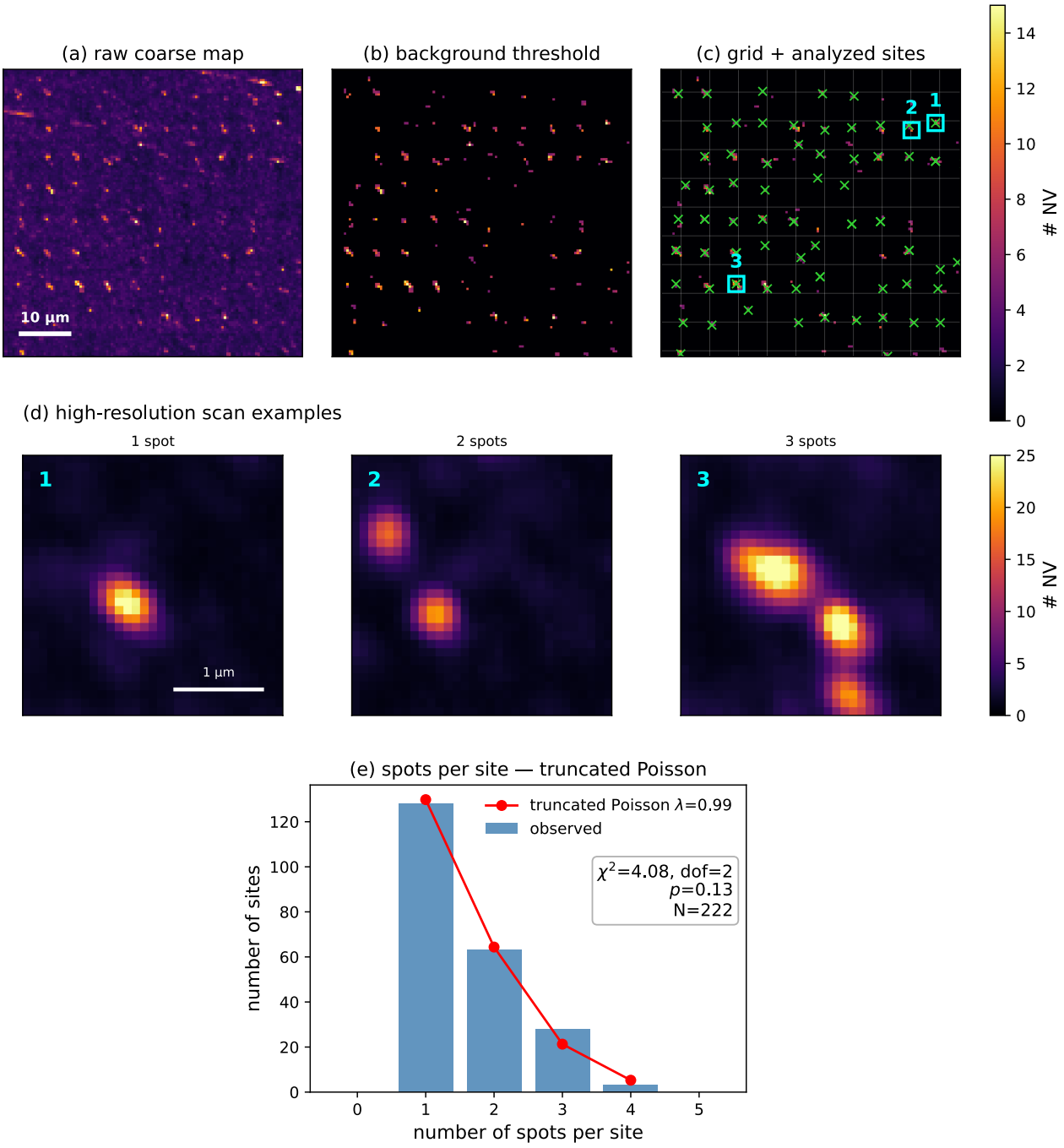


Figure S1: **NV-count extraction pipeline and multi-ion statistics**, illustrated for a representative 800 keV implantation square (6 μm pitch). **(a)** Raw coarse confocal map (500 nm pixels), in units of calibrated single-NV PL. **(b)** After region-of-interest and scratch masking and intensity thresholding, isolating the implantation-site PL. **(c)** Elliptical grid mask ($\pm 3\sigma$ of the beam profile) overlaid on the thresholded map; green crosses mark the analyzed sites and cyan boxes the three example sites shown in (d). **(d)** High-resolution nanovolume scans of the three boxed sites, resolving one, two, and three PL spots. **(e)** Distribution of the number of resolved spots per site over all $N = 222$ analyzed 800 keV sites (bars), with a maximum-likelihood truncated-Poisson fit (red; $\lambda = 0.99$, $\chi^2 = 4.1$, 2 d.o.f., $p = 0.13$).

A.2 Poisson analysis of multi-ion impacts

At sites that received more than one ion, the high-resolution scans resolve multiple, spatially separated PL spots (Fig. S1d), each attributed to a single ion impact. The distribution of the number of resolved spots per site, pooled over all 222 analyzed sites, is shown in Fig. S1e. Because sites with zero spots are not observed, we model the counts with a Poisson distribution truncated to $n \geq 1$,

$$P(n) = \frac{\lambda^n e^{-\lambda}}{n!\left(1 - e^{-\lambda}\right)}, \qquad n = 1, 2, 3, \ldots, \tag{S2}$$

whose single parameter λ is the mean number of ions delivered per grid node. A maximum-likelihood fit gives $\lambda = 0.99$. A Pearson χ^2 goodness-of-fit test, with the counts grouped into bins $n = 1$, 2, 3, and ≥ 4 so that every expected count exceeds five, yields $\chi^2 = 4.1$ for 2 degrees of freedom ($p = 0.13$); the data are thus statistically consistent with a truncated Poisson distribution. This confirms that each resolved PL spot corresponds to a single, independent ion impact, since the arrival of ions at a given grid node is expected to follow Poisson statistics.

B SIMS measurement of nitrogen concentration

The kinetic Monte Carlo (KMC) annealing simulations described in the main text take the substitutional nitrogen concentration $[N_s^0]$ as an input. We determined this value for the implanted sample ("Sandia 1") using two independent methods: Time-of-Flight Secondary Ion Mass Spectrometry (TOF-SIMS) and an estimate based on the NV background PL of the annealed sample. Both methods are anchored to a commercial SIMS measurement of a second HPHT diamond sample ("Sandia 2") from the same manufacturer (Element Six). We use the average of the two estimates as the simulation input, treating the spread between them as a conservative uncertainty range; as shown in Section C.2 (Fig. S6), the predicted NV yield is only weakly dependent on $[N_s^0]$ over this range, so the residual calibration uncertainty does not affect the conclusions of the main text.

B.1 TOF-SIMS depth profiling

Negative high mass-resolution depth profiles were acquired with a TOF-SIMS NCS instrument, which combines a TOF.SIMS5 spectrometer (ION-TOF GmbH, Münster, Germany) and an in-situ

scanning probe microscope (NanoScan, Switzerland) at the Shared Equipment Authority from Rice University. The analysis field of view was $80 \times 80\,\mu m^2$ (Bi_3^+ at 30 keV, 0.3 pA) with a raster of 256 by 256 along the depth profile. Charge compensation was provided by an electron flood gun, with surface potential adjusted to 20 V. The cycle time was 70 μs, corresponding to a mass range of m/z = 0–911 amu. Sputtering was performed with Cs^+ at 2 keV, 105 nA, rastered over $100 \times 100\,\mu m^2$. The two beams were operated in non-interlaced mode, alternating one analysis cycle with ten sputtering frames followed by a 5 s charge-compensation pause. Depth calibration was performed using the interface tool in SurfaceLab v7.3 (ION-TOF GmbH), with sputter rates referenced to crater depths measured ex-situ with a DekTak stylus profilometer. For each region of interest, the analyzed area was further cropped to remove crater edges and ensure homogeneity of the extracted signal.

Quantification of the CN^- ion signal into an absolute nitrogen concentration requires an external calibration. To this end, we sent a second HPHT diamond with similar properties (Sandia 2) to EAG for commercial SIMS analysis at multiple regions of interest, and then measured the same regions on Sandia 2 with the Rice TOF-SIMS. The resulting linear EAG-Rice calibration curve (Fig. S2b) was applied to all subsequent Rice TOF-SIMS depth profiles acquired on Sandia 1. A representative depth profile is shown in Fig. S2a. This SIMS procedure was performed across six different regions on Sandia 1 adjacent to implanted regions, yielding an average value of $[N_s^0]_{SIMS} \approx 318$ ppm.

B.2 NV background PL as an independent estimate

As an independent cross-check, we exploit the correlation between substitutional nitrogen concentration and grown-in NV-related PL in HPHT type Ib diamond (*2*). NV background PL intensities were measured in implanted regions of Sandia 1 (i.e., the regions surrounding the NV clusters at the ion impact sites) and at multiple regions of the EAG-measured Sandia 2 sample, the latter providing calibration. Plotting confocal PL intensity against the corresponding EAG-reported $[N_s^0]$ values yields a linear correlation (Fig. S2c) which confirms the validity of this proxy for our HPHT material. Applying this calibration to the averaged background PL in implanted regions of Sandia 1 gives $[N_s^0]_{PL} \approx 188$ ppm.

B.3 Combined estimate and propagation into the simulations

The two methods yield estimates that differ by approximately ±25% from their mean: ≈ 318 ppm from TOF-SIMS and ≈ 188 ppm from NV background PL. We adopt the average, $[N_s^0]$ = 253 ppm, as the nominal input to the KMC annealing simulations, and treat the spread between the two methods as a conservative uncertainty range of ∼180–320 ppm. As shown in Section C.2 (Fig. S6), the predicted NV yield varies only weakly with $[N_s^0]$ across this range and indeed across nearly two orders of magnitude in concentration. The dominant determinants of NV yield are instead the vacancy-interstitial and vacancy-vacancy reactions discussed in Section D, and the calibration uncertainty on $[N_s^0]$ does not propagate appreciably into the conclusions of the main text.

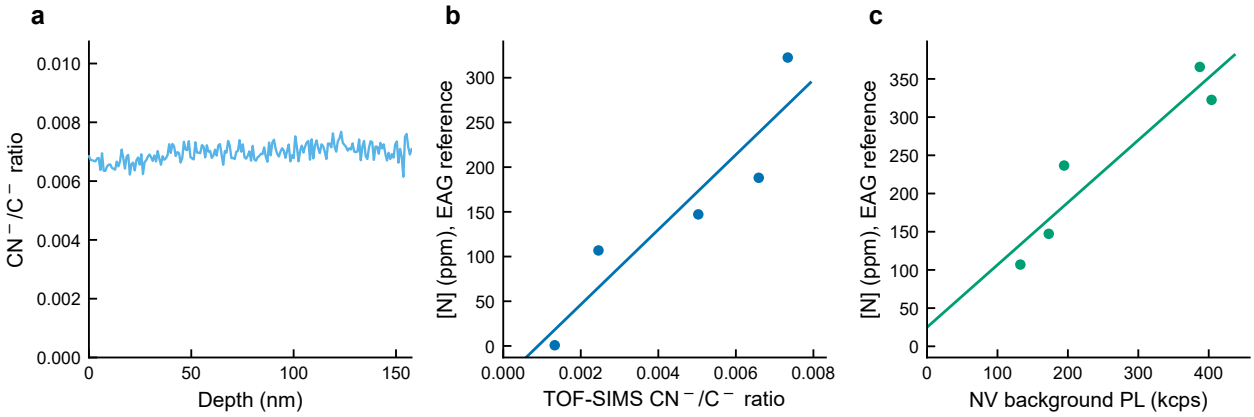


Figure S2: **Nitrogen-concentration measurement. (a)** Representative TOF-SIMS depth profile of the CN^-/C^- ion ratio, showing a uniform signal over the profiled depth. **(b)** EAG–Rice TOF-SIMS calibration: CN^-/C^- ratio measured at Rice University plotted against the EAG reference nitrogen concentration for five regions of Sandia 2 (solid line: linear fit). **(c)** NV background PL calibration: confocal PL intensity plotted against the EAG reference nitrogen concentration for the same five regions (solid line: linear fit).

C Annealing simulation details

C.1 Cluster decomposition and KMC framework

The defect distribution – consisting of both vacancies and carbon interstitials – produced by ion implantation forms an extended but spatially discontinuous damage track. The defect–defect interaction probability decreases with initial separation r (assuming defects are mobile), and in

a three-dimensional continuous limit, this interaction (or "hitting") probability is described as $P(r) = a_0/r$, where a_0 is the interaction radius. Based on this scaling, we partition each damage track into smaller clusters and simulate each cluster independently, effectively assuming negligible interaction between clusters.

To ensure the validity of this approximation, defects belonging to different clusters are required to be separated by more than 154 Å. Under the assumption that defect reactions occur at nearest-neighbor (1NN) separation (i.e. $a_0 = 1.54$ Å), this corresponds to an interaction probability of $\leq 1\%$ between any defect pairs from different clusters in the absence of other species. The presence of immobile substitutional nitrogen impurities can further suppress defect–defect interactions between clusters by providing an additional reaction channel. We quantify this effect for vacancy–vacancy interactions by directly simulating two-vacancy systems with varying initial separations r and nitrogen concentrations $[N_s^0]$, as shown in Fig. S3(a). The interaction probability as a function of r is fitted to a power-law dependence $1.54\,\text{Å}/r^p$. For our experimentally measured nitrogen concentration of $[N_s^0] = 253$ ppm, the expected vacancy-vacancy interaction probability at a separation of 154 Å is reduced to ~0.3%. In practice, vacancy–vacancy interactions and vacancy–interstitial recombination within each cluster further reduce the likelihood of cross-cluster interactions, supporting the validity of independent cluster simulations. An example of an 800 keV damage track and its decomposition into clusters is shown in Fig. S3(b). The corresponding carbon interstitial distributions are also shown; due to their close spatial proximity to vacancies, interstitial clusters closely follow the vacancy distribution, with approximately equal numbers of vacancies and interstitials within each cluster.

After partitioning the damage track into clusters, a diamond lattice is constructed around each cluster. The simulation domain must be sufficiently large such that vacancies are unlikely to diffuse out of the box and instead undergo reactions within the domain. To ensure this condition, the box size is determined by requiring that trapping by substitutional nitrogen impurities alone is sufficient to capture diffusing vacancies.

For a vacancy undergoing a random walk on the diamond lattice, the number of unique lattice sites visited after N hops is given by (*3, 4*)

$$N_{\text{uni}} \approx 0.56N + 0.63\sqrt{N}. \tag{S3}$$

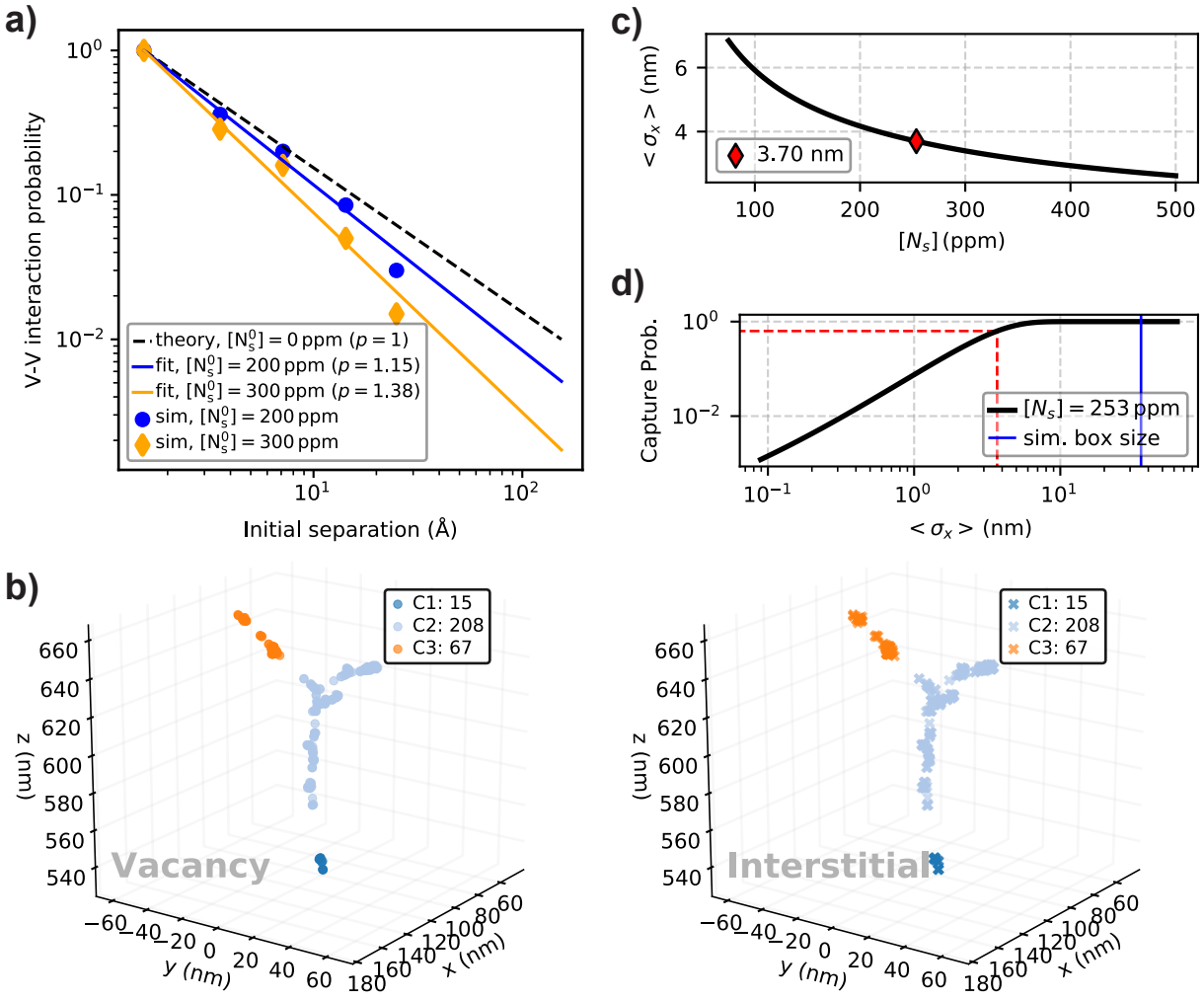


Figure S3: Cluster decomposition and lattice domain size selection for annealing simulations. (a) Probability of vacancy–vacancy interaction as a function of initial separation r, obtained from two-vacancy simulations at different nitrogen concentrations. The dashed theoretical line is computed using the 3D hitting probability, assuming an effective interaction radius of 1.54 Å. Solid lines show power-law fits of the form 1.54 Å$/r^p$. For the experimental nitrogen concentration [N_s^0] = 253 ppm, the interaction probability at 154 Å is ~0.3%, supporting negligible cross-cluster interactions. (b) Example decomposition of an 800 keV implantation damage track into spatially separated clusters. The minimum pairwise defect spacing between clusters exceeds 154 Å. Vacancy (left) and carbon interstitial (right) clusters are shown. The legend indicates the number of defects in each cluster. (c) Root-mean-square (RMS) displacement of a diffusing vacancy as a function of nitrogen concentration, evaluated at the point where the number of unique sites visited equals the number of available capture sites ($N_{uni} = N_{cap}$), corresponding to a capture probability of ~64%. (d) Vacancy capture probability as a function of RMS displacement for a nitrogen concentration of [N_s^0] = 253 ppm. For each cluster, the simulation domain is defined by padding 100 unit cells in each direction from all vacancy defects. This ensures that the probability of vacancy escape is $\lesssim$ 0.01%.

For a given nitrogen concentration $[N_s^0]$ (in ppm), the expected number of distinct lattice sites that must be visited before encountering a nitrogen capture site is

$$N_{\text{cap}} = \frac{10^6}{4[N_s^0]}, \tag{S4}$$

where the factor of 4 reflects the four nearest-neighbor sites adjacent to each substitutional nitrogen, each of which is a possible capture site for NV formation. The cumulative probability χ that a vacancy is captured by nitrogen after N hops is

$$\chi = 1 - \exp\left(-\frac{N_{\text{uni}}}{N_{\text{cap}}}\right), \tag{S5}$$

and the root-mean-square (RMS) displacement along any Cartesian direction is given by

$$\langle \sigma_{x/y/z} \rangle = \sqrt{\frac{N}{3}} \times 1.54\,\text{Å}. \tag{S6}$$

Figure S3(c) shows the RMS displacement as a function of nitrogen concentration at the point where $N_{uni} = N_{cap}$, corresponding to a capture probability of ~64%. To ensure that vacancies remain within the simulation domain, the lattice is padded by at least 100 unit cells in each direction from any vacancy defects in the simulated cluster. At this scale, the probability of a vacancy reaching the boundary without being captured is $\lesssim 0.01\%$, as shown in Fig. S3(d).

During the annealing simulation, the kinetic Monte Carlo (KMC) algorithm (*5*) samples allowed processes – including defect diffusion and reactions – according to Arrhenius rates of the form $\gamma = \nu_0 \exp(-\frac{E_a}{k_B T})$, which determine the relative frequency of each event. Here, ν_0 is taken as the Debye frequency of diamond (*4*), and the annealing temperature is set to 800 °C. After each event is selected and applied, the global lattice configuration is updated and the simulation time advances, resulting in non-uniform time steps that depend on the selected processes. The simulation is terminated when no further events are available, ensuring that all diffusion and reaction processes are exhausted. Additional damage tracks before and after annealing are shown in Fig. S4, with corresponding head-tail asymmetry values computed using the defect counting method. Diffusion-induced broadening during annealing reduces the asymmetry of the NV distribution relative to the initial vacancy track.

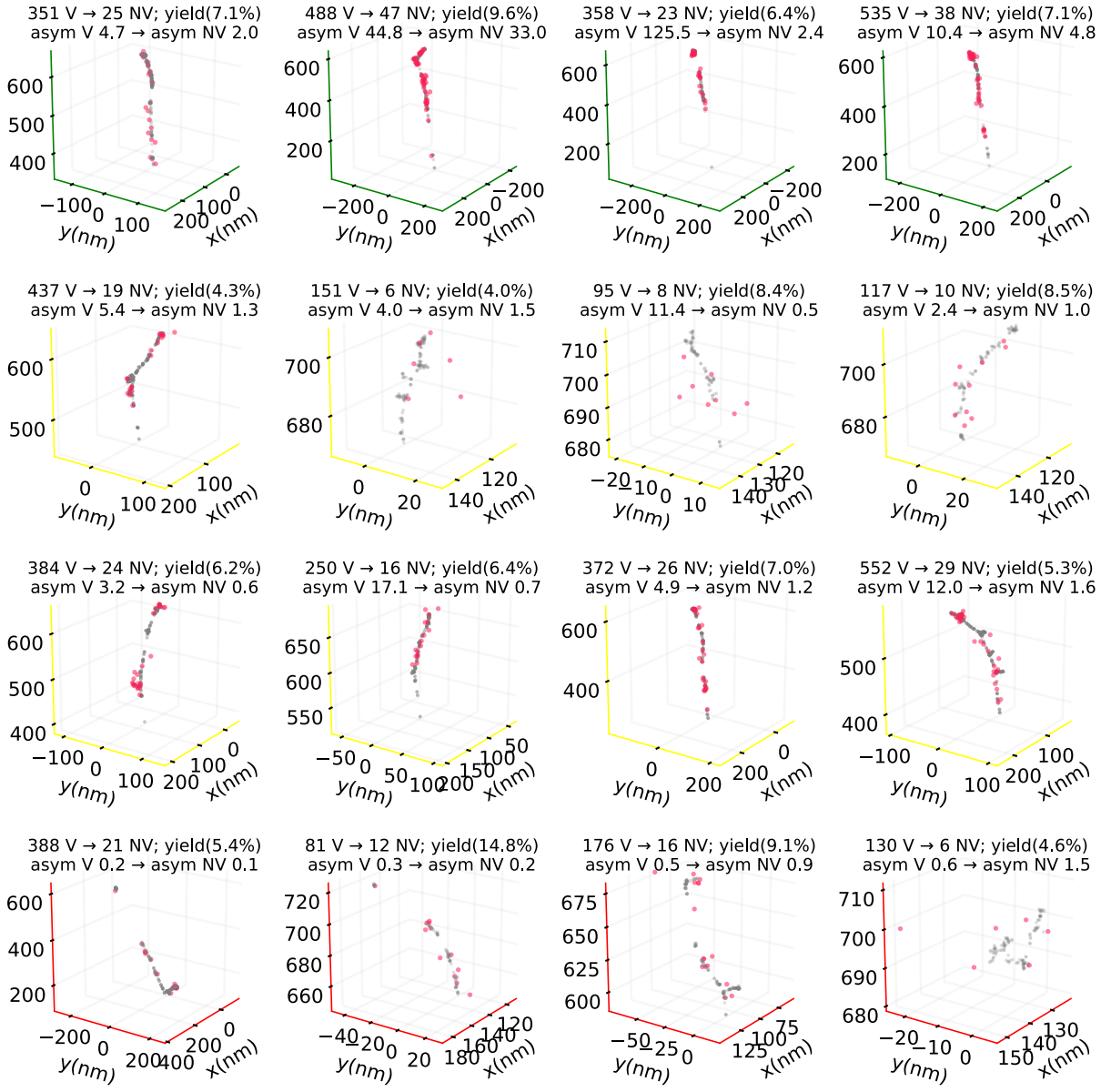


Figure S4: Representative damage tracks before and after annealing simulations. Gray dots indicate the initial vacancy distribution, and red dots indicate the resulting NV distribution after annealing. The vacancy-to-NV conversion yield and head-tail asymmetry values, computed using a defect counting method, are shown for both vacancy and NV tracks.

C.2 Additional factors influencing NV yield

Here we consider two sources of uncertainty that may affect the predicted NV yield: the total vacancy production in the implantation simulation and the measured nitrogen concentration. As discussed below, neither is sufficient to account for the discrepancy of NV yield between model predictions and experimental measurements. This further supports the need for DFT-informed corrections to defect interactions as an important mechanism governing NV formation.

Displacement energy. The threshold displacement energy E_d used in the SIIMPL simulation controls the number of vacancies generated in the damage track. In diamond, E_d is known to be directionally dependent (*6*), with reported values ranging from ~37.5 eV for [100] to ~47.6 eV for [110]. Similar to SRIM, SIIMPL employs a single scalar value of E_d, effectively averaging over crystallographic directions during the recoil cascade. In our analysis, we use $E_d = 43$ eV. To evaluate the sensitivity of vacancy production to this parameter, we vary E_d over a broad range. As shown in Fig. S5, the mean vacancy number decreases monotonically with increasing E_d. Within the experimentally relevant range of ~37–48 eV, this corresponds to a ~10–20% uncertainty in the number of resulting vacancies. Assuming a fixed vacancy-to-NV conversion efficiency, this rescales the KMC-predicted NV yield from ~23 to ~21 at the upper end of E_d, and therefore does not fully account for the experimentally measured value of ~17 NV centers.

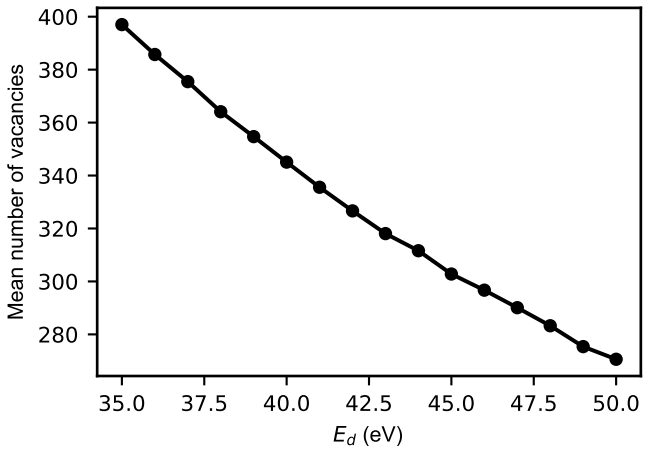


Figure S5: Sensitivity of vacancy production to displacement energy E_d. Mean number of vacancies generated per 800 keV carbon recoil as a function of the threshold displacement energy used in SIIMPL simulations. Each data point represents the average over 5000 simulated damage tracks.

Nitrogen concentration. The measured nitrogen concentration carries a relatively large uncertainty, ranging from ~180–320 ppm as described above. To assess its impact on NV yield, we perform annealing simulations across a range of nitrogen concentrations, as shown in Fig. S6a. For each concentration, a single damage track is simulated, and the nitrogen lattice configuration is randomized five times to estimate the mean NV yield. We find that the mean NV yield changes only weakly over the range from 150 ppm to 5000 ppm. Increasing nitrogen concentration primarily leads to NV formation occurring closer to the original damage track, as illustrated in Fig. S6b. This weak dependence can be understood from the spatial scales involved: vacancy and interstitial defects are typically separated by less than ~0.5 nm, whereas the average nitrogen–nitrogen spacing is ~3 nm at 200 ppm and scales with the inverse cubic root of concentration. As a result, vacancies predominantly react with interstitials and other vacancies before diffusing away from the damage track to interact with nitrogen, making the influence of nitrogen concentration on overall NV yield much less pronounced.

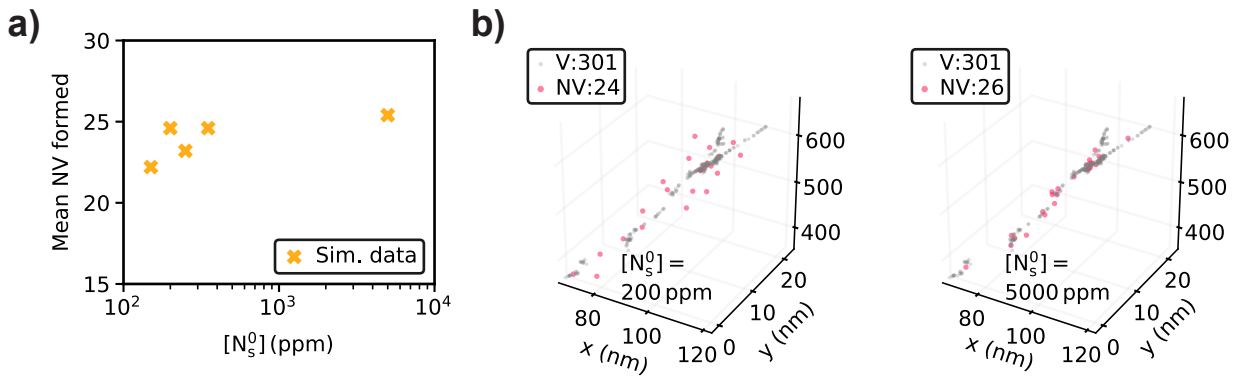


Figure S6: Impact of nitrogen concentration on annealing simulations. (a) Mean NV yield as a function of nitrogen concentration $[N_s^0]$, showing weak dependence over a broad range (150–5000 ppm). (b) Representative spatial distribution of vacancies (gray) and resulting NV centers (red) at $[N_s^0]$ = 200 ppm and 5000 ppm. Higher nitrogen concentration leads to NV formation closer to the original damage track, while the overall NV yield remains largely unchanged.

D DFT calculations of defect interactions

D.1 Methods

First principles calculations of defect formation in diamond are performed with the Vienna *Ab initio* Simulation Package (VASP) (*7, 8*), which implements plane-wave basis sets using a projector-augmented wave (PAW) scheme (*9*). We examined two elementary formation mechanisms of vacancy defects in diamond:

$$C_C \leftrightarrow V_C + C_i \quad \text{(Reaction 1)} \tag{S7}$$

$$V_{2C} \leftrightarrow 2\,V_C \quad \text{(Reaction 2)} \tag{S8}$$

where Reaction 1 describes the formation of a carbon monovacancy (V_C) and carbon interstitial (C_i) via ejection of a lattice-site carbon atom (C_C), and Reaction 2 describes the ejection of carbon monovacancies from a divacancy site (V_{2C}).

Defects were embedded within a $3 \times 3 \times 3$ (216 atoms) supercell. Ionic geometries of vacancy defects were relaxed at the Perdew-Burke-Ernzerhof (PBE) level of theory according to the generalized gradient approximation (GGA) (*10, 11*). Sampling of the Brillouin zone for defect structures was done at the Γ point with a cutoff energy of 520 eV. Convergence criteria for electronic self-consistency and ionic relaxation were set to 10^{-10} eV and 10^{-3} eV Å^{-1}, respectively.

Defect formation energies (E_f) were calculated according to

$$E_f = E_{\text{def}} - E_{\text{pristine}} + \sum_i n_i \mu_i, \tag{S9}$$

where E_{def} and E_{pristine} denote the energies of the defective and pristine lattices, respectively, and μ_i and n_i represent the chemical potential and number of atoms corresponding to species i. Because the kinetic Monte Carlo (KMC) simulations for annealing do not explicitly account for charge states of defects, only neutral configurations were considered, and no charge correction term is applied.

The energy barriers (E_B) associated with Reactions 1 and 2 were determined using a climbing image-modified nudged elastic band (NEB) approach (*12*). The NEB calculations were carried out using the PBE functional with a spring constant of 5.0 eV Å^{-2} between images. Here, E_f and E_B are distinct quantities, where E_B corresponds to the energy difference between the transition state and the initial configuration.

D.2 Results

We performed first principles calculations to qualitatively improve the prediction of NV center formation in KMC annealing simulations. Using the NEB method, we elucidate the transition state geometries and associated energy barriers for Reactions 1 and 2. The MEPs corresponding to V_C formation and V_C ejection from V_{2C} are shown in Figures S7 and S8, respectively; we have included images of the initial state, transition state, and final state of each reaction pathway, as well.

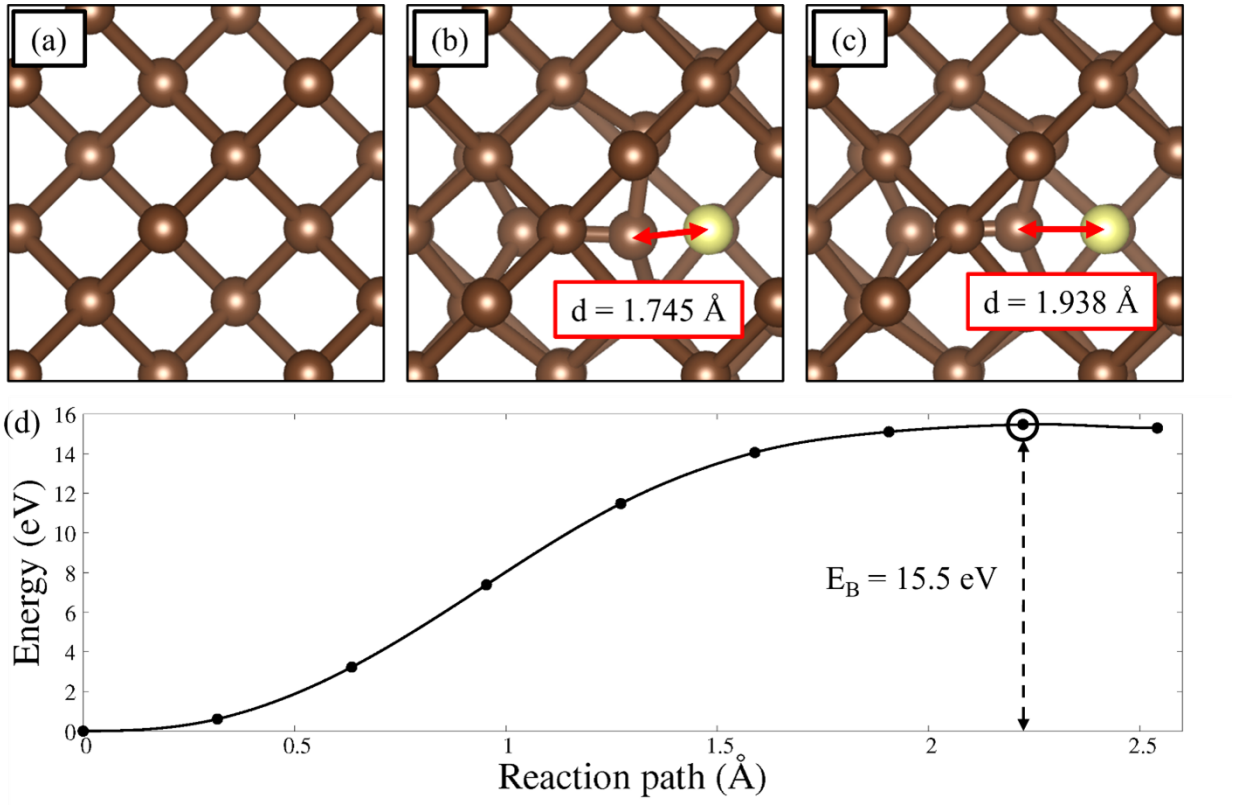


Figure S7: NEB-DFT calculations of Reaction 1. Geometries of the (a) initial, (b) transition geometry, and (c) final states along the reaction path. V_C and C_i defects are formed by the ejection of a C_C atom from its lattice site. Carbon atoms are shown in brown, while the resulting vacancy is shown in yellow. (d) MEP of Reaction 1. The transition state occurs at a V_C-C_i spacing of 1.75 Å corresponding to E_B = 15.5 eV.

As shown in Figure S7, we predict that formation of a V_C defect and accompanying C_i defect from a pristine diamond structure is achieved via the formation of a split interstitial located two lattice sites away from the vacancy position. By examining the reaction pathway, we can calculate the radial distance between the ejected C_i atom and the resulting V_C defect at both the transition state and final state geometries. We have found that at its stable endpoint, the C_i defect lies about

1.94 Å away from V_C, whereas the C_i and V_C defects are separated by only 1.75 Å at the transition state. The transition state is predicted to have a formation energy of approximately 15.5 eV, and the endpoint has a formation energy of about 15.2 eV.

The MEP for the dissociation of a V_C in a V_{2C} defect is shown in Figure S8. Unlike Reaction 1, we find that the constituent V_C defects require a separation of three lattice sites (~3.88 Å) to stably form two distinct vacancies. Instead, the transition state occurs when the V_C defects are separated by two lattice sites (~2.53 Å). The dissociation of a V_C defect from a V_{2C} defect is predicted to have an energy barrier of 5.95 eV, which is in good agreement with previously reported values (*13*).

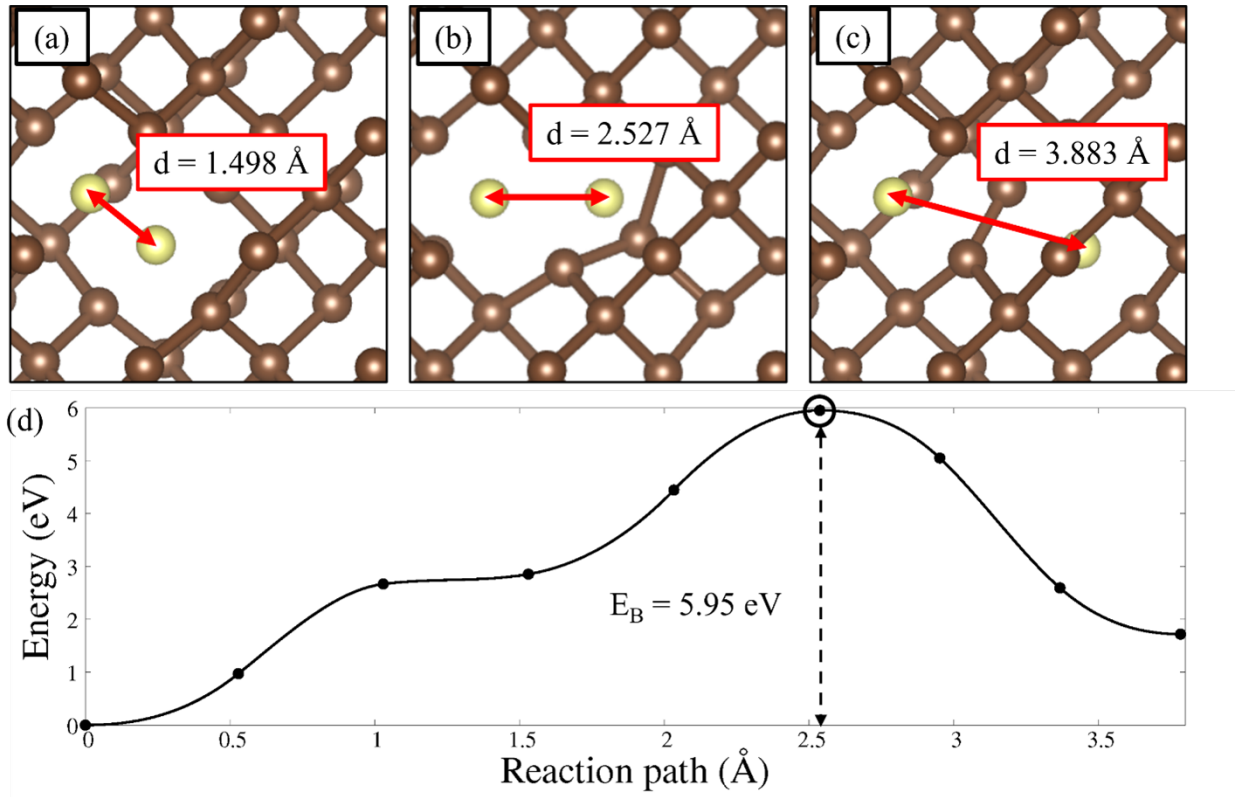


Figure S8: NEB-DFT calculations of Reaction 2. Geometries of the (a) initial, (b) transition geometry, and (c) final states along the reaction path. Two V_C defects are formed by the dissociation of constituent V_C from the V_{2C} defect. Carbon atoms are shown in brown, while the resulting vacancy is shown in yellow. (d) MEP of Reaction 2. The transition state occurs at a V_C-V_C spacing of 2.53 Å corresponding to E_B = 5.95 eV.

These NEB results provide a basis for refining the defect–defect interaction distances used in the KMC simulations, which are restricted to first-nearest neighbor (1NN) separation. Vacancy–vacancy interactions can occur spontaneously when defects are within 2NN separation. For

vacancy–interstitial pairs, although the transition state occurs at a separation of 1.75 Å, the relaxed configuration lies only ~0.2 eV lower in energy. The apparent atom-vacancy separation in this configuration is 1.94 Å; however, in the KMC lattice picture, the effective interstitial defect position is defined by the arithmetic mean of the two carbon atoms forming the split interstitial, corresponding to a 2NN separation from the vacancy. In the current KMC framework, defect migration barriers are treated as environment-independent, with isolated hopping barriers of 1.5 eV for interstitials and 2.59 eV for vacancies. In contrast, the NEB results indicate a barrier of only ~0.2 eV for V–I recombination from the 2NN configuration. At the annealing temperature of 800 °C, this reduced barrier leads to a rate enhancement on the order of $\exp([1.5 - 0.2]\,\mathrm{eV}/k_B T) \sim 10^6$ compared with isolated defect hopping, indicating that 2NN V–I pairs will preferentially approach and recombine.

As a first-order correction to the 1NN assumption in the KMC model, we modify the initial population of free vacancies available for NV formation – here, "free" denotes vacancies that can diffuse rather than react immediately at the beginning of the annealing simulation. Specifically, in the current model, free vacancies are defined as those without 1NN defect neighbors. In the corrected estimate, we further exclude vacancies that have 2NN vacancy or interstitial neighbors, accounting for enhanced spontaneous divacancy formation and vacancy–interstitial recombination. The resulting reduced vacancy population is then rescaled using the same conversion factor (free-vacancy-to-NV yield) as in the uncorrected model to estimate the NV yield.

Finally, we note that NEB results indicate that vacancy–nitrogen interactions can also occur at approximately 2NN separation. However, on average, only ~0.2% of the initial vacancies have a nitrogen partner within 1NN or 2NN distance; therefore, this effect is neglected in the first-order correction.

E Machine learning-based directionality analysis

E.1 Model and training details

To extract directional information beyond counting-based metrics, we employ simulation-based inference (SBI) (*14*), a Bayesian framework for inverse problems in which the likelihood $P(X|\theta)$ is not available in closed form. In our case, the inversion maps the defect distribution X to the

initial recoil parameters θ (e.g., energy and head-tail parity). Due to the stochastic nature of the SIIMPL collision cascade and NV formation during annealing, direct evaluation of the likelihood is intractable. Instead, the model learns an approximation to the posterior $P(\theta|X)$. We implement a Mixed Neural Posterior Estimation (MNPE) architecture, which enables joint inference over continuous (energy) and discrete (parity) variables. The model combines a Neural Spline Flow (*15*) for the continuous parameter with a categorical softmax head for parity classification, implemented using the `sbi` toolkit (*14*). We note that energy parameters are included when learning the general inversion from simulated damage tracks across different ion implantation energies; however, for predicting head-tail parity of the annealed 800 keV NV tracks considered here, these parameters are effectively fixed. Parity classification is binary, i.e. the $+z$ or $-z$ direction in the global coordinate system, as the injected carbon ion is along the z axis in the SIIMPL simulation.

Rather than using raw defect coordinates, the model learns from a set of summary features derived from defect spatial distributions. Each track is encoded as a 13-dimensional feature vector (see Table S1) designed to capture directional information. The model is trained on SIIMPL-simulated vacancy tracks with ground-truth parity labels using the Adam optimizer (learning rate 5×10^{-4}, batch size 128) and a 90/10 train-validation split, with early stopping applied when the validation loss does not improve for 20 epochs. For inference, the trained model is directly applied to NV distributions after annealing using the same feature representation. For each track, 200 samples are drawn from $P(\theta|X)$ to estimate the probabilities P of the $+z$ direction, with $1-P$ corresponding to the $-z$ direction. A probability threshold symmetric around 0.5 can then be applied to classify tracks as correct, incorrect, or indeterminate.

E.2 Classification performance comparison using NV tracks

To compare the counting-based asymmetry metric and the machine learning model across implantation energies, we present cumulative plots of NV track classification performance (Fig. S9). NV tracks are simulated for both 800 keV and 100 keV ion implantations. Analogous to Fig. 3c,d in the main text, thresholds are chosen for the asymmetry metric A and posterior probability P such that the false positive rate is fixed at $\lesssim 5\%$. For the machine learning analysis, the model applied to 800 keV NV tracks is trained on 4000 800 keV vacancy tracks, while the model applied to 100 keV

Table S1: The 13 summary features computed for each track and passed to the MNPE model. Here N is the number of defects and z_i is the position of the i-th defect along the track axis, measured relative to the track's centroid (each track is re-centered to $z = 0$). $N_\pm$ and $z_\pm$ denote the counts and positions of defects with $z_i > 0$ and $z_i < 0$, respectively.

#	Feature	Definition	Interpretation
1	Mean depth	$\frac{1}{N}\sum_i \lvert z_i \rvert$	The average extent of the track. Longer mean depth usually means a higher-energy ion.
2	Max depth	$\max_i \lvert z_i \rvert$	Distance of the farthest defect from track center.
3	Vacancy count	N	The total number of defects.
4 – 9	Depth-profile fractions ($k = 1, \ldots, 6$)	Fraction of defects in each of six log-spaced bins along the normalized track depth	Defect distribution binned along z axis.
10	Centered asymmetry	$(N_+ - N_-)/N$	Whether more defects sit on one side of the centroid than the other.
11	Asymmetry ratio defined in Ref. (*16*)	$N_{\text{last third}} \big/ N_{\text{first third}}$	The asymmetry ratio from Ref. (*16*).
12	Longitudinal skewness	Skewness of $\{z_i\}$ along the track axis	Whether the defects lean more toward the head or the tail of the track.
13	Log-variance difference	$\log(1 + \mathrm{Var}(z_+)) - \log(1 + \mathrm{Var}(z_-))$	Compares how spread out the defects are on each side of the centroid.

NV tracks is separately trained on 20,000 100 keV vacancy tracks. For the 100 keV case, despite the lower average number of NV centers formed per track, the machine learning approach improves the classification efficiency ϵ from 37% to 96%.

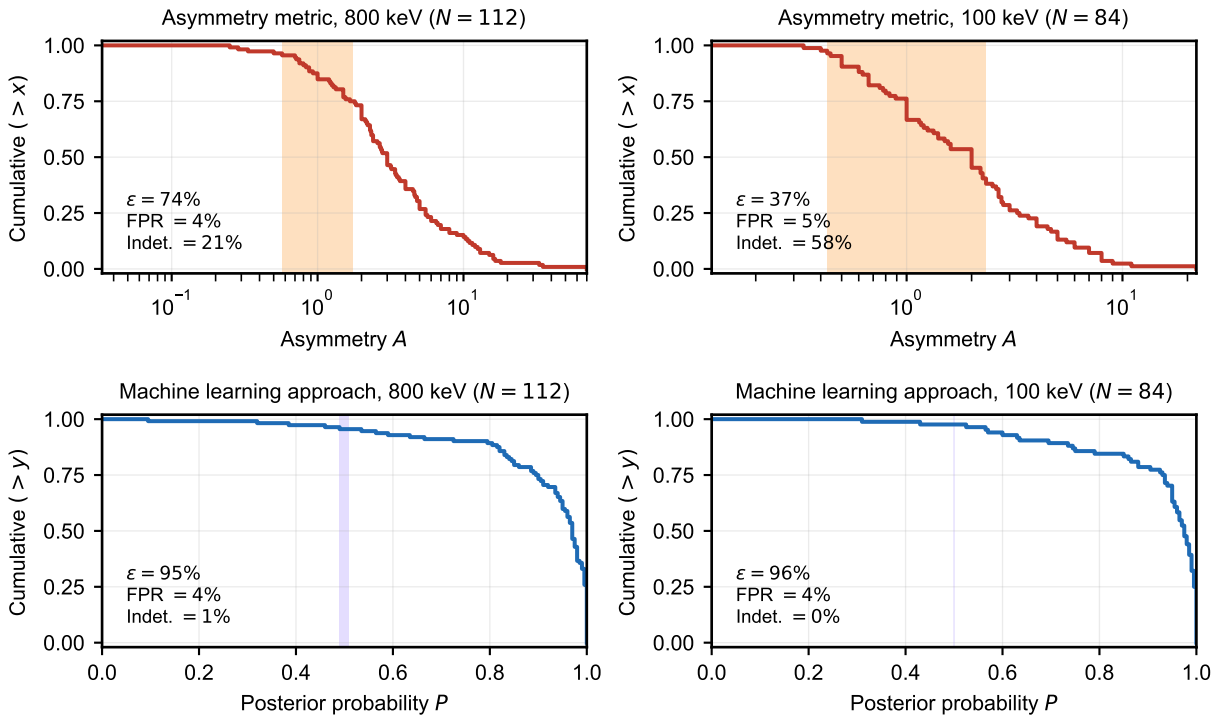


Figure S9: Cumulative plots of NV track classification using the counting-based asymmetry metric and machine learning approach. Top row: cumulative distributions of the asymmetry metric for 800 keV and 100 keV NV tracks. Bottom row: cumulative distributions of the machine learning posterior probability for the same tracks. Shaded regions indicate indeterminate ranges chosen to give a $\lesssim$5% false positive rate. The machine learning approach increases the classification efficiency relative to the asymmetry metric for both implantation energies.

E.3 Classification performance at low recoil energy

We further evaluated classification performance at a much lower recoil energy of 1 keV. In this regime, each track produces only about 11 vacancies on average, and the expected number of NV centers after annealing is < 1, below the optical detection threshold. Therefore, NV-based readout is not expected to be viable at this energy under the current annealing-based scheme. Nevertheless, vacancy tracks provide a useful benchmark for testing how directional classification degrades at very low recoil energy and for comparing the machine learning (ML) approach with the counting-based

asymmetry metric. Training the model on 5000 tracks, at a $\lesssim$5% false positive rate, the machine learning classification efficiency decreases to 26%, but remains substantially higher than the 6% efficiency obtained using the asymmetry metric. This suggests that our ML analysis can still extract directional information from sparse damage tracks, and may be useful for future defect candidates that are optically bright, single-defect resolvable, and localizable with high spatial precision without requiring annealing activation.

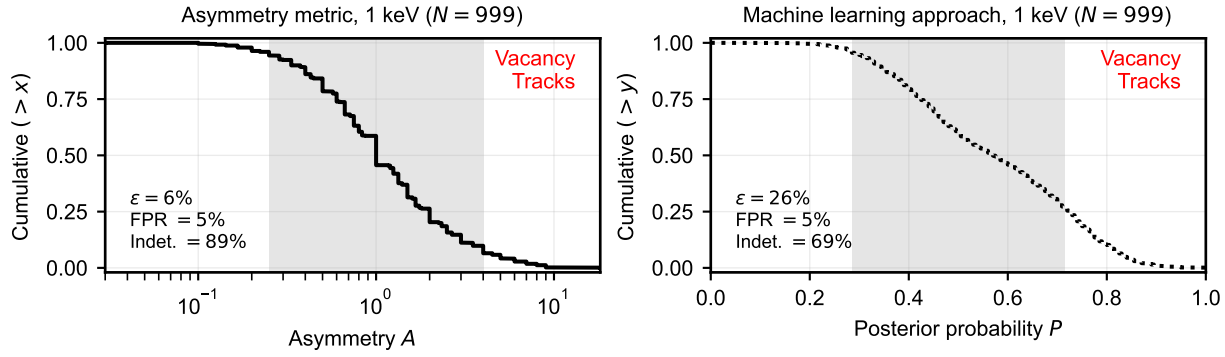


Figure S10: Cumulative plots of vacancy track classification using the counting-based asymmetry metric and machine learning approach. Vacancy tracks are used instead of NV tracks because the expected number of NV centers after annealing is < 1. At a $\lesssim$5% false positive rate, the machine learning approach gives higher classification efficiency than the asymmetry metric.

F Additional spin echo measurement data

Additional spin echo measurements are performed on 14 implantation sites in the implanted HPHT sample to investigate potential correlations among spin contrast, coherence properties, and the number of NV centers per site. Individual decay traces and the corresponding NV number distribution are shown in Fig. S11. These data were acquired with reduced averaging compared to the main text and were fitted using a stretched exponential model with the exponent fixed at $p = 1.55$, informed by the main text results, i.e. $C_0 \exp[-(\tau/T_2)^{1.55}]$.

The extracted spin echo parameters and NV numbers are summarized in Fig. S12. The coherence time T_2 remains relatively constant across measurement sites, with an average value of $\sim$1.4 μs. In contrast, the PL contrast exhibits a larger spread ($\sim$1–3%). Neither shows a clear dependence on NV count. This variation in contrast may arise from differences in the distribution of NV orientations

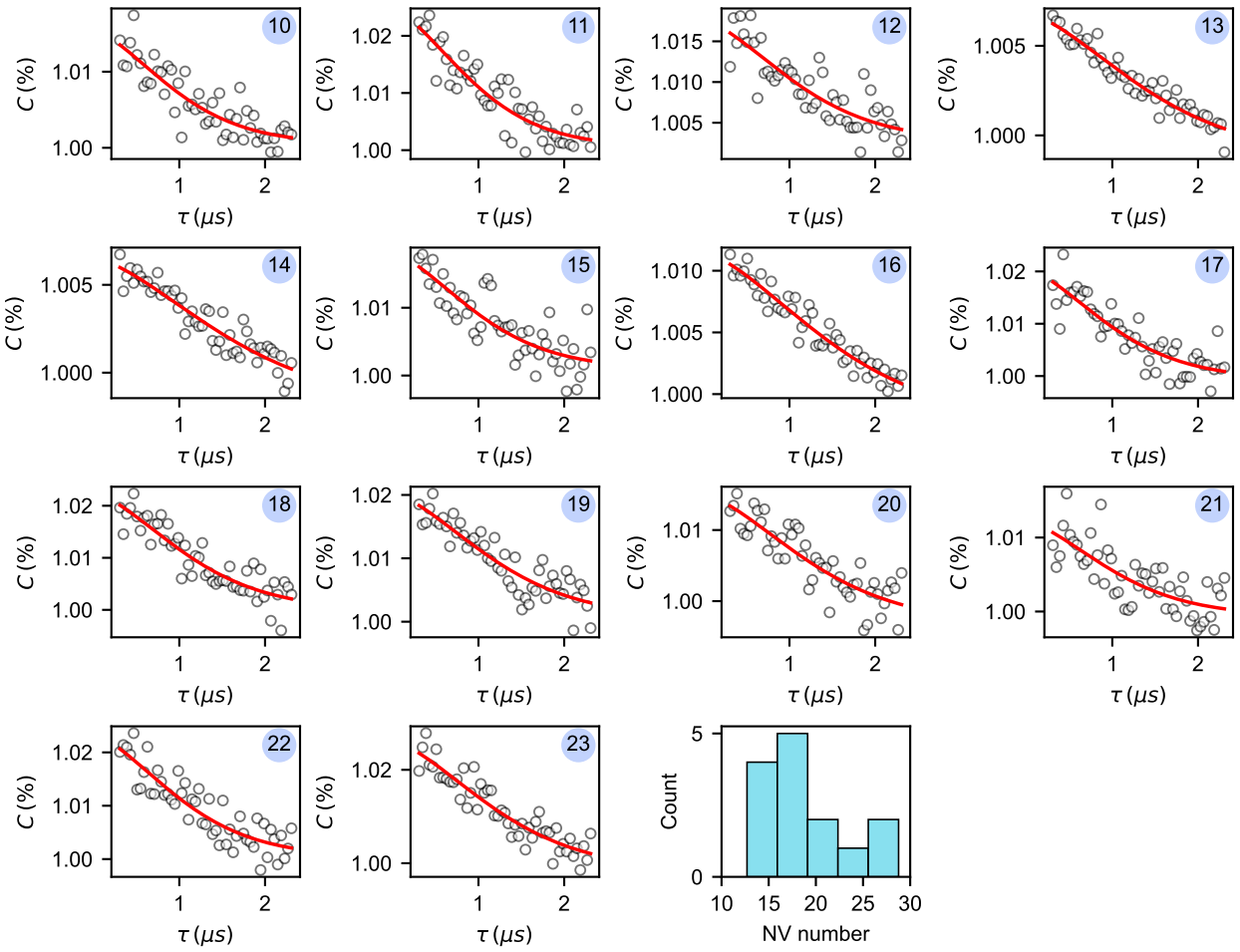


Figure S11: Additional NV spin echo measurements. Spin echo measurements from 14 additional implantation sites in the HPHT sample. Each trace is fitted to a stretched exponential decay (red curve) with fixed exponent $p = 1.55$. The final panel shows the distribution of NV PL counts for the characterized sites.

within each site, as the measurement selectively probes a single NV axis while the remaining three crystal orientations contribute to background PL.

G Improving diamond surface preparation

The NV count analysis of Section A required manually excluding sub-regions with high NV backgrounds from residual subsurface polishing damage. These backgrounds arise from the diamond polishing process. Synthetic HPHT diamond is typically laser-cut into individual plates and mechanically polished to achieve optically smooth surfaces. However, mechanical polishing is known to introduce subsurface damage (*17*), generating lattice vacancies that can form NV centers after annealing. In our ion implantation studies, where damage tracks are formed near the surface, such polishing-induced defects can obscure or interfere with the implantation signal. As shown in Fig. S13a, a scaife-polished diamond surface (with roughness $R_a \lesssim 1\,\mathrm{nm}$) exhibits strong and highly non-uniform photoluminescence after annealing. Although regions with low background can still be found after annealing, their spatial distribution is difficult to control. In contrast, prior to annealing, the same material typically exhibits fewer than ~2 NV centers per confocal spot across the sample surface, except for the non-$\{100\}$ sectors (e.g., the four "petal" features near the diamond center).

We performed several experiments to suppress this background through improved surface preparation. These procedures were developed after the implanted sample measured in this work and are intended for future studies. The primary approach we investigated was argon–chlorine (Ar/Cl_2) inductively coupled plasma reactive ion etching (ICP-RIE) (*18–20*), which removes polishing-induced damage from the near-surface region. While such etching processes have shown success in low-nitrogen CVD diamond, their application to high-nitrogen HPHT diamond ($[N_s^0] \sim 200\,\mathrm{ppm}$) requires further evaluation, as etching-induced damage may be more pronounced. We performed etching with varying RF power (Table S2), followed by annealing and PL imaging at different depths to quantify the resulting NV density. In each case, more than $\sim 10\,\mu\mathrm{m}$ of material is removed to eliminate the original polishing-damaged layer. For high RF power (250 W), the NV density peaks at ~5–7 per confocal spot at a depth of ~500 nm (Fig. S13b), whereas reducing the power to 20 W lowers the density to ~1 per spot (Fig. S13c), comparable to the as-received HPHT diamond prior to annealing. Further reduction in RF power yields diminishing improvement while reducing

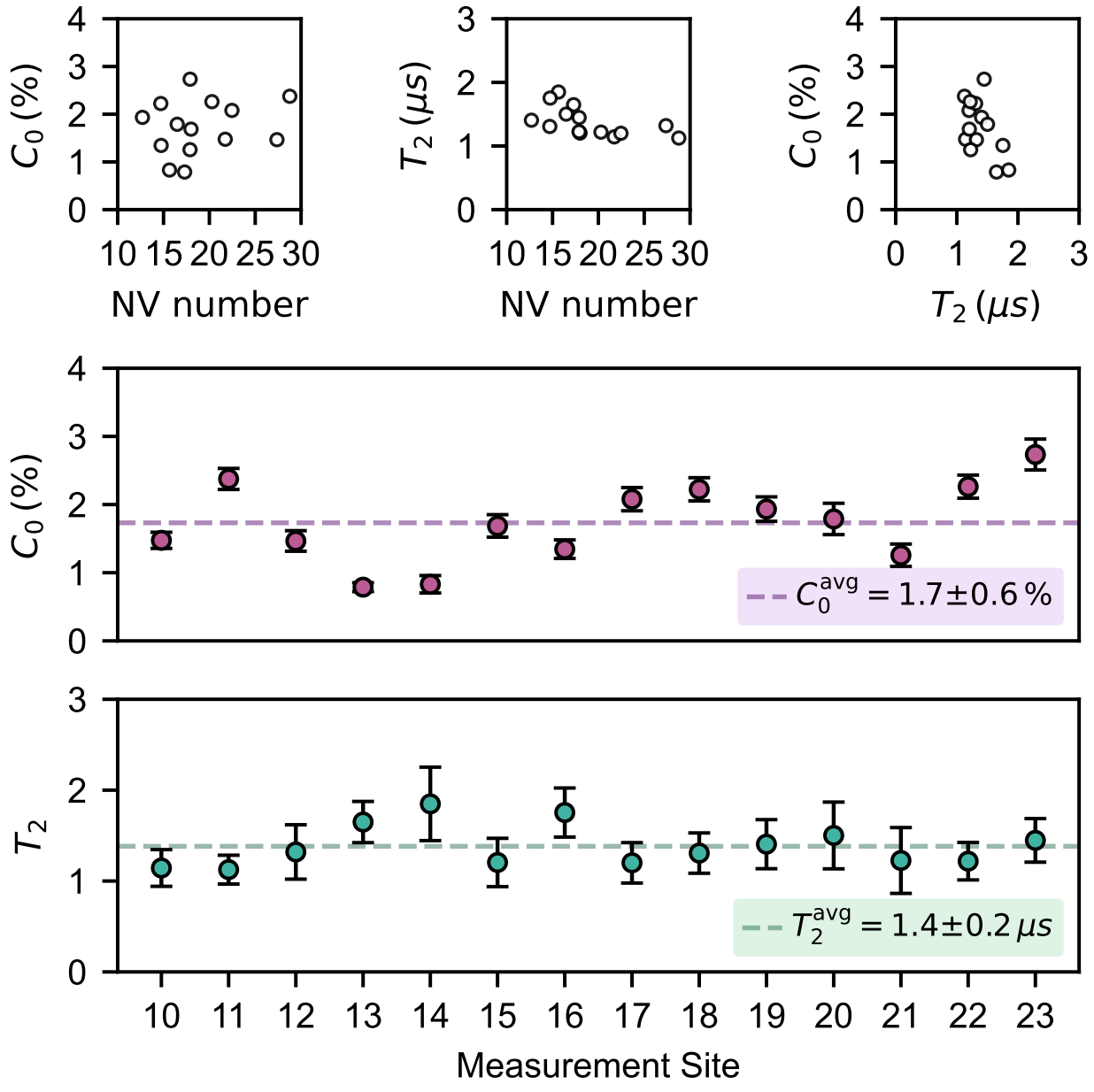


Figure S12: Spin properties versus NV number. (a) Scatter plots of spin-state-dependent PL contrast C_0, coherence time T_2, and NV number per site, showing no clear correlation between spin properties and NV count. (b) Distribution of C_0 and T_2 across measurement sites. T_2 remains relatively constant around 1.4 μs, while C_0 exhibits a larger spread (~1-3%).

the etch rate. We note that similar surface NV density can also be achieved with shorter etching time by combining an initial high-power etch with a subsequent low-power step to remove 3–4 μm of fast-etch-induced damage.

Table S2: Ar/Cl_2 ICP-RIE etching recipes and O_2 surface finishing step.

Step	**Gas**	**Flow (sccm)**	**RF Power (W)**	**ICP Power (W)**	**Etch Rate (nm/hr)**	**NVs / confocal spot**
Etch Recipe 1	Ar/Cl_2	25 / 40	250	400	3000	~5-7
Etch Recipe 2	Ar/Cl_2	25 / 40	20	400	480	$\lesssim 1$
Etch Recipe 3	Ar/Cl_2	25 / 40	8	400	117	$\lesssim 1$
O_2 finish	O_2	30	20	700	–	–

References and Notes

1. T. Yamamoto, *et al.*, Isotopic identification of engineered nitrogen-vacancy spin qubits in ultra-pure diamond. *Physical Review B* **90** (8), 081117 (2014), doi:10.1103/PhysRevB.90.081117, https://link.aps.org/doi/10.1103/PhysRevB.90.081117.

2. W. J. P. van Enckevort, H. G. M. Lochs, Photoluminescence tomography as a method to image point-defect distributions in crystals: Nitrogen-vacancy pairs in synthetic diamonds. *Journal of Applied Physics* **64** (1), 434–437 (1988), doi:10.1063/1.341212.

3. R. Fastenau, *et al.*, *Diffusion limited reactions in crystalline solids*, Tech. rep., Technische Hogeschool Delft (Netherlands) (1982).

4. S. T. Alsid, *et al.*, Photoluminescence Decomposition Analysis: A Technique to Characterize N-*V* Creation in Diamond. *Physical Review Applied* **12**, 044003 (2019), doi:10.1103/PhysRevApplied.12.044003, https://link.aps.org/doi/10.1103/PhysRevApplied.12.044003.

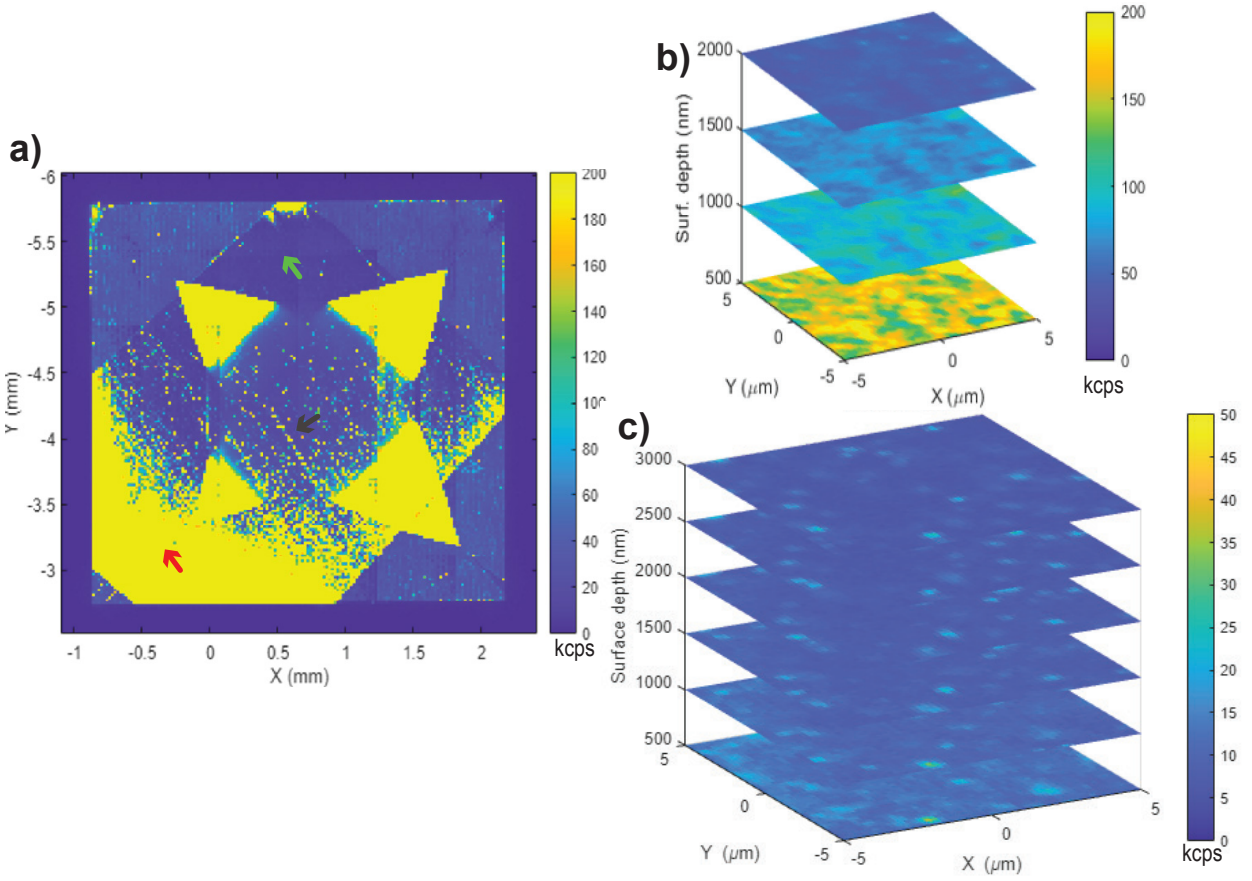


Figure S13: Mitigation of polishing-induced NV background. (a) Confocal PL map of a scaife-polished HPHT diamond measured at a depth of ~1 μm after annealing. The diamond is predominantly {100}-oriented, with four high-PL "petal" regions arising from growth sectors. Strong and highly non-uniform background PL is observed due to polishing-induced damage. Red arrows indicate regions of severe damage that can overwhelm implantation signals, black arrows highlight streak-like features consisting of isolated high-NV spots, and green arrows indicate regions with relatively low background. Three corners have been previously etched, showing indented features. (b) Confocal PL maps at increasing surface depths following high-power (250 W) Ar/Cl_2 ICP-RIE etching. The near-surface region ($\lesssim$ 1 μm) exhibits elevated PL corresponding to ~5-7 NV centers per confocal spot, indicating etch-induced damage. (c) Confocal PL maps at increasing depths following low-power (20 W) Ar/Cl_2 ICP-RIE etching. The near-surface PL is significantly reduced, with NV densities ~1 per confocal spot, comparable to as-received HPHT diamond without annealing.

5. J. A. Mitchell, *et al.*, Parallel simulation via SPPARKS of on-lattice kinetic and Metropolis Monte Carlo models for materials processing. *Modelling and Simulation in Materials Science and Engineering* **31** (5), 055001 (2023), doi:10.1088/1361-651X/accc4b, https://doi.org/10.1088/1361-651X/accc4b.

6. J. Koike, D. Parkin, T. Mitchell, Displacement threshold energy for type IIa diamond. *Applied Physics Letters* **60** (12), 1450–1452 (1992).

7. G. Kresse, J. Hafner, Ab initio molecular dynamics for liquid metals. *Physical Review B* **47** (1), 558 (1993).

8. G. Kresse, J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical Review B* **54** (16), 11169 (1996).

9. P. E. Blöchl, Projector augmented-wave method. *Physical Review B* **50** (24), 17953 (1994).

10. J. P. Perdew, Generalized gradient approximations for exchange and correlation: A look backward and forward. *Physica B: Condensed Matter* **172** (1-2), 1–6 (1991).

11. J. P. Perdew, K. Burke, M. Ernzerhof, Generalized Gradient Approximation Made Simple. *Physical Review Letters* **77**, 3865–3868 (1996), doi:10.1103/PhysRevLett.77.3865, https://link.aps.org/doi/10.1103/PhysRevLett.77.3865.

12. D. Sheppard, P. Xiao, W. Chemelewski, D. D. Johnson, G. Henkelman, A generalized solid-state nudged elastic band method. *The Journal of Chemical Physics* **136** (7) (2012).

13. B. Slepetz, M. Kertesz, Divacancies in diamond: a stepwise formation mechanism. *Physical Chemistry Chemical Physics* **16** (4), 1515–1521 (2014).

14. A. Tejero-Cantero, *et al.*, SBI – A toolkit for simulation-based inference (2020), doi:10.48550/arXiv.2007.09114, http://arxiv.org/abs/2007.09114, arXiv:2007.09114 [cs.LG].

15. C. Durkan, A. Bekasov, I. Murray, G. Papamakarios, Neural spline flows. *Advances in Neural Information Processing Systems* **32** (2019).

16. S. Rajendran, N. Zobrist, A. O. Sushkov, R. Walsworth, M. Lukin, A method for directional detection of dark matter using spectroscopy of crystal defects. *Physical Review D* **96** (3), 035009 (2017), doi:10.1103/PhysRevD.96.035009.

17. T. Schuelke, T. A. Grotjohn, Diamond polishing. *Diamond and Related Materials* **32**, 17–26 (2013).

18. P. Appel, *et al.*, Fabrication of all diamond scanning probes for nanoscale magnetometry. *Review of Scientific Instruments* **87** (6) (2016).

19. T. X. Zhou, R. J. Stöhr, A. Yacoby, Scanning diamond NV center probes compatible with conventional AFM technology. *Applied Physics Letters* **111** (16) (2017).

20. C. Lee, E. Gu, M. Dawson, I. Friel, G. Scarsbrook, Etching and micro-optics fabrication in diamond using chlorine-based inductively-coupled plasma. *Diamond and Related Materials* **17** (7-10), 1292–1296 (2008).