

# Optical properties of a diamond NV color center from capped embedded multiconfigurational correlated wavefunction theory

John Mark P. Martirez

Applied Materials and Sustainability Sciences, Princeton Plasma Physics Laboratory,

Princeton, New Jersey, 08540

[martirez@pppl.gov](mailto:martirez@pppl.gov), [martirez@princeton.edu](mailto:martirez@princeton.edu)

## Abstract

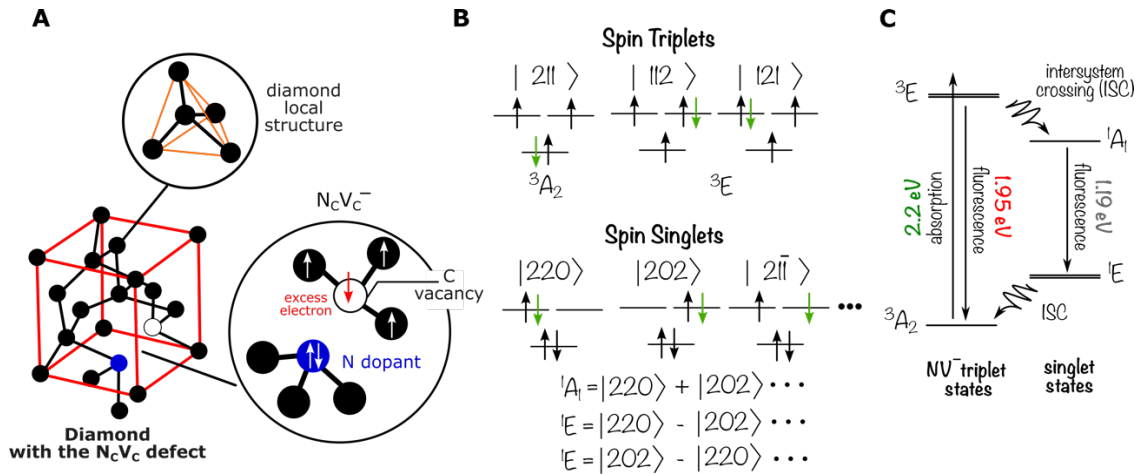
Diamond defects are among the most promising qubits. Modelling their properties through accurate quantum mechanical simulations can further their development into robust units of information. We use the recently developed capped density functional embedding theory (capped-DFET) with the multiconfigurational  $n$ -electron valence second-order perturbation theory to characterize the electronic excitation energies for different spin manifolds of the well-characterized negatively charged substitutional N defect adjacent to a vacancy ( $V_C$ ) in diamond ( $N_C V_C^-$ ). We successfully reproduce vertical excitation energies for both triplet and singlet states of  $N_C V_C^-$  with errors  $< 0.1$  eV. Unlike other embedding methods, capped-DFET exhibits robust predictions that are approximately independent of the embedded cluster size: it only requires a cluster to contain the defect atoms and their nearest neighbors (as small as a 40-atom capped cluster). Furthermore, our method is free from slowly converging Coulomb interactions between charged defects, and thus also only weakly dependent on supercell size.

## Introduction

Quantum coherence, e.g., of electron and nuclear spins, is key to establishing quantum bits (qubits) suitable for quantum computing.<sup>1-3</sup> Quantum coherence in qubits is the core property that classical bits do not possess, in that the written value of a classical bit can only be either "0" or "1" at a given moment, whereas a quantum bit can be a superposition of both "0" and "1" states, only resolving its value upon measurement.<sup>3</sup> Qubits can be made from isolated atoms or defects in a material, whose spin and charge are well isolated from its host, so as not to interfere with qubit preparation, manipulation, and read-out.<sup>1-3</sup> Quantum coherence is not only potentially revolutionary for computing, but also for communication, data security, and sensing.<sup>4</sup>

Defect properties, such as spin, charge, and optical response, are dictated by: the atoms that compose them; their local environment, e.g., coordination number, identity of neighbors, strain, etc.; and, especially in the case of charged defects, the presence of other (charged) defects. A quantum defect host meant for quantum computing applications, ideally, has a wide band gap (so that the host does not absorb the light used to probe the defects), can be made with high purity, is diamagnetic, induces small spin-orbit

coupling, and has naturally occurring isotopes with no nuclear spin.<sup>3,5</sup> Diamond is among the most promising (and therefore, among most studied) materials for solid-state quantum applications because it can host an array of magnetically (via electron spin) and optically active point defects, namely, vacancies, substitutions (e.g., H, N, B, O, etc.), interstitials (e.g., Si, Ge, P, Sn, Pb, etc.), and combinations thereof that are either neutral or positively or negatively charged.<sup>6,7</sup> Take, for example, a well-known solid-state qubit: the negatively charged substitutional N defect ( $N_C$ ) with an adjacent vacancy ( $V_C$ ) in diamond ( $N_C V_C^-$ ). The removal of two adjacent C atoms from the diamond lattice creates six C atoms with only three neighbors, rendering them coordinatively saturated (leaves a “dangling bond” per C) with nominally one unpaired electron (“open shell”) on each. However, when an N atom occupies one of the vacancy sites, three of the six C atoms form bonds with the N and the N, having one more electron than C, will be coordinatively saturated (“closed shell”), forming three covalent bonds and a lone pair of electrons. This leaves three open shell C atoms, each with one ( $sp^3$ ) orbital containing only one electron (**Fig. 1A**). This defect is called an  $NV^0$  center. If this defect is negatively charged ( $N_C V_C^-$ ), the additional electron can pair with one of the three unpaired  $sp^3$  electrons (**Fig. 1A**). Indeed, in the ground state,  $NV^-$  center has two unpaired electrons – a spin triplet.<sup>8</sup> **Fig. 1B** shows the electronic configuration of the spin triplet ground- and first excited-state states labeled  $^3A_2$  and  $^3E$  (the superscript 3 refers to their spin multiplicity and the alphanumeric labels originate from the defect symmetry – *vide infra*), respectively, showing how the three C  $sp^3$ -derived states at the vacancy are populated (one lower energy and two degenerate higher energy orbitals).



**Fig. 1. Atomic and electronic structures of a prototypical defect in diamond.** **A**, diamond is composed of C atoms each with four nearest neighbors that are  $\sim 1.5$  Å away in a tetrahedral local geometry (top inset). The cubic diamond unit cell contains eight unique C atoms with cell side lengths  $\sim 3.6$  Å (red box). An NV center consists of a N substitutional defect adjacent to a C vacancy. **B**,  $N_C V_C^-$  center has a triplet spin ground state (i.e., two unpaired electrons). Electron occupation of the three orbitals derived from the dangling  $sp^3$  orbitals are shown for both triplet ( $^3A_2$  and  $^3E$ ) and singlet states ( $^1E$  and  $^1A_1$ ). For the spin singlet, only three of the six possible electronic configurations are shown. The bra-ket notation, e.g.,  $|211\rangle$ , refers to the wavefunction associated with a specific occupation of the three

defect orbitals: either doubly (2), spin-up singly (1), or spin-down singly ( $\bar{1}$ ) occupied or unoccupied (0). **C**, diamond absorbs in the deep UV (5.5 eV), whereas the ground-state spin triplet  $\text{NV}^-$  center absorbs and fluoresces within the visible ( $\sim 2.0$  eV) - hence referred to as a "color center". The energy diagram schematically shows the relative energetic positions of the four electronic states in **B** and the various optical processes that link them.

In the  $^3A_2 \rightarrow ^3E$  excitation, a spin-down electron (**Fig. 1B**, green pointing down arrow) in the lowest energy frontier orbital gets excited to one of the two singly occupied highest energy orbitals. While  $^3A_2$  is singly degenerate,  $^3E$  is doubly degenerate due to the two energetically equivalent ways of adding an excited electron to the two (degenerate) higher energy orbitals. Moreover, higher-energy spin singlets result from low-spin pairing of all the electrons. Among the primary electron configurations are where one of the two highest defect frontier orbitals is empty – there are two such configurations, and where the degenerate states are singly occupied but of opposite spins (**Fig. 1B**). These are only three of the six possible spin singlet configurations (*vide infra*). The spin singlet ground- and first excited-states are labeled  $^1E$  (doubly degenerate) and  $^1A_1$  (non-degenerate), respectively. The  $^1E$  states are “destructive” combinations of the  $|220\rangle$  and  $|202\rangle$  wavefunctions (and some other terms), whereas the  $^1A_1$  state is a “constructive” combination of the aforementioned wavefunctions (and some other contributions). **Fig. 1C** shows an energy diagram for all four states and the electronic processes that connect them: absorption, fluorescence, and intersystem crossing (ISC). The excitation and fluorescence energies for  $^3A_2 \rightarrow ^3E$  (2.2 and 1.95 eV, visible)<sup>9</sup> and  $^1E \rightarrow ^1A_1$  (1.19 eV, infrared)<sup>10</sup> are much smaller than diamond’s band gap (5.5 eV, ultraviolet).<sup>11</sup>

By assigning Boolean logic elements “1” and “0” to different spin magnetic quantum states ( $m_s$ ) sub-levels, (a qubit of) information can be “written.” The relative energy of the different  $m_s$  of the spin triplet can be manipulated, e.g., with a static magnetic field (along the defect axis) where parallel:  $m_s = +1$  ( $|211\rangle$ ) and anti-parallel:  $m_s = -1$  ( $|2\bar{1}\bar{1}\rangle$ ) spins have splitting energy proportional to the external magnetic field strength, with the latter being higher in energy (Zeeman shift). Under zero field, the splitting between the  $m_s = 0$  ( $|21\bar{1}\rangle$ ) and  $m_s = \pm 1$  sub-levels (zero-field splitting) are in the microwave region (ground-state zero-field magnetic resonance of 2.88 GHz).<sup>12</sup> Therefore, microwave pulses can be used to manipulate coherence between the spin states that are then optically probed (read out) via the  $^3A_2 \rightarrow ^3E$  excitation and its subsequent spin-dependent fluorescence (emission).<sup>13</sup> Spin-forbidden relaxation via ISC of triplets to singlets, e.g., from  $^3E$  to the non-magnetic  $^1A_1$  state (**Fig. 1C**) will render the qubit magnetically inactive (“shelving”).<sup>14</sup>

Fundamental research opportunities abound in diamond-based quantum materials, notably establishing rules for discovery of viable new defects and alternative host materials is in its infancy.<sup>1,2,5,7,15</sup> The quantum mechanical (QM) aspect of the mode of operation of qubits (and multiconfigurational nature of the defect electronic states) motivates the importance of using accurate multireference QM theories to predict the electronic structure of qubits considering different electronic spins and charge states.

Understanding these fundamental properties will be key to qubit characterization, operation, design, and eventually optimization. QM simulations within accurate electronic structure theory to understand and characterize the magnetic and optical properties of existing and candidate defects would be valuable for addressing these challenges. QM simulations of qubits have been performed largely within (single-determinant, i.e., single-configuration) density functional theory (DFT)<sup>16,17</sup> and studies that used excited-state theories to understand them have only been conducted for a few known defects, notably  $N_C V_C^-$ .<sup>18-24</sup> In cases where experiments are available, comparisons between DFT simulations and experiments are at best approximate and incomplete, although DFT may offer a qualitative picture on the nature of the magnetic and optical properties of a qubit.<sup>25</sup> When exploring a new phase space for defects and hosts, a more rigorous electronic structure theory would be required for a reliable prediction to be made to ascertain that novel systems are worth experimental investigation.

Here, we computationally characterize the electronic structure of the  $N_C V_C^-$  color center via capped density functional embedding theory (capped-DFET)<sup>26</sup> and for the first time test the capped-DFET method to characterize localized defects in condensed matter. Specifically, the goal of this study is to test the ability of this method to reproduce experimental magnetic and electronic properties of the  $N_C V_C^-$  defect when combined with a rigorous multiconfigurational correlated wavefunction theory, namely, the complete active space self-consistent field (CASSCF) theory<sup>27,28</sup> to incorporate *static electron correlation*, together with a multireference second order perturbation (MRPT2) method, specifically, the  $n$ -electron valence second-order perturbation theory (NEVPT2),<sup>29</sup> to include *dynamic electron correlation*. Static correlation is particularly important to accurately describe electronic state degeneracies as well as excited states.

## Results

*Principles of capped-DFET theory.* In standard DFET by Huang, Pavone, and Carter, one divides the system into fragments, typically, into two: the region of interest (typically, a small cluster) and the remaining atoms (cluster's environment).<sup>30</sup> One then obtains the inter-fragment interaction via the optimized effective potential (OEP) method within DFT, where one maximizes the extended Wu-Yang functional<sup>31</sup>

$$W = E_{DFT}^{cl}[\rho^{cl}, V_{emb}] + E_{DFT}^{env}[\rho^{env}, V_{emb}] - \int V_{emb} \rho^{full} dr \quad \text{Eq. 1}$$

The above equation is a constrained maximization of the sum of the fragment DFT energies ( $E_{DFT}^{cl}[\rho^{cl}, V_{emb}] + E_{DFT}^{env}[\rho^{env}, V_{emb}]$ ), where, *cl*, *env*, and *full* refer to the cluster, environment, and the full system. The embedding potential ( $V_{emb}$ ) is the Lagrange multiplier to be variationally optimized and the electron densities ( $\rho$ ) define the gradient:

$$\frac{\delta W}{\delta V_{emb}(\mathbf{r})} = \frac{\delta E_{DFT}^{cl}}{\delta V_{emb}(\mathbf{r})} + \frac{\delta E_{DFT}^{env}}{\delta V_{emb}(\mathbf{r})} - \frac{\delta \{ \int V_{emb}(\mathbf{r}) \rho^{full} dr \}}{\delta V_{emb}(\mathbf{r})} \quad \text{Eq. 2}$$

$$\frac{\delta W}{\delta V_{emb}(\mathbf{r})} = \rho^{cl}(\mathbf{r}) + \rho^{env}(\mathbf{r}) - \rho^{full}(\mathbf{r}) \quad \text{Eq. 3}$$

with  $\rho^{cl}(\mathbf{r})$  and  $\rho^{env}(\mathbf{r})$  respectively the self-consistent densities of the cluster and environment fragments, whereas  $\rho^{full}(\mathbf{r})$  is the self-consistent density of the full system. At maximum, naturally, **Eq. 3** equals 0 for all  $\mathbf{r}$ . In other words, the action of the optimized local potential  $V_{emb}(\mathbf{r})$  on the cluster and environment leads to the reproduction of the full electron density from the sum of their densities. DFET in conjunction with correlated wavefunction methods has been used extensively to simulate reactions on metal surfaces and in solution both in ground (e.g., thermal<sup>32-39</sup> and electrochemical<sup>40-43</sup>) and electronically excited (e.g., plasmonic catalysis<sup>44,45</sup>) states. Various QM-based partitioning and embedding methods exist in the literature that either relies on density or orbital partitioning. For a recent review on the subject, see ref. <sup>46</sup>.

When a chemical system is fragmented, electrons are also partitioned along with the atoms, in doing so, a user assumes how electrons distribute among the fragments. When partitioning across metallic, ionic, and non-covalent (hydrogen or van der Waals) bonds, both the charge and atom partitioning are rather straightforward. The *sp*-valence electrons are free-electron-like for metals, and ions and molecules generally are closed shell therefore fragments may simply be assigned electrons according to the total number of electrons of their constituent elements. For covalent bonds, following the same rule (method 1) will likely lead to open-shell fragments; if one considers the polarity of the bond, one can also heterolytically break the bonds, where one assigns all shared electrons between a bond to one fragment (method 2).<sup>47,48</sup> Alternatively, one can partition using method 1 and “cap” the atoms at the fragmentation points using elements of complementary valence (method 3), e.g., a broken single bond will be capped with a monovalent element. Martirez and Carter subsequently introduced a modification of DFET<sup>26</sup> following method 3 to make the method suitable for when covalent or ionic bonds are broken due to partitioning by introducing “capping” elements. This method is akin to the link-atom method implemented in QM/Molecular Mechanics (QM/MM) methods<sup>49</sup> and capping moieties schemes implemented with frozen density embedding theory,<sup>50</sup> which are typically used to model large organic and bio-organic molecules. It is therefore similar in spirit to method 2 charge partitioning, without *a priori* input from the user on how charge is to be distributed across bonds, granted a valence complementary atom is used to cap the dangling bonds. In so-called capped-DFET, Martirez and Carter proposed that in addition to the introduction of capping elements to all fragments, an auxiliary fragment composed of all capping elements may be used to project out the densities of the capping elements, which are not native to the original system.<sup>26</sup> In doing so, a completely local representation of the  $V_{emb}(\mathbf{r})$  is preserved, charge partitioning is simplified, and most importantly, use of projector orbitals or implementation of an orbital localization scheme is avoided. **Eq. 1** is thus rewritten as

$$W = E_{DFT}^{cl+cap1}[\rho^{cl}, V_{emb}] + E_{DFT}^{env+cap2}[\rho^{env}, V_{emb}] - \int V_{emb} (\rho^{full} + \rho^{cap1+cap2}) dr \quad \text{Eq. 4}$$

**Eq. 3** (the gradient) is thus rewritten as:

$$\frac{\delta W}{\delta V_{emb}(\mathbf{r})} = \rho^{cl+cap1}(\mathbf{r}) + \rho^{env+cap2}(\mathbf{r}) - \rho^{full}(\mathbf{r}) - \rho^{cap1+cap2} \quad \text{Eq. 5}$$

where  $\rho^{cl+cap1}(\mathbf{r})$ ,  $\rho^{env+cap2}(\mathbf{r})$ , and  $\rho^{cap1+cap2}$  are the self-consistent electron densities of the cluster with capping elements 1 (*cap1*), environment with capping elements 2 (*cap2*), and the combined capping elements *cap1+cap2*, *i.e.*, the auxiliary fragment.  $V_{emb}(\mathbf{r})$  therefore, furnishes the missing influence of the environment on the cluster and of the cluster on the environment not provided by their respective capping elements.

DFET and capped-DFET can enable regional accurate treatment of the cluster, *e.g.*, an accurate correlated wavefunction (CW) description locally, while also incorporating the influence of its environment quantum mechanically via the optimized  $V_{emb}$ . Accordingly, we modify the one-electron Hamiltonian ( $H^0$ ) of the cluster by adding the local  $V_{emb}$  operator to yield the embedded Hamiltonian

$$H_{emb}^0 = H^0 + V_{emb} \quad \text{Eq. 6}$$

We refer to such a method as capped embedded CW or capped emb-CW theory. Further discussion on the nature of the approximation and implementation of capped-DFET can be found in Ref. <sup>26</sup>. Emb-CW and capped emb-CW are used to provide a regional (within the cluster) correction to the total energy of the chemical system, enabling the calculation of reaction energies and barriers or even full energy landscapes for electronic ground and excited states, the latter being beyond the ability of time-independent DFT. When evaluating excited-state energies using emb-CW, one approximates the influence of the environment as fixed and the excited state as local: a strongly coupled electron-hole pair (exciton) is generated and only the orbitals of the cluster atoms are involved. Thus, only localized excited states will be properly described. For a fixed number of electrons, the excitation energy from the ground ( $n = 0$ ) to the  $n$ th excited state within capped emb-CW is simply

$$\Delta E(0 \rightarrow n) = E_{CW}^{cl+cap1}(n) - E_{CW}^{cl+cap1}(n = 0) \quad \text{Eq. 7}$$

where  $E_{CW}^{cl+cap1}(n)$  and  $E_{CW}^{cl+cap1}(n = 0)$  are the energies of the  $n$ th excited and ground electronic states of the  $V_{emb}$ -embedded capped cluster within CW, respectively. From **Eq. 6** and **7**, it becomes clear that the contributions of  $V_{emb}$  to the excitation energy ( $\Delta E_{emb}(0 \rightarrow n)$ ) is not only through the self-consistent optimization of the orbitals and electron densities, but also as a first-order correction to the excitation energy in the form of

$$\Delta E_{emb}(0 \rightarrow n) = \int V_{emb} [\rho_n^{cl+cap1}(\mathbf{r}) - \rho_0^{cl+cap1}(\mathbf{r})] d\mathbf{r} \quad \text{Eq. 8}$$

where  $\rho_n^{cl+cap1}(\mathbf{r}) - \rho_0^{cl+cap1}(\mathbf{r})$  is the difference in the electron densities of state  $n$  and the ground state. Indeed, we found that this difference to be highly localized within the cluster with little to no contributions from the capping elements (*vide infra*).

Martirez and Carter demonstrated the utility of capped emb-CW to calculate the metal-to-ligand charge-transfer spectrum of a ruthenium aqua bipyridine complex attached to a small titanium oxide cluster<sup>51</sup> meant to model a piece of a dye-sensitized solar cell, within complete-active-space self-consistent field (CASSCF)<sup>27,28</sup> with second order perturbation theory (CASPT2).<sup>52-54</sup>

*Bulk and defect electronic structure within DFT.* **Fig. 2** shows the periodic supercell models used in this study and **Table 1** demonstrates that the de-orbitalized re-regularized strongly constrained appropriately normed (r<sup>2</sup>-SCAN-L)<sup>55,56</sup> exchange-correlation (XC) functional exhibits exceptional lattice constant prediction compared to experiment.

**Table 1.** Predicted lattice constant and indirect band gap<sup>a</sup> for a pristine diamond (C<sub>8</sub> supercell) from different DFT approximations compared to the experimental structure and fundamental band gap.

| Method                                 | Cubic $a$ (Å)           | $E_g$ (eV)          |
|----------------------------------------|-------------------------|---------------------|
| Experiment                             | $3.56712 \pm 0.00005^b$ | $5.480 \pm 0.004^c$ |
| DFT-r <sup>2</sup> -SCAN-L             | 3.5677                  | $4.22 \pm 0.02$     |
| DFT-HSE06                              | -                       | $5.30 \pm 0.02$     |
| G <sub>0</sub> W <sub>0</sub> @DFT-PBE | -                       | $5.50 \pm 0.02^d$   |

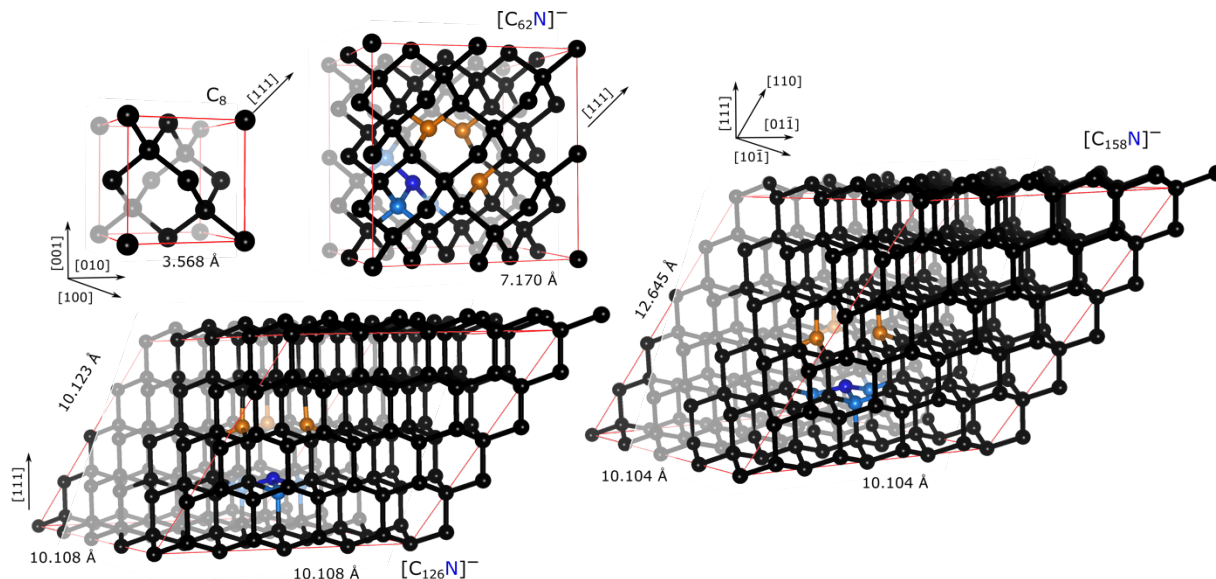
<sup>a</sup>Difference between the energies of the conduction band minimum and valence band maximum;

<sup>b</sup>X-ray diffraction at room temperature, ref. <sup>57</sup>; <sup>c</sup>photoemission-inverse photoemission, ref. <sup>11</sup>

<sup>d</sup>formally corresponds to the fundamental gap

However, as with most nonhybrid, local, semi-local, and even meta-GGA XC functionals, DFT- r<sup>2</sup>-SCANL severely underpredicts the fundamental gap ( $E_g$ ) by more than 1 eV when using the eigenenergies of the frontier band states (**Table 1**). Using the DFT- r<sup>2</sup>-SCAN-L-optimized structure, we recalculated the electronic structure using the hybrid-DFT Heyd-Scuseria-Ernzerhof XC functional with the 2006 parameterization (HSE06).<sup>58,59</sup> We also performed a quasi-particle calculation via a low-scaling implementation<sup>60</sup> of single-shot GW<sup>61-63</sup> (G<sub>0</sub>W<sub>0</sub>) using the one-electron orbitals from DFT with Perdew-Burke-Ernzerhof GGA XC functional<sup>64</sup> (G<sub>0</sub>W<sub>0</sub>@DFT-PBE). These two higher levels of theory predict  $E_g$ s in much better agreement with the experiment (**Table 1**). **Fig. S1** shows the densities of states for diamond from all three levels of theory. Based on these results, hereafter, we use DFT- r<sup>2</sup>-SCAN-L for *atomic* structure optimizations and subsequently use DFT-HSE06 to calculate the ground-state *electronic* structures. Despite higher computational cost, we deemed DFT-HSE06 to be a much more reliable baseline

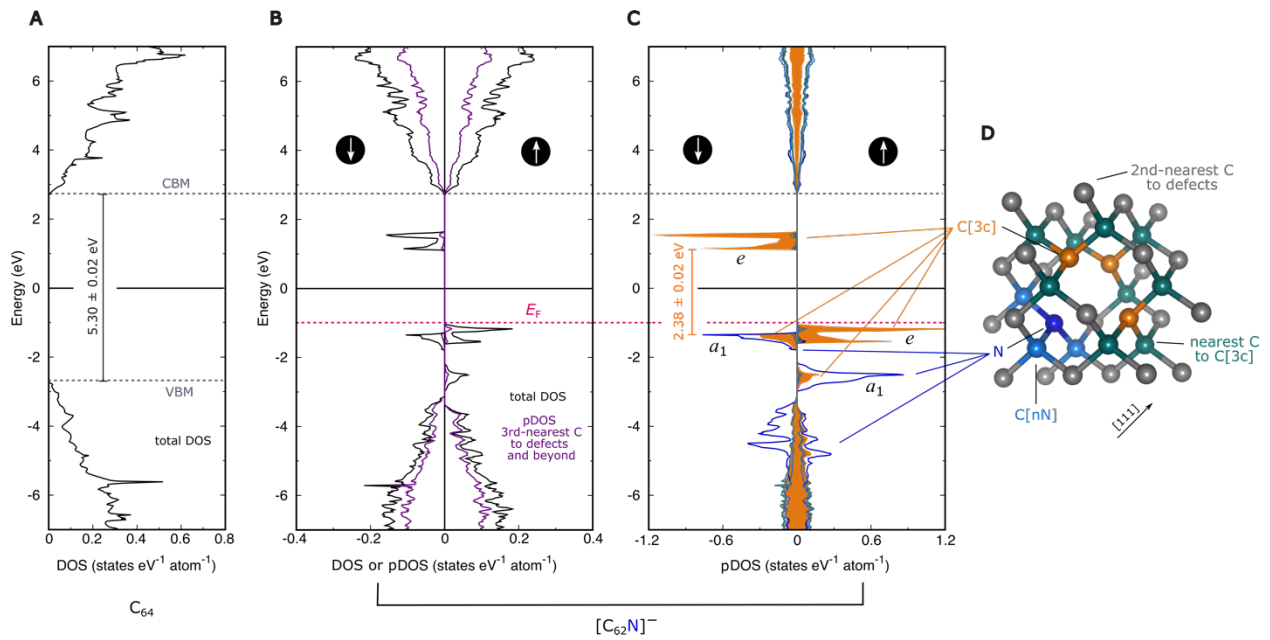
theory than DFT- $r^2$ -SCAN-L to perform  $V_{emb}$  optimization, based on the accuracy of their  $E_g$  predictions (**Table 1**). The optimized  $V_{emb}$  is subsequently used in capped emb-CW theory (*vide infra*).



**Fig. 2. Bulk atomic structures of diamond** – flawless cubic ( $C_8$ ) and with  $N_C V_C^-$  defect:  $C_{62}N^-$ ,  $C_{126}N^-$ , and  $C_{158}N^-$  (as labelled), optimized within DFT- $r^2$ -SCAN-L. The dimensions of the supercell are given (if only one is given, then all sides are equal). N – dark blue; C – black. For clarity, the three C atoms adjacent to  $V_C$  are orange whereas the atoms bound to N are light blue. The red lines mark the supercell edges. The lattice vectors according to the primitive cubic cell are shown. **Table S1** summarizes the optimized cell parameters.

**Fig. 3A** shows the calculated DFT-HSE06 densities of states (DOS) for a flawless diamond using a  $C_{64}$  supercell. **Fig. 3B** shows the corresponding DOS for a diamond with  $N_C V_C^-$  defect for a  $C_{62}N^-$  supercell (**Fig. 2**). **Fig. S1** shows the DFT- $r^2$ -SCAN-L, DFT-HSE06, and  $G_0W_0 @DFT-PBE$  DOS for the flawless  $C_8$  diamond structure for comparison. For pure diamond, the DOS plot is centered, i.e., zeroed, midway between the valence band (VB) and conduction band (CB) edges – the band gap center, which is also the Fermi level ( $E_F$ ). The DOS of the pristine diamond clearly show the expected band gap, closely matching the experimental fundamental gap (*vide supra*). The pristine diamond frontier band states are simply the C-C bonding (VB) and anti-bonding states (CB). For the defective structure DOS (**Fig. 3B**), we match the position of the CB minimum (CBM) with the CBM of the pure diamond of the same supercell size (**Fig. 3A**). Furthermore, for the defective diamond, the spin majority (up) and minority (down) states are separately plotted. The  $E_F$  shifts from the band gap center to the highest occupied defect state in the spin-up channel (**Fig. 3B**). The DOS profile appears to be only weakly sensitive to supercell size from  $C_{62}N^-$  to  $C_{158}N^-$  supercell - we performed this analysis only within DFT- $r^2$ -SCAN-L (**Fig. S2**) because of the tremendous computational cost of a dense  $k$ -point mesh that would be required to generate a corresponding DFT-HSE06 DOS for the  $C_{126}N^-$  and  $C_{158}N^-$  supercells.

**Fig. 3B** also shows the projected DOS (pDOS) of the third-nearest-neighbor C atoms of the defects and beyond, which exhibit an almost pristine-diamond-like profile. To further resolve atomic contributions in the frontier states, **Fig. 3C** plots the pDOS of the C atoms adjacent to  $V_C$  (three-fold coordinated C or C[3c], orange) and their nearest neighbors (teal), C atoms next to the substitutional  $N_C$  (C[nN], light blue), and  $N_C$  (dark blue). In the same plot, the pDOS of the second-nearest neighbors of the defects (C[3c] and N) are also shown (grey). **Fig. 3D** illustrates the locations of the different types of C atoms relative to C[3c] and  $N_C$ , showing atoms in the same color-coding scheme as their pDOS. **Fig. 3C** shows that the diamond with  $N_C V_C^-$  defect exhibits pristine-like pDOS for the C atoms away from the defect (including the second nearest-neighbor C atoms).



**Fig. 3. Bulk electronic DOS of diamond ( $C_{64}$ ) and diamond with a  $N_C V_C^-$  defect ( $C_{62}N^-$ ) from DFT-HSE06.** **A**, pure diamond DOS. The DOS are normalized by the total number of atoms in the supercell. The DOS of the defective structure (**B** and **C**) shows separately the spin down (left) and up (right) manifolds. **B** also shows the pDOS for the C atoms far from the defect (as labelled). The pDOS are normalized over the number of atoms whose atomic contributions are summed. **C** shows the contributions of the defect atoms along with their nearest neighbors (as shown in **D**). **D** shows a partial structure of the supercell annotating the position of the defect atoms and their neighbors. Optimized atomic structures from DFT-r<sup>2</sup>-SCAN-L were used. Note the difference in DOS scale between plots.

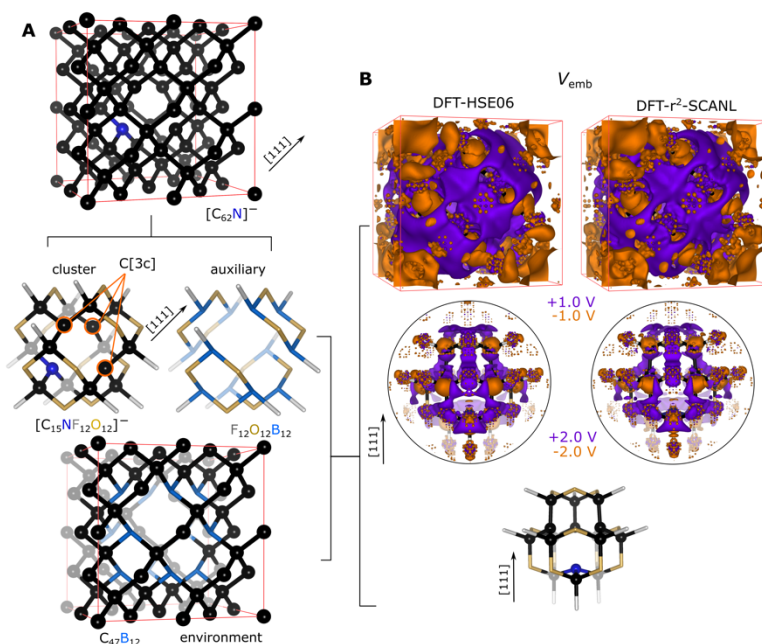
The highest occupied state of the N dopant (corresponding its lone pair) is near the VB edge. The mid-gap states are defect states associated with the “dangling bonds” of the three C[3c]s with some contributions from a N 2p state. The lower energy mid-gap state is doubly occupied (both up and down spin states are occupied) – an  $a_1$  state (*vide infra*), which has contributions from C[3c]s and N. The higher energy doubly degenerate gap states are singly occupied (their corresponding spin down states are above  $E_F$ ) –  $e$  states (*vide infra*), which has contributions exclusively from C[3c]s. The two unpaired electrons localized at three

C[3c]s give rise to the  $V_C$ -centered triplet spin multiplicity (**Fig. S3**). The approximate spin triplet defect excitation energy is the difference in the eigenenergies of the  $e$  and  $a_1$  states in the spin down manifold with the  $e$  orbitals being empty in the ground state for this spin channel, which is 2.38 eV (**Fig. 3**). The defect excitation energy depends weakly, although non-negligibly, on supercell size from  $C_{62}N^-$  to  $C_{158}N^-$ , with the value decreasing by 0.08 eV from the smallest to the biggest supercell within DFT-  $r^2$ -SCAN-L (**Fig. S2**). We will show later that excitation energies from our embedded capped cluster method are even less sensitive to supercell size.

This (Kohn-Sham DFT) orbital-energy-based method of approximating the defect excitation energies in solids is widely adopted in the literature, with mixed success, and is naturally dependent on the XC functional used.<sup>65-67</sup> Excitation energies based on one-electron orbital energies are at best an approximation of the energy difference between adding an electron and taking away an electron from and to the vacuum, *i.e.*, ionization energy minus electron affinity (IP-EA) known as the fundamental gap.<sup>68</sup> The IP approximation is based on Koopmans’ theorem that the energy of the highest occupied one-particle (Hartree-Fock) orbital of a molecule is its negative IP,<sup>69,70</sup> which was later extended to include the approximation that the energy of the lowest energy empty one-particle orbital is the molecule’s EA (extended Koopmans’ theorem) – both at the limit of no orbital relaxation upon electron addition or removal.<sup>71</sup> These approximations were later generalized to one-particle orbitals from Kohn-Sham density functional theory for molecules and solids.<sup>72-74</sup> A local defect-centered excitation, on the other hand, corresponds to the energy needed to form an exciton (bound electron and hole) – optical gap.<sup>68</sup> The optical (excitonic) gap should be lower than the fundamental gap by the exciton binding energy.<sup>68</sup> The experimental vertical excitation energy (VEE) of the  $N_C V_C^-$  defect was reported to be around 2.18 eV – lattice relaxation in the excited state preceding an electronic relaxation (emission) lowers the energy of emitted light, corresponding to the sharp zero phonon line (ZPL) in the fluorescence spectrum: 1.95 eV for  $N_C V_C^-$ .<sup>9,75</sup> The electronic structures predicted by DFT-HSE06 closely match the experimental and quasi-particle method ( $G_0W_0$ ) band gap of pure diamond (5.30 vs. 5.50 eV; **Table 1, Figs. S1 and 3A**) and provides a reasonable upper bound to the VEE of the spin triplet state of the defect (2.38 eV, **Fig. 3C**) when using DFT-  $r^2$ -SCAN-L-optimized atomic structures. We wish to evaluate the VEE of this defect within a more accurate and formally excited-state theory that explicitly solves for the excited-state wavefunction and incorporates the multiconfigurational nature of the spin singlet state of the defect. We next describe the partitioning scheme we used and the embedded capped cluster calculations it enabled.

*System partitioning and  $V_{emb}$  optimization.* The charge and spin density difference plots in **Fig. S3** and pDOS in **Fig. 3C** offer guidance on how to construct a suitable cluster. **Fig. S3** shows that the excess charge

and spin are highly localized at the three C[3c]s. Further, the atomic pDOS (**Fig. 3C**) illustrates the degree of the contributions to the defect states (labelled  $a_1$  and  $e$ ) of the C atoms surrounding the defect atoms (C[3c]s and N) and suggests that the C atoms nearest to the defect atoms (teal and light blue C atoms in **Fig. 3D**) adequately capture the spatial extent of the electronic influence of the defect. Therefore, starting from the full  $C_{62}N^-$  supercell, we carved out a  $C_{15}N$  cluster where the  $C_{15}$  includes the three C[3c]s, three C[nN], and nine nearest neighbors of C[3c]s. We designate the remaining C atoms ( $C_{47}$ ) as the environment fragment (**Fig. 4A**). The  $C_{15}N$  yields 12 coordinatively unsaturated C atoms at the edges, with each having to share a missing common neighbor *individually* with two other edge C atoms. We replaced the unshared missing C atoms with F atoms (terminal or monovalent capping) – a total of 12, and the doubly shared missing C atoms with O atoms (bridge or divalent capping) – also a total of 12, to yield a  $C_{15}NF_{12}O_{12}$  capped cluster (**Fig. 4A**). On the other hand, the environment fragment contains 24 edge C atoms that are missing



**Fig. 4. Partitioning and embedding schemes.** **A**, partitioning and capping scheme for the  $C_{62}N^-$  periodic supercell. Cluster, environment, and auxiliary fragments' structures shown as labelled. **B**, optimized  $V_{emb}$  within DFT-HSE06 and DFT-r<sup>2</sup>-SCANL as labelled at two isosurface levels.  $V_{emb}$ s feature repulsive (purple) and attractive (orange) potentials.

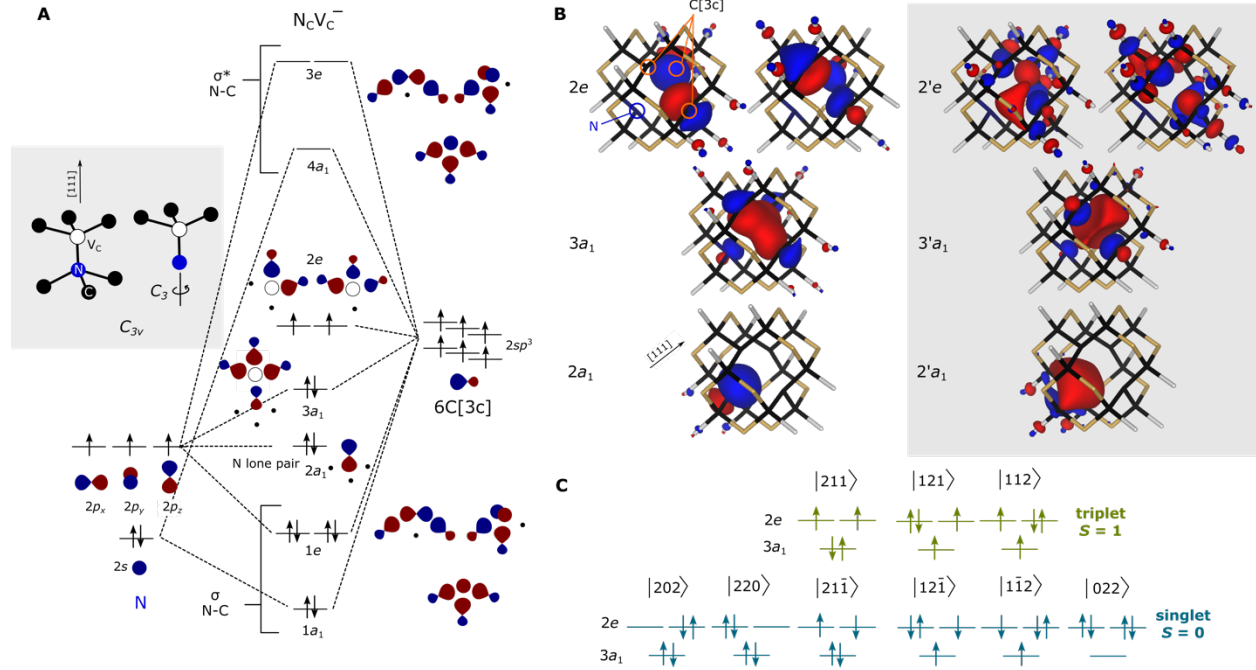
a C neighbor, however, each of them shares a missing neighbor *jointly* with two other edge C atoms. We replaced the triply shared missing C atoms with B atoms (facial or trivalent capping) – a total of 12, to yield a  $C_{47}B_{12}$  capped environment (**Fig. 4A**). We do not relax the positions of the capping elements, which were placed at the lattice sites of the atoms they replace. This capping scheme results in fragments that are coordinatively saturated at the fragmentation points. The auxiliary fragment composed of combined capping elements from both the cluster and environment ( $F_{12}O_{12}B_{12}$ ) also results in coordinatively saturated

closed-shell hypothetical molecular fragments (**Fig. 4A**), which is a consequence of valence complementarity principle we abide by in deciding what elements to cap the cluster and environment fragments. Specifically, the auxiliary fragment contains four closed-shell  $F_3O_3B_3$  units, where each B atom bonds with two O and one F atom. We also chose capping elements that would form bonds with the edge C atoms with optimal bond lengths that would be similar to the C-C bonds ( $\sim 1.54$  to  $1.55$  Å) that they replace (H would thus be too small to satisfy this criterion, in addition to its inability to form more than one covalent bond).

We optimize the  $V_{emb}(\mathbf{r})$  (see **Methods**) following the recipe of the capped-DFET scheme by Martirez and Carter<sup>26</sup> using  $C_{15}NF_{12}O_{12}$  with an excess electron ( $[C_{15}NF_{12}O_{12}]^-$ ),  $C_{47}B_{12}$ , and  $F_{12}O_{12}B_{12}$  to reconstruct the charge density of  $C_{62}N^-$  within DFT-HSE06, as well as DFT- $r^2$ -SCAN-L for comparison. The optimized  $V_{emb}(\mathbf{r})$ s shown in **Fig. 4B** features a repulsive potential shell (purple) surrounding the defect cluster with a complementary attractive potential (orange) within the environment that limits the spillover of the electron density of the environment into the defect region. Conversely, the positive potential surrounding the defect cluster partially delocalizes the electron density from the periphery of the cluster into the environment region. At higher absolute potentials, the potential repolarizes the C–O and C–F bonds to offset the electron pull of O and F (notice the attractive potential at the boundary C and positive potential surrounding O). The  $V_{emb}(\mathbf{r})$ s from the two XC functionals share the same features (**Fig. S4** shows their difference). Using these optimized  $V_{emb}$ , we performed MRPT2 calculations for the  $[C_{15}NF_{12}O_{12}]^-$  cluster to calculate excitation energies for both spin triplet and singlet. Note that we use the same  $V_{emb}(\mathbf{r})$  optimized for the ground-state spin triplet for the embedded capped cluster spin singlet to retain the same Hamiltonian for both spins – we use after all also the same spin-triplet optimized structure. The use of the same  $V_{emb}(\mathbf{r})$  enables us to directly compare energies of the two spins without the risk of artifacts that may originate from two different  $V_{emb}$ s. We thus assume that the dielectric screening of the environment in response to defect transitions (electronic excitation and spin crossover) will not change significantly as long as the atomic structure remains the same.

*Defect electronic structures within emb-capped-CW.* To initiate an MRPT2 calculation, we begin with a series of CASSCF calculations to determine the dominant electronic configurations for both ground and excited states for the two spin states, as well as to construct a set of self-consistently optimized many-body wavefunctions that will be the basis for the multi-configurational second-order perturbation energy correction. CASSCF requires one to specify an active space (AS) size, *i.e.*, number of valence electrons and a set of orbitals to generate a series of electronic configurations to define a set of determinants to be used to expand the wavefunction(s). Because the (canonical) orbitals are self-consistently optimized along with the contributions of each electronic configuration to the overall wavefunction (configuration interaction,

CI), CASSCF captures self-consistently the (static) electronic correlation that is important to describe degenerate/near degenerate states as well as to robustly represent excited states. Because we are interested in calculating the electronic transitions involving the  $N_C V_C^-$  defect, we consult the molecular orbital (MO) diagram resulting from the interaction of N atom with six C[3c]s at two neighboring  $V_C$ s (**Fig. 5A**). As the N atom fills one of the vacancies, N uses its valence orbitals ( $2s2p$ ) to mix with three ( $2sp^3$ -like) dangling orbitals of the C[3c]s, yielding three N-C bonding ( $\sigma$ ) and antibonding ( $\sigma^*$ ) covalent single bonds, and a



**Fig. 5. Defect MO, calculated orbitals, and dominant electronic configurations.** **A**, grey box: simplification of the structure of the defect site emphasizing the  $C_{3v}$  symmetry. Corresponding molecular orbital diagram of the  $N_C V_C^-$  defect. The defect's  $C_{3v}$  symmetry dictates orbital and degeneracy of the electronic states. The valence  $3a_1$  and  $2e$  orbitals originate from the non-bonding  $sp^3$  states of the three carbon atoms at the vicinity of the vacancy. **B**, natural frontier orbitals  $a_1$  and doubly degenerate  $e$  calculated from the state-averaged density matrix of the three-state triplet embedded SA-CASSCF with CAS(6e,8o) from apvdz basis set. Isosurface value of  $0.06 \text{ (e/bohr}^3)^{1/2}$ . **C**, electronic configurations for the spin triplet and singlet states showing only the occupation of the valence  $3a_1$  and  $2e$  orbitals with spin-up and spin-down electrons (up and down arrows). In the bra-ket notation, 2, 1,  $\bar{1}$ , or 0 mean that an orbital is either doubly, spin-up singly, spin-down singly, or unoccupied, respectively. Note that these electronic configurations are not necessarily degenerate and in fact are not.

N-centered lone pair. Because the  $N_C V_C^-$  defect exhibits a local symmetry of  $C_{3v}$ , we label non-degenerate  $\sigma$  and  $\sigma^*$  as  $1a_1$  and  $4a_1$  (invariant with respect to  $C_3$  rotation and reflection), respectively, whereas we label the doubly degenerate  $\sigma$  and  $\sigma^*$  as  $1e$  and  $3e$ , respectively. The N lone pair, also possess an  $a_1$  symmetry, and thus, labelled as  $2a_1$ . The remaining three  $2sp^3$ -like C[3c] orbitals linearly mix where all three are in phase ( $3a_1$ ) and a doubly degenerate state where at least two  $2sp^3$  orbitals are out of phase ( $2e$ ). For the ground state  $N_C V_C^-$ ,  $3a_1$  is doubly occupied, whereas  $2e$  are each singly occupied, hence the spin triplet

(two unpaired electron) magnetization. If one considers all the aforementioned orbitals (and electrons) in the AS, the AS size will be 12 electrons in 10 orbitals or (12e,10o). Furthermore, for the excited states, higher energy virtual orbitals may correlate with the lone-pair  $2a_1$  and vacancy-centered  $3a_1$  and  $2e$  orbitals non-trivially. If these four orbitals are also to be included, labelled  $2'a_1$ ,  $3'a_1$ , and  $2'e$ , then the full AS size would be (12e,14o) – a gargantuan task even with the reduced size of the system enabled by embedding. We therefore systematically increase the AS in the CASSCF calculations to determine the sensitivity of the calculated excitation energies: (4e,6o) with 6o = ( $3a_1$ ,  $2e$ ,  $2'e$ ,  $3'a_1$ ); (6e,8o) with 8o = ( $2a_1$ ,  $3a_1$ ,  $2e$ ,  $2'e$ ,  $2'a_1$ ,  $3'a_1$ ); (10e,12o) with 12o = ( $1e$ ,  $2a_1$ ,  $3a_1$ ,  $2e$ ,  $2'e$ ,  $2'a_1$ ,  $3'a_1$ ,  $3e$ ) and the full (12e,14o). We found (6e,8o) to be the optimal AS that balances precision and computational cost (*vide infra*). We calculated three states for each spin (based on local  $C_{3v}$  symmetry and frontier orbital symmetry and occupation, **Fig. 5A**): for  $S = 1$ , the ground state is labelled  $^3A_2$  and the first excited state is a doubly degenerate  $^3E$ ; for  $S = 0$ , the doubly degenerate ground state is labelled  $^1E$  and the first excited state is  $^1A_1$ .

We confirm via continuous symmetry measure<sup>76</sup> implemented within the VASPKIT code<sup>77</sup> that our  $[C_{15}NF_{12}O_{12}]^-$  cluster possesses a  $C_{3v}$  symmetry, i.e., invariant toward  $E$ , two  $C_3$ , and three  $\sigma_v$  symmetry operations.<sup>76</sup> **Fig. 5B** shows the optimized state-averaged natural orbitals for a (6e,8o) AS from a three-state  $S = 1$  CASSCF calculation using the apvdz basis set. **Fig. S5** shows the corresponding state-averaged natural orbitals for a (6e,8o) AS from a three-state  $S = 0$  CASSCF calculation, which resemble those of  $S = 1$ . Despite not explicitly imposing symmetry in the calculations, the natural orbitals in **Fig. 5B** exhibit the expected symmetry of the ( $2a_1$ ,  $3a_1$ ,  $2e$ ) orbitals (linear combination of vacancy  $2p$  states) in **Fig. 5A**, although the natural orbitals go beyond the single-particle picture provided in MO diagrams. The virtual ( $2'a_1$ ,  $3'a_1$ ,  $2'e$ ) as expected exhibit the same approximate symmetry of the ( $2a_1$ ,  $3a_1$ ,  $2e$ ) orbitals but have more nodes (linear combination of  $3p$ -like orbitals). **Figs. S6** and **S7** show the state-averaged natural orbitals from three-state  $S = 1$  with AS of (10e,12o) and (12e,14o) that now feature some or all the C–N  $\sigma$  ( $1a_1$ ,  $1e$ ) and  $\sigma^*$  ( $4a_1$ ,  $3e$ ) orbitals. For the (6e,8o) AS, all the active (natural) orbitals participate, i.e., their occupation number changes relative to the dominant ground-state configuration, in at least one electronic configuration whose absolute CI coefficient ( $c$ ) is  $\geq 0.05$  for at least one of the three calculated states. The dominant configurations are shown in **Fig. 5C** in the basis of the ( $2a_1$ ,  $3a_1$ ,  $2e$ ) natural orbitals shown in **Figs. 5B** and **S4** for both  $S = 1$  and 0. These configurations are used to build the ground and excited state wavefunctions and account for 93-96 % configuration state functions used in the expansion of the wavefunction (**Table 2**).

**Table 2.** CI coefficients ( $c$ ) in the state-averaged natural orbital basis from SA-CASSCF (6e,8o) AS.<sup>a</sup>

| State                       | c (3a <sub>1</sub> ,2e) |        |       |        |        | Σ  c  <sup>2</sup> |
|-----------------------------|-------------------------|--------|-------|--------|--------|--------------------|
| S = 1                       |                         |        |       |        |        |                    |
|                             |                         | 211⟩   | 121⟩  | 112⟩   |        |                    |
| <sup>3</sup> A <sub>2</sub> |                         | 0.978  | 0.000 | 0.000  |        | 0.956              |
| <sup>3</sup> E              |                         | 0.000  | 0.174 | 0.949  |        | 0.931              |
|                             |                         | 0.000  | 0.949 | -0.174 |        | 0.931              |
| S = 0                       |                         |        |       |        |        |                    |
|                             | 202⟩                    | 220⟩   | 211̄⟩ | 121̄⟩  | 112̄⟩  | 022⟩               |
| <sup>1</sup> E              | -0.468                  | 0.468  | 0.564 | -0.104 | -0.425 | 0.000              |
|                             | 0.399                   | -0.399 | 0.662 | 0.425  | -0.104 | 0.000              |
| <sup>1</sup> A <sub>1</sub> | 0.570                   | 0.570  | 0.000 | 0.000  | 0.000  | -0.532             |
|                             |                         |        |       |        |        | 0.933              |

<sup>a</sup>Electronic configurations illustrated in **Fig. 5C**, SA-CASSCF natural orbitals are shown in **Fig. 5B** ( $S = 1$ ) and **Fig. S5** ( $S = 0$ )

For  $S = 1$ , a single determinant ( $|211\rangle$ ) accounts for 96 % of the  $^3A_2$  state. A single excitation from the  $3a_1$  to  $2e$  orbital (some constructive and destructive linear combination of the  $|121\rangle$  and  $|112\rangle$  configurations) – accounting for 93 % electron correlation – yields the  $^3E$  states. For  $S = 0$ , the ground and excited states are more truly multiconfigurational with four and three configurations, respectively, almost equally dominating the wavefunction expansion. This easily explains the failure of single-determinant methods (e.g., DFT) to describe even the ground state of the spin singlet  $N_C V_C^-$ . Of note, the ground  $^1E$  states are two destructive combinations of the  $|202\rangle$  and  $|220\rangle$  configurations, whereas  $^1A_1$  is the constructive combination of the two configurations. The  $^1E$  states also include  $|21\bar{1}\rangle$ ,  $|12\bar{1}\rangle$ , and  $|1\bar{1}\bar{2}\rangle$ , which are spin flipped ground and excited states configurations of  $S = 1$ . Also, of great significance is that  $^1A_1$  includes a non-trivial contribution ( $|c|^2 \sim 0.28$ ) from a double excitation  $|022\rangle$ , therefore, methods that only account for single excitations (e.g., time-dependent DFT) to simulate the first excited state of the spin singlet  $N_C V_C^-$  will most certainly lead to incorrect predictions.<sup>78</sup> Furthermore, (ignoring spin-orbit coupling) the absence of contributions from the spin-flipped excited  $S = 1$  configurations in  $^1A_1$  state indicate that spin crossover from  $^3E$  to  $^1A_1$  is not simply a spin flip relaxation.

*Defect excitation energies within emb-capped-CW.* We followed the SA-CASSCF calculations with state-specific NEVPT2 (SS-NEVPT2 or simply NEVPT2) to calculate the first vertical excitation energies (VEE) for  $S = 1$  and  $S = 0$ , as well as the excited-state singlet-triplet splitting energy, i.e., intersystem crossing (ISC) between  $^1A_1$  and  $^3E$ , using **Eq. 7**. Note that VEE is an upper bound of the so-called zero phonon line (ZPL) because it ignores structural relaxation following the electronic transition (fixed nuclei approximation). The first VEE for  $N_C V_C^-$  are also experimentally known for both  $S = 1$  and 0 along with the ZPL, the former from optical absorption spectroscopy whereas the latter from photoluminescence

measurements.<sup>9,10,75,79,80</sup> The  $^1A_1 \rightarrow ^3E$  ISC energy has also been reported in the literature, which is approximated from ISC rates.<sup>81-83</sup>

**Table 3** summarizes the  $^3A_2 \rightarrow ^3E$ ,  $^1E \rightarrow ^1A_1$ , and  $^1A_1 \rightarrow ^3E$  VEEs calculated here from different basis set sizes, AS sizes, DFT approximations to derive an optimized  $V_{emb}$ , and MRPT2 methods. Even without imposing symmetry, the two highest energy spin triplet states are (numerically) degenerate (within 0.36 meV of each other), thus, aptly assigned as  $^3E$ . Likewise, the two lowest-energy spin singlet states are degenerate (difference of 0.23 meV), thus, aptly assigned as  $^1E$ . The same table shows the best available experimental values along with values from the computational literature that went beyond DFT eigenvalue differences in calculating excitation energies. We first focus on our emb-NEVPT2 calculations along with

**Table 3. Calculated vertical excitation energies (VEE) for low-lying transitions in  $NcV_6^-$  center from emb-MRPT2 using an  $[C_{15}NF_{12}O_{12}]^-$  capped cluster carved from  $C_{62}N^-$  periodic supercell.**

| Exc.                                         | Exp. VEE <sup>a</sup><br>(ZPL) <sup>b</sup>    | Basis set                    | AS size   | Calculated VEE [eV]    |                                     |                          | Recent literature<br>(beyond DFT methods)<br><div><div> error  ≤ 0.10 eV</div><div> error  ≤ 0.15 eV</div><div> error  &gt; 0.15 eV</div></div> |
|----------------------------------------------|------------------------------------------------|------------------------------|-----------|------------------------|-------------------------------------|--------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|
|                                              |                                                |                              |           | V <sub>emb</sub> (DFT) |                                     |                          |                                                                                                                                                 |
|                                              |                                                |                              |           | DFT-HSE06              |                                     | DFT-r <sup>2</sup> SCANL |                                                                                                                                                 |
|                                              |                                                |                              |           | Emb-NEVPT2             | Emb-CASPT2 <sub>w</sub> /IPEA shift | Emb-NEVPT2               |                                                                                                                                                 |
| <sup>3</sup> A <sub>2</sub> → <sup>3</sup> E | <b>2.18<sup>c</sup></b><br>(1.95) <sup>d</sup> | apvdz                        | (4e,6o)   | 2.32                   | 2.36                                | 2.35                     | 2.32 GW+BSE <sup>i</sup>                                                                                                                        |
|                                              |                                                |                              | (6e,8o)   | 2.25                   | 2.31                                | 2.29                     | 2.05 CI cRPA <sup>j</sup>                                                                                                                       |
|                                              |                                                |                              | (10e,12o) | 2.23                   | 2.30                                | 2.27                     | 2.00 QET-beyond-RPA <sup>k</sup>                                                                                                                |
|                                              |                                                | apvtz+<br>apvdz <sup>h</sup> | (4e,6o)   | 2.30                   | -                                   | -                        | 2.31 DMET-NEVPT2 <sup>l</sup>                                                                                                                   |
|                                              |                                                |                              | (6e,8o)   | 2.24                   | -                                   | -                        | 2.15 QDET EDC@G <sub>0</sub> W <sub>0</sub> <sup>m</sup>                                                                                        |
|                                              |                                                |                              | (6e,8o)   | 2.24                   | -                                   | -                        | 1.98 DMFET-FCIQMC <sup>n</sup>                                                                                                                  |
| <sup>1</sup> E → <sup>1</sup> A <sub>1</sub> | <b>1.26<sup>e</sup></b><br>(1.19) <sup>f</sup> | apvdz                        | (4e, 6o)  | 1.28                   | 1.25                                | 1.27                     | 2.14 cluster CASSCF <sup>o</sup>                                                                                                                |
|                                              |                                                |                              | (6e,8o)   | 1.28                   | 1.26                                | 1.28                     | 0.59 GW+BSE <sup>i</sup>                                                                                                                        |
|                                              |                                                |                              | (10e,12o) | 1.29                   | 1.27                                | 1.29                     | 0.89 CI cRPA <sup>j</sup>                                                                                                                       |
|                                              |                                                | apvtz+<br>apvdz <sup>h</sup> | (6e,8o)   | 1.27                   | -                                   | -                        | 1.20 QET-beyond-RPA <sup>k</sup>                                                                                                                |
|                                              |                                                |                              | (6e,8o)   | 1.27                   | -                                   | -                        | 1.02 DMET-NEVPT2 <sup>l</sup>                                                                                                                   |
|                                              |                                                |                              | (6e,8o)   | 1.27                   | -                                   | -                        | 0.81 QDET EDC@G <sub>0</sub> W <sub>0</sub> <sup>m</sup>                                                                                        |
| <sup>1</sup> A <sub>1</sub> → <sup>3</sup> E | (0.32-0.41) <sup>g</sup>                       | apvdz                        | (4e, 6o)  | 0.43                   | 0.38                                | 0.47                     | 0.99 DMFET-FCIQMC <sup>n</sup>                                                                                                                  |
|                                              |                                                |                              | (6e,8o)   | 0.35                   | 0.33                                | 0.39                     | 1.35 cluster CASSCF <sup>o</sup>                                                                                                                |
|                                              |                                                |                              | (10e,12o) | 0.36                   | 0.32                                | 0.40                     | - GW+BSE <sup>i</sup>                                                                                                                           |
|                                              |                                                | apvtz+<br>apvdz <sup>h</sup> | (6e,8o)   | 0.35                   | -                                   | -                        | 0.69 CI cRPA <sup>j</sup>                                                                                                                       |
|                                              |                                                |                              | (6e,8o)   | 0.35                   | -                                   | -                        | 0.24 QET-beyond-RPA <sup>k</sup>                                                                                                                |
|                                              |                                                |                              | (6e,8o)   | 0.35                   | -                                   | -                        | 0.80 DMET-NEVPT2 <sup>l</sup>                                                                                                                   |

<sup>a</sup>VEE corresponds to the excitation from the structural ground state to an electronically excited state that is in a higher vibronic state. This transition corresponds to the needed energy to excite the electrons with the nuclei frozen at the ground state

<sup>b</sup>Zero phonon line (ZPL) accounts for phonon relaxation following vertical excitation when the structural minima of the lower and higher states are not the same

<sup>c</sup>Approximated by ref. <sup>75</sup> based on the sideband maximum of the visible absorption spectra measured at liquid N<sub>2</sub> temperature after 2 MeV electron irradiation at room temperature followed by annealing from ref. <sup>9</sup>. Furthermore, subtracting ref. <sup>18</sup>'s calculated VEE-to-ZPL energy correction of -0.23 eV (derived from a 63-atom GW with Bethe-Salpeter equation (BSE) calculation) from the experimental ZPL of 1.95 eV also yields an approximate VEE of 2.18 eV.

<sup>d</sup>Sharp peak in the visible absorption spectra measured at liquid N<sub>2</sub> temperature after 2 MeV electron irradiation at room temperature followed by annealing from Ref. <sup>9</sup>. Fluorescence spectra at 4 K with 532 nm excitation laser from ref. <sup>79</sup>

<sup>e</sup>Supercontinuum IR absorption spectra at 10 K with 35 mW pump laser, 0.071 eV blue shift from ZPL absorption associated with the first phonon sideband peak of the  $^1E \rightarrow ^1A_1$  excitation from ref. <sup>79</sup>

<sup>f</sup>IR emission after 532 nm laser excitation from ref. <sup>10</sup>, Fluorescence IR from ref. <sup>80</sup>

<sup>g</sup>Energy window within which  $^1A_1 \rightarrow ^3E$  ISC rate was evaluated in ref. <sup>81,82</sup>, corrected in an erratum ref. <sup>83</sup>

<sup>h</sup>apvtz for C and N atoms; apvdz for the capping atoms, F and O (Table S2)

<sup>i</sup>GW with Bethe-Salpeter equation (BSE) for a C<sub>254</sub>N<sup>-</sup> supercell from ref. <sup>18</sup>. Only ZPL is available for  $^1E \rightarrow ^1A_1$ .

<sup>j</sup>Configuration interaction constrained RPA (CI-cRPA) for a C<sub>510</sub>N<sup>-</sup> supercell from ref. <sup>19</sup>

<sup>k</sup>Defect orbital embedding via quantum embedding theory (QET) with “beyond” random phase approximation (RPA) for a  $C_{214}N^-$  supercell ref. <sup>20</sup>

<sup>l</sup>Density matrix embedding theory (DMET) with NEVPT2 AS(10e,9o) in restricted open-shell Hartree-Fock for a  $C_{15}N$  impurity cluster within  $C_{214}N^-$  supercell from ref. <sup>21</sup>

<sup>m</sup>Localized (defect) orbital embedding via quantum defect embedding theory (QDET) with full CI and double counting correction (EDC) from  $G_0W_0$  approximation for a  $C_{510}N^-$  supercell from ref. <sup>22</sup>

<sup>n</sup>Density matrix functional embedding theory (DMFET) with full CI quantum Monte Carlo (FCIQMC).  $C_3H_{12}N^-$  embedded fragment within  $C_{42}H_{42}N^-$  cluster from ref. <sup>23</sup>.

<sup>o</sup>Non-periodic  $C_{85}H_{76}N^-$  cluster with (6e,6o) AS from ref. <sup>24</sup>.

using the DFT-HSE06-derived  $V_{emb}$ . For the same basis set size, the  $^3A_2 \rightarrow ^3E$  VEE decreases by as much as 0.07 eV from an AS of (4e,6o) to (6e,8o), which demonstrates the importance of including the N lone pair ( $2a_1$ ) and its correlating orbital in the wavefunction expansion to accurately describe the excited state. The spatial reach of the lone pair into the vacancy site and similarity of its energy to the vacancy states can explain its importance. Indeed, **Fig 5B** shows that the  $3a_1$  state exhibits some N  $2p$  character and DFT-HSE06-calculated pDOS (**Fig. 3C**) illustrates the contribution of N to the defect  $a_1$  state. Expanding the AS to (10e,12o) only lowers the  $^3A_2 \rightarrow ^3E$  VEE by 0.01 to 0.02 eV, indicating good convergence for (6e,8o) AS. While it is possible to perform CASSCF calculations with an AS of (12e,14o), the memory requirements made it impossible to perform the subsequent MRPT2 calculation. The  $^1E \rightarrow ^1A_1$  VEE is less sensitive to the AS size, whereas  $^1A_1 \rightarrow ^3E$  ISC energy exhibits the same sensitivity as the triplet excitation energy presumably due to the sensitivity of the relative energy of the  $^3E$  state with the AS size. The effect of the basis set size going from apvdz to apvtz (additional  $s,p,d$  diffuse and polarization, *i.e.*, two  $f$ , functions are included, **Table S2**) for C and N (with the capping elements F and O remaining apvdz) led to  $\sim 0.01$  eV (within the round-off error) reduction in VEE for the same AS.

To compare our emb-NEVPT2 results to another widely used MRPT2 method, we also performed emb-CASPT2 within the apvdz basis set. The method deviates from the experiment more than emb-NEVPT2 and required an empirical IP-EA shift of 0.25 eV to achieve an acceptable level of agreement. Due to the rather arbitrary selection for the IP-EA shift, we consider the emb-CASPT2 to be less reliable. Finally, to investigate the effect of the baseline electronic structure theory, we also optimized the  $V_{emb}$  within DFT-r<sup>2</sup>-SCAN-L. Although it appears that the DFT-HSE06- and DFT-r<sup>2</sup>-SCAN-L-derived  $V_{emb}$ s are similar (**Fig. 4B** and their difference in **Fig. S4**), DFT-r<sup>2</sup>-SCAN-L-derived  $V_{emb}$  yields a higher  $^3A_2 \rightarrow ^3E$  VEE by only 0.04 eV than when using a DFT-HSE06-derived potential. The  $^1E \rightarrow ^1A_1$  VEE is again less sensitive to the choice of the DFT XC approximation for  $V_{emb}$  optimization, whereas  $^1A_1 \rightarrow ^3E$  ISC energy exhibits the same sensitivity as  $^3A_2 \rightarrow ^3E$  VEE. The somewhat similar electron densities from DFT-r<sup>2</sup>-SCAN-L and DFT-HSE06, almost agreeing to within 0.005  $e/\text{bohr}^3$  (**Fig. S8**), is consistent with this behavior, because  $V_{emb}$  depends directly only on electron density and thus on the (occupied) one-electron orbitals, not on their orbital energies. However, care must be taken when defect states incorrectly overlap with the diamond’s VB and CB states due to the more than 1 eV underprediction of diamond’s band gap

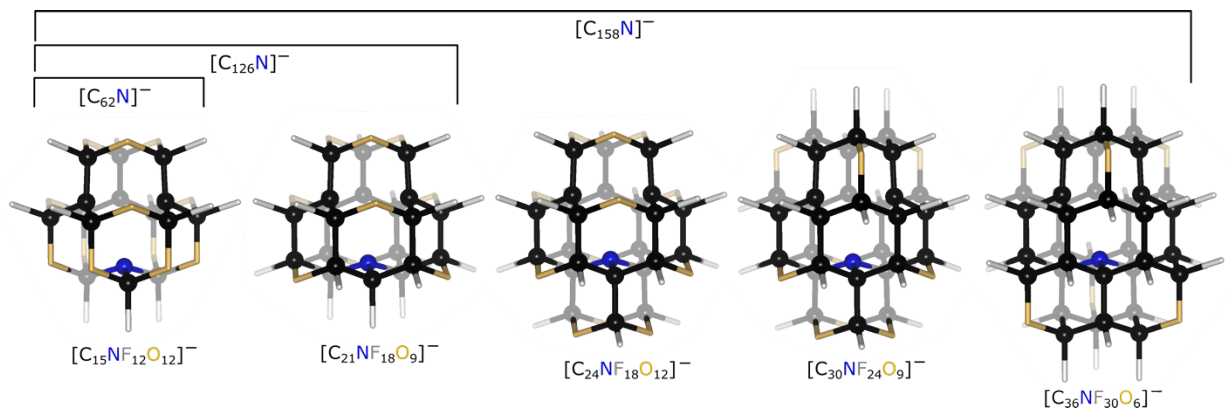
within DFT-r<sup>2</sup>-SCAN-L (**Table 1**). Such a superfluous energy placement of the defect state(s) relative to diamond’s frontier bands will likely yield a completely incorrect baseline electron density from which  $V_{emb}$  will be optimized. We therefore trust more the results derived using the DFT-HSE06  $V_{emb}$ .

The VEE and ISC energy predictions from emb-NEVPT2 with  $V_{emb}$  from DFT-HSE06 compare very well with the experimental values, with  $^1A_1 \rightarrow ^3E$  VEE overestimated by  $\sim 0.05$  to  $0.07$  eV. The  $^1E \rightarrow ^1A_1$  VEE exhibits better agreement (within  $0.03$  eV) whereas the predicted  $^1A_1 \rightarrow ^3E$  ISC energy is within the experimental range. At least for  $N_C V_C^-$ , the use of DFT-r<sup>2</sup>-SCAN-L method appears to also lead to expected errors for the VEEs ( $\sim 0.09$  to  $0.11$  eV). Our calculations are much better than other excited-state-theory-based methods applied to full periodic cells in the literature, e.g.,  $GW$  with the Bethe-Salpeter equation ( $GW+BSE$ )<sup>18</sup> and CI with constrained RPA (CI cRPA),<sup>19</sup> which concurrently require tremendously large supercell sizes for convergence – 255- and 511-atom supercells, respectively (**Table 3**). Our calculations also perform similarly or better than other more complex embedding schemes in the recent literature that requires an orbital localization procedure (orbital embedding), e.g., density matrix embedding theory (DMET) and quantum defect embedding theory (QDET), with random-phase approximation,<sup>20</sup> NEVPT2,<sup>21</sup> and full CI<sup>22,23</sup> as the CW methods, while also requiring large supercells (215- and 511-atom supercells) – **Table 3**. A recent H-capped cluster-only simulation demonstrated the need for 85-C-atom cluster ( $C_{85}H_{76}N^-$ ) to achieve  $0.2$  eV convergence relative to a 33-C-atom cluster ( $C_{33}H_{36}N^-$ ) for CASSCF excitation energies.<sup>24</sup> None of the alternative methods in **Table 3** exhibited the same accuracy as our prediction across the board. In the next section, we will discuss the relative insensitivity with cell and cluster sizes of our predictions in contrast to the other aforementioned methods.

*Bulk- and cluster-size dependence of calculated VEEs.* In **Methods (Embedding potential optimization)**, we discuss the slow convergence of the Coulomb interaction of the charged defects under periodic boundary conditions even when subject to a neutralizing uniform background charge. We argue that although the periodic supercells we chose in this study would have non-vanishing defect-defect Coulomb interactions, the existence of such interactions in both the full system and the periodic embedded capped cluster would render the optimized  $V_{emb}$  free of such interactions, therefore,  $V_{emb}$  more closely mimics a dielectric response of the environment at the dilute defect limit. Furthermore, we subsequently calculated the excitation energies from a nonperiodic embedded capped cluster that is, by construction, free of defect-defect interactions and captures only the excitation of one defect site (instead of a site and its neighbors simultaneously as in periodic calculations). To test our assertion that  $V_{emb}$  is free of slowly convergent Coulomb interactions, we optimized  $V_{emb}$ s for the same  $[C_{15}NF_{12}O_{12}]^-$  capped cluster carved out from  $C_{126}N^-$  and  $C_{158}N^-$  supercells (**Fig. 2**). To also test the effect of cluster size, from the  $C_{126}N^-$  supercell, we construct a bigger  $[C_{21}NF_{18}O_9]^-$  cluster where we added six C atoms that are nearest neighbors of C[nN],

*i.e.*, second-nearest neighbors of the dopant N. Accordingly, the additional C atoms replace six capping O atoms in  $[\text{C}_{15}\text{NF}_{12}\text{O}_{12}]^-$ , which required additional six capping F atoms (total of 18) and three new capping O atoms (total of nine) – structure in **Fig. 6**. **Fig. S9** shows the corresponding environment and auxiliary fragments and the optimized  $V_{emb}$  for capped clusters  $[\text{C}_{15}\text{NF}_{12}\text{O}_{12}]^-$  and  $[\text{C}_{21}\text{NF}_{18}\text{O}_9]^-$  from the  $\text{C}_{126}\text{N}^-$  supercell.

**Table 4** summarizes the calculated VEEs for the  $[\text{C}_{15}\text{NF}_{12}\text{O}_{12}]^-$  and  $[\text{C}_{21}\text{NF}_{18}\text{O}_9]^-$  capped clusters from the three bulk sizes with and without DFT-HSE06-derived  $V_{emb}$ . In the presence of  $V_{emb}$ , the spin triplet VEE is independent of the cluster and bulk size of origin, whereas the bare (embedding-free) case exhibits some sensitivity. Indeed, this is the hallmark of an effective embedding scheme for a localized defect. Note that although the bare capped cluster appears to compare better with the experiment (**Tables 3** and **4**), it falls short in predicting the spin-singlet VEE. The embedded capped cluster performs better for



**Fig. 6. Various clusters explored carved from three periodic supercells).** The clusters' bulk origins are shown above, and their sizes are indicated below. Atom colors match the colors of their element symbol in their labels.

the spin singlet than the bare cluster and exhibits weaker cluster size dependence (**Table 4**). The underestimation of the excited spin singlet state in the bare cluster results in a higher  $^1A_1 \rightarrow ^3E$  ISC energy compared to the embedded capped cluster, although is less sensitive to system size. The results in **Table 4** further support (in addition to **Note S1** and **Table S4**) our postulation that our model approaches the dilute defect limit even when using an optimized  $V_{emb}$  from a finite periodic model subject to slowly convergent

**Table 4. Calculated vertical excitation energies (VEE) for low-lying transitions in  $N_C V_C^-$  center from SS-NEVPT2 (6e,8o) AS using apvdz basis set and DFT-HSE06-derived  $V_{emb}$ s for different bulk and cluster sizes.**

| Exc.                    | Cluster      |                           | Calculated VEE [eV]      |      |
|-------------------------|--------------|---------------------------|--------------------------|------|
|                         | Bulk origin  | Size                      | $V_{emb}$<br>(DFT-HSE06) | Bare |
| $^3A_2 \rightarrow ^3E$ | $C_{62}N^-$  | $[C_{15}NF_{12}O_{12}]^-$ | 2.25                     | 2.21 |
|                         | $C_{126}N^-$ | $[C_{15}NF_{12}O_{12}]^-$ | 2.25                     | 2.23 |
|                         |              | $[C_{21}NF_{18}O_9]^-$    | 2.24                     | 2.12 |
|                         | $C_{158}N^-$ | $[C_{15}NF_{12}O_{12}]^-$ | 2.25                     | 2.23 |
| $^1E \rightarrow ^1A_1$ | $C_{62}N^-$  | $[C_{15}NF_{12}O_{12}]^-$ | 1.28                     | 1.18 |
|                         | $C_{126}N^-$ | $[C_{15}NF_{12}O_{12}]^-$ | 1.27                     | 1.18 |
|                         |              | $[C_{21}NF_{18}O_9]^-$    | 1.22                     | 1.10 |
|                         | $C_{158}N^-$ | $[C_{15}NF_{12}O_{12}]^-$ | 1.28                     | 1.18 |
| $^1A_1 \rightarrow ^3E$ | $C_{62}N^-$  | $[C_{15}NF_{12}O_{12}]^-$ | 0.35                     | 0.44 |
|                         | $C_{126}N^-$ | $[C_{15}NF_{12}O_{12}]^-$ | 0.37                     | 0.45 |
|                         |              | $[C_{21}NF_{18}O_9]^-$    | 0.44                     | 0.46 |
|                         | $C_{158}N^-$ | $[C_{15}NF_{12}O_{12}]^-$ | 0.36                     | 0.45 |

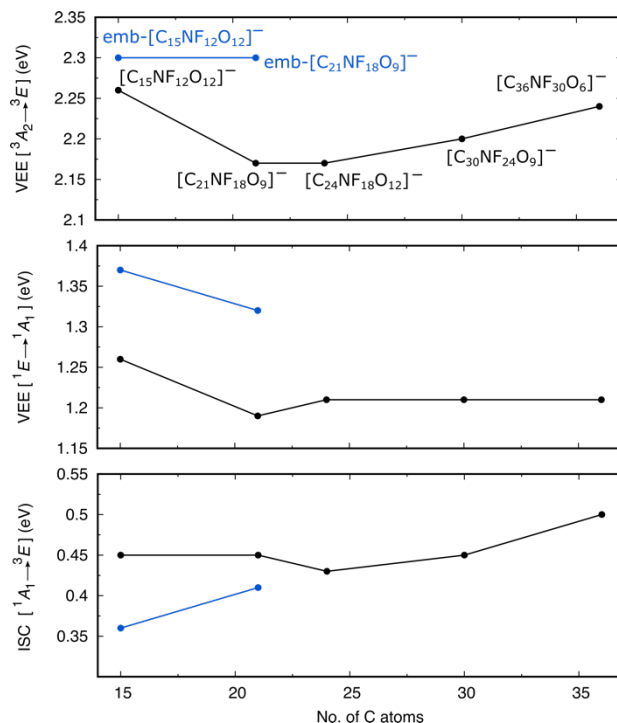
defect-defect Coulomb interactions. Furthermore, while there is some cluster-size dependence on VEEs, they are  $\leq 0.05$  eV, whereas it is 0.07 for the excited-state ISC energy. The weak dependence of VEEs to the embedded cluster size also becomes apparent when taking the difference between the excited and ground state electron densities. The density differences show that the  $^3A_2 \rightarrow ^3E$  transitions are indeed localized mainly within the unsaturated C[3c]s and dopant N (**Fig. S10**), which illustrates the perturbative nature of the correction of  $V_{emb}$  to the excitation energy (**Eq. 8**) and further justifies the appropriateness even of the smallest cluster considered here ( $[C_{15}NF_{12}O_{12}]^-$ ). **Table 4** also shows that the embedded capped cluster is overall superior to the bare capped cluster. That said, the capping procedure we introduced here alone yields acceptable levels of (in)accuracy even without embedding.

Other embedding methods based on orbital projections or orbital localization to partition the electrons between a region of interest and its environment (e.g., the ones listed in **Table 3**) still often require large supercells to reach convergence due to the inclusion of the slowly convergent Coulomb interaction in the full embedding Hamiltonian, which remains periodic (a large supercell also forgoes the need for  $k$ -point sampling, which simplifies implementation of orbital embedding for a periodic nonmetallic system). Naturally, fully periodic excited-state methods suffer from the same slowly vanishing charged defect-charged defect interactions and thus also require large supercell sizes, compounding the computational cost. Here we demonstrate that for the same cluster size,  $[C_{15}NF_{12}O_{12}]^-$ , the calculated VEEs and excited-state ISC energy is invariant with supercell size:  $C_{62}N^-$  vs.  $C_{126}N^-$  vs.  $C_{158}N^-$  (greatest sensitivity is expected going from small to medium supercell sizes, which eventually plateaus for large supercells – roughly  $L^{-1} + L^{-3}$  dependence).<sup>84,85</sup> From the results above, the embedding scheme appears to be converged with respect

to the capped cluster and bulk supercell sizes for the chosen model sizes. That said, if one wishes to calculate energy differences between systems that have different excess charge, e.g., when calculating for ionization potentials, electron affinities, and defect formation energies subject to PBCs, then an electrostatic correction to the energy or potential, or an explicit decomposition of the interaction terms between the embedded cluster and its environment that removes the periodic charge interactions (see for example work by Tölle et al.<sup>86</sup>), would still be necessary.

The fact that we perform the embedded CW calculation with an embedded capped cluster subject to an open boundary and an optimized  $V_{emb}$  free of the periodic excess charge interactions (the reason for size convergence tests), is an advantage. **Table 4** however begs the question of whether there would be a bare capped-cluster size that will approach the embedded-capped-cluster values that is computationally tenable.

*Embedding-free cluster model convergence.* **Fig. 6** also shows three bigger capped cluster sizes in addition to  $[C_{15}NF_{12}O_{12}]^-$  and  $[C_{21}NF_{18}O_9]^-$ , namely,  $[C_{24}NF_{18}O_{12}]^-$ ,  $[C_{30}NF_{24}O_9]^-$ , and  $[C_{36}NF_{30}O_6]^-$ , which we generated from an even bigger  $C_{158}N^-$  supercell (**Fig. 2**). Note that we maintain the  $C_{3v}$  symmetry of the capped cluster even with the new atomic additions to yield bigger clusters. We systematically increase the number of C atoms in the capped clusters first by completing the first nearest neighbors of  $C[nN]$  to be all C atoms, as in  $[C_{24}NF_{18}O_{12}]^-$ . For  $[C_{30}NF_{24}O_9]^-$ , we replace the six O capping atoms of  $[C_{24}NF_{18}O_{12}]^-$  with C atoms in the upper half of the cluster (vacancy side), thus recovering in the process six second-nearest C neighbors of  $C[3c]$ s. Finally, for  $[C_{36}NF_{30}O_6]^-$ , we replace the six O capping atoms of  $[C_{30}NF_{24}O_9]^-$  with C atoms in the bottom half of the cluster (N side), thus recovering in the process six second-nearest C neighbors of  $C[nN]$ s. After these C additions, we follow the same capping principles as before: we place an F or O atom at the “missing” C sites when terminal (monovalent) or bridge (divalent) capping is required, respectively. We used these capped cluster structures to investigate the convergence of the VEEs with respect to the size of the cluster without embedding. However, the capped clusters with C atoms  $\geq 24$  ( $\geq 55$  total atoms) require a smaller basis set to be computationally feasible. Capped clusters with C atoms  $\geq 30$  ( $\geq 64$  total atoms) demand jointly a smaller basis set and AS to accommodate NEVPT2 memory requirements. To make possible direct comparison between different cluster sizes, we used the largest possible basis set size and AS we can do with  $[C_{36}NF_{30}O_6]^-$ , namely, pvdz for C and N with a smaller valence double  $\zeta$  (vdz) for F and O - **Table S2**, and an AS of (4e,3o) with  $3o = (3a_1, 2e)$ , which is the smallest possible correlation space to include at least  $\sim 93\%$  of all CI contributions relative to the (6e,8o) AS for all of the states of interest (**Table 2**).



**Fig. 7. NEVPT2 calculated VEEs and ISC energy as a function of bare capped cluster size.** Clusters were carved from a bulk  $C_{158}N^-$  supercell with Emb-NEVPT2 performed using pvdz+vdz basis set with (4e,3o) AS. Emb-NEVPT2 benchmarks from a bulk  $C_{126}N^-$  are shown for comparison (in blue). Data summarized in **Table S3**.

**Fig. 7** plots the spin triplet and singlet VEEs along with the excited-state ISC energy as a function of the bare capped cluster size (black data points and curves). Note that due to the small basis set and AS the description of the excited state will be poorer than the ground state, and thus, the VEEs increase relative to bigger basis set and AS (**Table S3** shows the values for when using the apvdz basis set whereas **Table 4** shows values for apvdz with (6e,8o) AS for the two smallest clusters). We also calculated embedded capped cluster benchmarks for the two smallest clusters using structures and  $V_{emb}$  from the  $C_{126}N^-$  supercell along with the smaller basis set (pvdz+vdz) and (4e,3o) AS. **Fig. 7** also plots these embedded capped cluster benchmarks (light blue data points and curves).

The calculated spin triplet VEE from bare capped clusters dips for  $[C_{21}NF_{18}O_9]^-$  (as seen in the previous section), which gradually increases back to almost the value for  $[C_{15}NF_{12}O_{12}]^-$  at  $[C_{36}NF_{30}O_6]^-$ . Based on the trend in **Fig. 7**, the bare cluster may reach the embedded  $[C_{15}NF_{12}O_{12}]^-$  but for a cluster size too large to perform an MRPT2 calculation (or any other CW method) with the proper basis set and adequate correlation space size, which thus highlights the advantage of DFET embedding. As for the spin singlet VEE, the bare cluster reaches an apparent convergence at  $[C_{24}NF_{18}O_{12}]^-$  but remains below embedded  $[C_{21}NF_{18}O_9]^-$  by 0.07 eV. The ISC energy between the excited-state spin triplet and singlet is less sensitive to system size, likely due to error cancellation between the two excited spin states for a given cluster size.

Can a larger cluster without embedding achieve the same quality of predictions as embedded small clusters, i.e., MRPT2 or a comparable method with a satisfactory basis set size, at tenable cost? It appears that the answer is no.

## Discussion

Our study presents a detailed benchmarking of the capped-DFET method using embedded NEVPT2 to describe a complex electronic defect in a crystal, specifically, the negatively charged “NV center” in diamond. We found that using capped-DFET with emb-NEVPT2, we can reproduce VEEs and the ISC energy for spin triplet and singlet  $N_C V_C^-$  with errors less than 0.1 eV. Furthermore, we show that although we derive an embedding potential from models subject to PBCs, which suffer from slowly vanishing charged defect - charge defect Coulomb interactions as a function of cell size, our embedded capped-cluster VEEs are free from such interactions, yielding predictions that are converged even for  $V_{emb}$  derived for a 63-atom defective supercell. Therefore, we show that unlike other embedding methods, capped-DFET exhibits robust prediction that is approximately independent of the embedded cluster size, making it especially useful in reducing computational cost, as it enables use of medium supercell sizes to predict properties of defects at the dilute limit, with the caveat that the defect must be electronically localized within a few atoms. Bare large cluster sizes that remain computationally viable with our method of choice are unable to reproduce the embedded smaller cluster predictions, which highlights the advantage of embedding over purely cluster type models. With our tests, we confirm and thus establish an accuracy-retaining method that will potentially enable fully ab initio QM characterization of a variety of isolated and localized defects in diamond that may be key materials as building blocks for quantum devices.

## Methods

*Periodic boundary calculations.* We performed all Kohn-Sham (KS) DFT calculations with periodic boundary conditions (PBCs) using the Vienna Ab-Initio Simulation Package (VASP) version 6.4.<sup>87,88</sup> We used a kinetic energy cut-off of 660 eV for the planewave basis set in conjunction with the projector augmented-wave (PAW) method.<sup>89</sup> We used the Gaussian smearing method with 0.01 eV width for the Brillouin zone (BZ) integration along with  $\Gamma$ -centered finite  $k$ -point sampling (*vide infra*), unless otherwise specified. We start with an eight-atom diamond ( $C_8$ ) supercell (**Fig. 2**), which we structurally optimized within DFT-  $r^2$ -SCAN-L<sup>55,56</sup> meta-GGA XC functional with a  $k$ -point mesh of (6×6×6). DFT-  $r^2$ -SCAN-L, in comparison to the original SCAN functional, is, by construction, faster to execute, reproduces the accuracy of the original with respect to bulk properties of solids, and improves prediction of magnetization of ferromagnetic metals.<sup>55</sup> We used a real-space fine fast Fourier transform (FFT) grid of 120×120×120 (corresponding to fine grid spacing of  $\sim 0.060$  Å).

To simulate a periodic negatively charged N-doped diamond with a vacancy, we first constructed 64-, 128-, and 160-atom diamond supercells from diamond’s unit cells. Specifically, we construct the 64-atom supercell by doubling all three lattice vectors of the eight-atom conventional cubic unit cell of diamond, i.e., a  $(2\times 2\times 2)$  conventional cubic supercell. To construct the 128- and 160-atom supercells, we multiply the lattice vectors of the two-atom primitive unit cell (face-centered cubic) of diamond by  $(4\times 4\times 4)$  and  $(4\times 4\times 5)$ , respectively. We then removed one C atom and replaced an adjacent C with N (**Fig. 2**). We added an electron to the total electron count without imposing constraint on its location and performed spin-polarized KS DFT. We use a uniform positive background charge to compensate for the negatively charged defect and thus yield an over-all neutral model with a converged (electrostatic) Ewald sum. We expect the larger supercells to be most reliable in terms of reducing errors associated with the uniform background charge.<sup>84,90,91</sup> We optimized these charged defective structures without imposing symmetry within DFT-  $r^2$ -SCANL using a  $\Gamma$ -point-centered  $(3\times 3\times 3)$   $k$ -point mesh. For the 128- and 160-atom diamond supercells, we used a real-space fine FFT grid of  $180\times 180\times 180$  and  $180\times 180\times 216$ , respectively (maintaining a fine grid spacing of  $\sim 0.06$  Å, *vide supra*).

*Periodic hybrid DFT HSE06 calculations.* To manage computational cost for the two largest supercells, namely,  $C_{126}N^-$  and  $C_{158}N^-$ , we down-sampled the  $k$ -point mesh for the exact exchange (EX) kernel of DFT-HSE06 by the lowest divisible integer not equal to 1, i.e., 3, yielding only the  $\Gamma$  point – a valid approximation for insulators and for these supercells with sides  $> 10$  Å in length because we use a short-range EX inverse screening length ( $\mu$ ) of  $0.2$  Å<sup>-1</sup>,<sup>58,59</sup> which corresponds to a maximum EX interaction distance of  $2/\mu = 10$  Å, at which the decay function in HSE06 XC functional approaches  $\sim 0$ .

*Periodic low-scaling  $G_0W_0$  calculation.* We performed a low-scaling  $G_0W_0$  @DFT-PBE calculation for a  $C_8$  diamond unit cell by first calculating the ground-state electron density from DFT-PBE explicitly solving for 40 bands (16 occupied + 24 unoccupied bands). We performed the BZ integration with a  $\Gamma$ -point-centered  $(10\times 10\times 10)$   $k$ -point mesh and Gaussian smearing of 0.01 eV. We followed this with a non-self-consistent calculation to obtain DFT-PBE KS eigen states and values, solving for 192 bands (16 occupied + 176 unoccupied bands) via exact diagonalization and the frequency-dependent dielectric matrix. Using the aforementioned 192 DFT-PBE KS one-electron orbitals, we performed a low-scaling single-shot GW (so-called  $G_0W_0R$  in VASP) with 12 imaginary frequencies and time grid points to calculate quasi- particle energies.

*DOS and pDOS simulation.* We performed DOS calculations from self-consistently optimized KS eigenstates within DFT-  $r^2$ -SCANL and DFT-HSE06. For all DOS and pDOS plots, we explicitly sampled states with energies between -16 and 24 eV (using default zero orbital energy) with energy increments of 0.02 eV. We performed the BZ sampling with  $\Gamma$ -point-centered  $10 \times 10 \times 10$  and  $6 \times 6 \times 6$   $k$ -point meshes for (pristine and defective)  $C_8$  and  $\geq C_{64}$  supercells, respectively, along with the linear tetrahedron method with Blöchl corrections.<sup>92</sup>

*Embedding potential optimization.* Upon electronic and structural relaxation, the excess electron localizes at the defect site, specifically at the three three-fold coordinated C atoms adjacent to the  $V_C$  ( $C[3c]$ ) as expected (**Fig. S3**). Note that a uniform background charge does not correspond to an applied electric field and is added as a spatially constant potential in the electronic Hamiltonian. The Hamiltonian with and without such uniform potential will therefore yield the same electronic eigenstates. However, despite the convergence of the Ewald sum when using a uniform compensating background charge, the Coulomb interactions between the periodic images of the charged defect converge slowly with system size: monopole-monopole (M-M) decays as  $1/L$  ( $L$  is distance), whereas dipole-dipole and quadrupole-monopole interactions decays more quickly as  $1/L^3$ .<sup>84</sup> The long-range M-M interactions render the total energy of charge defects also slowly convergent with system size. Although correction schemes to the total energy of charged supercells with compensating uniform background charge are available,<sup>84,85</sup> we avoid altogether the use of the total energy calculated from the PBC calculations and simply used PBC to obtain an optimized  $V_{emb}(\mathbf{r})$ . To optimize  $V_{emb}(\mathbf{r})$ , we select a suitable cluster (and environment) that we carve out from the periodic defective bulk diamond (see **Fig. 2** for the bulk structures). We capped the edges of the resulting  $C_xN$  cluster and  $C_y$  environment using F, O, and B atoms for when respectively monovalent, divalent, and trivalent capping is needed (more details in the **Results** section). We used a modified VASP v. 6.4.2 to perform  $V_{emb}(\mathbf{r})$  optimization (see *Principles of capped-DFET theory, vide supra*),<sup>93</sup> which utilizes a DFET implementation within the PAW method with its frozen-core approximation by Yu, Libisch, and Carter.<sup>47</sup>

In  $V_{emb}(\mathbf{r})$  optimization for the negatively charged defect, the periodically repeated capped cluster contains the excess charge and is solved for as a spin triplet within the spin-unrestricted (broken spin symmetry) formalism - the same as the full periodic bulk system. The auxiliary fragment and capped environment are charge-neutral and solved for as closed-shell within the spin-restricted formalism. The periodic charged capped embedded cluster (with compensating uniform background charge) is subject to the same slowly converging ( $1/L$ ) monopole-monopole interaction as the full system because they share the same supercell size and excess charge distribution: the excess charge is localized around the defect (**Fig. S3**) and at convergence of  $V_{emb}(\mathbf{r})$ , the regional matching of the electron density between the embedded cluster and full system, and by extension, the excess charge distribution, is ensured. Without loss of

generality, the same equality can be said about the local defect-induced dipole-dipole and monopole-quadrupole interactions when the capped cluster is chosen such that it preserves the symmetry of the defect in the full system (which we demonstrated to be true, see **Note S1** and **Table S4**). Therefore, the optimized  $V_{emb}(\mathbf{r})$  enacts the correct local dielectric response of the cluster to the environment (and environment to the cluster) not captured by the capping elements, approximately free of the slowly convergent Coulomb interactions.

*Cluster calculations.* We performed all bare and embedded cluster calculations using Molpro version 2024.1.1.<sup>94,95</sup> We employed atom-centered Gaussian-type orbital (GTO) basis sets. GTO bases include Dunning<sup>96</sup> augmented correlation-consistent polarizable double  $\zeta$  (aug-cc-pvdz or apvdz), which we benchmarked against the aug-cc polarizable triple  $\zeta$  (aug-cc-pvtz or apvtz) basis set. **Table S2** summarizes the basis sets we used throughout the study. We concurrently used the complementary auxiliary bases to evaluate the two-electron integrals (so-called resolution of identity or density fitting). Accordingly, we transformed the optimized  $V_{emb}$  from (real-space) uniform Cartesian grid (as produced within VASP) to the GTO bases space via the embedding integral generator code (EmbeddingIntegralGenerator).<sup>97</sup> We then add the  $V_{emb}$  in GTO to the one-electron Hamiltonian. Because the cluster is nonperiodic, a neutralizing background charge is not needed to converge the electrostatic energy, furthermore the distance-dependent Coulomb interaction between a charge defect and its periodic images are non-existent (*vide supra*). Furthermore, with or without the uniform charge-screening potential, the eigenstates will be the same, wherein the former all the eigenvalues (and hence the total energy) are simply shifted by a constant relative to the latter. Therefore, in calculating the excitation energies, the difference between the ground and excited states for the same number of electrons (**Eq. 7**) will lead to the perfect cancellation of the energy shift. Therefore, even though the cluster defect is charged, we choose not to apply the same uniform background charge (potential) as we have used in cluster PBC (where it is necessary).

We performed active space optimization using the molecular orbital theory as guide to determine the appropriate correlation space. We performed state-averaged CASSCF (SA-CASSCF), averaging over three states, to calculate for the ground and lowest valence excited states. We did not impose any symmetry. Three states are needed due to the double degeneracy of the first excited state of the spin triplet and the ground state of the spin singlet state. Because CASSCF is only able to capture static correlation within the limited correlation space (active space) size, we captured the dynamic correlation via an MRPT2 method, specifically, NEVPT2.<sup>29</sup> We also performed CASPT2<sup>52-54</sup> in some cases as an alternative method. Despite having no intruder states, we found an ionization potential-electron affinity (IP-EA) shift<sup>98</sup> of 0.25 ha is required in CASPT2 to avoid overprediction of the excitation energies (as is found in the literature). We

concluded NEVPT2 to be superior to CASPT2 because it does not require such a semi-empirical tuning parameter (*vide infra*). We therefore used NEVPT2 in the entirety of the paper.

We initialized the CASSCF calculations using restricted open-shell Hartree-Fock canonical orbitals for the ground-state spin triplet of  $\text{NcVc}^-$ , and subsequently used the optimized triplet CASSCF natural orbitals to initialize the spin singlet state. To preserve spin symmetry, we performed CASSCF either as a spin-pure triplet or singlet using only the symmetry-adapted configuration state functions (not the full determinant space).

### **Data Availability**

All raw data used to generate figures and tables that appear in main text and SI are available upon reasonable request to be sent to the corresponding author.

### **Code Availability**

VASP v. 6.4 and Molpro v. 2024.1.1 used in this work are commercial codes that require user licenses. VASP subroutines with embedding implementation and associated Python scripts, and the standalone embedding integral generator code used to transform the embedding potential from Cartesian grid to GTO bases, are publicly available via GitHub: <https://github.com/EACcodes/VASPEmbedding> and <https://github.com/EACcodes/EmbeddingIntegralGenerator>, respectively, both under the Mozilla Public License 2.0.

### **Acknowledgements**

The research described in this paper was conducted under the Laboratory Directed Research and Development (LDRD) Program at Princeton Plasma Physics Laboratory, a national laboratory operated by Princeton University for the U.S. Department of Energy under Prime Contract No. DE-AC02-09CH11466. The United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The simulations presented in this article were performed on Princeton University's Tiger HPC and Princeton Plasma Physics Laboratory's Stellar HPC resources managed and supported by Princeton Research Computing, a consortium of groups including the Princeton Institute for Computational Science and Engineering (PICSciE) and the Office of Information Technology's High Performance Computing Center and Visualization Laboratory at Princeton University. We thank Prof. Emily A. Carter, Dr. Phillips Hutchinson, and Dr. Vidushi Sharma for reviewing the first version of the manuscript.

## Author Contributions

JMPM conceptualized, generated and interpreted the data, and wrote the manuscript.

## Competing interests

The author declares no financial or non-financial competing interests

## References

- 1 Wolfowicz, G. *et al.* Quantum guidelines for solid-state spin defects. *Nature Reviews Materials* **6**, 906-925 (2021). <https://doi.org/10.1038/s41578-021-00306-y>
- 2 Chatterjee, A. *et al.* Semiconductor qubits in practice. *Nature Reviews Physics* **3**, 157-177 (2021). <https://doi.org/10.1038/s42254-021-00283-9>
- 3 Weber, J. R. *et al.* Quantum computing with defects. *Proceedings of the National Academy of Sciences* **107**, 8513-8518 (2010). <https://doi.org/doi:10.1073/pnas.1003052107>
- 4 Aharonovich, I., Englund, D. & Toth, M. Solid-state single-photon emitters. *Nature Photonics* **10**, 631-641 (2016). <https://doi.org/10.1038/nphoton.2016.186>
- 5 Ferrenti, A. M., de Leon, N. P., Thompson, J. D. & Cava, R. J. Identifying candidate hosts for quantum defects via data mining. *npj Computational Materials* **6**, 126 (2020). <https://doi.org/10.1038/s41524-020-00391-7>
- 6 Bradac, C., Gao, W., Forneris, J., Trusheim, M. E. & Aharonovich, I. Quantum nanophotonics with group IV defects in diamond. *Nature Communications* **10**, 5625 (2019). <https://doi.org/10.1038/s41467-019-13332-w>
- 7 Thiering, G. & Gali, A. in *Semiconductors and Semimetals* Vol. 103 (eds Christoph E. Nebel, Igor Aharonovich, Norikazu Mizuochi, & Mutsuko Hatano) 1-36 (Elsevier, 2020).
- 8 Doherty, M. W., Manson, N. B., Delaney, P. & Hollenberg, L. C. L. The negatively charged nitrogen-vacancy centre in diamond: the electronic solution. *New Journal of Physics* **13**, 025019 (2011). <https://doi.org/10.1088/1367-2630/13/2/025019>
- 9 Davies, G., Hamer, M. F. & Price, W. C. Optical studies of the 1.945 eV vibronic band in diamond. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* **348**, 285-298 (1976). <https://doi.org/doi:10.1098/rspa.1976.0039>
- 10 Rogers, L. J., Armstrong, S., Sellars, M. J. & Manson, N. B. Infrared emission of the NV centre in diamond: Zeeman and uniaxial stress studies. *New Journal of Physics* **10**, 103024 (2008). <https://doi.org/10.1088/1367-2630/10/10/103024>
- 11 Cheng, L., Zhu, S., Ouyang, X. & Zheng, W. Bandgap evolution of diamond. *Diamond and Related Materials* **132**, 109638 (2023). <https://doi.org/https://doi.org/10.1016/j.diamond.2022.109638>
- 12 Doherty, M. W. *et al.* The nitrogen-vacancy colour centre in diamond. *Physics Reports* **528**, 1-45 (2013). <https://doi.org/https://doi.org/10.1016/j.physrep.2013.02.001>
- 13 Liu, G.-Q., Feng, X., Wang, N., Li, Q. & Liu, R.-B. Coherent quantum control of nitrogen-vacancy center spins near 1000 kelvin. *Nature Communications* **10**, 1344 (2019). <https://doi.org/10.1038/s41467-019-09327-2>
- 14 Neumann, P. *et al.* Excited-state spectroscopy of single NV defects in diamond using optically detected magnetic resonance. *New Journal of Physics* **11**, 013017 (2009). <https://doi.org/10.1088/1367-2630/11/1/013017>
- 15 de Leon, N. P. *et al.* Materials challenges and opportunities for quantum computing hardware. *Science* **372**, eabb2823 (2021). <https://doi.org/doi:10.1126/science.abb2823>
- 16 Ivády, V., Abrikosov, I. A. & Gali, A. First principles calculation of spin-related quantities for point defect qubit research. *npj Computational Materials* **4**, 76 (2018). <https://doi.org/10.1038/s41524-018-0132-5>

- 17 Gali, Á. Ab initio theory of the nitrogen-vacancy center in diamond. *Nanophotonics* **8**, 1907-1943 (2019). <https://doi.org/doi:10.1515/nanoph-2019-0154>
- 18 Ma, Y., Rohlfing, M. & Gali, A. Excited states of the negatively charged nitrogen-vacancy color center in diamond. *Phys. Rev. B* **81**, 041204 (2010). <https://doi.org/10.1103/PhysRevB.81.041204>
- 19 Bockstedte, M., Schütz, F., Garratt, T., Ivády, V. & Gali, A. Ab initio description of highly correlated states in defects for realizing quantum bits. *npj Quantum Materials* **3**, 31 (2018). <https://doi.org/10.1038/s41535-018-0103-6>
- 20 Ma, H., Govoni, M. & Galli, G. Quantum simulations of materials on near-term quantum computers. *npj Computational Materials* **6**, 85 (2020). <https://doi.org/10.1038/s41524-020-00353-z>
- 21 Halder, S., Mitra, A., Hermes, M. R. & Gagliardi, L. Local Excitations of a Charged Nitrogen Vacancy in Diamond with Multireference Density Matrix Embedding Theory. *The Journal of Physical Chemistry Letters* **14**, 4273-4280 (2023). <https://doi.org/10.1021/acs.jpclett.3c00551>
- 22 Sheng, N., Vorwerk, C., Govoni, M. & Galli, G. Green's Function Formulation of Quantum Defect Embedding Theory. *Journal of Chemical Theory and Computation* **18**, 3512-3522 (2022). <https://doi.org/10.1021/acs.jctc.2c00240>
- 23 Chen, Y. *et al.* Multiconfigurational nature of electron correlation within nitrogen vacancy centers in diamond. *Phys. Rev. B* **108**, 045111 (2023). <https://doi.org/10.1103/PhysRevB.108.045111>
- 24 Bhandari, C., Wysocki, A. L., Economou, S. E., Dev, P. & Park, K. Multiconfigurational study of the negatively charged nitrogen-vacancy center in diamond. *Phys. Rev. B* **103**, 014115 (2021). <https://doi.org/10.1103/PhysRevB.103.014115>
- 25 Jin, Y. *et al.* Photoluminescence spectra of point defects in semiconductors: Validation of first-principles calculations. *Physical Review Materials* **5**, 084603 (2021). <https://doi.org/10.1103/PhysRevMaterials.5.084603>
- 26 Martirez, J. M. P. & Carter, E. A. Projector-Free Capped-Fragment Scheme within Density Functional Embedding Theory for Covalent and Ionic Compounds. *Journal of Chemical Theory and Computation* **17**, 4105-4121 (2021).
- 27 Roos, B. O. The complete active space SCF method in a Fock-matrix-based super-CI formulation. *International Journal of Quantum Chemistry* **17**, 175-189 (1980).
- 28 Roos, B. O., Taylor, P. R. & Siegbahn, P. E. M. A complete active space SCF method (CASSCF) using a density-matrix formulated super-CI approach. *Chemical Physics* **48**, 157-173 (1980). [https://doi.org/10.1016/0301-0104\(80\)80045-0](https://doi.org/10.1016/0301-0104(80)80045-0)
- 29 Angeli, C., Cimiraglia, R., Evangelisti, S., Leininger, T. & Malrieu, J. P. Introduction of n-electron valence states for multireference perturbation theory. *J. Chem. Phys.* **114**, 10252-10264 (2001). <https://doi.org/10.1063/1.1361246>
- 30 Huang, C., Pavone, M. & Carter, E. A. Quantum mechanical embedding theory based on a unique embedding potential. *Journal of Chemical Physics* **134**, 154110 (2011). <https://doi.org/10.1063/1.3577516>
- 31 Wu, Q. & Yang, W. A Direct Optimization Method for Calculating Density Functionals and Exchange-Correlation Potentials from Electron Densities. *Journal of Chemical Physics* **118**, 2498-2509 (2003).
- 32 Cheng, J., Libisch, F. & Carter, E. A. Dissociative Adsorption of O<sub>2</sub> on Al(111): The Role of Orientational Degrees of Freedom. *Journal of Physical Chemistry Letters* **6**, 1661-1665 (2015). <https://doi.org/10.1021/acs.jpclett.5b00597>
- 33 Zhao, Q., Zhang, X., Martirez, J. M. P. & Carter, E. A. Benchmarking an Embedded Adaptive Sampling Configuration Interaction Method for Surface Reactions: H<sub>2</sub> Desorption from and CH<sub>4</sub> Dissociation on Cu(111). *J. Chem. Theory Comput.* **16**, 7078-7088 (2020).
- 34 Wei, Z., Martirez, J. M. P. & Carter, E. A. Introducing the embedded random phase approximation: H<sub>2</sub> dissociative adsorption on Cu(111) as an exemplar. *The Journal of Chemical Physics* **159** (2023). <https://doi.org/10.1063/5.0181229>

- 35 Martirez, J. M. P. & Carter, E. A. Solvent Dynamics Are Critical to Understanding Carbon Dioxide Dissolution and Hydration in Water. *Journal of the American Chemical Society* **145**, 12561-12575 (2023). <https://doi.org/10.1021/jacs.3c01283>
- 36 Boyn, J.-N. & Carter, E. A. Characterizing the Mechanisms of Ca and Mg Carbonate Ion-Pair Formation with Multi-Level Molecular Dynamics/Quantum Mechanics Simulations. *The Journal of Physical Chemistry B* **127**, 10824-10832 (2023). <https://doi.org/10.1021/acs.jpcb.3c05369>
- 37 Boyn, J.-N. & Carter, E. A. Probing pH-Dependent Dehydration Dynamics of Mg and Ca Cations in Aqueous Solutions with Multi-Level Quantum Mechanics/Molecular Dynamics Simulations. *Journal of the American Chemical Society* **145**, 20462-20472 (2023). <https://doi.org/10.1021/jacs.3c06182>
- 38 Boyn, J.-N. & Carter, E. A. Elucidating and contrasting the mechanisms for Mg and Ca sulfate ion-pair formation with multi-level embedded quantum mechanics/molecular dynamics simulations. *The Journal of Chemical Physics* **161** (2024). <https://doi.org/10.1063/5.0235460>
- 39 Bobell, B., Boyn, J.-N., Martirez, J. M. P. & Carter, E. A. Modeling bicarbonate formation in an alkaline solution with multi-level quantum mechanics/molecular dynamics simulations. *Molecular Physics*, e2375370 (2024). <https://doi.org/10.1080/00268976.2024.2375370>
- 40 Zhao, Q. & Carter, E. A. Revisiting Competing Paths in Electrochemical CO<sub>2</sub> Reduction on Copper via Embedded Correlated Wavefunction Theory. *Journal of Chemical Theory and Computation* **16**, 6528-6538 (2020). <https://doi.org/10.1021/acs.jctc.0c00583>
- 41 Zhao, Q., Martirez, J. M. P. & Carter, E. A. Revisiting Understanding of Electrochemical CO<sub>2</sub> Reduction on Cu(111): Competing Proton-Coupled Electron Transfer Reaction Mechanisms Revealed by Embedded Correlated Wavefunction Theory. *Journal of the American Chemical Society* **143**, 6152-6164 (2021). <https://doi.org/10.1021/jacs.1c00880>
- 42 Zhao, Q., Martirez, J. M. P. & Carter, E. A. Charting C-C coupling pathways in electrochemical CO<sub>2</sub> reduction on Cu(111) using embedded correlated wavefunction theory. *submitted* (2022).
- 43 Zhao, Q., Martirez, J. M. P. & Carter, E. A. Electrochemical Hydrogenation of CO on Cu(100): Insights from Accurate Multiconfigurational Wavefunction Methods. *Journal of Physical Chemistry Letters*, 10282-10290 (2022). <https://doi.org/10.1021/acs.jpcllett.2c02444>
- 44 Martirez, J. M. P., Bao, J. L. & Carter, E. A. First-Principles Insights into Plasmon-Induced Catalysis. *Annual Review of Physical Chemistry* **72**, 99-119 (2021). <https://doi.org/10.1146/annurev-physchem-061020-053501>
- 45 Wen, X. & Carter, E. A. Plasmon-driven ammonia decomposition on Pd(111) surface: ground- and excited-state reaction energy curves from embedded correlated wavefunction theory. *in preparation* (2022).
- 46 Jacob, C. R. & Neugebauer, J. Subsystem density-functional theory (update). *WIREs Computational Molecular Science* **14**, e1700 (2024). <https://doi.org/https://doi.org/10.1002/wcms.1700>
- 47 Yu, K., Libisch, F. & Carter, E. A. Implementation of density functional embedding theory within the projector-augmented-wave method and applications to semiconductor defect states. *Journal of Chemical Physics* **143**, 102806 (2015). <https://doi.org/10.1063/1.4922260>
- 48 Yu, K. & Carter, E. A. Extending density functional embedding theory for covalently bonded systems. *Proceedings of the National Academy of Sciences of the United States of America* **114**, E10861-E10870 (2017). <https://doi.org/10.1073/pnas.1712611114>
- 49 Chung, L. W. *et al.* The ONIOM Method and Its Applications. *Chemical Reviews* **115**, 5678-5796 (2015). <https://doi.org/10.1021/cr5004419>
- 50 Jacob, C. R. & Visscher, L. A subsystem density-functional theory approach for the quantum chemical treatment of proteins. *The Journal of Chemical Physics* **128** (2008). <https://doi.org/10.1063/1.2906128>
- 51 Martirez, J. M. P. & Carter, E. A. Metal-to-Ligand Charge-Transfer Spectrum of a Ru-Bipyridine-Sensitized TiO<sub>2</sub> Cluster from Embedded Multiconfigurational Excited-State Theory. *Journal of Physical Chemistry A* **125**, 4998-5013 (2021).

- 52 Andersson, K., Malmqvist, P. A. & Roos, B. O. Second-order perturbation-theory with a complete active space self-consistent field reference function. *Journal of Chemical Physics* **96**, 1218-1226 (1992). <https://doi.org/10.1063/1.462209>
- 53 Andersson, K., Malmqvist, P. A., Roos, B. O., Sadlej, A. J. & Wolinski, K. Second-order perturbation-theory with a CASSCF reference function. *J. Phys. Chem.* **94**, 5483-5488 (1990). <https://doi.org/10.1021/j100377a012>
- 54 Celani, P. & Werner, H. J. Multireference perturbation theory for large restricted and selected active space reference wave functions. *Journal of Chemical Physics* **112**, 5546-5557 (2000). <https://doi.org/10.1063/1.481132>
- 55 Mejía-Rodríguez, D. & Trickey, S. B. Meta-GGA performance in solids at almost GGA cost. *Phys. Rev. B* **102**, 121109 (2020). <https://doi.org/10.1103/PhysRevB.102.121109>
- 56 Kaplan, A. D. & Perdew, J. P. Laplacian-level meta-generalized gradient approximation for solid and liquid metals. *Physical Review Materials* **6**, 083803 (2022). <https://doi.org/10.1103/PhysRevMaterials.6.083803>
- 57 Yamanaka, T. & Morimoto, S. Isotope effect on anharmonic thermal atomic vibration and [kappa] refinement of <sup>12</sup>C and <sup>3</sup>C diamond. *Acta Crystallographica Section B* **52**, 232-238 (1996). <https://doi.org/doi:10.1107/S0108768195010810>
- 58 Heyd, J., Scuseria, G. E. & Ernzerhof, M. Hybrid functionals based on a screened Coulomb potential. *Journal of Chemical Physics* **118**, 8207-8215 (2003). <https://doi.org/10.1063/1.1564060>
- 59 Krukau, A. V., Vydrov, O. A., Izmaylov, A. F. & Scuseria, G. E. Influence of the exchange screening parameter on the performance of screened hybrid functionals. *Journal of Chemical Physics* **125**, 224106 (2006). <https://doi.org/10.1063/1.2404663>
- 60 Liu, P., Kaltak, M., Klimeš, J. & Kresse, G. Cubic scaling *GW*: Towards fast quasiparticle calculations. *Phys. Rev. B* **94**, 165109 (2016). <https://doi.org/10.1103/PhysRevB.94.165109>
- 61 Hedin, L. New Method for Calculating the One-Particle Green's Function with Application to the Electron-Gas Problem. *Physical Review* **139**, A796-A823 (1965).
- 62 Shishkin, M. & Kresse, G. Implementation and performance of the frequency-dependent *SGW* method within the PAW framework. *Phys. Rev. B* **74**, 035101 (2006). <https://doi.org/10.1103/PhysRevB.74.035101>
- 63 Shishkin, M. & Kresse, G. Self-consistent *SGW* calculations for semiconductors and insulators. *Phys. Rev. B* **75**, 235102 (2007). <https://doi.org/10.1103/PhysRevB.75.235102>
- 64 Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **77**, 3865-3868 (1996).
- 65 Scuseria, G. E. Advancing solid-state band gap predictions. *Proc Natl Acad Sci U S A* **118** (2021). <https://doi.org/10.1073/pnas.2113648118>
- 66 Wing, D. *et al.* Band gaps of crystalline solids from Wannier-localization-based optimal tuning of a screened range-separated hybrid functional. *Proceedings of the National Academy of Sciences* **118**, e2104556118 (2021). <https://doi.org/doi:10.1073/pnas.2104556118>
- 67 Morales-García, Á., Valero, R. & Illas, F. An Empirical, yet Practical Way To Predict the Band Gap in Solids by Using Density Functional Band Structure Calculations. *The Journal of Physical Chemistry C* **121**, 18862-18866 (2017). <https://doi.org/10.1021/acs.jpcc.7b07421>
- 68 Bredas, J. L. Mind the gap! *Materials Horizons* **1**, 17-19 (2014). <https://doi.org/10.1039/c3mh00098b>
- 69 Koopmans, T. Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den Einzelnen Elektronen Eines Atoms. *Physica* **1**, 104-113 (1934). [https://doi.org/https://doi.org/10.1016/S0031-8914\(34\)90011-2](https://doi.org/https://doi.org/10.1016/S0031-8914(34)90011-2)
- 70 Richards, W. G. The use of Koopmans' Theorem in the interpretation of photoelectron spectra. *International Journal of Mass Spectrometry and Ion Physics* **2**, 419-424 (1969). [https://doi.org/https://doi.org/10.1016/0020-7381\(69\)80040-9](https://doi.org/https://doi.org/10.1016/0020-7381(69)80040-9)

- 71 Day, O. W., Smith, D. W. & Garrod, C. A generalization of the hartree-fock one-particle potential. *International Journal of Quantum Chemistry* **8**, 501-509 (1974).  
<https://doi.org/https://doi.org/10.1002/qua.560080855>
- 72 Janak, J. F. Proof that  $\epsilon_F = \epsilon$  in density-functional theory. *Phys. Rev. B* **18**, 7165-7168 (1978).  
<https://doi.org/10.1103/PhysRevB.18.7165>
- 73 Schlüter, M. & Sham, L. J. in *Advances in Quantum Chemistry* Vol. 21 (ed Per-Olov Löwdin) 97-112 (Academic Press, 1990).
- 74 Zhang, G. & Musgrave, C. B. Comparison of DFT Methods for Molecular Orbital Eigenvalue Calculations. *The Journal of Physical Chemistry A* **111**, 1554-1561 (2007).  
<https://doi.org/10.1021/jp061633o>
- 75 Gali, A., Janzén, E., Deák, P., Kresse, G. & Kaxiras, E. Theory of Spin-Conserving Excitation of the  $N^0$  Center in Diamond. *Physical Review Letters* **103**, 186404 (2009). <https://doi.org/10.1103/PhysRevLett.103.186404>
- 76 Zabrodsky, H., Peleg, S. & Avnir, D. Continuous symmetry measures. 2. Symmetry groups and the tetrahedron. *Journal of the American Chemical Society* **115**, 8278-8289 (1993).  
<https://doi.org/10.1021/ja00071a042>
- 77 Wang, V., Xu, N., Liu, J.-C., Tang, G. & Geng, W.-T. VASPKIT: A user-friendly interface facilitating high-throughput computing and analysis using VASP code. *Computer Physics Communications* **267**, 108033 (2021). <https://doi.org/https://doi.org/10.1016/j.cpc.2021.108033>
- 78 Jin, Y., Govoni, M. & Galli, G. Vibrationally resolved optical excitations of the nitrogen-vacancy center in diamond. *npj Computational Materials* **8**, 238 (2022). <https://doi.org/10.1038/s41524-022-00928-y>
- 79 Kehayias, P. *et al.* Infrared absorption band and vibronic structure of the nitrogen-vacancy center in diamond. *Phys. Rev. B* **88**, 165202 (2013). <https://doi.org/10.1103/PhysRevB.88.165202>
- 80 Acosta, V. M., Jarmola, A., Bauch, E. & Budker, D. Optical properties of the nitrogen-vacancy singlet levels in diamond. *Phys. Rev. B* **82**, 201202 (2010).  
<https://doi.org/10.1103/PhysRevB.82.201202>
- 81 Goldman, M. L. *et al.* State-selective intersystem crossing in nitrogen-vacancy centers. *Phys. Rev. B* **91**, 165201 (2015). <https://doi.org/10.1103/PhysRevB.91.165201>
- 82 Goldman, M. L. *et al.* Phonon-Induced Population Dynamics and Intersystem Crossing in Nitrogen-Vacancy Centers. *Physical Review Letters* **114**, 145502 (2015).  
<https://doi.org/10.1103/PhysRevLett.114.145502>
- 83 Goldman, M. L. *et al.* Erratum: State-selective intersystem crossing in nitrogen-vacancy centers [Phys. Rev. B 91, 165201 (2015)]. *Phys. Rev. B* **96**, 039905 (2017).  
<https://doi.org/10.1103/PhysRevB.96.039905>
- 84 Makov, G. & Payne, M. C. Periodic boundary conditions in ab-initio calculations. *Phys. Rev. B* **51**, 4014-4022 (1995). <https://doi.org/10.1103/PhysRevB.51.4014>
- 85 Freysoldt, C., Neugebauer, J. & Van de Walle, C. G. Fully Ab Initio Finite-Size Corrections for Charged-Defect Supercell Calculations. *Physical Review Letters* **102**, 016402 (2009).  
<https://doi.org/10.1103/PhysRevLett.102.016402>
- 86 Tölle, J., Severo Pereira Gomes, A., Ramos, P. & Pavanello, M. Charged-cell periodic DFT simulations via an impurity model based on density embedding: Application to the ionization potential of liquid water. *International Journal of Quantum Chemistry* **119**, e25801 (2019).  
<https://doi.org/https://doi.org/10.1002/qua.25801>
- 87 Kresse, G. & Fürthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169-11186 (1996).  
<https://doi.org/10.1103/PhysRevB.54.11169>
- 88 Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758-1775 (1999). <https://doi.org/10.1103/PhysRevB.59.1758>
- 89 Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953-17979 (1994).  
<https://doi.org/10.1103/PhysRevB.50.17953>

- 90 Schultz, P. A. Theory of Defect Levels and the "Band Gap Problem" in Silicon. *Physical Review Letters* **96**, 246401 (2006). <https://doi.org/10.1103/PhysRevLett.96.246401>
- 91 Schultz, P. A. Charged Local Defects in Extended Systems. *Physical Review Letters* **84**, 1942-1945 (2000). <https://doi.org/10.1103/PhysRevLett.84.1942>
- 92 Blöchl, P. E., Jepsen, O. & Andersen, O. K. Improved tetrahedron method for Brillouin-zone integrations. *Phys. Rev. B* **49**, 16223-16233 (1994). <https://doi.org/10.1103/PhysRevB.49.16223>
- 93 Yu, K. *et al.* *VASPEmbedding*, <<https://github.com/EACcodes/VASPEmbedding>> (2024).
- 94 Werner, H.-J., Knowles, P. J., Knizia, G., Manby, F. R. & Schütz, M. Molpro: a general-purpose quantum chemistry program package. *WIREs Comput. Mol. Sci.* **2**, 242-253 (2012).
- 95 Werner, H.-J. *et al.* The Molpro quantum chemistry package. *The Journal of Chemical Physics* **152**, 144107 (2020). <https://doi.org/10.1063/5.0005081>
- 96 Dunning, T. H. Gaussian-basis sets for use in correlated molecular calculations. 1. The atoms boron through neon and hydrogen. *Journal of Chemical Physics* **90**, 1007-1023 (1989). <https://doi.org/10.1063/1.456153>
- 97 Krauter, C. M., Zhang, X. & Carter, E. A. *Embedding Integral Generator*, <<https://github.com/EACcodes/EmbeddingIntegralGenerator>> (
- 98 Ghigo, G., Roos, B. O. & Malmqvist, P. A. A modified definition of the zeroth-order Hamiltonian in multiconfigurational perturbation theory (CASPT2). *Chemical Physics Letters* **396**, 142-149 (2004). <https://doi.org/10.1016/j.cplett.2004.08.032>

**Fig. 1. Atomic and electronic structures of a prototypical defect in diamond.** **A**, diamond is composed of C atoms each with four nearest neighbors that are  $\sim 1.5$  Å away in a tetrahedral local geometry (top inset). The cubic diamond unit cell contains eight unique C atoms with cell side lengths  $\sim 3.6$  Å (red box). An NV center consists of a N substitutional defect adjacent to a C vacancy. **B**,  $\text{NcVc}^-$  center has a triplet spin ground state (i.e., two unpaired electrons). Electron occupation of the three orbitals derived from the dangling  $sp^3$  orbitals are shown for both triplet ( $^3A_2$  and  $^3E$ ) and singlet states ( $^1E$  and  $^1A_1$ ). For the spin singlet, only three of the six possible electronic configurations are shown. The bra-ket notation, e.g.,  $|211\rangle$ , refers to the wavefunction associated with a specific occupation of the three defect orbitals: either doubly (2), spin-up singly (1), or spin-down singly ( $\bar{1}$ ) occupied or unoccupied (0). **C**, diamond absorbs in the deep UV (5.5 eV), whereas the ground-state spin triplet  $\text{NV}^-$  center absorbs and fluoresces within the visible ( $\sim 2.0$  eV) - hence referred to as a "color center". The energy diagram schematically shows the relative energetic positions of the four electronic states in **B** and the various optical processes that link them.

**Fig. 2. Bulk atomic structures of diamond** – flawless cubic ( $\text{C}_8$ ) and with  $\text{NcVc}^-$  defect:  $\text{C}_{62}\text{N}^-$ ,  $\text{C}_{126}\text{N}^-$ , and  $\text{C}_{158}\text{N}^-$  (as labelled), optimized within DFT-r<sup>2</sup>-SCAN-L. The dimensions of the supercell are given (if only one is given, then all sides are equal). N – dark blue; C – black. For clarity, the three C atoms adjacent to  $\text{Vc}$  are orange whereas the atoms bound to N are light blue. The red lines mark the supercell edges. The lattice vectors according to the primitive cubic cell are shown. **Table S1** summarizes the optimized cell parameters.

**Fig. 3. Bulk electronic DOS of diamond ( $\text{C}_{64}$ ) and diamond with a  $\text{NcVc}^-$  defect ( $\text{C}_{62}\text{N}^-$ ) from DFT-HSE06.** **A**, pure diamond DOS. The DOS are normalized by the total number of atoms in the supercell. The DOS of the defective structure (**B** and **C**) shows separately the spin down (left) and up (right) manifolds. **B** also shows the pDOS for the C atoms far from the defect (as labelled). The pDOS are normalized over the number of atoms whose atomic contributions are summed. **C** shows the contributions of the defect atoms along with their nearest neighbors (as shown in **D**). **D** shows a partial structure of the supercell annotating the position of the defect atoms and their neighbors. Optimized atomic structures from DFT-r<sup>2</sup>-SCAN-L were used. Note the difference in DOS scale between plots.

**Fig. 4. Partitioning and embedding schemes.** **A**, partitioning and capping scheme for the  $\text{C}_{62}\text{N}^-$  periodic supercell. Cluster, environment, and auxiliary fragments' structures shown as labelled. **B**, optimized  $V_{\text{emb}}$  within DFT-HSE06 and DFT-r<sup>2</sup>-SCANL as labelled at two isosurface levels.  $V_{\text{emb}}$ s feature repulsive (purple) and attractive (orange) potentials.

**Fig. 5. Defect MO, calculated orbitals, and dominant electronic configurations.** **A**, grey box: simplification of the structure of the defect site emphasizing the  $\text{C}_{3v}$  symmetry. Corresponding molecular orbital diagram of the  $\text{NVcC}_6$  defect. The defect's  $\text{C}_{3v}$  symmetry dictates orbital and degeneracy of the electronic states. The valence  $3a_1$  and  $2e$  orbitals originate from the non-bonding  $sp^3$  states of the three carbon atoms at the vicinity of the vacancy. **B**, natural frontier orbitals  $a_1$  and doubly degenerate  $e$  calculated from the state-averaged density matrix of the three-state triplet embedded SA-CASSCF with CAS(6e,8o) from apvdz basis set. Isosurface value of 0.06 ( $e/\text{bohr}^3$ )<sup>1/2</sup>. **C**, electronic configurations for the spin triplet and singlet states showing only the occupation of the valence  $3a_1$  and  $2e$  orbitals with spin-up and spin-down electrons (up and down arrows). In the bra-ket notation, 2, 1,  $\bar{1}$ , or 0 mean that an orbital is either doubly, spin-up singly, spin-down singly, or unoccupied, respectively. Note that these electronic configurations are not necessarily degenerate and in fact are not.

**Fig. 6. Various clusters explored carved from three periodic supercells).** The clusters' bulk origins are shown above, and their sizes are indicated below. Atom colors match the colors of their element symbol in their labels.

**Fig. 7. NEVPT2 calculated VEEs and ISC energy as a function of bare capped cluster size.** Clusters were carved from a bulk  $\text{C}_{158}\text{N}^-$  supercell with Emb-NEVPT2 performed using pvdz+vdz basis set with (4e,3o) AS. Emb-NEVPT2 benchmarks from a bulk  $\text{C}_{126}\text{N}^-$  are shown for comparison (in blue). Data summarized in **Table S3**.

# Supplementary Information

## Optical properties of a diamond NV color center from capped embedded multiconfigurational correlated wavefunction theory

John Mark P. Martirez

Applied Materials and Sustainability Sciences, Princeton Plasma Physics Laboratory,  
Princeton, New Jersey, 08540  
martirez@pppl.gov, martirez@princeton.edu

**Table S1. DFT-r<sup>2</sup>-SCAN-L-calculated optimized cell parameters for the defective diamond structures (Fig. 2, main text).**

| Structure                       | $ a ,  b ,  c $ [Å]                 | $\alpha, \beta, \gamma$ [°] |
|---------------------------------|-------------------------------------|-----------------------------|
| C <sub>62</sub> N <sup>-</sup>  | 7.170, 7.170, 7.170 <sup>a</sup>    | 90.17, 90.17, 89.83         |
| C <sub>126</sub> N <sup>-</sup> | 10.108, 10.108, 10.123 <sup>b</sup> | 60.05, 60.05, 60.00         |
| C <sub>158</sub> N <sup>-</sup> | 10.104, 10.104, 12.645 <sup>b</sup> | 60.04, 60.04, 60.00         |

<sup>a</sup> $\mathbf{a} \parallel [100]$ ;  $\mathbf{b} \parallel [001]$ ;  $\mathbf{c} \parallel [0\bar{1}0]$ : vectors were arbitrarily defined to match visualization in Fig. 2

<sup>b</sup> $\mathbf{a} \parallel [10\bar{1}]$ ;  $\mathbf{b} \parallel [01\bar{1}]$ ;  $\mathbf{c} \parallel [110]$ : vectors defined so that the vector orthogonal to the plane defined by vectors  $\mathbf{a}$  and  $\mathbf{b}$  plane is  $[111]$  and vector  $\mathbf{c}$  is  $[110]$

**Table S2. Summary of the size of the GTO-type all-electron correlation consistent Dunning basis sets<sup>1</sup> used.<sup>a,b</sup>**

| Basis set   | C                       | N                       | F                       | O                       |
|-------------|-------------------------|-------------------------|-------------------------|-------------------------|
| apvdz(all)  | 10s,5p,2d →<br>4s,3p,2d | 10s,5p,2d →<br>4s,3p,2d | 10s,5p,2d →<br>4s,3p,2d | 10s,5p,2d →<br>4s,3p,2d |
| apvtz(C,N)  | 11s,6p,3d,2f →          | 11s,6p,3d,2f            | 10s,5p,2d →             | 10s,5p,2d →             |
| +apvdz(F,O) | 5s,4p,3d,2f             | → 5s,4p,3d,2f           | 4s,3p,2d                | 4s,3p,2d                |
| pvdz(C,N)   | 9s,4p,1d →              | 9s,4p,1d →              | 9s,5p →                 | 9s,5p →                 |
| +vdz(F,O)   | 3s,2p,1d                | 3s,2p,1d                | 3s,2p                   | 3s,2p                   |

<sup>a</sup>The number of basis functions for each angular momentum before and after (→) contraction is shown.

<sup>b</sup>Basis sets from the Basis Set Exchange repository (<https://www.basissetexchange.org>)<sup>2,3</sup>

**Table S3. Calculated vertical excitation energies (VEE) for low-lying  $\text{NcVc}^-$  center transitions in embedded clusters carved from bulk  $\text{C}_{126}\text{N}^-$  and bare clusters carved out from bulk  $\text{C}_{158}\text{N}^-$  using SS-NEVPT2 AS(4e,3o) with apvdz and pvdz+vdz basis sets.**

| Exc.                    | Cluster                                                 | Calculated VEE [eV] |          |
|-------------------------|---------------------------------------------------------|---------------------|----------|
|                         |                                                         | apvdz               | pvdz+vdz |
| $^3A_2 \rightarrow ^3E$ | emb-[ $\text{C}_{15}\text{NF}_{12}\text{O}_{12}$ ] $^-$ | 2.30                | 2.30     |
|                         | emb-[ $\text{C}_{21}\text{NF}_{18}\text{O}_9$ ] $^-$    | 2.27                | 2.30     |
|                         | [ $\text{C}_{15}\text{NF}_{12}\text{O}_{12}$ ] $^-$     | 2.24                | 2.26     |
|                         | [ $\text{C}_{21}\text{NF}_{18}\text{O}_9$ ] $^-$        | 2.14                | 2.17     |
|                         | [ $\text{C}_{24}\text{NF}_{18}\text{O}_{12}$ ] $^-$     | -                   | 2.17     |
|                         | [ $\text{C}_{30}\text{NF}_{24}\text{O}_9$ ] $^-$        | -                   | 2.20     |
|                         | [ $\text{C}_{36}\text{NF}_{30}\text{O}_6$ ] $^-$        | -                   | 2.24     |
| $^1E \rightarrow ^1A_1$ | emb-[ $\text{C}_{15}\text{NF}_{12}\text{O}_{12}$ ] $^-$ | 1.34                | 1.37     |
|                         | emb-[ $\text{C}_{21}\text{NF}_{18}\text{O}_9$ ] $^-$    | 1.28                | 1.32     |
|                         | [ $\text{C}_{15}\text{NF}_{12}\text{O}_{12}$ ] $^-$     | 1.22                | 1.26     |
|                         | [ $\text{C}_{21}\text{NF}_{18}\text{O}_9$ ] $^-$        | 1.15                | 1.19     |
|                         | [ $\text{C}_{24}\text{NF}_{18}\text{O}_{12}$ ] $^-$     | -                   | 1.21     |
|                         | [ $\text{C}_{30}\text{NF}_{24}\text{O}_9$ ] $^-$        | -                   | 1.21     |
|                         | [ $\text{C}_{36}\text{NF}_{30}\text{O}_6$ ] $^-$        | -                   | 1.21     |
| $^1A_1 \rightarrow ^3E$ | emb-[ $\text{C}_{15}\text{NF}_{12}\text{O}_{12}$ ] $^-$ | 0.53                | 0.36     |
|                         | emb-[ $\text{C}_{21}\text{NF}_{18}\text{O}_9$ ] $^-$    | 0.44                | 0.41     |
|                         | [ $\text{C}_{15}\text{NF}_{12}\text{O}_{12}$ ] $^-$     | 0.50                | 0.45     |
|                         | [ $\text{C}_{21}\text{NF}_{18}\text{O}_9$ ] $^-$        | 0.48                | 0.45     |
|                         | [ $\text{C}_{24}\text{NF}_{18}\text{O}_{12}$ ] $^-$     | -                   | 0.43     |
|                         | [ $\text{C}_{30}\text{NF}_{24}\text{O}_9$ ] $^-$        | -                   | 0.45     |
|                         | [ $\text{C}_{36}\text{NF}_{30}\text{O}_6$ ] $^-$        | -                   | 0.50     |

### Note S1. Slowly converging Coulomb interactions in charged supercells

Makov and Payne introduced total energy corrections for  $L \rightarrow \infty$  extrapolation of a periodic charged supercell (with a localized charged defect) subject to a uniform compensating background charge.<sup>4</sup> They approximate the monopoles as point charges, which have a simple interaction energy expression. The energy correction due to monopole-monopole (M-M) interaction is the negative of this interaction. For a cubic cell, this correction is

$$E_{\text{M-M}} = \frac{e^2 q^2 \alpha}{2L\varepsilon} \quad \text{Eq. 1}$$

$q$  is the excess electron in the cell,  $e$  is electron charge,  $L$  is the length of the side of the (cubic) cell,  $\alpha$  is Madelung constant (2.8373 for simple cubic),<sup>5</sup> and  $\varepsilon$  is the material's dielectric constant ( $4\pi\varepsilon_0 = 0.069 \text{ e}^2 \text{ eV}^{-1} \text{ \AA}^{-1}$ ;  $\varepsilon = 4\pi\varepsilon_0\varepsilon_r$ ;  $\varepsilon_r$  is the relative dielectric constant). On the other hand, the correction for dipole-dipole (D-D) interaction is

$$E_{\text{D-D}} = \frac{2\pi}{3L^3\varepsilon} \left| \int \rho(\mathbf{r}) (\mathbf{r} - \mathbf{r}_0) d^3\mathbf{r} \right|^2 = \frac{2\pi}{3L^3\varepsilon} |\mu|^2 \quad \text{Eq. 2}$$

with the integrand within the volume of supercell.  $\rho(\mathbf{r})$  is the electron density at point  $\mathbf{r}$  and  $\mathbf{r}_0$  is an arbitrary origin, noting that for non-zero monopole, the above expression for a dipole (and quadrupole below) becomes origin dependent. Finally, the correction for monopole-quadrupole (M-Q) interaction is

$$E_{\text{M-Q}} = \frac{2\pi e q (Q_q - Q_{q=0})}{3L^3\varepsilon} \quad \text{Eq. 3}$$

where

$$Q = \int \rho(\mathbf{r}) |\mathbf{r} - \mathbf{r}_0|^2 d^3\mathbf{r} \quad \text{Eq. 4}$$

evaluated for  $q$ -charged and corresponding neutral ( $q = 0$ ) cells. Here,  $(Q_q - Q_{q=0})$  approximates the charged defect induced quadrupole moment. The total energy of a charged supercell with a neutralizing jellium is

$$E(L) = E(L \rightarrow \infty) - \frac{e^2 q^2 \alpha}{2L\varepsilon} - \frac{2\pi}{3L^3\varepsilon} |\mu|^2 - \frac{2\pi e q (Q_q - Q_{q=0})}{3L^3\varepsilon} \quad \text{Eq. 5}$$

In Makov and Payne's formulation,  $\mathbf{r}_0$  can be chosen so that  $E_{\text{D-D}} \cong 0$  (also by symmetry exactly 0 for a cubic cell), which is therefore ignored

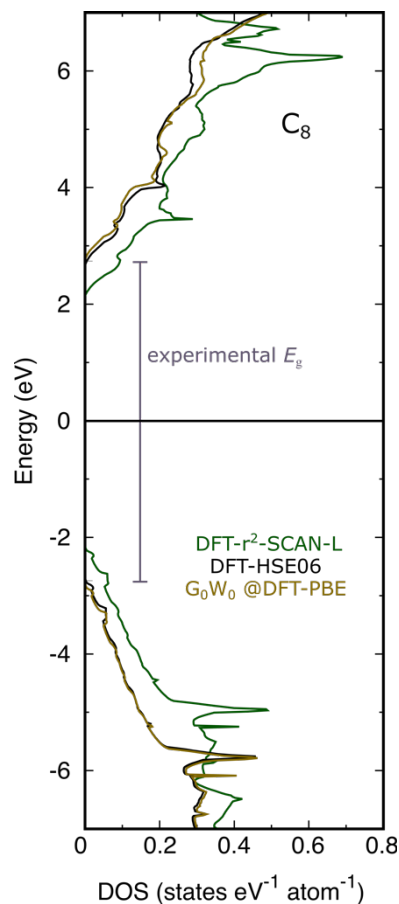
$$E(L) = E(L \rightarrow \infty) - \frac{e^2 q^2 \alpha}{2L\varepsilon} - \frac{2\pi e q (Q_q - Q_{q=0})}{3L^3\varepsilon} \quad \text{Eq. 6}$$

We evaluate these approximate energy corrections for  $q = -1$ , using  $\mu$  and  $Q$  that we calculated from VASP by setting IDPIOL = 4. **Table S4** summarizes these energy corrections along with the dipole and quadrupole moments for the approximately cubic  $\text{C}_{62}\text{N}^-$  and the embedded periodic  $[\text{C}_{15}\text{NF}_{12}\text{O}_{12}]^-$  cluster. These corrections are by no means perfect<sup>5</sup> nor used directly in this study. For instance,  $E_{\text{M-M}}$  ignores the spatial distribution of the excess charge. However, **Table S4** does demonstrate that the charged full system and embedded capped cluster exhibit similar built-in charge-induced slowly converging Coulomb interactions.

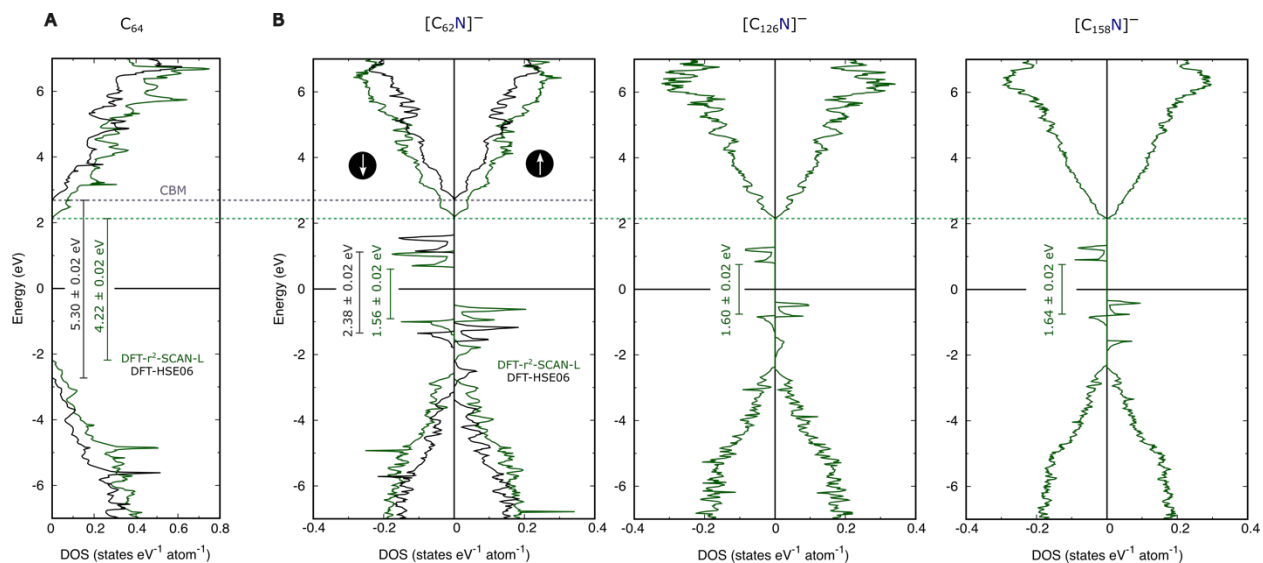
**Table S4. Calculated dipole ( $\mu$ ) and induced quadrupole ( $Q_q - Q_{q=0}$ ) moments and approximate energy correction to extrapolate to  $L \rightarrow \infty$  for the bulk supercell and embedded cluster within periodic DFT-HSE06.<sup>a</sup>**

| System                                              | $ \mu $<br>[eÅ] <sup>b</sup> | $(Q_q - Q_{q=0})$<br>[eÅ <sup>2</sup> ] <sup>b,c</sup> | Energy corrections using ref. <sup>4</sup> [eV] |                    |                    |                    |
|-----------------------------------------------------|------------------------------|--------------------------------------------------------|-------------------------------------------------|--------------------|--------------------|--------------------|
|                                                     |                              |                                                        | $E_{\text{M-M}}$                                | $E_{\text{(D-D)}}$ | $E_{\text{(M-Q)}}$ | $E_{\text{total}}$ |
| $\text{C}_{62}\text{N}^-$                           | 0.16                         | 10.07                                                  | 0.533                                           | 0.000              | -0.154             | 0.379              |
| emb- $[\text{C}_{15}\text{NF}_{12}\text{O}_{12}]^-$ | 0.15                         | 10.46                                                  | 0.533                                           | 0.000              | -0.160             | 0.373              |

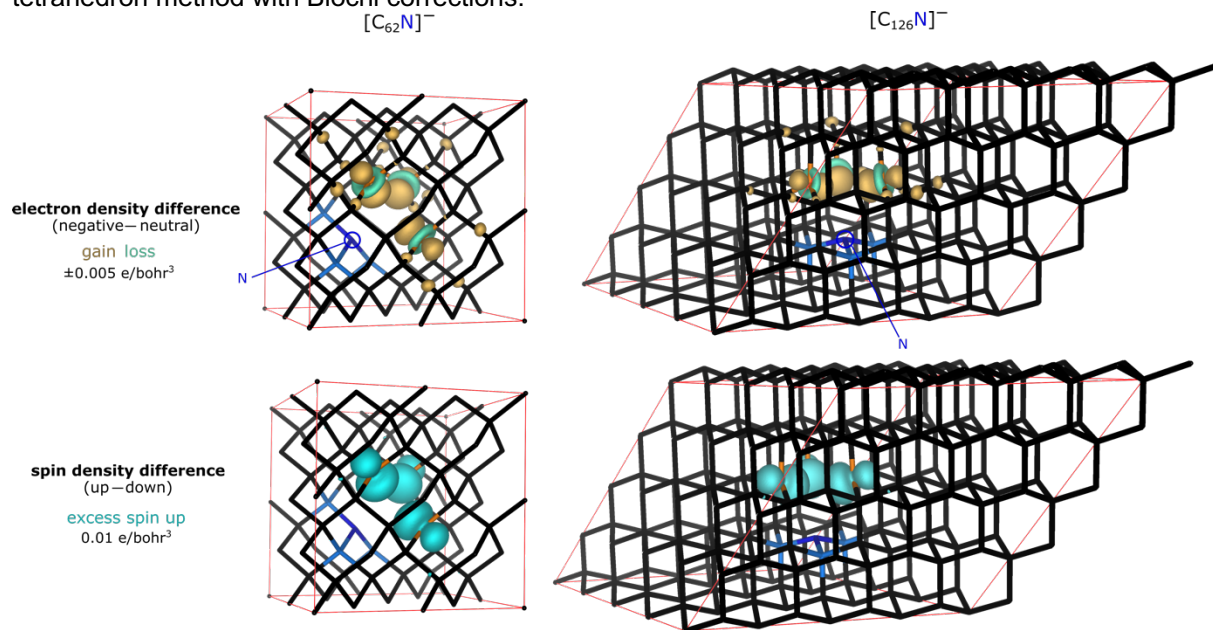
<sup>a</sup>The origin was set to (0.5, 0.5, 0.5) in reduced lattice coordinates. Because the monopole is non-vanishing for charged cells, their dipole and quadrupole moments are origin dependent.<sup>4</sup> Here, the origin is approximately at the center of the vacancy. <sup>b</sup>We use the experimental  $\varepsilon_r$  of carbon diamond, which is 5.35 from ref. <sup>6</sup>



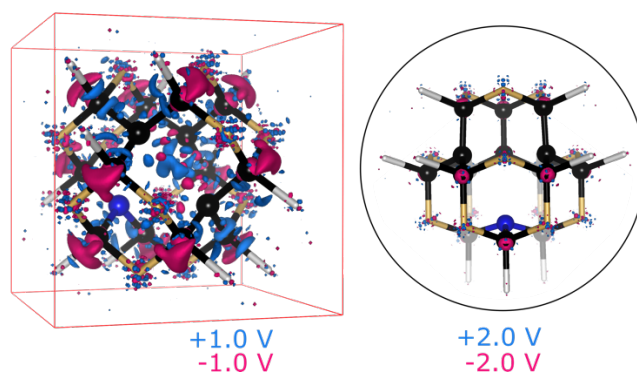
**Fig. S1. Calculated DOS for pure diamond** using a  $C_8$  cubic supercell within various levels of theory (as labelled). The band gap center is set to zero. All three yield almost identical DOS profiles except that their band edges are shifted relative to each other, thus, yielding different band gaps. BZ integration:  $\Gamma$ -point-centered ( $10 \times 10 \times 10$ )  $k$ -point mesh and the linear tetrahedron method with Blöchl corrections. See **Table 1** in the main text for the exact values of the calculated band gaps. Electronic convergences were done self-consistently except for  $G_0W_0$  @DFT-PBE (see **Methods**). DFT-HSE06 and  $G_0W_0$  @DFT-PBE are almost overlapping. The DFT-HSE06 valence and conduction bands are simply slightly shifted up and down by 0.1 eV, respectively, relative to  $G_0W_0$  @DFT-PBE (thus, a 0.2 eV higher calculated gap for the latter).



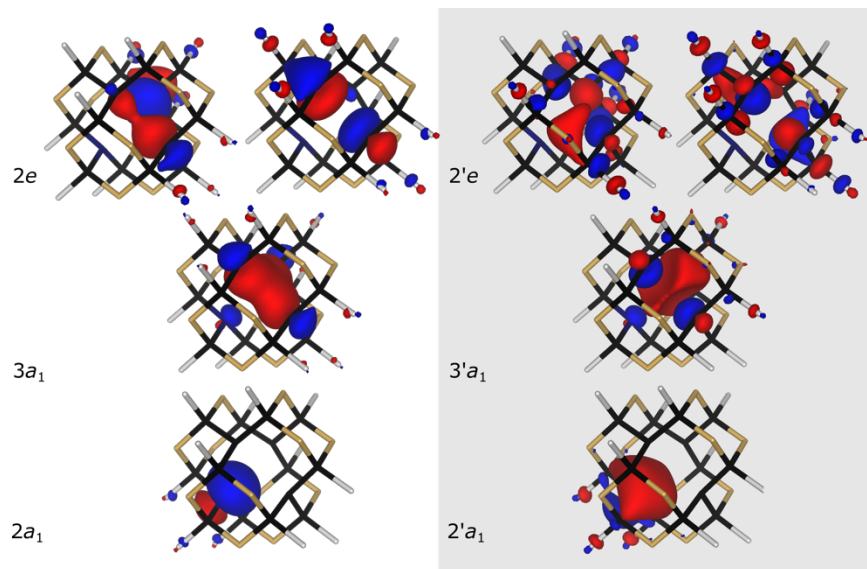
**Fig. S2. Bulk electronic structures of diamond ( $C_{64}$ ) and diamond with a  $NcVc^-$  center:  $C_{62}N^-$ ,  $C_{126}N^-$ , and  $C_{158}N^-$  (as labelled).** **A**, pure from DFT- $r^2$ -SCAN-L and DFT-HSE06 with respective band gaps annotated. **B**, defective diamond. Optimized atomic structures from DFT- $r^2$ -SCAN-L were used. Note the difference in DOS scale between plots. BZ sampling:  $\Gamma$ -point-centered  $6 \times 6 \times 6$   $k$ -point mesh and the linear tetrahedron method with Blöchl corrections.



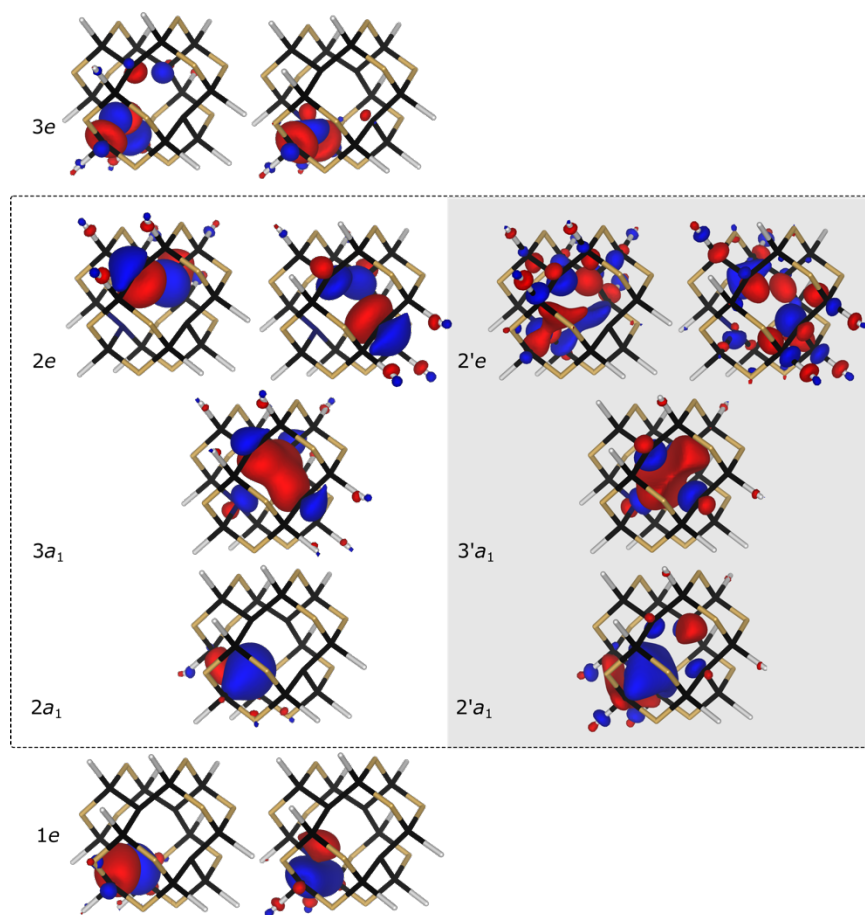
**Fig. S3. Electron density:  $\rho[-1] - \rho[0]$  and spin triplet density:  $\rho[S_z(\uparrow)] - \rho[S_z(\downarrow)]$  difference plots for  $C_{62}N^-$  and  $C_{126}N^-$  supercells from DFT-HSE06.** Structure: N – dark blue; C – black. For clarity, the three C atoms adjacent to  $V_c$  are orange whereas the atoms bound to N are light blue (see main text Fig. 2 for comparison). The red lines mark the supercell edges. The neutral cell, with one unpaired electron (spin doublet), was also calculated using the DFT-  $r^2$ -SCAN-L-optimized structure for the negatively charged cell.



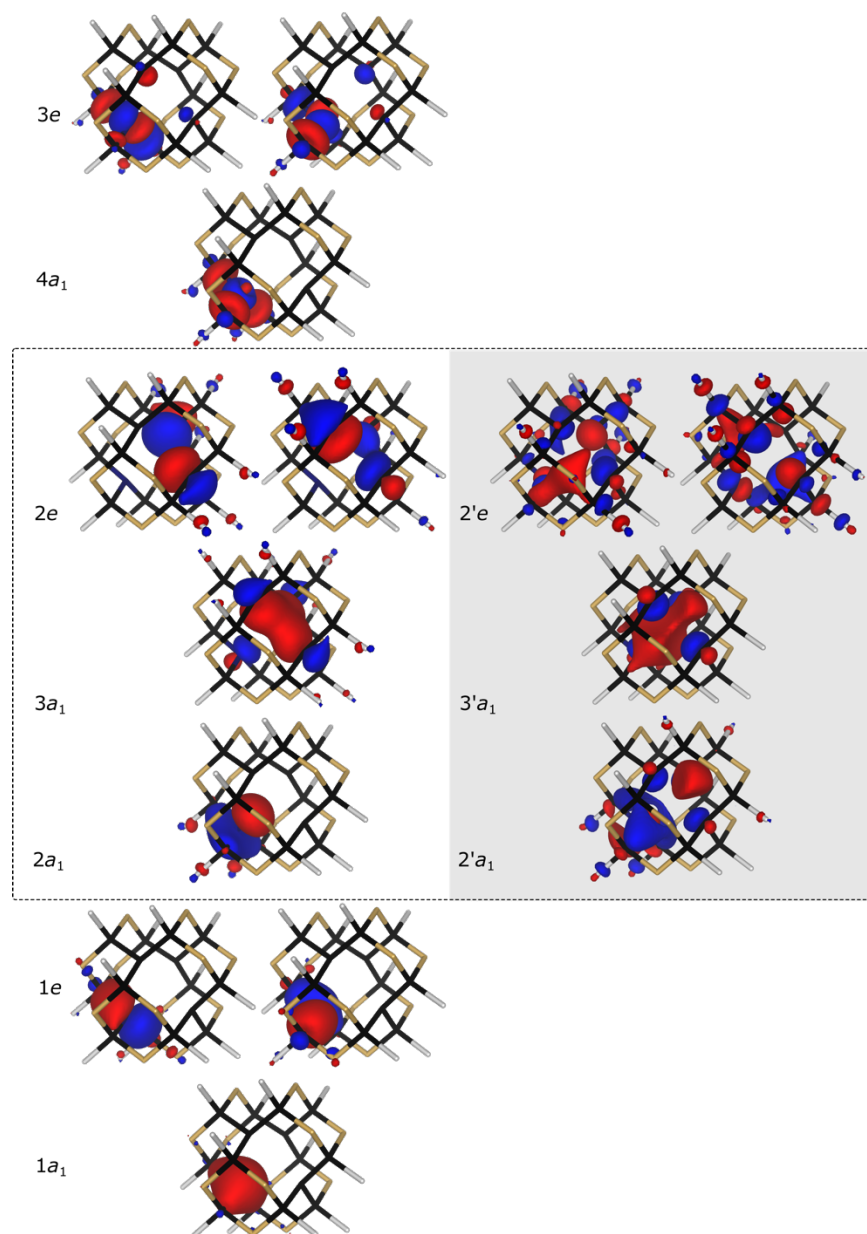
**Fig. S4.** Difference between the  $V_{\text{emb}}$  calculated for the  $\text{C}_{62}\text{N}^-$  periodic supercell from DFT-HSE06 and DFT-r<sup>2</sup>-SCANL as shown in **Fig. 4** (former minus the latter). Isosurface levels as labelled (the same as in **Fig. 4**).



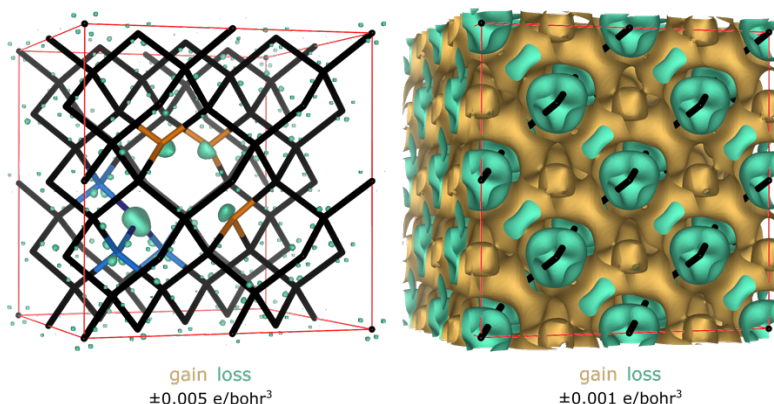
**Fig. S5.** Active space natural orbitals from the diagonalization of the state-averaged density matrix of the **three-state spin singlet** calculated from embedded SA-CASSCF with **AS(6e,8o)** with apvdz basis set. Approximate orbital symmetry labels shown. Isosurface of  $\pm 0.06$  a.u.



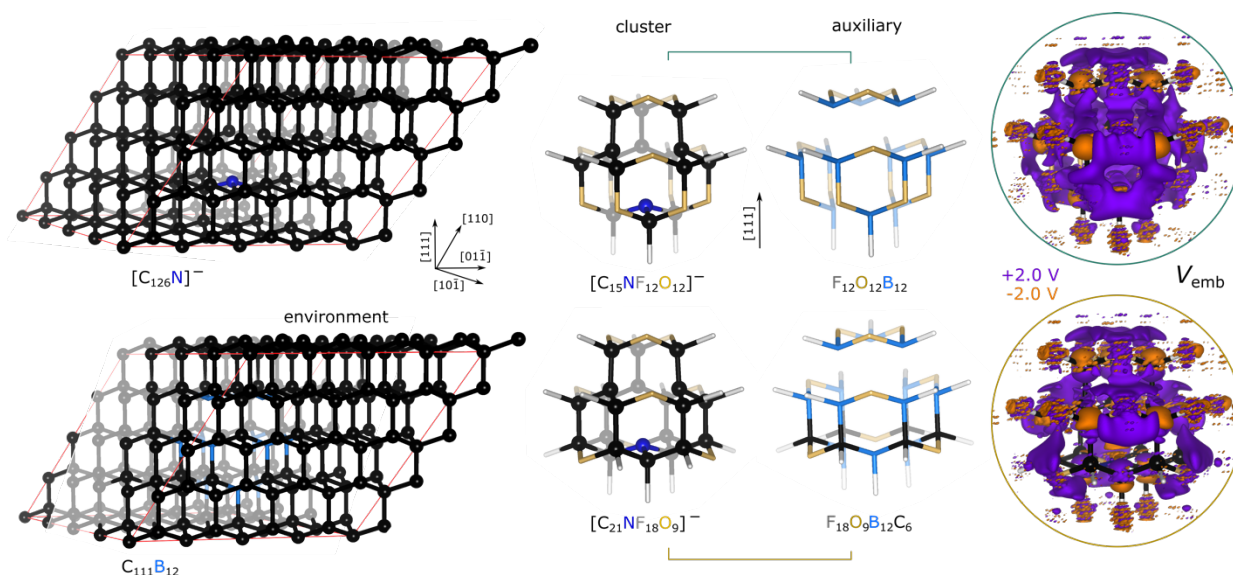
**Fig. S6.** Active space natural orbitals from the diagonalization of the state-averaged density matrix of the **three-state spin triplet** calculated from embedded SA-CASSCF with **AS(10e,12o)** with apvdz basis set. Approximate orbital symmetry labels shown. Isosurface of  $\pm 0.06$  a.u.



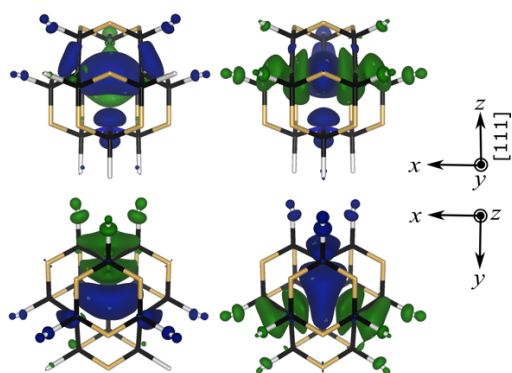
**Fig. S7.** Active space natural orbitals from the diagonalization of the state-averaged density matrix of the **three-state spin triplet** calculated from embedded SA-CASSCF with **AS(12e,14o)** with apvdz basis set. Approximate orbital symmetry labels shown. Isosurface of  $\pm 0.06$  a.u.



**Fig. S8.** Electron density difference plots  $\rho[\text{DFT-HSE06}] - \rho[\text{DFT-}r^2\text{-SCAN-L}]$  for  $\text{C}_{62}\text{N}^-$  supercell using the DFT-  $r^2$ -SCAN-L-optimized structure at two difference iso-surface levels. Structure: N – dark blue; C – black. For clarity, the three C atoms adjacent to  $\text{V}_\text{C}$  are orange whereas the atoms bound to N are light blue (see main text **Fig. 2** for comparison). The red lines mark the supercell edges.



**Fig. S9.** Partitioning scheme and DFT-HSE06-derived optimized  $V_\text{emb}$ S for  $[\text{C}_{15}\text{NF}_{12}\text{O}_{12}]^-$  and  $[\text{C}_{21}\text{NF}_{18}\text{O}_9]^-$  capped clusters from  $\text{C}_{126}\text{N}^-$  supercell. For  $[\text{C}_{21}\text{NF}_{18}\text{O}_9]^-$ , the environment requires six capping elements after the addition of six C atoms into the cluster. However, the new empty sites are four-fold coordinated sites, which thus require a capping atom able to form four bonds, consequently we also use C atoms as capping atoms at those sites. Therefore, the environment fragment for  $[\text{C}_{21}\text{NF}_{18}\text{O}_9]^-$  is the same as in  $[\text{C}_{15}\text{NF}_{12}\text{O}_{12}]^-$ , *i.e.*,  $\text{C}_{111}\text{B}_{12}$ . The auxiliary fragment of  $[\text{C}_{21}\text{NF}_{18}\text{O}_9]^-$ , however, now includes the six capping C atoms along with 18 F, 9 O, and 12 B atoms. The optimized  $V_\text{emb}$  for  $[\text{C}_{15}\text{NF}_{12}\text{O}_{12}]^-$  is, as expected, similar to the smaller supercell (**Fig. 4** in the main text). The optimized  $V_\text{emb}$  for  $[\text{C}_{21}\text{NF}_{18}\text{O}_9]^-$  is the same as in  $[\text{C}_{15}\text{NF}_{12}\text{O}_{12}]^-$  except in the region next to where additional C atoms were introduced. Less embedding potential amplitude is thus around the C atoms next to N ( $\text{C}[\text{nN}]$  sites).



**Fig. S10. Electron density difference ( $\rho_n - \rho_0$ ) plots between the  $n$ th excited and ground states.** State-specific electron densities taken from SA-CASSCF for the two  ${}^1A_1 \rightarrow {}^3E$  excitations for the  $[\text{C}_{15}\text{NF}_{12}\text{O}_{12}]^-$  cluster carved from the  $\text{C}_{126}\text{N}^-$  supercell (**Fig. S9**). Isosurface: green (gain)/blue (loss) =  $\pm 0.002$  e/bohr<sup>3</sup>. Side (top panels) and top (bottom panels) views: the vectors are defined as:  $x = [10\bar{1}]$ ,  $y = [\bar{1}2\bar{1}]$ ,  $z = [111]$  in the basis of the cubic lattice vectors of diamond.

### Supplementary References

- 1 Dunning, T. H. Gaussian-basis sets for use in correlated molecular calculations. 1. The atoms boron through neon and hydrogen. *Journal of Chemical Physics* **90**, 1007-1023 (1989). <https://doi.org/10.1063/1.456153>
- 2 Pritchard, B. P., Altarawy, D., Didier, B., Gibson, T. D. & Windus, T. L. New Basis Set Exchange: An Open, Up-to-Date Resource for the Molecular Sciences Community. *J. Chem. Inf. Model.* **59**, 4814-4820 (2019). <https://doi.org/10.1021/acs.jcim.9b00725>
- 3 Schuchardt, K. L. *et al.* Basis set exchange: A community database for computational sciences. *J. Chem. Inf. Model.* **47**, 1045-1052 (2007). <https://doi.org/10.1021/ci600510j>
- 4 Makov, G. & Payne, M. C. Periodic boundary conditions in ab-initio calculations. *Phys. Rev. B* **51**, 4014-4022 (1995). <https://doi.org/10.1103/PhysRevB.51.4014>
- 5 Shim, J., Lee, E.-K., Lee, Y. J. & Nieminen, R. M. Density-functional calculations of defect formation energies using supercell methods: Defects in diamond. *Phys. Rev. B* **71**, 035206 (2005). <https://doi.org/10.1103/PhysRevB.71.035206>
- 6 Groves, L. G. & Martin, A. E. The dielectric constant of diamond. *Transactions of the Faraday Society* **35**, 575-581 (1940). <https://doi.org/10.1039/TF9403500575>