

Transcorrelated wave-function framework for solids: An application to bulk and defected silicon

Kristoffer Simula¹,¹ Johannes Hauskrecht,¹ Evelin Martine Corvid Christlmaier¹,¹ Pablo Lopez-Rios,¹
Daniel Kats¹,¹ Denis Usvyat²,² and Ali Alavi^{1,3}

¹*Max Planck Institute for Solid State Research, Heisenbergstr. 1, 70569 Stuttgart, Germany*

²*Institut für Chemie, Humboldt-Universität zu Berlin, Brook-Taylor-Str. 2, Berlin 12489, Germany*

³*Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom*



(Received 9 December 2025; revised 20 March 2026; accepted 6 April 2026; published 15 May 2026)

Accurate wave-function descriptions of pristine and defected solids remain challenging because of the simultaneous presence of finite-size, basis-set, and correlation errors. While embedding techniques alleviate finite-size effects and correlated wave-function approaches systematically improve correlation, basis-set incompleteness continues to limit practical accuracy. Here, we present a study of transcorrelated (TC) many-body wave-function methods on properties of solid state systems. We augment the existing xTC theory to periodic systems, and establish an unified transcorrelated embedding framework that integrates periodic TC theory with fragment-based correlated solvers. Using silicon as a test case, we validate the method against coupled-cluster, FCIQMC, and diffusion Monte Carlo benchmarks for bulk. Then, we apply transcorrelated embedding to the calculation of formation energies of two silicon self-interstitials. The TC Hamiltonian yields rapid basis convergence and quantitatively reliable defect formation energies at the triple- ζ level, substantially reducing the basis-set bottleneck for wave-function treatments of crystalline defects.

DOI: [10.1103/d65l-5865](https://doi.org/10.1103/d65l-5865)

I. INTRODUCTION

The accurate description of electronic structure in solids, especially in the presence of point defects, remains one of the central challenges of theoretical condensed-matter physics. The complexity arises from the simultaneous need to control three sources of error: (i) finite-size effects (FSE) in supercell models, (ii) basis-set incompleteness, and (iii) incomplete treatment of electron correlation. Wave-function methods offer systematic improvability in (iii), but their steep scaling with system size and basis dimension makes it difficult to reach convergence in (i) and (ii) [1–7], particularly for point defects that require large simulation cells and often high basis-set resolution.

Because of these difficulties, wave-function-based methods that directly solve the many-electron Schrödinger equation have seen limited use in solid-state physics, despite their systematic improvability and their success in molecular systems. Density functional theory (DFT) remains the workhorse of condensed-matter simulation because of its favorable scaling and ability to treat large periodic systems. However, DFT often lacks the accuracy required for quantitative predictions, particularly for systems in which correlation effects are strong, motivating the search for alternatives that combine high accuracy with computational feasibility.

Diffusion Monte Carlo (DMC) [8] solves the exact Schrödinger equation within the fixed-node approximation and has been successfully applied to solid-state systems [1,3]. Its favorable N^3 – N^4 scaling with system size N and small basis-set errors allow the treatment of supercells containing hundreds of atoms, establishing DMC as a standard benchmark for solids. However, the fixed-node approximation introduces a trial-wave-function-dependent bias. Because of the variational principle, this bias is in principle controllable, but in practice the available computational resources often limit the extent to which the nodes can be optimized with advanced wave-function forms, such as backflow corrections or multideterminant expansions. This practical limitation is particularly severe in cases in which the convergence to thermodynamic limit requires large supercells.

Auxiliary-field quantum Monte Carlo (AFQMC) [9] is a second-quantized method that, like DMC, relies on a trial wave function to control the sign problem. In addition to beneficial scaling comparable to that of DMC, AFQMC offers a natural framework for all-electron calculations that are often problematic for DMC. Also embedding methods are straightforward to implement within AFQMC. However, basis-set errors are typically more severe in AFQMC than in DMC, and the presence of the sign problem in AFQMC presents similar practical challenges to those of DMC.

Second-quantized methods such as coupled-cluster (CC) theory and full configuration-interaction quantum Monte Carlo (FCIQMC) [10–12] provide a systematically improvable hierarchy toward the exact many-electron solution. However, their scaling with system size and basis set is far worse than that of DMC, making their use for realistic defect descriptions with high resolution infeasible.

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Open access publication funded by Max Planck Society.

Embedding techniques have therefore emerged as a powerful strategy to alleviate finite-size effects [13–19] and enable the use of second-quantized methods for realistic systems with defects. In these approaches, a chemically motivated *fragment* containing the defect is treated with a high-level correlated solver, while the surrounding crystal is described at the mean-field level with sufficiently large supercells or even with an entirely nondefective matrix using the aperiodic defect method [19]. Embedding based on localized Wannier and projected atomic orbitals (PAOs) enables systematically enlargable fragments that capture essential defect physics at greatly reduced cost.

Even within such frameworks, basis-set incompleteness remains a severe obstacle. Plane-wave bases offer periodicity and systematic convergence but require prohibitively large cutoffs for correlated solvers, whereas localized Gaussian-type orbitals, though compact, introduce incompleteness errors that are difficult to eliminate in periodic environments. As the Gaussian basis sets are enlarged, linear dependencies arise, often making convergence with respect to the basis set impossible [20,21]. Recent attempts have been made to cure the linear dependency issue and introduce heavier quadruple-zeta Gaussian basis sets for solids that match plane-wave accuracy [22]. Within a periodic MP2 framework, the basis-set incompleteness issues have also been addressed using the explicitly correlated theories [23,24].

Yet, among the three principal sources of error, the basis-set incompleteness may remain the main practical limitation for descriptions of defects under second-quantized theories. While embedding mitigates the finite-size effects and correlated solvers control the electron correlation, achieving a reliable and transferable basis convergence continues to limit attainable accuracy. These persistent limitations motivate the development of new theoretical frameworks with systematic treatment of correlation and small basis-set errors.

The transcorrelated (TC) framework [25–34] represents a conceptually distinct route to accurate theoretical description of electronic structure, directly addressing the basis-convergence problem by performing a similarity transformation of the Hamiltonian with respect to a Jastrow factor J that captures the dominant cusp and dynamical correlation physics [35–37]. This transformation yields an effective Hamiltonian with compactified wave functions and greatly accelerated basis convergence, at the cost of introducing three-body interaction terms and non-Hermiticity. The recently developed xTC approximation replaces these three-body contributions with an effective two-body operator, enabling efficient and accurate many-body calculations [38]. When combined with norm-conserving pseudopotentials (PPs) [28], the Jastrow can focus on valence correlations while avoiding nuclear cusps, further improving efficiency for large atoms [29], molecules, or solids.

In this work, we introduce a xTC-PP framework for periodic solids and formulate a unified *transcorrelated embedding* framework. We demonstrate the potential of these ideas with case studies in pristine and defected silicon. Starting from periodic Hartree–Fock (HF) orbitals and norm-conserving pseudopotentials, we optimize Drummond–Towler–Needs-type Jastrow factors in variational Monte Carlo (VMC) [35,37] and construct the periodic xTC-PP Hamiltonian. In

defect simulations, we downfold the xTC-PP Hamiltonian to fragment subspaces spanned by localized occupied Wannier functions and projected virtual orbitals. The resulting fragment Hamiltonians retain Jastrow-induced correlations between fragment and environment while remaining tractable for post-HF solvers.

The bulk silicon calculations are done in an eight-atom cell, comparing CCSD, DCSD, CCSD(T), and CCSDT results with FCIQMC and DMC benchmarks across multiple Jastrow cutoffs and basis levels. We show that the xTC-PP formalism achieves near-complete basis-set convergence already at the triple- ζ level without linear dependence issues. This shows that TC substantially improves the Gaussian-basis frameworks for solids. Our results are comparable to DMC benchmarks and we find coupled cluster with triple excitations to reach very close to FCI result.

After bulk Si validation, we apply the xTC-PP embedding scheme to the traditional problem of defect formation energy estimation, a task that often requires large cells, accurate basis resolution, and correlation description. We use as test systems the hexagonal (H) and split (X) silicon self-interstitials in 65-atom supercells, analyzing the convergence of defect formation energies with fragment size and basis level. We find convergent fragment sizes and obtain estimates of formation energies that fall within the experimental range measured for the H interstitial. The final formation energies are hence obtained with relatively large simulation cells, and they are likely to be very close to convergence in both basis-set resolution and correlated treatment.

The remainder of the paper outlines the periodic xTC-PP formalism, details the embedding construction, and presents bulk and defect benchmarks demonstrating the accuracy and efficiency of the TC theory for pristine and defected solids.

II. THEORY

Here, we provide an overview of the periodic xTC-PP method and introduce xTC-PP embedding for applications to lattice defects. For a detailed description of TC theory, we refer the reader to Refs. [28,36–38]. For the HF embedding, we follow the approach described in Ref. [15].

Throughout this section, orbital indices p, q, r, s, t, u denote the full one-particle basis. Occupied orbitals are labeled i, j, k, l . In the embedding decomposition, we partition the full orbital space into *fragment orbitals* p_f, q_f, r_f, s_f and *environment orbitals* p_e, q_e, r_e, s_e , using the same notation for occupied orbitals within each subspace. For the integrals over orbitals p, q or p, q, r, s and an operator $\hat{O}$ we write $\hat{O}_{pq} = \langle p|\hat{O}|q\rangle$ and $\hat{O}_{pqrs} = \langle pq|\hat{O}|rs\rangle$, with the bras and kets being the one-electron basis orbitals, $|p\rangle = \phi_p(\mathbf{r})$.

The similarity transformation of the Hamiltonian $\hat{H}$ (under PP approximation) with a Jastrow factor J_α [of particle positions, $J_\alpha(\mathbf{r}_1, \dots, \mathbf{r}_N)$] with at most two-electron terms leads to the following second-quantized TC-PP Hamiltonian $\hat{H}_{\text{TC-PP}}$:

$$\begin{aligned} \hat{H}^{\text{xTC-PP}} = & e^{-J_\alpha} \hat{H} e^{J_\alpha} \xrightarrow[\text{xTC-PP}]{[30,38,40]} E_0^{\text{xTC}} \\ & + \sum_{pq} h_{pq}^{\text{xTC}} a_p^\dagger a_q + \frac{1}{2} \sum_{pqrs} w_{pqrs}^{\text{xTC-PP}} a_p^\dagger a_q^\dagger a_s a_r \quad (1) \end{aligned}$$

with $W^{\text{xTC-PP}}$ being the two-body xTC-PP interaction term defined as

$$\begin{aligned} W_{pqrs}^{\text{xTC-PP}} &= V_{pqrs} + \Delta V_{pqrs}, \\ \Delta V_{pqrs} &= -K_{pqrs} + P_{pqrs} + \Delta W_{pqrs}, \\ \Delta W_{pqrs} &= -\sum_{ij} (L_{priqs} - L_{prijs} - L_{prijs}) \gamma_{ij} \end{aligned} \quad (2)$$

and the one-body xTC term as

$$\begin{aligned} h_{pq}^{\text{xTC}} &= h_{pq} + \Delta h_{pq}, \\ \Delta h_{pq} &= -\frac{1}{2} \sum_{ij} (\Delta W_{piqj} - \Delta W_{pijq}) \gamma_{ij}, \end{aligned} \quad (3)$$

and the operators $a_p^\dagger$ and a_p are the creation and annihilation operators of the orbitals $\phi_p(\mathbf{r})$. The one-body reduced density matrix γ_{ij} is that of Hartree-Fock (HF) reference. We have defined ΔW_{pqrs} and Δh_{pq} to be the corrections to transcorrelated one- and two-body TC terms owing to the xTC approximation. Because of the xTC, we also introduce a correction to the constant energy term [38],

$$\begin{aligned} E_0^{\text{xTC}} &= E_0 + \Delta E_0^{\text{xTC}}, \\ \Delta E_0^{\text{xTC}} &= -\frac{1}{3} \sum_{ij} \Delta h_{ij} \gamma_{ij}. \end{aligned} \quad (4)$$

In Eqs. (1)-(3), the one-electron operator $\hat{h}$ contains the kinetic energy and electron-nucleus pseudopotential parts. $\hat{V}$ is the bare two-electron Coulomb interaction $\hat{V}(\mathbf{r}_1, \mathbf{r}_2) = 1/|\mathbf{r}_2 - \mathbf{r}_1|$. The $\hat{K}$, $\hat{L}$, and $\hat{P}$ operators originate from the Baker-Campbell-Hausdorff expansion of the similarity transformed Hamiltonian, giving rise to the kinetic energy commutators $\hat{K} + \hat{L} = [\frac{1}{2} \sum_p \nabla_p^2, J_\alpha] + [[\frac{1}{2} \sum_p \nabla_p^2, J_\alpha], J_\alpha]$, and the pseudopotential commutators $\hat{P} = [\sum_p \hat{V}_{ecp}^p, J_\alpha] + [[\sum_p \hat{V}_{ecp}^p, J_\alpha], J_\alpha]$. $\hat{K}$ is a two-body operator, while $\hat{L}$ and $\hat{P}$ are three-body operators. Based on earlier findings [28,29], we neglect the three-body contribution of $\hat{P}$, which has proved to be a very good approximation.

We use a Jastrow factor $J_{\alpha_u, \alpha_\chi, \alpha_f} = \sum_{i \neq j} u_{\alpha_u}(\mathbf{r}_i, \mathbf{r}_j) + \sum_i \chi_{\alpha_\chi}(\mathbf{r}_i, \mathbf{R}_i) + \sum_i \sum_{j \neq k} f_{\alpha_f}(\mathbf{r}_i, \mathbf{R}_i, \mathbf{r}_j)$, defined by parameters $\alpha_u, \alpha_\chi, \alpha_f$, of Drummond-Towler-Needs type [35] with two-body (u), one-body (χ), and three-body (f) terms. In the Jastrow terms, $\mathbf{r}_i$ and $\mathbf{r}_j$ are the positions of the electrons, and $\mathbf{R}_i$ is the position of the nuclei. The parameters $\alpha_u, \alpha_\chi, \alpha_f$ are optimized with VMC. Each Jastrow term is truncated at a cutoff length, denoted by (L_u, L_χ, L_f) . The u term captures the electron-electron cusp condition. With periodic boundary conditions (PBC), we employ the periodic version of the Jastrow factor [35]. The orbitals are obtained from a periodic HF calculation with a bare Hamiltonian.

We call the HF energy of the nontranscorrelated periodic system the reference energy, or the non-TC reference energy. With TC, we call the expectation value of the TC Hamiltonian with respect to the HF wave function the transcorrelated, or xTC-PP reference energy.

When we study the fully periodic bulk silicon system, without embedding, we use both non-TC $\hat{H}$ and transcorrelated $\hat{H}_{\text{TC-PP}}$ evaluated in the full Hilbert space of the chosen

basis set for coupled cluster (CC) theory and full configuration interaction quantum Monte Carlo (FCIQMC) to get the correlation energy of the supercell. We also evaluate the total energy of the system using diffusion Monte Carlo (DMC), with both Slater-Jastrow (SJ) and Slater-Jastrow-backflow (SJB) trial wave functions.

A periodic mean-field embedding implies separation of the full system into a fragment and an environment. The fragment is treated with a correlated wave-function method, while the environment is left at the mean-field level. The embedding technique is described in detail in Ref. [15]. The fragment Hamiltonian is

$$\begin{aligned} \hat{H}_{\text{frag}} &= \sum_{p_f q_f} h_{p_f q_f}^{\text{frag}} a_{p_f}^\dagger a_{q_f} \\ &+ \frac{1}{2} \sum_{p_f q_f r_f s_f} V_{p_f q_f r_f s_f} a_{p_f}^\dagger a_{q_f}^\dagger a_{s_f} a_{r_f} + E_0^{\text{frag}}, \end{aligned} \quad (5)$$

with

$$\begin{aligned} h_{p_f q_f}^{\text{frag}} &= h_{p_f q_f}^{\text{per}} + \sum_{i_e} [2V_{p_f i_e q_f i_e} - V_{p_f i_e i_e q_f}] \\ &= f_{p_f q_f}^{\text{per}} - \sum_{i_f} [2V_{p_f i_f q_f i_f} - V_{p_f i_f i_f q_f}], \end{aligned} \quad (6)$$

where the one-electron Hamiltonian $\hat{h}^{\text{per}}$ and the Fock operator $\hat{f}^{\text{per}}$ correspond to the complete periodic system. Finally, in order to reproduce the periodic HF energy per cell $E_{\text{HF}}^{\text{per}}$ in $E_{\text{HF}}^{\text{frag}}$, we define E_0^{frag} as

$$E_0^{\text{frag}} = E_{\text{HF}}^{\text{per}} - 2 \sum_{i_f} h_{i_f i_f}^{\text{frag}} - \sum_{i_f j_f} (2V_{i_f j_f i_f j_f} - V_{i_f j_f j_f i_f}). \quad (7)$$

Before defining the fragment, the HF orbitals of the full periodic system are localized. The localization of the occupied Wannier orbitals is carried out using the method of Refs. [39,40]. For the virtual manifold, we employ projected atomic orbitals (PAOs) [41]. The fragment is then defined by a set of “seed” atoms, defining which Wannier orbitals belong to the fragment on the basis of their Mulliken populations [15]. As PAOs are nonorthogonal and even redundant, the PAOs belonging to the fragment are canonicalized with the cutoff threshold for the eigenvalues the PAO-overlap matrix of 10^{-4} .

To incorporate the transcorrelated embedding in the fragment’s one-electron Hamiltonian, we append it with the environment’s mean-field contributions from the Jastrow commutator operators,

$$h_{p_f q_f}^{\text{xTC-PP-frag}} = h_{p_f q_f}^{\text{frag}} + \Delta h_{p_f q_f} + \sum_{i_e} [2\Delta V_{p_f i_e q_f i_e} - \Delta V_{p_f i_e i_e q_f}]. \quad (8)$$

The two-electron part of the fragment Hamiltonian takes the form

$$W_{p_f q_f r_f s_f}^{\text{xTC-PP-frag}} = V_{p_f q_f r_f s_f} + \Delta V_{p_f q_f r_f s_f}. \quad (9)$$

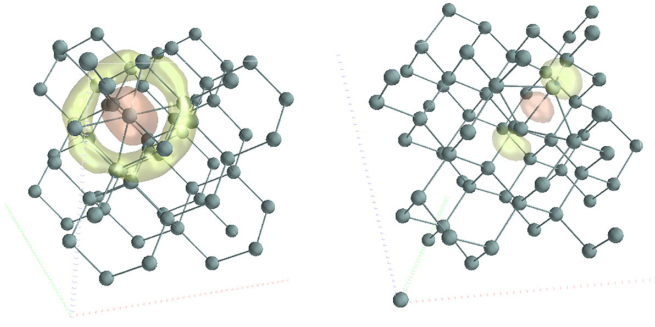


FIG. 1. The periodic simulation cells of the relaxed hexagonal (H, left) and split (X, right) silicon self-interstitial defects used in the formation energy calculations, with the highest occupied fragment orbitals plotted.

Finally, the constant energy term is redefined as

$$E_0^{\text{xTC-PP-frag}} = E_0^{\text{frag}} + 2 \sum_{i_e} \Delta h_{i_e i_e} + \sum_{i_e j_e} [2\Delta V_{i_e j_e i_e j_e} - \Delta V_{i_e j_e j_e i_e}] + \Delta E_0^{\text{xTC}} \quad (10)$$

such that the expectation value of the xTC-Hamiltonian $\hat{h}^{\text{xTC-PP-frag}} + \hat{W}^{\text{xTC-PP-frag}} + E_0^{\text{xTC-PP-frag}}$ within the fragment's occupied space reproduces the periodic xTC-PP reference energy, defined as the expectation value of the transcorrelated Hamiltonian of the fully periodic system with respect to the periodic HF wave function. This energy, which we denote as $E_{\text{HF}}^{\text{xTC-PP-per}}$, serves as the reference energy for the transcorrelated fragment treatment within the xTC-PP embedding model.

We also note that in the above-formulated xTC-PP embedding model, the transcorrelated environment is coupled to the fragment not only via the mean field of the $\Delta\hat{V}$ operator included in $\hat{h}^{\text{xTC-PP-frag}}$, but also through the $\hat{L}$ operator contributions to $\Delta W_{p_f q_f r_f s_f}$ and $\Delta h_{p_f q_f}$, as the sums over ij in Eqs. (2) and (3) run over *all* occupied orbitals, including those in the environment. The integrals of the $\hat{K}$, $\hat{P}$, and $\hat{L}$ operators are calculated numerically under the minimum image convention using the Γ -point Bloch sums of the Wannier functions i_e and fragment orbitals p_f, q_f . Yet, the Coulomb integrals $V_{p_f q_f r_f s_f}$ used in Eqs. (6), (7), and (9) are evaluated in direct space using periodic local density fitting as described in Ref. [15].

With the obtained second-quantized transcorrelated fragment Hamiltonians, we use CC theory to obtain fragment correlation energies. The total energy of the full system is then obtained as the HF energy of the periodic cell plus the CC correlation energy of the fragment.

We estimate the defect formation energies as

$$E_f = E_{\text{HF}}^{\text{defect}} - \frac{N_a^d}{N_a^b} E_{\text{HF}}^{\text{bulk}} + E_{\text{corr}}^{\text{defect}} - \frac{N_e^d}{N_e^b} E_{\text{corr}}^{\text{bulk}}, \quad (11)$$

where $E_{\text{HF}}^{\text{defect}}$ and $E_{\text{HF}}^{\text{bulk}}$ are the fully periodic HF total energies of the defect and bulk supercells, respectively (see defect cell in Fig. 1); $E_{\text{corr}}^{\text{defect}}$ and $E_{\text{corr}}^{\text{bulk}}$ are the corresponding fragment correlation energies; N_a^d (N_a^b) is the number of atoms in the defect (bulk) supercell; and N_e^d (N_e^b) is the number of electrons in the defect (bulk) fragment. This definition of the formation

energy asymptotically approaches the full simulation cell formation energy with increasing fragment size.

III. COMPUTATIONAL DETAILS

We separate the study of transcorrelated solid-state theory into two cases: bulk silicon in the periodic eight-atom conventional cell, and the silicon self-interstitial defects in periodic 65-atom supercells. For the bulk system, we model the full periodic conventional cell at all stages of our workflow. For the defect, we perform a periodic HF calculation, localize the resulting orbitals, and define a periodic fragment to be subjected to the correlation treatment. Figure 2 below summarizes the workflow for these two cases.

Both workflows begin with a periodic HF calculation. In the bulk case, the occupied HF orbitals are used to construct the Slater-Jastrow trial wave function for the VMC optimization. The subsequent evaluation of xTC-PP integrals employs the optimized Jastrow factor together with the HF orbitals and the Hamiltonian matrix elements obtained at the HF stage.

In the embedding workflow, the HF orbitals are first localized and then used to build the trial wave function for VMC. The xTC-PP integral calculation receives the fragment orbitals, the occupied environment orbitals, and the corresponding Hamiltonian elements produced during the fragment-construction phase.

In bulk, we study the convergence of total energy with respect to basis size and level of correlation treatment. In defected systems, we calculate formation energies and compare them against theoretical and experimental benchmark values. The defects studied are hexagonal (H) and split (X) defects. The defect simulation cells are constructed by adding one interstitial Si atom to the bulk 64-atom supercell. The lattice constant is taken to be the experimental value of 5.43 Å in all cases.

The H defect is constructed by inserting a single atom in the center of a hexagonal ring of silicon atoms. The split defect consists of two symmetrically equivalent atomic positions on both sides of an atomic site of a pristine lattice. The atomic positions of the periodic defect supercells are relaxed using DFT with the HSE06 hybrid functional [42]. We used a plane-wave basis and PAW treatment with the VASP package [43] for relaxations. The VASP relaxations use a plane-wave cutoff of 400 eV and a $3 \times 3 \times 3$ Monkhorst-Pack k-point grid. The relaxed defect structures are then used in all subsequent calculations. The relaxed defect structures are shown in Fig. 1.

We use the periodic HF module of PySCF [44] with a single $\mathbf{k}$ -vector at Γ to get the bulk eight-atom cell orbitals. The HF orbitals of the 64/65-atom supercells (64 for bulk and 65 for defects) are obtained with the Crystal17 package [45], using a $3 \times 3 \times 3$ Monkhorst-Pack k-point grid. The orbitals are generated using ccECP pseudopotentials and corresponding cc-pVDZ (DZ) and cc-pVTZ (TZ) basis sets for silicon [46]. Both PySCF and Crystal17 HF calculations use the same ccECPs and basis sets. To avoid linear dependencies, we remove long-tailed Gaussians with exponents smaller than 0.08 a.u.

With the HF orbitals, we construct a trial wave function with a single Slater determinant to optimize the Jastrow parameters with VMC using the CASINO package [47]. The

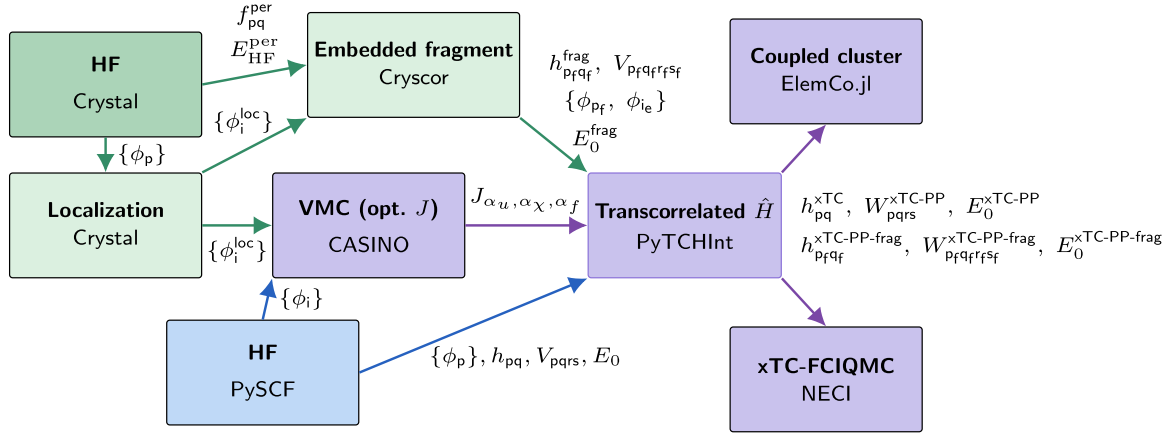


FIG. 2. Computational workflows used in this work. Boxes represent the calculation stages and the codes employed. Directed arrows indicate the flow of data between stages, and their labels specify the quantities transferred (see Sec. II). The orbitals ϕ_p denote periodic HF orbitals, ϕ_i^{loc} their localized counterparts, and ϕ_{pf} and ϕ_{ie} the Γ -point Bloch sums of the Wannier functions i_e and fragment orbitals p_f, q_f , respectively. E_0 is the nuclear repulsion energy. Green indicates steps specific to the embedding workflow, blue those specific to fully periodic calculations, and lavender the stages shared by both workflows. For clarity, we show the xTC data passed to CC and FCIQMC only once.

optimization is done by minimizing the variance of the transcorrelated reference energy [37]. We use (8,8,24) parameters for the u , χ , and f terms, respectively. For the eight-atom bulk cell we test two sets of Jastrow cutoffs, $(L_u, L_\chi, L_f) = (4, 1, 1)$ and $(5, 3, 3)$ bohr, denoted as 4 1 1 and 5 3 3 in the following sections. In the TZ basis, we show that these two cutoffs lead to very similar total energies for the eight-atom cells. In the scope of the present study, for the larger defect cells, we use only the 4 1 1 Jastrow.

The optimized Jastrow parameters are used together with the orbital and Hamiltonian information to construct the xTC-PP integrals numerically according to Eqs. (2), (3), (8), and (9) with our in-house code, PyTCHInt, adapted to periodic calculations. For formation energies, we use the Jastrow optimized in the defect supercell for the bulk fragments for balanced energy comparison.

When computing numerical xTC-PP integrals of the Si conventional cell, we use all of the orbitals obtained in the HF phase with a given Gaussian basis. In fragment calculations, we include all fragment orbitals and the occupied environment orbitals in the integral evaluation. This folds the electron correlation at the xTC level between the fragment and the environment into the fragment Hamiltonian. After evaluating the xTC-PP integrals for the fragment and occupied environment orbitals, we freeze the environment electrons and hence fold the xTC-PP two- and one-body contributions from the environment into the fragment Hamiltonian. The periodic HF orbitals from PySCF and the Γ -point orbital Bloch sums from Cryscor are passed via a molden-format interface (together with the simulation-cell vectors) to PyTCHInt.

We use the FCIDUMP-format [48] to pass the h_{pq} and V_{pqrs} of the eight-atom conventional cell and the direct space fragment integrals h_{pqf} and $V_{pqf\tau f\tau f}$ evaluated for the defect and corresponding bulk supercells. The fragment Hamiltonian construction is done as implemented in the CRYSCOR package [15].

Instead of using atom-centered grids for integral evaluation as in molecular TC [28,29,37,38], we use a uniform grid in the

periodic cell. We found the uniform grids to converge with fewer grid points than atom-centered grids for the eight-atom silicon cell, based on a series of tests with xTC-PP-CCSD done in the eight-atom bulk cell. Figure 7 and Table IV in Appendix B show that with atom-centered Becke grids 227 000 grid points are needed to reach within 0.1 mE_h of the energy per primitive cell obtained with 64 000 uniform grid points. The computational cost of the xTC-PP integrals scales as N_{grid}^2 , and hence the use of uniform grids can lead to significant computational savings. We believe it is the use of PPs which allows the use of relatively sparse uniform grids for the evaluation of the xTC-PP matrix elements, as the strong cusps at the nuclei are removed, and the need to evaluate tight core orbitals as well as highly oscillatory valence orbitals (both of which require dense grids) is avoided.

To study the fragment-size dependence of the embedding model, we define a series of fragments increasing in size around the defect interstitial atoms. We take the first (and second) nearest-neighbor atoms around the H (X) defect into the smallest fragment, and then keep adding shells of next-nearest-neighbor atoms to form larger fragments. The fragments used in this work are illustrated in Fig. 3, where, for clarity, we only depict the fragment atoms and part of the surrounding environment atoms in the periodic simulation cell. Fragment-atom counts range from 7 to 27 atoms for the H-interstitial series and from 8 to 22 atoms for the X-interstitial series. For each defect fragment we define a corresponding fragment in the pristine bulk 64-atom periodic supercell, with one atom less than in the defect fragment.

Finally, with both the bare and xTC-PP second-quantized Hamiltonians we perform correlation calculations. In the conventional eight-atom bulk periodic silicon simulation cell, we use the full Hamiltonian in xTC-PP-CC and -FCIQMC calculations, and in periodic defect supercells we use the embedding Hamiltonian for xTC-PP-CC. Although xTC-PP-FCIQMC calculations for embedded fragments would be technically feasible, the defect calculations employ only xTC-PP-CC, which is sufficient for the desired accuracy. The

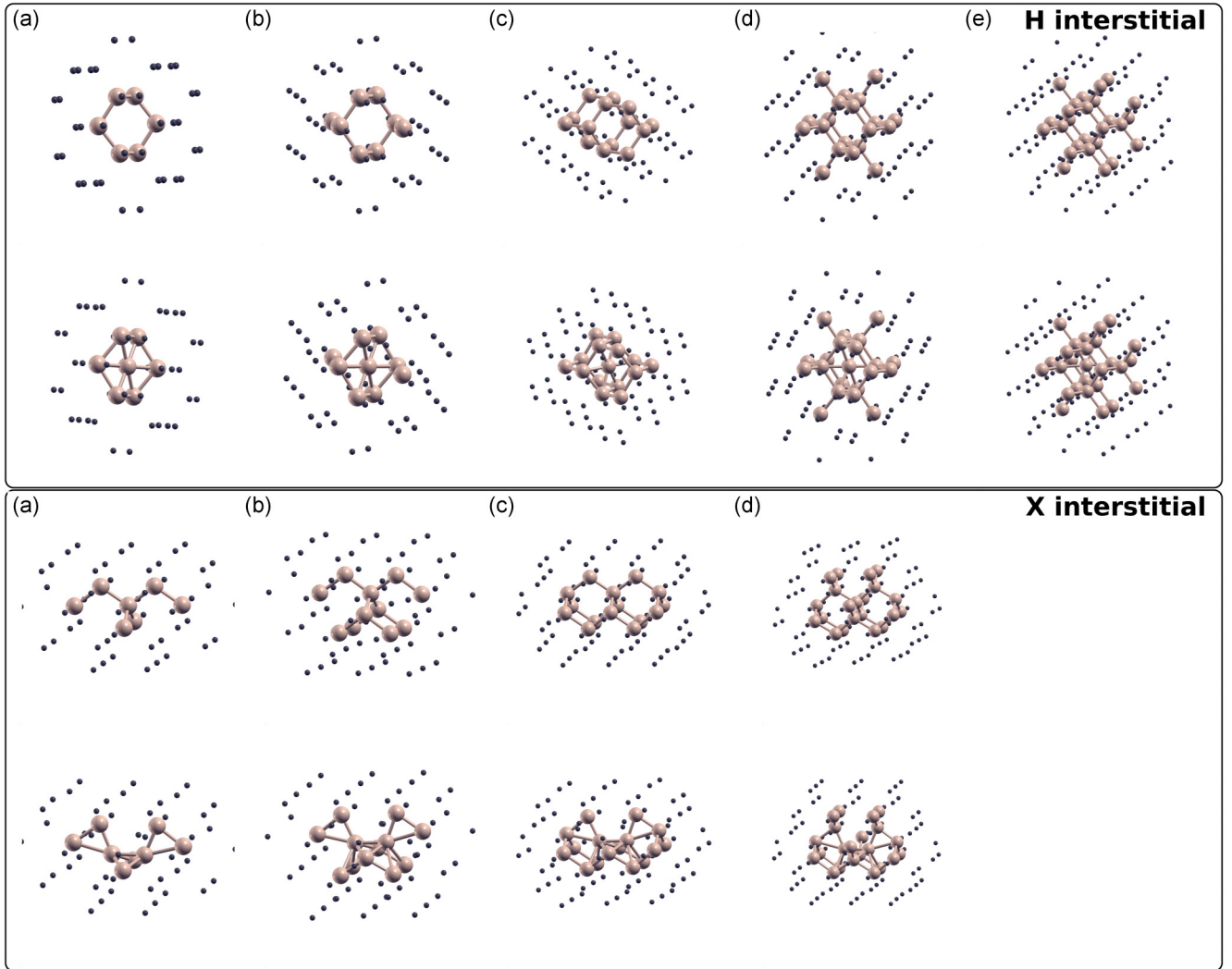


FIG. 3. Embedding fragments used in this work. Each letter denotes a bulk–defect fragment pair: for the H-interstitial series (a)–(e) correspond to pairs with (7), (9), (15), (21), and (27) atoms in the defect fragments; for the X-interstitial series (a)–(d) correspond to pairs with (8), (12), (16), and (22) atoms. In bulk fragments corresponding to a certain defect fragment, there is always one atom less. Fragment Si atoms are shown in orange; environment Si atoms are shown as smaller midnight-blue balls. For clarity, we only show a subset of the atoms in the periodic simulation cells of Fig. 1.

FCIQMC calculations are done using the NECI code [49], and coupled-cluster calculations are done using the ElemCo.jl package [50]. As CC variants we employ coupled cluster with singles and doubles (CCSD), CCSD with perturbative triples (CCSD(T)), full triples (CCSDT) and distinguishable cluster with singles and doubles (DCSD) calculations [51–53].

For the non-TC CCSD(T), we present the complete-basis-set (CBS) estimate of the non-TC CCSD(T) energy, based on a two-point extrapolation from the DZ and TZ results, in Fig. 4 and Table I. We have assumed an exponential decay of the basis-set error in the HF energy as $E_n = E_\infty + B \exp(-Cn)$, with $C = 1.65$, as suggested by Jensen [54]. For the correlation energy, we used the standard extrapolation, assuming an inverse-cubic dependence of the basis set incompleteness error on the basis level.

NECI calculations are done with the adaptive-shift initiator FCIQMC (AS-FCIQMC) [11,12] using an initiator threshold of 10, which yields faster convergence than using a threshold

of 5, but we tested both to provide the same total energies to within 1 mE_h. We increased the walker populations in FCIQMC runs until convergence in total energies, with final populations ranging from 2×10^7 to 4×10^8 walkers depending on the basis and TC case.

In DMC calculations, we use cutoffs of $(L_u, L_\chi, L_f) = (5, 3, 3)$ bohr for the electron–electron, electron–nucleus, and electron–electron–nucleus Jastrow and backflow terms. We do two separate DMC runs with time steps of 0.02 and 0.005 a.u., with corresponding walker populations of 2000 and 8000 walkers, respectively, and extrapolate linearly and simultaneously to zero time step and infinite population. The trial wave function for DMC is the HF determinant obtained in a TZ Gaussian basis, using the same orbitals as the second-quantized methods used for the conventional eight-atom simulation cell.

To analyze the basis-set incompleteness errors in the silicon interstitial problem between different core treatments,

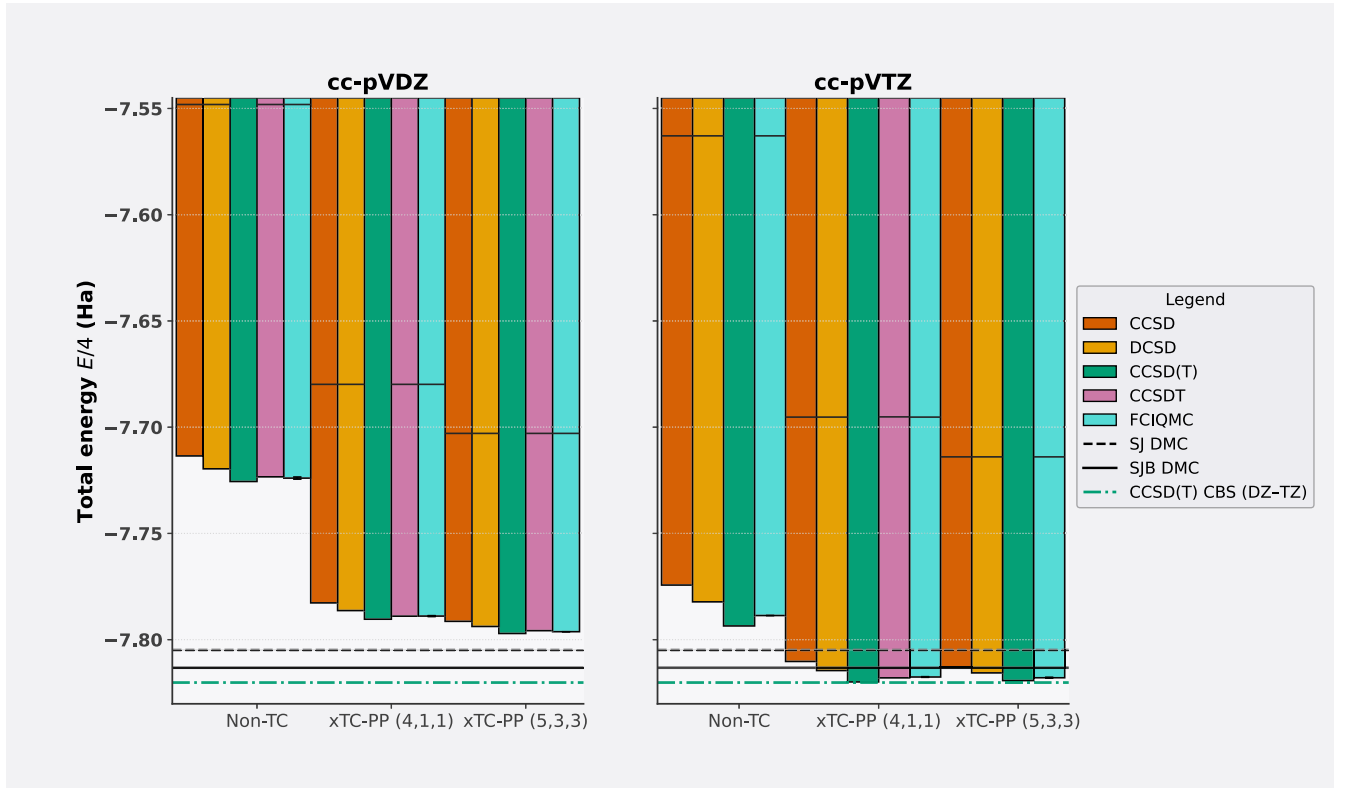


FIG. 4. Total electronic energies per primitive unit cell for bulk silicon obtained from eight-atom simulation-cell calculations. Bars show CCSD, DCSD, CCSD(T), and CCSDT results, together with the highest-population FCIQMC (init 10) energies, evaluated using Gaussian DZ and TZ basis sets for the xTC-PP Jastrow cutoff combinations 4 1 1 and 5 3 3, as well as non-TC cases. Horizontal lines indicate fixed-node diffusion Monte Carlo benchmarks with single-determinant (SJ, black dashed) and backflow-corrected (SJB, black solid) trial wave functions; stochastic DMC uncertainty is not visible because it is small. We have also drawn a black solid line across the bars in the histogram groups to denote the reference energy and the amount of captured correlation energy below the lines. We also show the extrapolated CSB CCSD(T) estimate for non-TC results as horizontal green dash-dotted line.

basis sets, and Hamiltonians in the HF level, we in addition describe in Appendix A a number of tests on the total and formation energies of silicon interstitials in 17-atom simulation cells. Also the reference energies of the 65-atom simulation cells used in the Results-section, both with and without TC, are presented in Appendix A.

IV. RESULTS

A. Bulk silicon

Figure 4 and Table I report the total electronic energies per primitive unit cell for bulk silicon, obtained from calcu-

lations in the conventional eight-atom periodic supercell. We compare coupled-cluster results (CCSD, DCSD, CCSD(T), and CCSDT) with the highest-population FCIQMC energies, and evaluate all of these using Gaussian DZ and TZ basis sets. For the transcorrelated calculations there are two Jastrow-cutoff parameter sets (411 and 533), alongside results obtained with the non-TC Hamiltonian. In addition, we include diffusion Monte Carlo benchmarks, obtained with both Slater-Jastrow (SJ) and backflow-corrected (SJB) forms in the TZ basis. The reference energies are drawn as well. We also show a CBS-extrapolated CCSD(T) estimate.

TABLE I. Total energies per primitive two-atom cell (Ha) based on an eight-atom bulk silicon simulation cell (see also Fig. 4).

Basis	TC type	CCSD	DCSD	CCSD(T)	CCSDT	FCIQMC	SJ-DMC	SJB-DMC
DZ	411	-7.782690	-7.786310	-7.790374	-7.788940	-7.7889(3)		
DZ	533	-7.791376	-7.793789	-7.797095	-7.795766	-7.79625(3)		
DZ	non-TC	-7.713525	-7.719594	-7.725563	-7.723366	-7.7240(5)		
TZ	411	-7.810257	-7.814538	-7.819887	-7.818081	-7.8176(2)		
TZ	533	-7.812840	-7.815637	-7.819259		-7.8179(2)	-7.8049(2)	-7.8136(3)
TZ	non-TC	-7.774302	-7.782201	-7.793540		-7.78864(3)		
CBS	non-TC			-7.820187				

Figure 4 shows that the CC energy decreases through the CCSD-DCSD-CCSD(T) hierarchy. The CCSDT and FCIQMC methods add a small positive correction to CCSD(T) of ~ 1 – 2 mE_h in both DZ and TZ. CCSDT agrees with FCIQMC benchmarks within statistical error except for the 5 3 3 DZ Hamiltonian that shows a minor deviation of ~ 0.5 mE_h.

CCSD and DCSD have larger discrepancy with the FCI result than CC with triples, but xTC-PP-CC methods are generally closer to FCIQMC than non-TC CC because of the wave-function compactification. Non-TC CCSD differs from FCIQMC by 4–14 mE_h, xTC-PP-CCSD by 3–8 mE_h. The DCSD is a small over- and CCSD(T) a small underestimation of FCIQMC energy, with deviations up to 3.5 and -3 mE_h, respectively, for xTC-PP Hamiltonians. In the non-TC case, the respective deviations are 6.5 and -5 mE_h. As a conclusion, we find that xTC-PP-CC with triples, either perturbative or full, provides very accurate results against xTC-PP-FCIQMC.

Examining the reference energies in Fig. 4, we observe that the primary advantage of the transcorrelated (TC) approach is the substantial improvement in the xTC reference energy over the HF reference energy. The xTC-PP reference energies are systematically lower, reducing the amount of residual correlation energy that the post-HF methods must recover compared to the non-TC Hamiltonian. This improvement becomes more pronounced as the Jastrow factor is enlarged. Nevertheless, the coupled-cluster and FCIQMC methods ultimately bring the total energies into close agreement across different Jastrow choices, as discussed in the next paragraph.

Figure 4 shows that in the DZ basis the two Jastrow factors still yield noticeably different xTC-PP energies (6–7 mE_h). In contrast, the TZ basis removes this sensitivity for high-level methods: xTC-PP CCSD(T) energies agree to within <0.6 mE_h between the two Jastrows. The xTC-PP-FCIQMC energies obtained with the different Jastrows agree virtually exactly. Since any FCI-quality method can differ between similarity-transformed Hamiltonians only through basis-set incompleteness, this agreement demonstrates that TC essentially eliminates basis-set error in TZ.

Because CCSDT is of the FCI-quality here, and CCSD(T) reproduces it extremely well, their consistency across Jastrow choices provides a strong indicator that the xTC-PP results are at (or extremely near) the CBS limit. This is confirmed by the non-TC CBS CCSD(T) benchmark, which matches both TZ xTC-PP-CCSD(T) energies.

These findings make the xTC-PP-CCSD(T) the cheapest reliable FCI-quality method in this context. Larger Jastrow factors may further compactify the wave function, as revealed by the xTC-PP-CCSD and -DCSD results with different Jastrow factors, and lower the CC level needed for FCI-quality accuracy. We can see that the xTC-PP-CCSD(T) with 533 Jastrow has a discrepancy with FCIQMC of 1.6 mE_h, just within the chemical-accuracy range, while with 411 the discrepancy is 0.6 mE_h larger. We leave a systematic study of further wave-function compactification with larger Jastrows for future work.

Finally, we compare the xTC-PP results to DMC. The basis-set error in DMC can be expected to be much smaller than in non-TC second quantized DZ and TZ calculations, since the simulation is done in real space using a continuum

representation. We can see that the backflow parametrization yields a notable energy lowering of ~ 9 mE_h compared to SJ DMC. The SJB energy is ~ 4.3 mE_h above the xTC-PP-FCIQMC and xTC-PP-CCSDT results in the TZ basis and corresponds roughly to the xTC-PP-CCSD energy in TZ. The fact that backflow introduces a notable correction means that the DMC energy is not converged with respect to wave-function complexity. Introducing more variational parameters into the DMC calculation, such as multideterminant expansions, is expected to lower the DMC energy further. Also using DFT-based orbitals instead of HF could improve the DMC energy [55]. It is difficult to know how much the DMC energy would decrease with a more involved trial wave function, but SJB-DMC is often capable of recovering about half of the correlation energy missing at the SJ-DMC level [47,56], which would mean SJB-DMC is in agreement with our xTC-PP-FCIQMC energies with an uncertainty of a few mE_h.

In conclusion, we find strong indication of basis-set convergence in the transcorrelated calculations with TZ basis sets: CBS extrapolated non-TC CCSD(T) energy and xTC-PP-CCSD(T) energies with two different Jastrow factors mutually agree. Furthermore, the SJB-DMC energy is also very close to those with the remaining small discrepancy possibly because of the missing correlations caused by the residual fixed node error. We have systematically increased the method accuracy up to FCIQMC and CCSDT levels, finding reliable benchmarks for the correlation treatment. The computationally feasible xTC-PP-CCSD(T) method in TZ basis is found to provide very accurate total energies, with nearly an exact match to FCIQMC and CCSDT benchmarks.

B. Interstitial defects

In this section, we present the formation energies of the silicon self-interstitials, obtained in the 65-atom supercells using embedded fragments with non-TC and xTC-PP coupled cluster treatment. For the sake of completeness, we also report analysis on the magnitudes of basis-set and core-exclusion errors, evaluated at the HF level with smaller supercells, in Appendix A. The data in Appendix A also contains the reference energies of the larger supercells, to which the CC correlation energies from embedded fragments are added to obtain the final formation energy estimates.

Figure 5 summarizes the formation energies of H and X interstitials obtained from embedded CC calculations across fragment sizes and basis levels, using both non-TC and xTC-PP Hamiltonians. Also, prior benchmark values from theory and experiments are shown, see also Fig. 6 and Table II. In all cases, the formation energies decrease systematically with increasing fragment size, reflecting the improved description of correlation effects.

Without transcorrelation, the formation energies are substantially overestimated. For the H defect, DZ values of 6.2–7.2 eV lie way above the experimental range of ~ 4.2 – 4.7 eV [57–61], while TZ reduces them by about 1 eV but still leaves a noticeable discrepancy. The X defect follows a similar pattern, with DZ values of 6.0–6.8 eV lowered by ~ 1 eV at the TZ level. Notably, the relative ordering of the two defects reverses with increasing fragment size: X has the higher for-

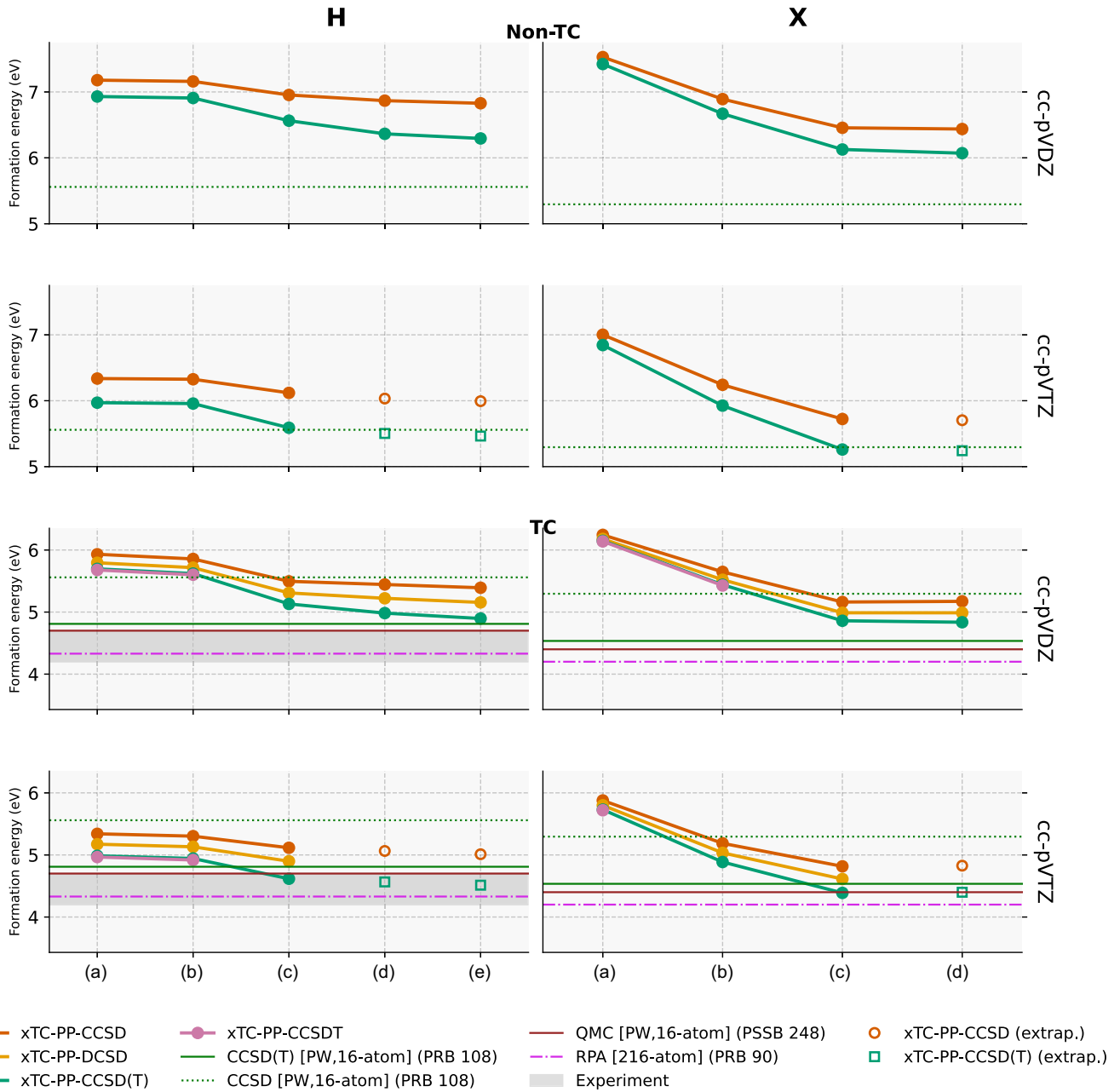


FIG. 5. Formation energies of H- and X-interstitials obtained from correlated wave-function methods at different basis levels. The upper panels show nontranscorrelated (Non-TC) results, and the lower panels show their transcorrelated (xTC-PP) counterparts. Horizontal reference lines indicate periodic benchmark values. Experimental references fall within the gray shaded region for H-interstitial. The x axis labels (a–e for H, a–d for X) correspond to the fragment pairs defined in Fig. 3. The Horizontal reference line references are: CCSD [62], CCSD(T) [62], QMC [63] (stochastic error bar of QMC is omitted for clarity), RPA [64], HSE [62], and PBE [64]. The experimental range of results from different sources [57–61] is shown by gray shaded region.

mation energy for small fragments but becomes lower than H for larger ones, consistent with periodic benchmarks [62–64].

The non-TC results capture the trends in convergence with respect to fragment size: Smaller fragments systematically overestimate formation energies, which approach the periodic supercell limit as fragment size increases. For H, fragments d and e (Fig. 3) give nearly identical values, and for X, fragments c and d are already converged within a few tens of meV. Because the largest fragments are computationally

prohibitive in TZ, we employ an extrapolation scheme that extends each CC flavor using fragment-size corrections from the lower basis set. Because we can do TZ-basis calculations up to fragments (c) for H and X defects, which are found to be very close to convergence in DZ level, the approximation in TZ is expected to be very good. The extrapolated TZ formation energies are reported in Fig. 6 and Table II.

Finally, within the non-TC series, CCSD(T) yields formation energies lower than CCSD by about 0.5 eV for H

TABLE II. Largest-fragment formation energies (eV) at the TZ level, and the reference benchmarks. Values with $\dagger$ are extrapolated from smaller fragments; letters in parentheses refer to fragment pairs in Fig. 3. QMC energy is shown with the stochastic error bar in the last decimal in parenthesis.

Defect	Series	CCSD	DCSD	CCSD(T)
H	Non-TC	5.993 † (e)		5.464 † (e)
xTC-PP	5.012 † (e)	4.899 (c)	4.513 † (e)	
Reference: CCSD [62] = 5.560, CCSD(T) = 4.810 [62], QMC = 4.7(1) [63], RPA = 4.33 [64], HSE = 4.82 [62], PBE = 3.626 [64], G_0W_0 = 4.40 [65]; experiment $\sim$ 4.2–4.7 eV.				
X	Non-TC	5.704 † (d)		5.242 † (d)
xTC-PP	4.827 † (d)	4.613 (c)	4.399 † (d)	
Reference: CCSD [62] = 5.295, CCSD(T) = 4.535 [62], QMC = 4.4(1) [63], RPA = 4.20 [64], HSE = 4.46 [62], PBE = 3.566 [64], G_0W_0 = 4.46 [65].				

and 0.3 eV for X, bringing them systematically closer to the benchmark values. This means that inclusion of triple excitations—explicitly or in approximate form—is important.

When we use xTC-PP Hamiltonians for the formation energies, we see a considerable improvement in results, with formation energy reducing from the non-TC case by 1.3–1.5 eV in DZ and by $\sim$ 1 eV in TZ. This brings the xTC-PP-CC energies into a very good agreement with experimental and theoretical benchmarks. In TZ we again employ the same 1.3–1.5 eV extrapolation for large-fragment results. In addition to xTC-PP-CCSD and xTC-PP-CCSD(T) results, we also show xTC-PP-DCSD results, which lie between xTC-PP-CCSD and xTC-PP-CCSD(T) values. Also, for smaller fragments, we performed full triples xTC-PP-CCSDT calculations, showing that the full triples correction on top of to xTC-PP-CCSD(T) energy is negligible. This confirms that xTC-PP-CCSD(T) is capturing all the relevant correlations.

The formation energy obtained with the reference energies alone is the same for all fragment sizes, as it is evaluated with the full periodic calculation. Comparison of the effect of TC on formation energies in Fig. 5 and Table III of Appendix A shows that the corrections introduced by xTC-PP are mostly because of correction in the reference energy, before simulating correlations with CC. This is expected as within the TC

treatment a large fraction of correlations is already included in the reference wave function, and the remaining part to be captured by CC is smaller.

The formation energy convergence with respect to fragment size is found to be similar to the non-TC case. As we are correlating the fragment explicitly with the environment via the xTC-PP Jastrow interaction, this finding is somewhat surprising. Yet, how general this effect is remains to be investigated on more comprehensive benchmarks. Furthermore, it should be noted that with the Jastrow cutoffs of 4 1 1 that we use here, the Jastrow correlation length is not very long-ranged (Si bond length is of the order of 4.35 Bohr), which limits the fragment-environment correlation effects to relatively short range. Use of longer-ranged Jastrows can be expected to accelerate convergence to the TDL, and will be pursued in future studies.

The extrapolated largest-fragment TZ results are compared with prior theoretical and experimental benchmarks in Fig. 6 and Table II. Among the methods considered, backflow-QMC [63] (17-atom), RPA [64] (217-atom), G_0W_0 [65] (64-atom), and xTC-PP-CCSD(T) (65-atom) yield H-interstitial formation energies within the experimental range, although, because of large stochastic error, the agreement of DMC with experiment is unclear. The transcorrelated hierarchy [xTC-PP-CCSD $\rightarrow$ xTC-PP-DCSD $\rightarrow$ xTC-PP-CCSD(T)] shows clear and systematic improvement.

As seen for the total energy in the previous section, a lower-level xTC-PP-CC method, xTC-PP-DCSD, has a relatively good match with SJB DMC. For total energy of the conventional eight-atom silicon cell, xTC-PP-CCSD and xTC-PP-DCSD align with SJB-DMC. These results hint that xTC-PP-CC methods with single and double excitations could serve as a cost-effective surrogate for backflow-DMC, while higher levels of transcorrelated coupled cluster or even FCIQMC can be used to improve the accuracy even further.

However, comparison to the theoretical benchmarks should be done with some caution, since most studies have been done in smaller, 17-atom simulation cells as opposed to our embedding calculations in 65-atom supercells. Notably, the G_0W_0 method was done in 64-atom supercells [65], and yields the best match with our results, although with reversed ordering for the H and X formation energies.

The effect of the supercell size on the formation energies in 16, 64, and 216 atom supercells has been studied with the random phase approximation (RPA). Their results with 17-

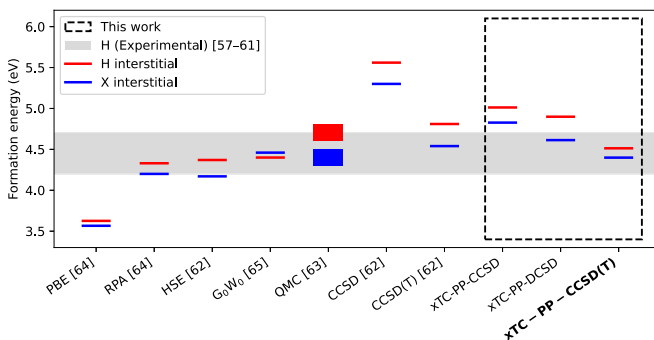


FIG. 6. Comparison of largest-fragment formation energies (eV) at the TZ level from this work to previous theoretical benchmarks and experiment. We plot the formation energy from this work and other references as red (H) and blue (X) horizontal lines. Experimental references fall within the grey shaded region for H interstitial. For the reference values we give the citation numbers in *x* labels. References for experimental range are [57–61].

TABLE III. Total and formation energies for different basis sets at the periodic Hartree-Fock level of theory and calculated from the expectation values of the respective xTC-PP-Hamiltonians with the HF wave function. AE stands for all-electron, ECP for effective-core-potential, PAW for projector-augmented wave.

Family	Basis	E_{bulk}	E_{hex}	E_X	$E_{\text{H}}^{\text{form}}$	E_X^{form}	Δ
16(17)-atom supercell $6 \times 6 \times 6$ k-mesh, $E_{\text{HF}}^{\text{per}}$							
Periodic POB	AE/POB-DZVP	−77.7500	−4910.5492	−4910.5603	8.87	8.56	0.30
Periodic POB	AE/POB-TZVP	−577.8461	−4911.3791	−4911.4096	8.53	7.70	0.83
Periodic POB	AE/POB-DZVP-REV2 ^a	−577.7374	−4910.4486		8.69		
Periodic POB	AE/POB-TZVP-REV2	−577.8777	−4911.6477	−4911.6650	8.51	8.04	0.47
Dunning	AE/cc-pVDZ ^b	−577.9152	−4911.9625	−4911.9747	8.63	8.30	0.33
Dunning	AE/cc-pVTZ ^c	−577.9345	−4912.1336	−4912.1433	8.44	8.18	0.26
Ahlrichs	AE/def2-SVP ^d	−577.7340	−4910.4197	−4910.4341	8.70	8.31	0.39
Ahlrichs	AE/def2-TZVP ^c	−577.9258	−4912.0650	−4912.0747	8.30	8.03	0.27
Ahlrichs	AE/def2-TZVPP ^c	−577.9259	−4912.0655	−4912.0754	8.30	8.03	0.27
ccECP	ECP/cc-pVDZ	−60.5148	−63.9716	−63.9859	8.86	8.46	0.40
ccECP	ECP/cc-pVTZ	−60.6044	−64.0867	−64.0962	8.31	8.05	0.26
ccECP	ECP/PW ($E_{\text{cut}} = 1088\text{eV}$)	−60.6281	−64.1137	−64.1194	8.26	8.11	0.15
Ref. [62]	PAW/PW ($E_{\text{cut}} = 400\text{eV}$)				8.16	7.93	0.23
64(65)-atom supercell, $3 \times 3 \times 3$ k-mesh, $E_{\text{HF}}^{\text{per}}$							
ccECP (nonl.)	ECP/cc-pVDZ	−242.0233	−245.4765	−245.4989	8.94	8.33	0.61
ccECP (nonl.)	ECP/cc-pVTZ	−242.3970	−245.8787	−245.8962	8.32	7.85	0.47
64(65)-atom supercell, $1 \times 1 \times 1$ k-mesh, $E_{\text{HF}}^{\text{xTC-PP-per}}$							
H Jastrow	ECP/cc-pVDZ	−246.2987	−249.8734		7.45		
X Jastrow	ECP/cc-pVDZ	−246.3144		−249.8994		7.17	(0.27)
H Jastrow	ECP/cc-pVTZ	−246.6819	−250.2836		6.88		
X Jastrow	ECP/cc-pVTZ	−246.6965		−250.3026		6.76	(0.12)

^aThe SCF for the X-interstitial did not converge.

^bA cutoff threshold for eigenvalues of the overlap matrix of 10^{-5} was used.

^cA cutoff threshold for eigenvalues of the overlap matrix of 10^{-3} was used.

^dA cutoff threshold for eigenvalues of the overlap matrix of 10^{-4} was used.

and 65-atom supercells are equal for the X-defect, and the larger cell brings the H formation energy down by 80 meV. This is not enough to match our results with those in [62]. The results of the 64-atom supercells were found to be within 70 meV of the results of the 216-atom supercells for both defects. This indicates that the finite-size errors on formation energies with 65-atom cells are small, but not insignificant. The use of our xTC-PP embedding method with even larger supercells will be a subject of following studies.

Based on the findings here and for the bulk in the previous section, our results are likely to have very small basis-set errors, they are essentially capturing the full wave-function correlation effects for these defects, and they are obtained in cells with modest finite-size errors.

V. CONCLUSIONS AND OUTLOOK

We have developed a transcorrelated wave-function framework for pristine and defected solids. For bulk silicon, fully periodic calculations with xTC-PP-CC and xTC-PP-FCIQMC using different Jastrow factors indicate that transcorrelation substantially accelerates basis convergence in Gaussian frameworks. xTC-PP-CCSD(T) in a TZ basis reaches accuracy comparable to xTC-PP-FCIQMC and seems to

improve upon fixed-node DMC benchmarks. For silicon self-interstitials, the transcorrelated embedding formulation yields formation energies that drastically improve upon non-TC calculations and agree well with established periodic references.

Methodologically, there were three most important developmental steps. First, the existing xTC-PP approach needed to be implemented for periodic systems, so that evaluation of Jastrow factors and pseudopotential commutators leveraged minimum-image convention. Second, we found that the numerical integration used to construct the xTC-PP Hamiltonian was greatly accelerated by the use of uniform real-space grids. And finally, the formulation of the transcorrelated embedding scheme was established. The possible implementation of k -point sampling in the xTC-PP Hamiltonian remains a topic for future work.

The present TC study has limitations that point to clear next steps. Our validation focuses on silicon and neutral self-interstitials; broader assessments on ionic, magnetic, and low-symmetry materials—including systems with stronger static correlation—are needed to probe generality. In particular, magnetic systems do not pose a fundamental obstacle to the transcorrelated framework: spin-polarized calculations are straightforward, opening the door to studies of spin gaps, excitations, and band gaps in magnetically ordered

materials. Strong static correlation can likewise be addressed within the present approach, by optimizing multideterminant Jastrow factors in the variational Monte Carlo step, and by pairing the TC Hamiltonian with methods such as FCIQMC that, unlike coupled cluster, do not rely on a single-reference expansion and can therefore capture multireference character. Charged defects and open-shell systems, by contrast, are limited by the periodic mean-field embedding employed here, as discussed further below.

The Gaussian basis sets used here are originally optimized for nonperiodic systems [46], and yet we found near convergent results without linear dependency issues. This sets the stage for future studies with basis sets built for periodic systems—possibly optimized directly for use with xTC-PP—for even faster convergence. Also, the Jastrow factors employed here have relatively short correlation lengths; exploring longer-ranged forms may enhance fragment–environment correlations and accelerate convergence to the thermodynamic limit for fully periodic systems. Also, there is an indication in this work that the use of larger Jastrow factors can help to compactify the wave function, allowing lower level CC methods—or even just xTC with the HF reference determinant—to reach good accuracy. The study of the use of much longer-range Jastrow factors is left for future.

The periodic mean-field embedding scheme employed in this study has limitations for open-shell, charged, or metallic systems, but the transcorrelated embedding framework itself is general and can be combined with alternative embedding strategies. Particularly promising in this respect, at least for nonconducting systems, is the recently introduced aperiodic defect model [19]. Not only does it eliminate the need for expensive large supercell calculations, but also, because of the absence of the unphysical defect replicas, treatment of open-shell and/or charged defects becomes straightforward.

A particular problem in this study was the need for large embedding fragments because of the displaced atoms that should be contained within the fragment. Yet, for simpler defects where the structural relaxation involves only the neighboring atoms, the xTC-PP embedding may converge already with rather modest fragments. Relatively small fragments may also be sufficient for studying local excitations on defects, as vertical excitation energies are obtained from calculations within a single structure.

Fully periodic, nonembedded simulations of solid-state systems with xTC-PP Hamiltonians represent another natural direction for further study. While we have shown that xTC-PP-CC with triples included can achieve FCI-quality total energies, the use of xTC-PP Hamiltonians at the MP2, CCSD, or DCSD levels can extend applicability to larger supercells, enabling systematic studies of finite-size effects under transcorrelation.

Overall, our results demonstrate that transcorrelation substantially mitigates basis-set incompleteness in correlated treatments of solids, and that embedding extends these gains to realistic defect cells. By maintaining systematic improvability of correlation treatment while leveraging compact TC Hamiltonians, the approach provides a practical route toward quantitatively reliable wave-function studies of pristine

and defected crystal structures at controlled computational cost.

DATA AVAILABILITY

The data supporting the findings in this article are available from the authors upon reasonable request.

APPENDIX A: BASIS SET COMPARISON

In Table III, we have listed the silicon interstitial (and bulk) total and formation energies in 17(16)-atom simulation cells obtained with different Gaussian basis sets and with plane-wave basis at the HF level. When looking at the ccECP results, we see a discrepancy of 0.06 eV between the triple- ζ and PW results. The discrepancy between the PW calculations, using PAW and ccECP core treatments, is of the order of 0.1 eV. If omitting the POB basis-set family, the discrepancies between all the Gaussian-type-orbital valence-triple- ζ —within both the all-electron and ccECP treatments—and PW results are about 0.2 eV for the formation energies and 0.1 eV for the relative stability. These values give an estimate of the magnitude of errors in HF energy because of basis-set incompleteness, approximate core treatment, etc. Although in the previous xTC-PP studies we have found ccECPs to provide very accurate total correlated energies for molecules and atoms [28,29], the mismatch between different PW calculations, which is of a comparable magnitude, hints that there might be nonnegligible errors owing to the approximate core electron representation. Nevertheless, for the interstitial formation energies of this study with larger simulation cells, this error estimate is likely an upper bound, since the transcorrelated treatment reduces the basis-set incompleteness error. We also note here that the uncertainty between different experimental estimates of the formation energy of the H-interstitial [57–61] is about 0.5 eV. Hence, with such relatively small discrepancies in the HF results and with the evidence of reduction of the basis-set errors with the TC treatment, we expect to be able to provide realistic estimates of the formation energies, at least with the TZ basis. As concerns the valence-double- ζ basis set level or the POB-basis sets altogether the deviation between the formation energies is much larger: up to 0.7 eV, which may be too large to be repaired by TC.

For the data in Table III, we performed a set of test periodic HF calculations for 16-(bulk) and 17-(defect)atom cells with structures taken from Ref. [62]. We used both Gaussian and plane-wave (PW) basis set with the ccECPs and calculated the formation energies E^{form} for both defects and their relative stabilities Δ . For these supercells, we used the $6 \times 6 \times 6$ Monkhorst-Pack $\mathbf{k}$ -point grids. The PW calculations were carried out with the Quantum Espresso package [66]. The tests indicated that the PW cutoff of 1088 eV allows for convergence within 1 meV/atom. The results with valence-double- ζ , valence-triple- ζ , and PW basis sets are presented in Table III with those of Ref. [62] that used PAW for the core treatment and a PW cutoff of 400 eV.

In Table III we also include the reference ccECP energies for larger 64/65-atom cells, calculated using $3 \times 3 \times 3$

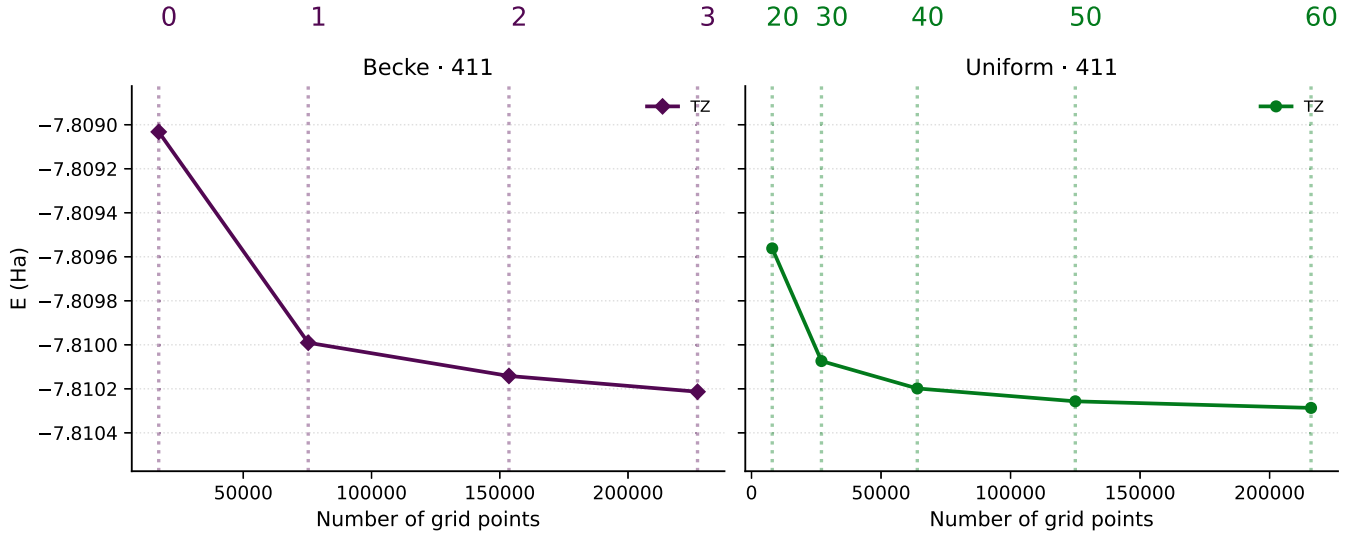


FIG. 7. Convergence of the xTC-PP-CCSD energy in an eight-atom bulk Si cell with respect to the number of integration grid points used for evaluating the xTC-PP integrals. The Jastrow cutoffs are set to $(L_u, L_\chi, L_f) = (4, 1, 1)$ bohr. In left, we show the convergence using Becke-type grids obtained from PySCF [44], using the grid density levels 0,1,2,3 as implemented in PySCF [44]. In right, we show the convergence using uniform grids. The Becke grid level is controlled by an integer from 0 to 3. The uniform grid is controlled by the number of points along the Cartesian axes of the cubic simulation cell. Results were evaluated in TZ basis.

Monkhorst-Pack $\mathbf{k}$ -point grids. The structures of these supercells were optimized at the DFT-HSE06/PW level with the fixed experimental lattice parameters. These HF orbitals were also used as a starting point for all the fragment formation energy calculations presented in this paper. In Table III, we also provide the expectation values for the xTC-PP Hamiltonians with the HF wave function for these supercells. The xTC-PP results are obtained with the Γ point only.

The presented results indicate that the deviation in the formation energies because of the larger cell size is rather small. At the same time, the xTC-PP Hamiltonian provides a

considerable improvement towards the benchmark values (see Table II) compared to bare HF.

APPENDIX B: XTC-PP INTEGRATION GRIDS

Figure 7 presents the convergence of the xTC-PP-CCSD energy as a function of number of grid points for both Becke and uniform grids. Table IV shows the same data in numbers and also splits the total energy into reference and correlation energy.

TABLE IV. TZ, Jastrow 4-1-1: xTC-PP-CCSD reference, correlation, and total energies, evaluated with Si eight-atom conventional cell, for Becke and uniform grids. N_{grid} is the total number of integration points; all energies per primitive cell in Hartree.

Scheme	Grid	N_{grid}	E_{ref}	$E_{\text{corr}}^{\text{CCSD}}$	$E_{\text{tot}}^{\text{CCSD}}$
Becke	0	17028	-7.692753	-1.222104	-7.809033
	1	75320	-7.694719	-1.221097	-7.809991
	2	153580	-7.695026	-1.220940	-7.810142
	3	227104	-7.695166	-1.220872	-7.810213
Uniform	10	$10^3 = 1,000$	-7.685756	-1.226860	-7.806791
	20	$20^3 = 8,000$	-7.693926	-1.221460	-7.809562
	30	$30^3 = 27,000$	-7.694908	-1.220991	-7.810074
	40	$40^3 = 64,000$	-7.695153	-1.220870	-7.810198
	50	$50^3 = 125,000$	-7.695254	-1.220827	-7.810257
	60	$60^3 = 216,000$	-7.695303	-1.220808	-7.810287

- [1] K. A. Simula and I. Makkonen, Calculation of the energies of the multideterminant states of the nitrogen vacancy center in diamond with quantum Monte Carlo, *Phys. Rev. B* **108**, 094108 (2023).
- [2] S. Verma, A. Mitra, Y. Jin, S. Haldar, C. Vorwerk, M. R. Hermes, G. Galli, and L. Gagliardi, Optical properties of neutral F centers in bulk MgO with density matrix embedding, *J. Phys. Chem. Lett.* **14**, 7703 (2023).
- [3] E. Ertekin, L. K. Wagner, and J. C. Grossman, Point-defect optical transitions and thermal ionization energies from quantum Monte Carlo methods: Application to the F-center defect in MgO, *Phys. Rev. B* **87**, 155210 (2013).
- [4] Y. Chen, T. Jiang, H. Chen, E. Han, A. Alavi, K. Yu, E. Wang, and J. Chen, Multiconfigurational nature of electron correlation within nitrogen vacancy centers in diamond, *Phys. Rev. B* **108**, 045111 (2023).
- [5] Y. Chen, H. Chen, N. Bogdanov, K. Yu, A. Alavi, E. Wang, and J. Chen, Towards an accurate electronic structure of single photon emitters in hexagonal boron nitride, *Phys. Rev. Res.* **7**, L012079 (2025).
- [6] T. Gruber, K. Liao, T. Tsatsoulis, F. Hummel, and A. Grüneis, Applying the coupled-cluster Ansatz to solids and surfaces in the thermodynamic limit, *Phys. Rev. X* **8**, 021043 (2018).
- [7] N. Masios, A. Irmeler, T. Schäfer, and A. Grüneis, Averting the infrared catastrophe in the gold standard of quantum chemistry, *Phys. Rev. Lett.* **131**, 186401 (2023).
- [8] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Quantum Monte Carlo simulations of solids, *Rev. Mod. Phys.* **73**, 33 (2001).
- [9] J. Zhang, M.-F. Chen, A. Rettig, T. Jiang, P. J. Robinson, H. Q. Dinh, A. Z. Ni, and J. Lee, *Ab initio* auxiliary-field quantum Monte Carlo in the thermodynamic limit, [arXiv:2602.16679](https://arxiv.org/abs/2602.16679).
- [10] G. H. Booth, A. J. W. Thom, and A. Alavi, Fermion Monte Carlo without fixed nodes: A game of life, death, and annihilation in Slater determinant space, *J. Chem. Phys.* **131**, 054106 (2009); D. Cleland, G. H. Booth, and A. Alavi, Communications: Survival of the fittest: Accelerating convergence in full configuration-interaction quantum Monte Carlo, *ibid.* **132**, 041103 (2010).
- [11] K. Ghanem, A. Y. Lozovoi, and A. Alavi, Unbiasing the initiator approximation in full configuration interaction quantum Monte Carlo, *J. Chem. Phys.* **151**, 224108 (2019).
- [12] K. Ghanem, K. Guther, and A. Alavi, The adaptive shift method in full configuration interaction quantum Monte Carlo: Development and applications, *J. Chem. Phys.* **153**, 224115 (2020).
- [13] O. Masur, M. Schütz, L. Maschio, and D. Usvyat, Fragment-based direct-local-ring-coupled-cluster doubles treatment embedded in the periodic Hartree–Fock solution, *J. Chem. Theory Comput.* **12**, 5145 (2016).
- [14] H.-H. Lin, L. Maschio, D. Kats, D. Usvyat, and T. Heine, Fragment-based restricted active space configuration interaction with second-order corrections embedded in periodic Hartree–Fock wave function, *J. Chem. Theory Comput.* **16**, 7100 (2020).
- [15] E. M. C. Christlmaier, D. Kats, A. Alavi, and D. Usvyat, Full configuration interaction quantum Monte Carlo treatment of fragments embedded in a periodic mean field, *J. Chem. Phys.* **156**, 154107 (2022).
- [16] Tobias Schäfer, F. Libisch, G. Kresse, and A. Grüneis, Local embedding of coupled cluster theory into the random phase approximation using plane waves, *J. Chem. Phys.* **154**, 011101 (2021).
- [17] B. T. G. Lau, G. Knizia, and T. C. Berkelbach, Regional embedding enables high-level quantum chemistry for surface science, *J. Phys. Chem. Lett.* **12**, 1104 (2021).
- [18] L. Muechler, D. I. Badrtdinov, A. Hampel, J. Cano, M. Rösner, and C. E. Dreyer, Quantum embedding methods for correlated excited states of point defects: Case studies and challenges, *Phys. Rev. B* **105**, 235104 (2022).
- [19] R. H. Lavroff, D. Kats, L. Maschio, N. Bogdanov, A. Alavi, A. N. Alexandrova, and D. Usvyat, Aperiodic fragments in periodic solids, *J. Chem. Phys.* **163**, 084105 (2025).
- [20] J. VandeVondele and J. Hutter, Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases, *J. Chem. Phys.* **127**, 114105 (2007).
- [21] M. F. Peintinger, D. V. Oliveira, and T. Bredow, Consistent Gaussian basis sets of triple-zeta valence with polarization quality for solid-state calculations, *J. Comput. Chem.* **34**, 451 (2013).
- [22] J. Lee, X. Feng, L. A. Cunha, J. F. Gonthier, E. Epifanovsky, and M. Head-Gordon, Approaching the basis set limit in Gaussian-orbital-based periodic calculations with transferability: Performance of pure density functionals for simple semiconductors, *J. Chem. Phys.* **155**, 164102 (2021).
- [23] D. Usvyat, Linear-scaling explicitly correlated treatment of solids: Periodic local MP2-F12 method, *J. Chem. Phys.* **139**, 194101 (2013).
- [24] Andreas Grüneis, J. J. Shepherd, A. Alavi, D. P. Tew, and G. H. Booth, Explicitly correlated plane waves: Accelerating convergence in periodic wavefunction expansions, *J. Chem. Phys.* **139**, 084112 (2013).
- [25] R. Sakuma and S. Tsuneyuki, Electronic structure calculations of solids with a similarity-transformed Hamiltonian, *J. Phys. Soc. Jpn.* **75**, 103705 (2006).
- [26] M. Ochi, K. Sodeyama, R. Sakuma, and S. Tsuneyuki, Efficient algorithm of the transcorrelated method for periodic systems, *J. Chem. Phys.* **136**, 094108 (2012).
- [27] J. P. Haupt, E. M. C. Christlmaier, P. Lopez Ríos, N. A. Bogdanov, D. Kats, and A. Alavi, Transcorrelated methods for multireference problems, *J. Chem. Phys.* **163**, 144113 (2025).
- [28] K. Simula, E. M. C. Christlmaier, M.-A. Filip, J. P. Haupt, D. Kats, P. Lopez-Rios, and A. Alavi, Transcorrelated theory with pseudopotentials, *J. Chem. Theory Comput.* **21**, 5155 (2025).
- [29] K. Simula, M.-A. Filip, and A. Alavi, Transcorrelated theory for transition-metal atoms, *Phys. Rev. A* **112**, 032805 (2025).
- [30] A. Ammar, A. Scemama, and E. Giner, Biorthonormal orbital optimization with a cheap core-electron-free three-body correlation factor for quantum Monte Carlo and transcorrelation, *J. Chem. Theory Comput.* **19**, 4883 (2023).
- [31] K. Liao, T. Schraivogel, H. Luo, D. Kats, and A. Alavi, Towards efficient and accurate *ab initio* solutions to periodic systems via transcorrelation and coupled cluster theory, *Phys. Rev. Res.* **3**, 033072 (2021).
- [32] T. Schraivogel, A. J. Cohen, A. Alavi, and D. Kats, Transcorrelated coupled cluster methods, *J. Chem. Phys.* **155**, 191101 (2021).
- [33] N. Lee and A. J. W. Thom, Studies on the transcorrelated method, *J. Chem. Theory Comput.* **19**, 5743 (2023).

- [34] K. Liao, H. Zhai, E. M. C. Christlmaier, T. Schraivogel, P. L. Ríos, D. Kats, and A. Alavi, Density matrix renormalization group for transcorrelated Hamiltonians: Ground and excited states in molecules, *J. Chem. Theory Comput.* **19**, 1734 (2023).
- [35] N. D. Drummond, M. D. Towler, and R. J. Needs, Jastrow correlation factor for atoms, molecules, and solids, *Phys. Rev. B* **70**, 235119 (2004).
- [36] A. J. Cohen, H. Luo, K. Guthier, W. Dobrautz, D. P. Tew, and A. Alavi, Similarity transformation of the electronic Schrödinger equation via Jastrow factorization, *J. Chem. Phys.* **151**, 061101 (2019).
- [37] J. P. Haupt, S. M. Hosseini, P. López Ríos, W. Dobrautz, A. Cohen, and A. Alavi, Optimizing Jastrow factors for the transcorrelated method, *J. Chem. Phys.* **158**, 224105 (2023).
- [38] E. M. C. Christlmaier, T. Schraivogel, P. López Ríos, A. Alavi, and D. Kats, xTC: An efficient treatment of three-body interactions in transcorrelated methods, *J. Chem. Phys.* **159**, 014113 (2023).
- [39] C. M. Zicovich-Wilson, R. Dovesi, and V. R. Saunders, A general method to obtain well localized wannier functions for composite energy bands in linear combination of atomic orbital periodic calculations, *J. Chem. Phys.* **115**, 9708 (2001).
- [40] C. M. Zicovich-Wilson and R. Dovesi, Localized functions in crystalline systems and their variational manifolds, in *Beyond Standard Quantum Chemistry: Applications From Gas to Condensed Phases*, edited by R. Hernandez-Lamonedá (Transworld Research Network, Trivandrum, Kerala, India, 2007), Vol. 125, pp. 140–169.
- [41] D. Usvyat, L. Maschio, C. Pisani, and M. Schütz, Second order local Møller-Plesset perturbation theory for periodic systems: The CRYSCor code, *Z. Phys. Chem.* **224**, 441 (2010).
- [42] J. Heyd, G. E. Scuseria, and M. Ernzerhof, Hybrid functionals based on a screened Coulomb potential, *J. Chem. Phys.* **118**, 8207 (2003).
- [43] G. Kresse and J. Furthmüller, Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set, *Phys. Rev. B* **54**, 11169 (1996).
- [44] Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui *et al.*, Recent developments in the PySCF program package, *J. Chem. Phys.* **153**, 024109 (2020).
- [45] R. Dovesi, A. Erba, R. Orlando, C. M. Zicovich-Wilson, B. Civalieri, L. Maschio, M. Rérat, S. Casassa, J. Baima, S. Salustro, and B. Kirtman, Quantum-mechanical condensed matter simulations with CRYSTAL, *WIREs Comput. Mol. Sci.* **8**, e1360 (2018).
- [46] A. N. Hill, A. J. H. M. Meijer, and J. G. Hill, Correlation consistent basis sets and core polarization potentials for Al–Ar with ccECP pseudopotentials, *J. Phys. Chem. A* **126**, 5853 (2022).
- [47] R. J. Needs, M. D. Towler, N. D. Drummond, P. López Ríos, and J. R. Trail, Variational and diffusion quantum Monte Carlo calculations with the CASINO code, *J. Chem. Phys.* **152**, 154106 (2020).
- [48] P. J. Knowles and N. C. Handy, A determinant based full configuration interaction program, *Comput. Phys. Commun.* **54**, 75 (1989).
- [49] K. Guthier, R. J. Anderson, N. S. Blunt, N. A. Bogdanov, D. Cleland, N. Dattani, W. Dobrautz, K. Ghanem, P. Jeszenszki, N. Liebermann *et al.*, NECI: N-electron configuration interaction with an emphasis on state-of-the-art stochastic methods, *J. Chem. Phys.* **153**, 034107 (2020).
- [50] D. Kats, T. Schraivogel, J. Hauskrecht, C. Rickert, and F. Wu, ElemCo.jl: Julia program package for electron correlation methods (2024), <https://elem.co.il/stable/>.
- [51] D. Kats and F. R. Manby, Communication: The distinguishable cluster approximation, *J. Chem. Phys.* **139**, 021102 (2013).
- [52] D. Kats, Communication: The distinguishable cluster approximation. II. The role of orbital relaxation, *J. Chem. Phys.* **141**, 061101 (2014).
- [53] D. Kats, E. M. C. Christlmaier, T. Schraivogel, and A. Alavi, Orbital optimisation in xTC transcorrelated methods, *Faraday Discuss.* **254**, 382 (2024).
- [54] F. Jensen, Estimating the Hartree-Fock limit from finite basis set calculations, *Theor. Chem. Acc.* **113**, 267 (2005).
- [55] A. Annaberdiyev, G. Wang, C. A. Melton, M. C. Bennett, and L. Mitás, Cohesion and excitations of diamond-structure silicon by quantum Monte Carlo: Benchmarks and control of systematic biases, *Phys. Rev. B* **103**, 205206 (2021).
- [56] P. López Ríos, A. Ma, N. D. Drummond, M. D. Towler, and R. J. Needs, Inhomogeneous backflow transformations in quantum Monte Carlo calculations, *Phys. Rev. E* **74**, 066701 (2006).
- [57] P. M. Fahey, P. B. Griffin, and J. D. Plummer, Point defects and dopant diffusion in silicon, *Rev. Mod. Phys.* **61**, 289 (1989).
- [58] A. Ural, P. B. Griffin, and J. D. Plummer, Self-diffusion in silicon: Similarity between the properties of native point defects, *Phys. Rev. Lett.* **83**, 3454 (1999).
- [59] H. Bracht, E. E. Haller, and R. Clark-Phelps, Silicon self-diffusion in isotope heterostructures, *Phys. Rev. Lett.* **81**, 393 (1998).
- [60] H. Bracht, N. A. Stolwijk, and H. Mehrer, Properties of intrinsic point defects in silicon determined by zinc diffusion experiments under nonequilibrium conditions, *Phys. Rev. B* **52**, 16542 (1995).
- [61] A. Ural, P. B. Griffin, and J. D. Plummer, Fractional contributions of microscopic diffusion mechanisms for common dopants and self-diffusion in silicon, *J. Appl. Phys.* **85**, 6440 (1999).
- [62] F. Salihbegovic, A. Gallo, and A. Grüneis, Formation energies of silicon self-interstitials using periodic coupled cluster theory, *Phys. Rev. B* **108**, 115125 (2023).
- [63] W. D. Parker, J. W. Wilkins, and R. G. Hennig, Accuracy of quantum Monte Carlo methods for point defects in solids, *Physica Status Solidi (b)* **248**, 267 (2011).
- [64] M. Kaltak, J. c. v. Klimeš, and G. Kresse, Cubic scaling algorithm for the random phase approximation: Self-interstitials and vacancies in Si, *Phys. Rev. B* **90**, 054115 (2014).
- [65] P. Rinke, A. Janotti, M. Scheffler, and C. G. Van de Walle, Defect formation energies without the band-gap problem: Combining density-functional theory and the GW approach for the silicon self-interstitial, *Phys. Rev. Lett.* **102**, 026402 (2009).
- [66] P. Giannozzi, Jr., O. Andreussi, T. Brumme, O. Bunau, M. B. Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni *et al.*, Advanced capabilities for materials modelling with quantum espresso, *J. Phys.: Condens. Matter* **29**, 465901 (2017).