

Supplementary Information of “A Comparative Study Using State-of-the-art Electronic Structure Theories on Solid Hydrogen Phases under High Pressures”

Ke Liao

Max Planck Institute for Solid State Research, Heisenbergstrasse 1, 70569 Stuttgart, Germany

Xin-Zheng Li

*School of Physics, State Key Laboratory for Artificial Microstructure and Mesoscopic Physics,
Peking University, 100871 Beijing, People’s Republic of China and
Collaborative Innovation Centre of Quantum Matter,
Peking University, 100871 Beijing, People’s Republic of China*

Ali Alavi

*Max Planck Institute for Solid State Research, Heisenbergstrasse 1, 70569 Stuttgart, Germany and
Department of Chemistry, University of Cambridge,
Lensfield Road, Cambridge CB2 1EW, United Kingdom*

Andreas Grüneis*

*Institute for Theoretical Physics, Vienna University of Technology,
Wiedner Hauptstrasse 8-10, 1040 Vienna, Austria*

* andreas.grueneis@tuwien.ac.at
k.liao@fkf.mpg.de

| Phases | C2/c-24 | Cmca-4 | Cmca-12 | P2 ₁ /c-24 | P6 ₃ /m-16 |
|--------------------------|---------|--------|---------|-----------------------|-----------------------|
| number of atoms in cell | 24 | 24 | 24 | 24 | 16 |
| number of bands per atom | / | 8 | 8 | 16 | 8 |

TABLE I: Summary of the number of bands per atom required to reach 1 meV/atom accuracy in the relative CCSD correlation energy for each phase at all pressures and the number of atoms in the simulation cells employed for this convergence test.

| | Cmca-4 | Cmca-12 | C2/c-24 | P2 ₁ /c-24 | P6 ₃ /m-16 |
|-----------------------|--------|---------|---------|-----------------------|-----------------------|
| Effective system size | 4096 | 8232 | 5184 | 5184 | 8192 |

TABLE II: Effective system size used to converge the HF energy of each phase to within 1 meV/atom at all pressures. The effective system size is defined as the product of the number of sampled k -points in the first BZ and the number of atoms in the supercell.

COMPUTATIONAL DETAILS

For the CCSD calculations we use the coupled cluster for solids (CC4S) code employing the cyclops tensor framework (CTF) [1] interfaced to the Vienna *ab-initio* simulation package (VASP). The projector augmented wave method, as implemented in VASP [2–4], is used for all calculations. We stress that HF energies and the correlation energies computed on the level of CCSD are converged to the complete basis set limit and the thermodynamic limit (TDL) separately using different schemes. We note that the total computational cost in obtaining the CCSD static enthalpy diagram is around 200 thousand CPU hours.

CCSD basis set convergence

We first outline the procedure to obtain the basis set converged CCSD correlation energies. Prior to the correlation energy calculations, we compute the occupied HF orbitals using a Γ -point sampling of the first Brillouin zone (BZ) for each phase at different volumes. We use a plane wave basis set corresponding to a kinetic energy cutoff of 700 eV. In addition, all occupied orbitals are constrained to be doubly occupied, avoiding fractional occupation numbers. This is achieved by utilizing the FERWE and ISMEAR flags in VASP. We note that this could prevent the HF orbitals (HFOs) from converging to the energetically lowest possible solution. However, the CCSD method can largely compensate for this constraint.

Following the HF ground state calculation, a full diagonalization of the Fock operator is carried out in the space of all employed basis functions, whose dimension is equal to the number of plane waves. We further construct MP2 natural orbitals (MP2NOs) by diagonalizing the virtual-virtual block of the one-body reduced density matrix in the HFO basis. As shown in Ref. [7], only a small fraction of the total MP2NOs can be used for the CCSD calculation without compromising the accuracy. We note that the Fock operator is not diagonal in the truncated MP2NO basis, thus it is necessary to recanonicalize in the subspace of the employed MP2NOs. The occupied HFO space is not affected by this procedure. In Fig. 1 we show the convergence of the correlation energies relative to phase C2/c-24 with respect to the number of natural orbitals per atom for Cmca-4, Cmca-12, P2₁/c-24 and P6₃/m-16. These calculations are carried out using the Γ -point sampling of the first BZ and the employed supercell sizes are summarized in Table I. Compared to 24 bands/atom, we find that 8 bands/atom suffice to achieve a basis set convergence for the relative correlation energy to within 1 meV/atom, except for P2₁/c-24, where the remaining error is approximately 4 meV/atom. Based on these findings, we conclude that for the 96-atom supercell, 800 bands (48 occupied HFOs + 752 MP2NOs) should yield a converged correlation energy difference to within a precision of approximately 1 meV/atom for all phases except for phase P2₁/c-24. We stress that the CCSD correlation energy differences converge much faster than the CCSD correlation energies of each phase alone due to error cancellations. The plane wave basis set cutoff for the overlap integrals is set to 600 eV in this step.

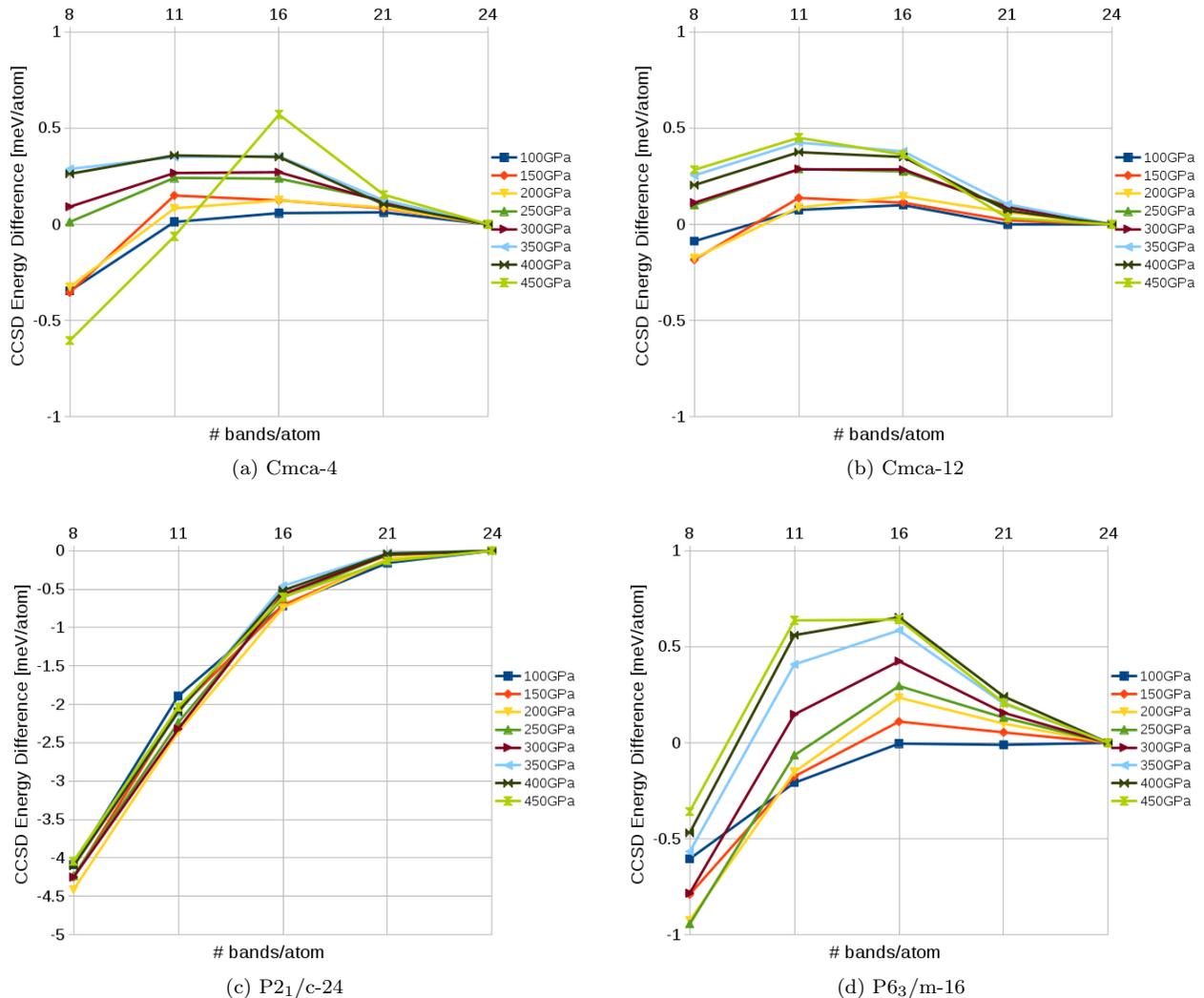


FIG. 1: (Color online) CCSD correlation energy difference convergence with respect to the number of bands per atom. Fig. 1a shows that the correlation energy difference between phase Cmca-4 and C2/c-24 converges to within 1 meV/atom with 8 bands per atom (1 occupied HFOs + 7 MP2NOs) at all pressures and the same accuracy is achieved with 8 bands per atom in Fig. 1b and Fig. 1d for phase Cmca-12 and P₆₃/m-16, respectively. Fig. 1c shows that the correlation energy difference between phase P₂₁/c-24 and C2/c-24 converges to within 1 meV/atom with 16 bands per atom (1 occupied HFOs + 15 MP2NOs) at all pressures. 24-atom cells are used for phase Cmca-4, Cmca-12 and P₂₁/c-24, while 16-atom cell is used for phase P₆₃/m-16.

Hartree–Fock finite size convergence

To obtain the converged Hartree–Fock energies per atom E^{HF} with respect to basis set, we use a kinetic energy cutoff for the plane wave basis of 700 eV. To be consistent with the HF step in the correlation energy calculations, we fix the occupancies in the manner described above. Moreover, sufficiently dense k -meshes are used to eliminate finite size errors. We use the effective system size, which is the product of the total number of k -points and the number of atoms in the supercell, as a measure of the system size. Table II summarizes the employed system sizes used to reach an accuracy of 1 meV/atom for E^{HF} .

CCSD finite size convergence

For the CCSD calculations, only one k -point is used to sample the first BZ of the 96-atom supercells. However, the twist-average (TA) technique [8] and finite size correction scheme [9, 10] are used to reach the thermodynamic limit and minimize finite size errors. The TA procedure corrects mostly for the one-body contribution to the correlation energy and the latter retrieves mainly the missing two-body contribution to the correlation energy at large distances that exceed the size of the employed supercell. We denote the total two-body finite size error corrected (FS) CCSD correlation energy as E^{FS} . The total twist-averaged correlation energy (TA) can be expressed as

$$E^{\text{TA+FS}} = \frac{1}{W} \sum_i^{N_t} E_i^{\text{FS}} w_i, \quad (1)$$

where $N_t \approx 13$ is the total number of twists that are generated using the irreducible wedge of a $3 \times 3 \times 3$ k -mesh. w_i is the number of equivalent k -points to the i th irreducible k -point and $W = \sum_j^{N_t} w_j = 27$ is the total weight. Applying a twist means that the first BZ is sampled with a shifted k -vector $\Delta \mathbf{k}_i$. We have checked the convergence with respect to the size of k -mesh used to generate the twists for Cmca-4 at 400 GPa. A $4 \times 4 \times 4$ k -mesh ($W = 64$) yields twist-averaged correlation energies that agree to within 1 meV/atom with the energies obtained using a $3 \times 3 \times 3$ k -mesh. We stress that TA accelerates convergence to the TDL mainly in small gapped systems. Therefore a single test on the Cmca-4 phase at high pressure ensures the convergence of the TA procedure with respect to the employed k -mesh in all other phases and pressures. We note in passing that 24 stochastic twists are used to sample the first BZ in Ref. [6] for carrying out TA. We stress that in CCSD theory, the total energy is partitioned into the mean field (HF) part and correlation energy part. Therefore relative fewer twists are needed for CCSD correlation energy calculations to reach convergence.

An acceleration scheme to reach the basis set convergence using MP2NOs is employed in conjunction with the twist-averaging technique: 400 bands (48 occupied HFOs + 352 MP2NOs) are used for each twist, and the basis set incomplete error is estimated as the energy difference between two calculations with 800 and 400 bands at the Γ -point,

$$\Delta E^{\text{bse}} = E_0^{\text{FS},800} - E_0^{\text{FS},400}. \quad (2)$$

The justification is that the contribution of the high energy bands is independent of the positions of the sampled k -points. This procedure saves a large amount of computational resources.

The total CCSD energy per atom is expressed as

$$E = \frac{1}{N} (E^{\text{TA+FS}} + \Delta E^{\text{bse}}) + E^{\text{HF}}, \quad (3)$$

where $N = 96$ is the total number of atoms in the supercell used in the CCSD calculations.

Post-CCSD error estimates

CCSD yields accurate results for systems composed of weakly interacting electron pairs. For the present study it is important to determine if the considered solid hydrogen phases belong to this class of systems and what accuracy can be expected from CCSD. We now estimate the error of the CCSD correlation energies by using more accurate post-CCSD wavefunction based theories; *e.g.* distinguishable cluster singles and doubles (DCSD) [11, 12], coupled cluster singles, doubles and perturbative triples (CCSD(T)) [13, 14] and full configuration interaction quantum Monte Carlo (FCIQMC). DCSD has been shown in Refs. [11, 12] to achieve qualitatively good ground state energies even in the dissociation limit of hydrogen systems by ignoring exchange interactions between electron pairs. In the dissociation limit, CCSD is not expected to provide even qualitatively correct results. Thus DCSD is a valuable tool to determine whether CCSD works reliably at high pressures, where the inter-atomic distances become comparable to the inter-molecular distances, especially in the potential metallic phase Cmca-4 where CCSD and DMC show a discrepancy. We note in passing that due to the perturbative nature of CCSD(T), reliable results in the thermodynamic limit for metallic periodic systems can not be obtained [15]. However, CCSD(T) results in a finite supercell are meaningful. FCIQMC[16, 17] and its initiator approximation [18] can obtain the exact ground state solution to the non-relativistic Schrödinger equation in a give basis set.

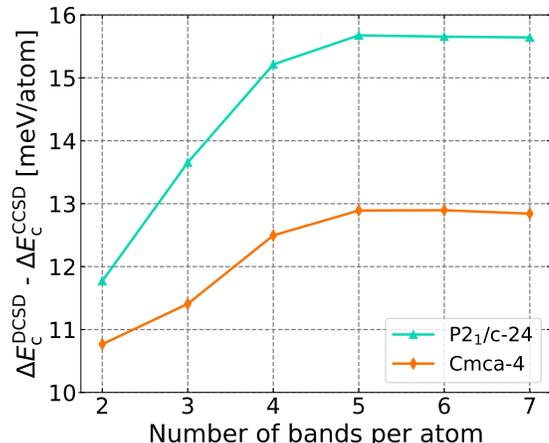


FIG. 2: (Color online) Difference of relative correlation energies (relative to phase C2/c-24) between DCSD and CCSD with respect to the number of bands (occupied HFOs + MP2NOs) per atom. 24-atom cells at 400 GPa DFT pressure are used in all three phases.

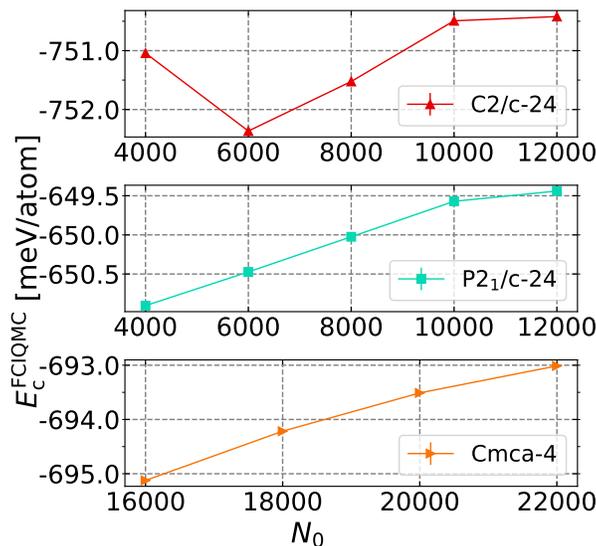


FIG. 3: (Color online) *i*-FCIQMC correlation energy E_c^{FCIQMC} convergence with respect to the number of walkers on the HF determinant N_0 . The statistical noise is smaller than the size of markers. 24-atom cells at 400 GPa DFT pressure are used and 60 bands (12 HFOs + 48 MP2NOs) are used for each phase.

DCSD, CCSD(T) and FCIQMC using Γ -point

We have performed calculations using DCSD, CCSD(T) and FCIQMC for phase C2/c-24, Cmca-4 and P2₁/c-24 at 400 GPa DFT pressure. Due to the unfavourable scaling of the computational complexity with system size, supercells that contain 24 atoms have been employed and the first BZs are sampled at the Γ -point only. Furthermore the computational cost of FCIQMC limits the number of MP2NOs. However, it is possible to obtain an error estimate of CCSD already using a very small basis set. As already shown in Fig. 1, CCSD correlation energies for various phases *relative to C2/c-24*, ΔE_c^{CCSD} , converge rapidly with respect to the employed number of bands. Fig. 2 depicts the difference between ΔE_c^{CCSD} and ΔE_c^{DCSD} , which corresponds to the *difference of the difference*, retrieved as a function of the number of bands per atom. We stress that the convergence of $\Delta E_c^{\text{CCSD}} - \Delta E_c^{\text{DCSD}}$ with respect to the number of bands is even faster than that of ΔE_c^{CCSD} . Already 2 bands per atom are sufficient to converge the difference between CCSD and DCSD to within 4 meV/atom. For the comparative studies between CCSD, DCSD,

| | E_c | E_c | E_c | ΔE_c | ΔE_c |
|------------------|------------|--------------------|------------|--------------------------------|--------------------|
| | C2/c-24 | P2 ₁ /c | Cmca-4 | (P2 ₁ /c - C2/c-24) | (Cmca-4 - C2/c-24) |
| CCSD | -707.658 | -627.121 | -665.296 | 80.537 | 42.362 |
| DCSD | -732.558 | -640.325 | -680.188 | 92.233 | 52.370 |
| CCSD(T) | -742.638 | -650.317 | -684.629 | 92.321 | 58.009 |
| <i>i</i> -FCIQMC | -750.42(1) | -649.44(0) | -693.01(3) | 100.98(1) | 57.40(8) |

TABLE III: Correlation energies obtained by CCSD, DCSD, CCSD(T) and *i*-FCIQMC on 24-atom cells for phase C2/c-24, P2₁/c and Cmca-4 at DFT pressure 400 GPa using 60 bands (12 HFOs + 48 MP2NOs) in total. A small number of bands are used because of the limited computational resources. All energies in meV/atom.

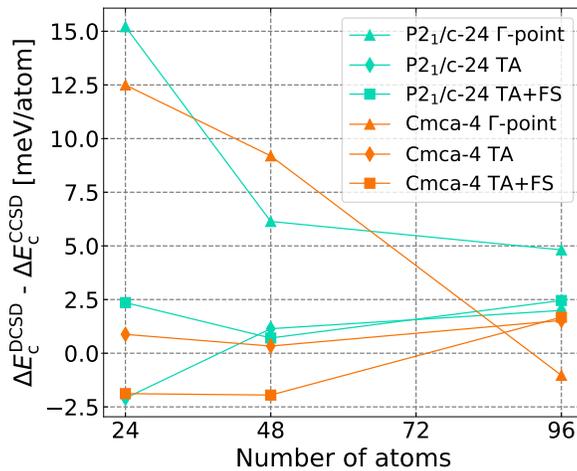


FIG. 4: (Color online) Relative correlation energy difference between DCSD and CCSD with respect to system size employing Γ -point sampling of the first Brillouin zone (Γ -point), twist-averaging (TA) and twist-averaging plus finite size correction (TA+FS), respectively, for phase P2₁/c and Cmca-4. Phase C2/c-24 is used as the reference. 200, 400 and 400 bands are used in total in 24-atom, 48-atom and 96-atom cells, respectively.

CCSD(T) and initiator FCIQMC in the 24-atom supercells, we will therefore use only 60 bands.

The initiator FCIQMC correlation energies, E_c^{FCIQMC} , are converged to within 1 meV/atom with respect to the number of walkers on the HF determinant, N_0 , using the recently developed auto-adaptive-shift method in the NECI code [19], which requires much fewer walkers to remove the initiator errors and converges to the exact ground state energy. The convergence of the correlation energies retrieved as a function of the number of walkers on the HF determinant for different phases are shown in Fig. 3.

Table III lists the correlation energies and their differences of phase C2/c-24, Cmca-4 and P2₁/c-24 using CCSD, DCSD, CCSD(T) and FCIQMC, respectively. We find differences in the correlation energies on the order of 10-20 meV/atom. We note that DCSD is closer to *i*-FCIQMC than CCSD and deviates by less than 10 meV/atom for the relative correlation energies. However, we stress that these error estimates are obtained sampling the Γ -point only. In the TA technique, we also have to account for correlation energy contributions from k -meshes centered at different k -points. In the following section we will assess the difference between CCSD and DCSD in the thermodynamic limit.

DCSD vs. CCSD in the thermodynamic limit

To estimate the difference between ΔE_c^{CCSD} and ΔE_c^{DCSD} in the thermodynamic limit, twist-averaging and finite size corrections for increasing supercell sizes are used. The convergence of the relative correlation energy differences between DCSD and CCSD retrieved as a function of the system size are shown in Fig. 4. We find that DCSD and CCSD agrees very well in the thermodynamic limit, which indicates that all considered phases at 400 GPa can be described accurately by CCSD theory. In the light of this and based on the deviation of DCSD from *i*-FCIQMC for the relative energies in the 24-atom cells, we estimate the post-CCSD corrections to be roughly 10 meV/atom.

-
- [1] E. Solomonik, D. Matthews, J. R. Hammond, J. F. Stanton, and J. Demmel, *Journal of Parallel and Distributed Computing* **74**, 3176 (2014), publisher: Academic Press.
- [2] P. E. Blchl, C. J. Frst, and J. Schimpl, *Bulletin of Materials Science* **26**, 33 (2003), arXiv: cond-mat/0201015 ISBN: 0250-4707.
- [3] G. Kresse and J. Hafner, *Journal of Physics: Condensed Matter* **6**, 8245 (1994), iISBN: 0953-8984.
- [4] G. Kresse and J. Furthmller, *Phys. Rev. B* **54**, 11169 (1996), iISBN: 1098-0121.
- [5] C. J. Pickard and R. J. Needs, *Nature Physics* **3**, 473 (2007), arXiv: 1609.07486 Publisher: Nature Publishing Group ISBN: 1745-2473 1745-2481.
- [6] N. D. Drummond, B. Monserrat, J. H. Lloyd-Williams, P. L. Ros, C. J. Pickard, and R. J. Needs, *Nature Communications* **6**, 7794 (2015).
- [7] A. Grneis, G. H. Booth, M. Marsman, J. Spencer, A. Alavi, and G. Kresse, *Journal of Chemical Theory and Computation* **7**, 2780 (2011).
- [8] C. Lin, F. H. Zong, and D. M. Ceperley, *Physical review. E, Statistical, nonlinear, and soft matter physics* **64**, 016702 (2001), arXiv: cond-mat/0101339.
- [9] K. Liao and A. Grneis, *Journal of Chemical Physics* **145**, 0 (2016).
- [10] T. Gruber, K. Liao, T. Tsatsoulis, F. Hummel, and A. Grneis, *Physical Review X* **8**, 21043 (2018), publisher: American Physical Society.
- [11] D. Kats and F. R. Manby, *The Journal of Chemical Physics* **139** (2013), 10.1063/1.4813481.
- [12] D. Kats, *The Journal of Chemical Physics* **141**, 061101 (2014).
- [13] K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, *Chemical Physics Letters* **157**, 479 (1989).
- [14] R. J. Bartlett and M. Musia, *Reviews of Modern Physics* **79**, 291 (2007).
- [15] J. J. Shepherd and A. Grneis, *Physical Review Letters* **110** (2013), 10.1103/PhysRevLett.110.226401, arXiv: 1310.6059.
- [16] G. H. Booth, A. J. Thom, and A. Alavi, *Journal of Chemical Physics* **131**, 054106 (2009), publisher: American Institute of Physics ISBN: 0021-9606.
- [17] G. H. Booth, A. Grneis, G. Kresse, and A. Alavi, *Nature* **493**, 365 (2013), publisher: Nature Publishing Group ISBN: 0028-0836.
- [18] D. Cleland, G. H. Booth, and A. Alavi, *Journal of Chemical Physics* **132** (2010), 10.1063/1.3302277, iISBN: 0021-9606.
- [19] K. Ghanem, A. Y. Lozovoi, and A. Alavi, To be published (2019).