Energy Gap Closure of Crystalline Molecular Hydrogen with Pressure

Vitaly Gorelov,[1] Markus Holzmann,[2,3] David M. Ceperley,[4] and Carlo Pierleoni[1,5,*]
[1]Université Paris-Saclay, UVSQ, CNRS, CEA, Maison de la Simulation, 91191, Gif-sur-Yvette, France.
[2]Univ. Grenoble Alpes, CNRS, LPMMC, 38000 Grenoble, France
[3]Institut Laue-Langevin, BP 156, F-38042 Grenoble Cedex 9, France
[4]Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA
[5]Department of Physical and Chemical Sciences, University of L'Aquila, Via Vetoio 10, I-67010 L'Aquila, Italy

(Received 15 November 2019; accepted 13 February 2020; published 16 March 2020)

We study the gap closure with pressure of crystalline molecular hydrogen. The gaps are obtained from grand-canonical quantum Monte Carlo methods properly extended to quantum and thermal crystals, simulated by coupled electron ion Monte Carlo methods. Nuclear zero point effects cause a large reduction in the gap (∼2 eV). Depending on the structure, the fundamental indirect gap closes between 380 and 530 GPa for ideal crystals and 330–380 GPa for quantum crystals. Beyond this pressure the system enters into a bad metal phase where the density of states at the Fermi level increases with pressure up to ∼450–500 GPa when the direct gap closes. Our work partially supports the interpretation of recent experiments in high pressure hydrogen.

DOI: 10.1103/PhysRevLett.124.116401

The metallization of crystalline hydrogen under pressure has attracted considerable attention over the last century. Predicted to be stable in an atomic bcc lattice around 25 GPa, the mechanism for molecular dissociation was first discussed by Wigner and Huntington [1]. The search for its metallization has driven high pressure research until the recent [2], still debated [3–6], observation of reflective samples at 495 GPa in a diamond anvil cell (DAC) apparatus. Even though it is the simplest element and H_2 the simplest homonuclear molecule in nature, the study of hydrogen under extreme conditions has uncovered rich and unexpected physics [7–9].

The mechanism by which the insulating crystal transforms into a conducting crystal is still unclear. Experiments have difficulty in determining the crystalline structure and its evolution with pressure because of the low cross section to x rays [10–12] and the small volume of the samples for neutron scattering. Structural information is obtained indirectly through vibrational spectroscopy while the electronic structure is probed by optical measurements [13]. Direct measurements of static conductivity in the DAC remain inconclusive [14–19]. A complex phase diagram comprising up to at least four different molecular phases (from I to IV) with different vibrational spectra has been traced experimentally [8]. Recent experiments [2,19–22] searched for metallization at low temperature ($\lesssim 100$ K) while raising pressure in phase III. Considerable attention has also been paid to the higher temperature phase IV since its discovery [14,23–27]. The emerging picture is that the transparent insulating molecular phase III transforms into a strongly absorbing (in the visible) molecular phase at ∼350–360 GPa with different IR frequencies, first named phase V [18] and later H_2-PRE or phase VI [13,22], with semiconducting characteristics [28]. Hydrogen finally reaches a metallic phase with the observation of reflective samples at ∼495 GPa [2], although disagreement concerning the pressure scale still remains [4,13,29]. New synchrotron infrared spectroscopy measurements [21] report a reversible collapse of the IR transmission spectrum at 427 GPa, interpreted as a first order transition to the metallic state.

In this Letter we investigate the closure of the electronic gap of candidate structures for phase III ($Cmca$-12 and $C2/c$-24) and phase IV (Pc48)[30,31] within a quantum Monte Carlo (QMC) framework [32]. For ideal structures, the fundamental gap decreases with pressure from ∼3–3.5 eV at ∼250 GPa to a vanishing value ∼380 GPa in the $Cmca$-12 structure and ∼530 GPa in the $C2/c$-24 structure. Using coupled electron-ion Monte Carlo (CEIMC) calculations, we then include zero point motion (ZPM) and finite temperature effects of the nuclei within a first principles, nonperturbative path integral approach. Extending the grand canonical method [32] to quantum crystal at finite temperature, we observe a strong gap reduction of ∼2 eV due to nuclear quantum effects (NQE) while temperature effects below 300 K are minor. At 200 K the fundamental indirect gap closes ∼330 GPa for $Cmca$-12 and ∼380 GPa for $C2/c$-24. Raising the temperature of $C2/c$-24 to 290 K reduces the closure pressure to 340 GPa while decreasing it to 100 K does not give any noticeable effect. For both structures the direct gap, as obtained by unfolding of the supercell bands [33], remains open up to ∼470–500 GPa. Values for the $C2/c$-24 structure are in agreement with recent experimental data [21], although we

cannot discuss the experimentally observed sudden closure at 427 GPa. Our new method for calculating energy gaps allows us to benchmark DFT functionals, not only for thermodynamics and structural properties, but also for excitation energies, important for predicting optical properties.

Method.—The primary information for theoretical investigations of solids are the crystalline structures. Candidate structures for high pressure phases have been obtained by *ab initio* random structural search methods [30,31,42,43]. For phase III we consider $C2/c$-24 and $Cmca$-12, which are among the lowest free energy structures in ground state QMC calculations assuming harmonic phonon corrections (with DFT-PBE) [44–46]. For phase IV we consider only $Pc48$, since the recently proposed $Pca21$ structure [43] is found to be rather similar to $Pc48$ after geometry relaxation. We first consider ideal crystal structures (protons fixed at lattice sites) relaxed at constant pressure with the DFT-vdW-DF [47] functional. Quantum crystals, with protons represented by path integrals at finite temperature, are addressed with CEIMC at constant volume [48]. All systems considered have 96 protons in nearly cubic supercells. Optimized Slater-Jastrow-Backflow trial wave functions have been used for the CEIMC calculations [49]; details of the CEIMC simulations are reported in Ref. [50]. Averages over ionic positions for gaps are obtained using 40 statistically independent configurations from the CEIMC trajectories.

For a given fixed nuclear configuration, the fundamental energy gap is obtained by considering systems with a variable number of electrons $n \in [-6,6]$, where $n = N_e - N_p$. For each system we perform reptation quantum Monte Carlo (RQMC) calculations with imaginary-time projection $t = 2.00$ and time step $\tau = 0.01$ Ha^{-1} for up to $6 \times 6 \times 6$ Monkhorst-Pack grid of twists. We check that those values are adequate for converging the band gaps within our resolution. The fundamental gap is obtained from grand-canonical twist-averaged boundary conditions (GCTABC) RQMC and corrected for finite size effects in leading and next-to-leading order [32].

Extending calculations of the fundamental gaps to quantum crystals, the trace over nuclear degrees of freedom must be taken with care. In the semiclassical approximation [33], the fundamental gap is the smallest electronic excitation energy that occurs from quantum or thermal fluctuations of the lattice. Strictly speaking this gap is always closed, since the probability of a proton configuration with a metallic character is never exactly zero. For dense molecular hydrogen phonon energies are ~0.1–0.5 eV [31]. ZPM dominates for $T \leq 1000$ K, so the semiclassical approach is not appropriate. Electronic energies should be averaged over the nuclear configurations according to their thermal distribution. The gap will be given by the minimum of the average excitation energies, always larger than the semiclassical gap. Figure 4 illustrates typical results for the integrated density of states as a function of (electronic) chemical potential. The gap of the quantum crystal can be directly read off from the width of the incompressible region. More details are given in Ref. [33].

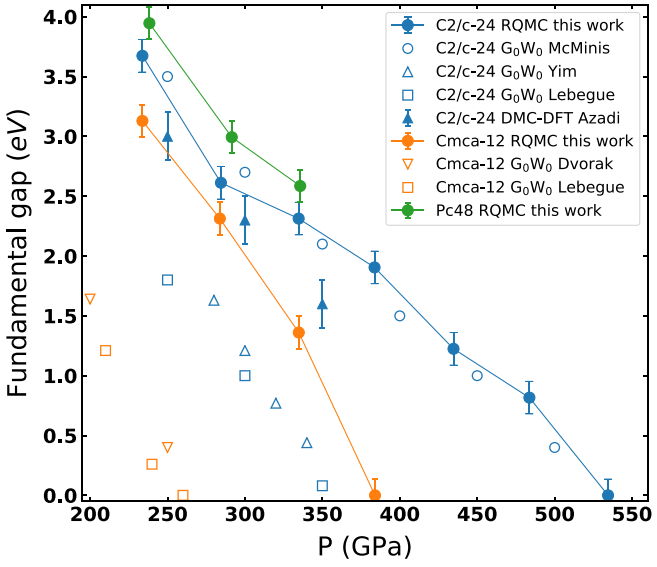

FIG. 1. Fundamental energy gap for ideal crystals. This work (closed circles): $C2/c$-24 (blue), $Cmca$-12 (orange), and $Pc48$ (green), GW results for $C2/c$-24 (open blue circles [44]). These structures were optimized with the vdW-DF functional. QMC for $C2/c$-24 optimized with the BLYP from Ref. [51] (closed blues triangles). GW results from Refs. [53–55] for $C2/c$-24 (blue) and $Cmca$-12 (orange) optimized with the PBE functional. Note that pressures from RQMC are 10–15 GPa lower than the nominal optimization pressure.

Results.—Figure 1 shows estimates of the fundamental gap for ideal crystals versus pressure. The gap decreases with pressure in a similar fashion for all structures: $Cmca$-12 has the smallest gap, followed by $C2/c$-24 and by $Pc48$. We find reasonable agreement with the QMC estimates of Refs. [51,52]. References [53–55] report smaller values of the gap based on GW. We believe this disagreement is primarily due to the lattice geometry that has been optimized at constant pressure with PBE in Refs. [53–55] and with vdW-DF in the present work. It has been previously observed that PBE optimized geometries have longer H_2 bonds and smaller gap values at the DFT level [56,57]. This propagates into G_0W_0. Indeed, GW results from structures optimized with vdw-DF [44] are in excellent agreement with our predictions.

Values of the fundamental gap from GCTABC for quantum crystals at various temperatures and pressures are shown in Fig. 2: they are smaller by ~2 eV with respect to the ideal crystal. ZPM is almost entirely responsible for this reduction. Note that the gap hardly changes from 300 to 200 K within our estimated errors. Similar to ideal crystals, the $Cmca$-12 gap is smaller than the $C2/c$-24 gap at $T = 200$ K, the former closing ~340 GPa, while the latter at higher pressures ~380 GPa. As for the $Pc48$ structure at

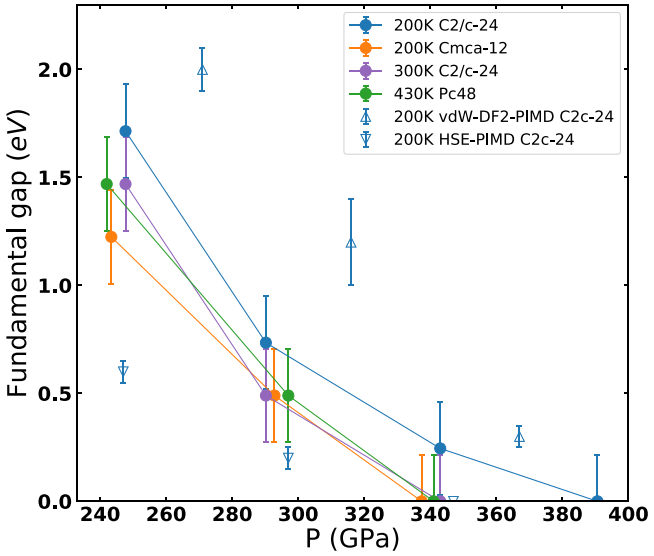

FIG. 2. The fundamental gap of quantum crystals at finite temperature. Closed circles indicate results from this work, for the three structures at various temperature as detailed in the legend. PIMD-DFT results at $T = 200$ K are obtained with two different XC approximations, namely, the HSE (downward open triangles) and the vdW-DF2 (upward open triangles) and the semiclassical averaging are reported for comparison [56].

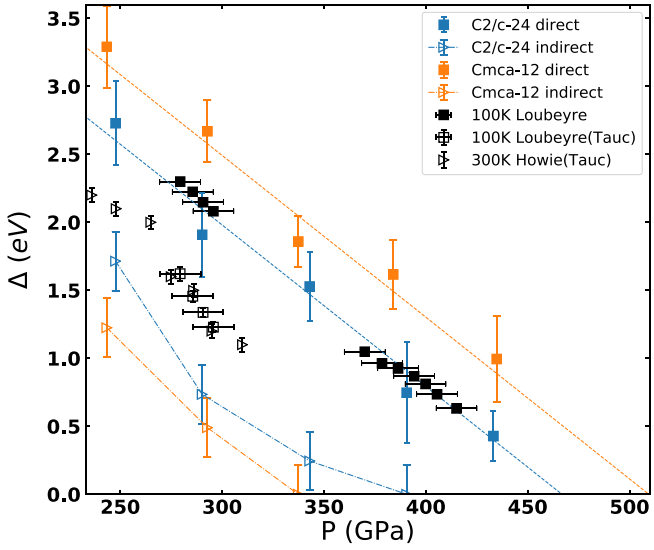

FIG. 3. Direct (closed symbols) and indirect (open symbols) gaps of quantum crystals. GCTABC-RQMC at $T = 200$ K: $C2/c$-24 indirect (blue triangles), direct (blues squares); *Cmca*-12 indirect (orange triangles), direct (closed squares). Experiments: indirect gap from the Tauc analysis at $T = 100$ K (phase III), (black squares) [64] and at 300 (phase IV), (black triangles) [17,24]; direct gap at 100 K (black squares) [21,64].

$T = 430$ K (phase IV) the gap is slightly below values for $C2/c$-24 at 200 K. Our results show that the electronic gap is fairly independent of the specific crystalline structure of the molecular quantum crystals. We also report gap values for $C2/c$-24 at $T = 200$ K from path integral molecular dynamics (PIMD) [56] with two different DFT functionals, namely, the HSE [58] and vdW-DF2 [59]. As vdW-DF2 underestimates the molecular bond lengths of the ideal crystalline structure [57], its PIMD configurations are expected to bias the electronic gap towards larger values. Our results do not agree with predictions of Ref. [60] (not shown) yielding a metallic state for $C2/c$-24 at 300 GPa and 300 K, and predict substantially larger gap reduction for $C2/c$-24 quantum crystals than Ref. [61]. However, those works are based on less controlled assumptions such as using "scissor corrected" BLYP [62,63] band structure and an *ad hoc* procedure for including nuclear motion.

For all structures considered the observed fundamental gap is indirect. An estimate of the direct gap can be obtained by unfolding the band structure of the supercell [33]. Figure 3 shows the direct gap for both $C2/c$-24 and *Cmca*-12 structures. While for the indirect gap *Cmca*-12 is always lower than $C2/c$-24, the direct gap is systematically larger. The difference between the direct and indirect gap is of $\sim$1 eV for $C2/c$-24, and of $\sim$2 eV for *Cmca*-12. Closure of the direct gaps, obtained by linear extrapolation, occurs $\sim$450 GPa in $C2/c$-24 and $\sim$500 GPa in *Cmca*-12. Hence for both structures we observe an intermediate pressure region where the fundamental indirect gap is closed but the direct vertical gap remains open and decreases linearly with pressure. In this region, we expect the density of states around the Fermi level to increase progressively with pressure, as qualitatively reported in Ref. [50]. This indicates the formation of a bad metal with properties similar to a semimetal upon closure of the indirect gap, a scenario strongly supporting the recently proposed experimental picture [28] (see also Refs. [13,22]). The nonvanishing direct gap naturally explains the reported observation of absorbing (black) hydrogen around 320–360 GPa (depending on the experimental pressure scale) [64].

Figure 3 also shows the experimental estimates of both indirect and direct gaps from optical absorption. Measuring indirect gaps is difficult in hydrogen since samples are very thin and the optical signal from phonon-assisted absorption is too low to be detected [16,19]. The indirect gap value has been extracted from a Tauc analysis of the absorption profiles at 300 K (phase IV) [17,24] and 100 K (phase III) [25,64] assuming the low energy absorption spectrum can be reliably extrapolated to zero energy. [65]. Conversely, the direct gap at 100 K (phase III) has been associated with the absorption edge at lower pressure [64] or with full absorption at higher pressure [21] and corresponds roughly to the energy where the absorption coefficient equals 30 000 cm^{-1}. The direct gap of $C2/c$-24 structure is in agreement with the experimental data up to 425 GPa, where experiments report a collapse of the gap value ascribed to the metallization transition [21]. Our results do not allow us to predict this transition, but rule out $C2/c$-24 and *Cmca*-12 for this new metallic phase [66]. For the indirect

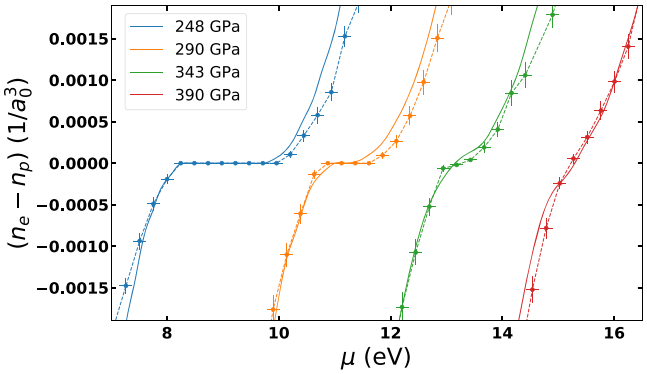

FIG. 4. Integrated density of states for $C2/c$-24 quantum crystals at 200 K from GCTABC-RQMC (points) and the HSE (smooth lines) at various pressures.

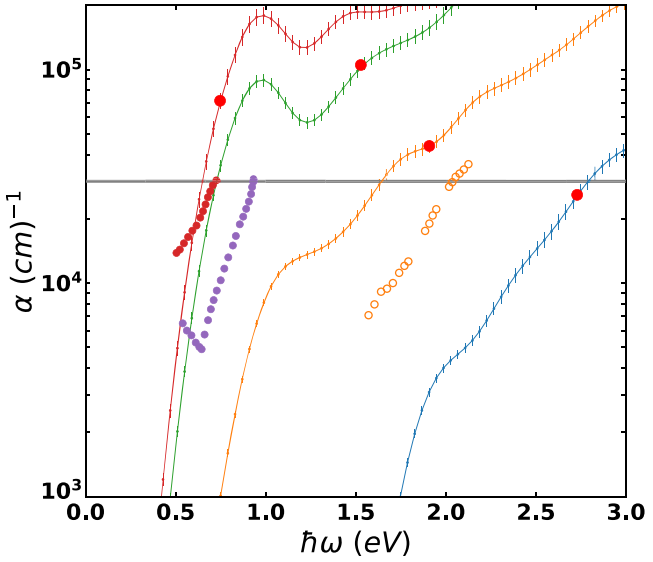

FIG. 5. Absorption spectra from the HSE band structure for $C2/c$-24 quantum crystals (solid lines) and comparison with the available experimental profiles (opened and filled circles). The spectra from the HSE has been shifted in energy by an amount equal to the difference between QMC and the HSE direct gap. The reported pressure are as in Fig. 4 (see the colors). The red dots indicate the location in energy of the direct gap of Fig. 3. Experimental pressures are 296 GPa, open orange circles [64] (corrected by 20 GPa [21]); 386 GPa, magenta filled circles and 406 GPa, red filled circles [21].

gap we predict ~0.3–0.5 eV smaller values than in experiments. However, the Tauc analysis of Refs. [17,24,64] does not consider the energy of the emitted or absorbed phonons, which should be comparable to the observed discrepancy. However, excitonic effects and exciton-phonon coupling, neglected within the present approach, need to be addressed for this level of precision. In agreement with our findings, the experimental indirect gap depends little on both temperature and structure [67].

Next, we explore optical properties computed using the Kubo-Greenwood (KG) framework with Kohn-Sham (KS) orbitals. To reduce the bias of the underlying DFT functional, we have benchmarked several XC approximations to reproduce the behavior of the QMC density of states close to the gap. In Fig. 4 for $C2/c$-24 at 200 K, we compare the electronic excess density $n_e - n_p$ as a function of electronic chemical potential μ from QMC and from DFT-HSE [68]. The observed plateau at $n_e - n_p = 0$ is the signature of the indirect gap. Deviations from the plateau on both sides characterize the density of states of the valence and conduction band close to the band edges. As shown in Fig. 4 the HSE approximation provides slightly smaller values of the fundamental gap and reproduces reasonably well the integrated density of states from GCTABC around the Fermi energy (more details are in Ref. [33]). We therefore employed the HSE to compute optical properties exploiting the KGEC code [69] in the QuantumEspresso suite [70]. For thermal and quantum crystals considered here, the William-Lax (WL) semiclassical (SC) approximation [71–75] is not appropriate as already discussed. Instead of a joint density of states based on excitation energies for each nuclear configuration entering the WL expression, we have used the corresponding one based on electronic energies averaged over ionic ZPM, more appropriate for low temperatures [33]. In Fig. 5 we compare the absorption profiles for $C2/c$-24 at $T = 200$ K and different pressures [76] to experimental profiles from Refs. [21,64] at $T = 100$ K. We observe a higher absorption than in experiments at comparable pressure, which cannot be explained by the temperature difference. We marked each predicted profile with a red dot at the energy corresponding to the observed direct gap and we report a thick horizontal line at 30 000 cm^{-1} the value of the absorption used in the experiments to extract the value of the direct gap. Our results at lower pressures are in reasonable agreement with this criterion. However at the higher pressure, absorption at the energy gap is about 2–3 times higher than 30 000 cm^{-1}.

Conclusions.—We have studied the closure of the fundamental gap with pressure of candidate structures of molecular hydrogen in phase III ($C2/c$-24 and $Cmca$-12) and phase IV ($Pc48$) entirely based on quantum Monte Carlo simulations. For ideal structures our gap values are in excellent agreement with GW prediction [44]. Considering quantum nuclei at finite temperature, we observe a strong reduction of the energy gap with respect to the ideal structures at the same pressure (~2 eV) caused by ZPM. At 200 K the fundamental (indirect) gap closes at ~370–380 GPa for $C2/c$-24 and at ~340 GPa for $Cmca$-12. We observe a reasonable agreement with experimental determinations of indirect gaps from optical absorption. The direct gap remains open until ~450 GPa for $C2/c$-24 and ~500 GPa for $Cmca$-12. In this range of pressure the system is a bad metal (or semimetal) suggesting a scenario that qualitatively supports recent experiments [19,20,22,28]. In Refs. [19,28] no discontinuities in the Raman vibrational spectrum are reported when entering

the semimetallic phase, while in Refs. [20,22] new IR vibron peaks are reported in this pressure range and ascribed to a structural phase transition. They have been tentatively assigned to a transition from the $C2/c$-24 to the $Cmca$-12 structure [22]. Our present results, supplemented by free energy calculations [77], do not disprove this hypothesis. Our predictions for direct gap are in good agreement with the experimental data at $T = 100$ K [21,64]. However our absorption profiles do not agree as well with the experimental behavior. This disagreement remains an open question.

We thank Paul Loubeyre, Mikhail Eremets, Mario Santoro, and Michele Nardone for useful suggestions. D. M. C. was supported by DOE Grant No. NA DE-NA0001789 and by the Fondation NanoSciences (Grenoble). V. G. and C. P. were supported by the Agence Nationale de la Recherche (ANR) France, under the program "Accueil de Chercheurs de Haut Niveau 2015" project: HyLightExtreme. Computer time was provided by the PRACE Project No. 2016143296, ISCRAB (IsB17_MMCRHY) computer allocation at CINECA Italy, the high-performance computer resources from Grand Equipement National de Calcul Intensif (GENCI) Allocation 2018-A0030910282, by an allocation of the Blue Waters sustained petascale computing project, supported by the National Science Foundation (Grant No. OCI 07- 25070) and the State of Illinois, and by the Froggy platform of CIMENT, Grenoble (Rhône-Alpes CPER07-13 CIRA and ANR-10-EQPX-29-01).

*carlo.pierleoni@aquila.infn.it

[1] E. Wigner and H. B. Huntington, J. Chem. Phys. **3**, 764 (1935).
[2] R. P. Dias and I. F. Silvera, Science **355**, 715 (2017).
[3] A. F. Goncharov and V. V. Struzhkin, Science **357**, eaam9736 (2017).
[4] P. Loubeyre, F. Occelli, and P. Dumas, arXiv:1702.07192.
[5] X.-D. Liu, P. Dalladay-Simpson, R. T. Howie, B. Li, and E. Gregoryanz, Science **357**, eaan2286 (2017).
[6] I. Silvera and R. Dias, Science **357**, eaan1215 (2017).
[7] H. K. Mao and R. Hemley, Rev. Mod. Phys. **66**, 671 (1994).
[8] J. M. McMahon, M. A. Morales, C. Pierleoni, and D. M. Ceperley, Rev. Mod. Phys. **84**, 1607 (2012).
[9] W. J. Nellis, Rep. Prog. Phys. **69**, 1479 (2006).
[10] P. Loubeyre, R. LeToullec, D. Hausermann, M. Hanfland, R. J. Hemley, H. K. Mao, and L. W. Finger, Nature (London) **383**, 702 (1996).
[11] C. Ji, B. Li, W. Liu, J. S. Smith, A. Majumdar, W. Luo, R. Ahuja, J. Shu, J. Wang, S. Sinogeikin, Y. Meng, V. B. Prakapenka, E. Greenberg, R. Xu, X. Huang, W. Yang, G. Shen, W. L. Mao, and H.-K. Mao, Nature (London) **573**, 558 (2019).
[12] L. Dubrovinsky, N. Dubrovinskaia, and M. I. Katsnelson, arXiv:1910.10772.
[13] I. Silvera and R. P. Dias, J. Phys. Condens. Matter **30**, 254003 (2018).
[14] M. I. Eremets and I. A. Troyan, Nat. Mater. **10**, 927 (2011).
[15] W. J. Nellis, A. L. Ruoff, and I. F. Silvera, arXiv:1201.0407.
[16] A. F. Goncharov and V. V. Struzhkin, https://arxiv.org/pdf/1207.6643.pdf.
[17] A. F. Goncharov, J. S. Tse, H. Wang, J. Yang, V. V. Struzhkin, R. T. Howie, and E. Gregoryanz, Phys. Rev. B **87**, 024101 (2013).
[18] M. I. Eremets, I. A. Troyan, and A. P. Drozdov, arXiv:1601.04479.
[19] M. Eremets, A. P. Drozdov, P. Kong, and H. Wang, arXiv:1708.05217.
[20] R. P. Dias, O. Noked, and I. F. Silvera, arXiv:1603.02162.
[21] P. Loubeyre, F. Occelli, and P. Dumas, Nature (London) **577**, 631 (2020).
[22] R. P. Dias, O. Noked, and I. F. Silvera, Phys. Rev. B **100**, 184112 (2019).
[23] R. T. Howie, T. Scheler, C. L. Guillaume, and E. Gregoryanz, Phys. Rev. B **86**, 214104 (2012).
[24] R. T. Howie, C. L. Guillaume, T. Scheler, A. F. Goncharov, and E. Gregoryanz, Phys. Rev. Lett. **108**, 125501 (2012).
[25] C. S. Zha, Z. Liu, and R. J. Hemley, Phys. Rev. Lett. **108**, 146402 (2012).
[26] P. Loubeyre, F. Occelli, and P. Dumas, Phys. Rev. B **87**, 134101 (2013).
[27] R. T. Howie, P. Dalladay-Simpson, and E. Gregoryanz, Nat. Mater. **14**, 495 (2015).
[28] M. I. Eremets, A. P. Drozdov, P. P. Kong, and H. Wang, Nat. Phys., https://doi.org/10.1038/s41567-019-0646-x (2019).
[29] M. I. Eremets and A. P. Drozdov, arXiv:1601.04479.
[30] C. J. Pickard and R. J. Needs, Nat. Phys. **3**, 473 (2007).
[31] C. J. Pickard, M. Martinez-Canales, and R. J. Needs, Phys. Rev. B **85**, 214114 (2012).
[32] Y. Yang, V. Gorelov, C. Pierleoni, D. M. Ceperley, and M. Holzmann, Phys. Rev. B **101**, 085115 (2020).
[33] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.124.116401 for more information about: computing band gap for thermal crystals, Tauc analysis of experimental indirect gap, benchmark of some XC functionals and Kubo-Greenwood optical calculations for thermal crystals, which includes Refs. [35–41].
[34] R. M. Martin, L. Reining, and D. M. Ceperley, *Interacting Electrons: Theory and Computational Approaches* (Cambridge University Press, Cambridge, England, 2016).
[35] J. P. Perdew, Int. J. Quantum Chem. **28**, 497 (1985).
[36] C. Lin, F. H. Zong, and D. M. Ceperley, Phys. Rev. E **64**, 016702 (2001).
[37] M. Holzmann, B. Bernu, V. Olevano, R. M. Martin, and D. M. Ceperley, Phys. Rev. B **79**, 041308(R) (2009).
[38] G. Grosso and G. Parravicini, *Solid State Physics*, 2nd ed. (Academic Press, Amsterdam, 2014).
[39] F. Wooten, *Optical Properties of Solids* (Academic Press, New York, 1972).
[40] M. A. Morales, R. Clay, C. Pierleoni, and D. M. Ceperley, Entropy **16**, 287 (2014).
[41] J. Tauc, R. Grigorovici, and A. Vancu, Physica Status Solidi (B) **15**, 627 (1966).
[42] C. Pickard, M. Martinez-Canales, and R. Needs, Phys. Rev. B **85**, 214114 (2012).
[43] B. Monserrat, N. D. Drummond, P. Dalladay-Simpson, R. T. Howie, P. L. Rios, E. Gregoryanz, C. J. Pickard, and R. J. Needs, Phys. Rev. Lett. **120**, 255701 (2018).

[44] J. McMinis, R. C. Clay, D. Lee, and M. A. Morales, Phys. Rev. Lett. **114**, 105305 (2015).
[45] S. Azadi, B. Monserrat, W. M. C. Foulkes, and R. J. Needs, Phys. Rev. Lett. **112**, 165501 (2014).
[46] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **78**, 1396 (1997).
[47] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist, Phys. Rev. Lett. **95**, 109902 (2005).
[48] We have checked that the stress tensor in the constant volume CEIMC run remains diagonal with the same diagonal elements within our statistical noise.
[49] C. Pierleoni, M. A. Morales, G. Rillo, M. Holzmann, and D. M. Ceperley, Proc. Natl. Acad. Sci. U.S.A. **113**, 4953 (2016).
[50] G. Rillo, M. A. Morales, D. M. Ceperley, and C. Pierleoni, J. Chem. Phys. **148**, 102314 (2018).
[51] S. Azadi, N. D. Drummond, and W. M. C. Foulkes, Phys. Rev. B **95**, 035142 (2017).
[52] The observed small difference, in particular at the higher pressure, is probably due to the different XC approximation used for geometry optimization, vdW-DF in our case, BLYP in Ref. [51] and different size extrapolation.
[53] S. Lebègue, C. Araujo, D. Kim, M. Ramzan, H. Mao, and R. Ahuja, Proc. Natl. Acad. Sci. U.S.A. **109**, 9766 (2012).
[54] Y. L. R. J. H. Wai-Leung Yim, H. Shi, and J. S. Tse, in *Correlations in Condensed Matter under Extreme Conditions*, edited by G. A. Editors and A. L. Magna (Springer International Publishing, AG, 2017), Chap. 9, pp. 107–126.
[55] M. Dvorak, X.-J. Chen, and Z. Wu, Phys. Rev. B **90**, 035103 (2014).
[56] M. A. Morales, J. M. McMahon, C. Pierleoni, and D. M. Ceperley, Phys. Rev. B **87**, 184107 (2013).
[57] R. C. Clay, J. Mcminis, J. M. McMahon, C. Pierleoni, D. M. Ceperley, and M. A. Morales, Phys. Rev. B **89**, 184106 (2014).
[58] J. Heyd, J. E. Peralta, G. E. Scuseria, and R. L. Martin, J. Chem. Phys. **123**, 174101 (2005).
[59] K. Lee, É. Murray, L. Kong, B. Lundqvist, and D. Langreth, Phys. Rev. B **82**, 081101 (2010).
[60] S. Azadi, R. Singh, and T. D. Kühne, J. Comput. Chem. **39**, 262 (2018).
[61] X.-Z. Li, B. Walker, M. I. J. Probert, C. J. Pickard, R. J. Needs, and A. Michaelides, J. Phys. Condens. Matter **25**, 085402 (2013).
[62] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
[63] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1998).
[64] P. Loubeyre, F. Occelli, and R. LeToullec, Nature (London) **416**, 613 (2002).
[65] We have re-analyzed the spectra of Ref. [64] to extract the value of the indirect gap from a Tauc plot [41], as was performed in Ref. [17] for the data from Ref. [24]. Details are given in Ref. [33].
[66] Our estimates of the direct gap could be biased by ∼0.3 eV due to the discreteness of our twist grid. Correcting for this bias will place the experimental data in between the $C2/c$-24 and $Cmca$-12 predictions.
[67] Note that the pressure values of Ref. [64] have been recently corrected [21].
[68] This quantity is closely related to the integrated density of states.
[69] L. Calderín, V. V. Karasiev, and S. B. Trickey, Comput. Phys. Commun. **221**, 118 (2017).
[70] P. Giannozzi *et al.*, J. Phys. Condens. Matter **29**, 465901 (2017).
[71] F. Williams, Phys. Rev. **82**, 281 (1951).
[72] M. Lax, J. Chem. Phys. **20**, 1752 (1952).
[73] C. E. Patrick and F. Giustino, J. Phys. Condens. Matter **26**, 365503 (2014).
[74] M. Zacharias, C. E. Patrick, and F. Giustino, Phys. Rev. Lett. **115**, 177401 (2015).
[75] M. Zacharias and F. Giustino, Phys. Rev. B **94**, 075125 (2016).
[76] To partially correct for the HSE inaccuracy, we shifted the energy scale by the difference between the QMC and HSE gap.
[77] C. Pierleoni (to be published).

Supplemental Material for: Energy gap closure of crystalline molecular hydrogen with pressure

Vitaly Gorelov[1], Markus Holzmann[2,3], David M. Ceperley[4], Carlo Pierleoni[1,5,*]
[1] *Maison de la Simulation, CEA-CNRS-UPS-UVSQ, Université Paris-Saclay, 91191 Gif-sur-Yvette, France*
[2] *Univ. Grenoble Alpes, CNRS, LPMMC, 3800 Grenoble, France*
[3] *Institut Laue-Langevin, BP 156, F-38042 Grenoble Cedex 9, France*
[4] *Department of Physics, University of Illinois Urbana-Champaign, USA*
[5] *Department of Physical and Chemical Sciences, University of L'Aquila, Via Vetoio 10, I-67010 L'Aquila, Italy*

A. Theory

Consider a neutral system with N_p protons and $N_e = N_p$ electrons in a simulation cell of volume V with periodic boundary conditions. The symmetry of the Hamiltonian (i.e. the protons) under translation by a supercell allows for the introduction of a "twist", i.e. a phase that the wave function acquires when any electron wraps around the supercell. The twist $\vec{\theta}$ (a 3-dimensional vector) is a quantum number that specifies the particular eigenfunction $\Phi_{\vec{\theta}}$ and eigenvalue $E(\vec{\theta})$ of the Hamiltonian. It can be reduced to values inside the first Brillouin zone of the supercell lattice.

Let us consider two more systems, both with identical nuclear configurations. In the first system we add $2n$ electrons (preserving the total spin equal to zero) and a rigid background of positive charge to maintain charge neutrality, in the second system we remove $2n$ electrons (again preserving the total spin) and a negative background. We define the two energy differences.

$$\Delta^{\pm}(\vec{\theta}) = \pm \left[E_{\pm 2n}(\vec{\theta}) - E_0(\vec{\theta}) \right] /2n. \tag{1}$$

The *fundamental energy gap* is then[1]

$$\Delta(\vec{\theta},\vec{\theta}') = \min_{\vec{\phi}} \left[\Delta^{+}(\vec{\phi}) \right] - \max_{\vec{\phi}'} \left[\Delta^{-}(\vec{\phi}') \right] \tag{2}$$

where $\vec{\theta}$ and $\vec{\theta}'$ correspond to the twists at which the *min* and the *max* are realized. If $\vec{\theta} \neq \vec{\theta}'$ the gap is indirect, while when $\vec{\theta} = \vec{\theta}'$ the gap can be direct. The *vertical* or *direct* gap can be accessed by

$$\widetilde{\Delta}(\vec{\theta}) = \min_{\vec{\phi}} \left[\Delta^{+}(\vec{\phi}) - \Delta^{-}(\vec{\phi}) \right] \tag{3}$$

if band folding in the supercell can be excluded, e.g. if Bloch orbitals of the primitive unit cell are used to build the wave functions (see section F below).

So far we have worked with systems at fixed number of electrons, i.e. in the canonical ensemble. Let us now switch to the grand-canonical (GC) ensemble: an open system that can exchange electrons with a reservoir [2] at a fixed chemical potential μ. There are several reasons to switch of ensemble (or to switch to fractional occupations in single electron theory): in DFT this is done to understand the deficiency of the theory to predict correct band gaps in insulators [1, 3], in QMC with twisted boundary conditions it is done to improve the convergence of results as the number of electrons goes to the thermodynamic limit [4, 5]. As shown in Ref.[6], for the ground state at given twists θ in the GC ensemble, the number of electrons assumes the value $\bar{N}_e(\mu,\theta)$ that minimize the grand potential $\Omega(\mu,\theta) = \min_{N_e}[E(N_e,\theta) - \mu N_e]$. Finite size effects are reduced by averaging over the twists:

$$\omega(\mu) = e(n_e(\mu)) - \mu n_e(\mu) \tag{4}$$

$$e(n_e(\mu)) = \frac{1}{M_\theta V} \sum_{\theta} E(\bar{N}_e,\theta) \tag{5}$$

$$n_e(\mu) = \frac{1}{M_\theta V} \sum_{\theta} \bar{N}_e(\mu,\theta) \tag{6}$$

where $\omega(\mu)$ is the free energy density, $e(\mu)$ the energy density and $n_e(\mu)$ the electron density. By eliminating μ from the last two equations we obtain the discontinuity in the energy derivative at n_e corresponding to the energy gap

$$\Delta_{gc} = \mu_+ - \mu_- = \left.\frac{de}{dn_e}\right|_{n_p^+} - \left.\frac{de}{dn_e}\right|_{n_p^-} \tag{7}$$

where the derivatives are computed at $n_e = n_p = N_p/V$. In the thermodynamic limit, this value of the energy gap coincides with the canonical ensemble gap of Eq. (2).

We can extend the formalism to compute the gap for crystals with quantum nuclei at finite temperature. Con-

sider again the system of N_p nuclei and $N_e = N_p + n$ electrons for a given twist angle θ. In the semigrand-ensemble the partition function is

$$\mathcal{Q}(\mu,\theta) = \sum_{N_e=0}^{\infty} e^{\beta\mu N_e} e^{-\beta F(N_e,\theta)} = e^{-\beta\Omega(\mu,\theta)} \quad (8)$$

where the dependence on temperature, volume and number of nuclei has been kept implicit and F and Ω are respectively the Helmholtz free energy and the grand-potential. For k_BT much smaller than the energies of electronic excitations (either at fixed N_e or at different N_e), electrons can be assumed to be in the ground state and the sum over N_e reduces to the values $\bar{N}_e(\mu,\theta)$ that minimizes the exponent

$$\Omega(\mu,\theta) = \min_{N_e} \left[F(N_e,\theta) - \mu N_e\right] = F(\bar{N}_e,\theta) - \mu\bar{N}_e. \quad (9)$$

The Helmholtz free energy density includes the average nuclear kinetic energy, the average potential energy over the ground state Born-Oppenheimer (BO) surface of the $\bar{N}_e$ electron system, and the nuclear entropy. To reduce finite size effects we can average over the twists as before to obtain

$$\tilde{\omega}(\mu) = f(n_e(\mu_e)) - \mu_e n_e(\mu) \quad (10)$$

$$f(n_e(\mu)) = \frac{1}{M_\theta V}\sum_\theta F(\bar{N},\theta) \quad (11)$$

$$n_e(\mu) = \frac{1}{M_\theta V}\sum_\theta \bar{N}(\mu,\theta) \quad (12)$$

As in the ideal crystal case, the fundamental energy gap is

$$\Delta_{gc} = \mu_+ - \mu_- = \left.\frac{df}{dn_e}\right|_{n_p^+} - \left.\frac{df}{dn_e}\right|_{n_p^-}. \quad (13)$$

In this work we analyze nuclear configurations generated during CEIMC calculations performed in the canonical ensemble at $N_e = N_p$ with twist averaged energies. Therefore we don't have access to free energies and it is tempting to replace them by total energies in eq.(13). We assume that both the average nuclear kinetic energy and the nuclear entropy are nearly independent of the specific number of electrons and can be replaced by their values at $N_e = N_p$. To justify this we write the partition function in the canonical ensemble as

$$\begin{aligned} Q(N_e,\theta) &= e^{-\beta F(N_e,\theta)} \\ &= Q(N_p,\theta)\langle e^{-\beta(E(N_e,\theta)-E(N_p,\theta))}\rangle_{N_p} \\ &= e^{-\beta F(N_p,\theta)}\langle e^{-\beta\Delta E(n,\theta)}\rangle_{N_p} \end{aligned} \quad (14)$$

where $\Delta E(n,\theta) = E(N_e,\theta) - E(N_p,\theta)$ and $n = N_e - N_p$, and $\langle\ldots\rangle_{N_p}$ indicates a nuclear average over the BO energy surface of the $N_e = N_p$ system ($n = 0$). Noticing that $\Delta E(n,\theta) << E(N_p,\theta)$ (the former is an intensive property while the latter is extensive) we can take the average into the exponent obtaining

$$Q(N_e,\theta) \simeq e^{-\beta[F(N_p,\theta)+\langle\Delta E(n,\theta)\rangle_{N_p}]} \quad (15)$$

which implies

$$F(N_e,\theta) \simeq F(N_p,\theta) + \langle\Delta E(n,\theta)\rangle_{N_p}. \quad (16)$$

Using this expression in eq.(9) we obtain

$$\begin{aligned} \Omega(\mu,\theta) &\simeq F(N_p,\theta) + \min_{N_e}\left[\langle\Delta E(n,\theta)\rangle_{N_p} - \mu N_e\right] \\ &= F(N_p,\theta) + \langle\Delta E(\bar{n},\theta)\rangle_{N_p} - \mu(N_p+\bar{n}) \end{aligned} \quad (17)$$

Following the same reasoning as in eqs. (9-13) we arrive at our final expression of the fundamental gap

$$\Delta_{gc} = \mu_+ - \mu_- \simeq \left.\frac{d\langle e(n_e(\mu))\rangle_{N_p}}{dn_e}\right|_{n_p^+} - \left.\frac{d\langle e(n_e(\mu))\rangle_{N_p}}{dn_e}\right|_{n_p^-} \quad (18)$$

As before $\langle..\rangle_{N_p}$ means that the averages are taken using nuclear states sampled on the BO energy surface with $N_e = N_p$. To ensure the convergence of the averages we consider 40 statistically independent nuclear configurations from the CEIMC trajectory.

Note that putting the average over nuclear configurations outside the derivatives in eq.(18) gives a different value for the fundamental gap. This is the usual procedure to compute electronic properties from nuclear trajectories. We call it the "semiclassical approximation" in the main text.

B. Tauc analysis of absorption profiles of ref[7]

We have reanalyzed the absorption profiles of ref.[7] using the indirect gap Tauc analysis[8], similar to the analysis performed in ref.[9] of the data of ref.[10]. In Fig. (1) we show the profiles at the four values of pressure reported in the original paper and the linear fit from which have extracted the indirect gap values reported in Fig. (3) of the main text. The pressure values have been adjusted according to the new scale reported in ref. [11]. We did not include the phononic shift, which should appear in the final expression for the onset of indirect phonon-assisted absorption [12, 13]. This shift will correct the gaps reported from the Tauc analysis, by lowering the estimated gap by typical phonon energies emitted in the indirect transition since at the temperatures of the experiments the protons are in their ground state with few thermally excited phonons.

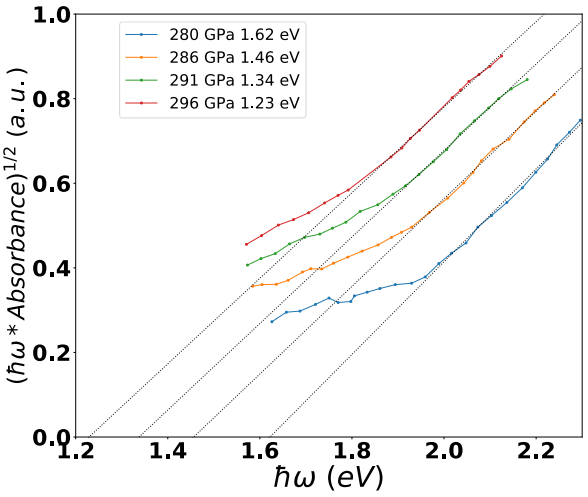

FIG. 1: Tauc analysis of the absorption profile reported in ref. [7] for an indirect gap system. Values of the gap extracted from the intercept of the linear fits and the horizontal axis are reported in the legend and in figure 3 of the main manuscript.

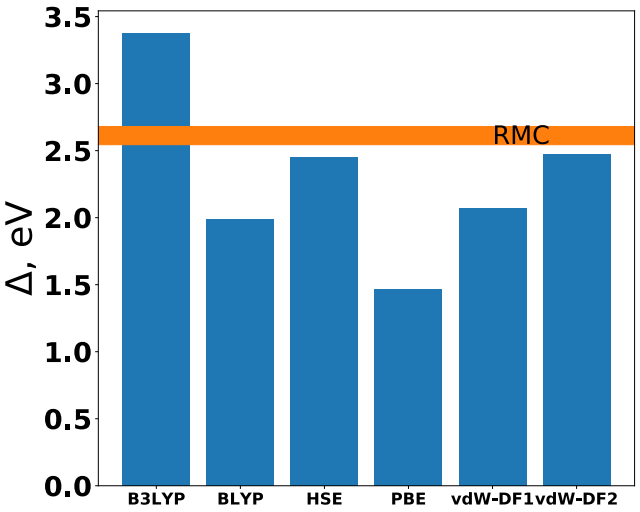

FIG. 2: The fundamental gap computed using various XC approximations for the ideal C2/c-24 structure at the nominal pressure of 300GPa. The orange horizontal bar reports the value of the RQMC-GCTABC gap and its thickness corresponds to its statistical uncertainty.

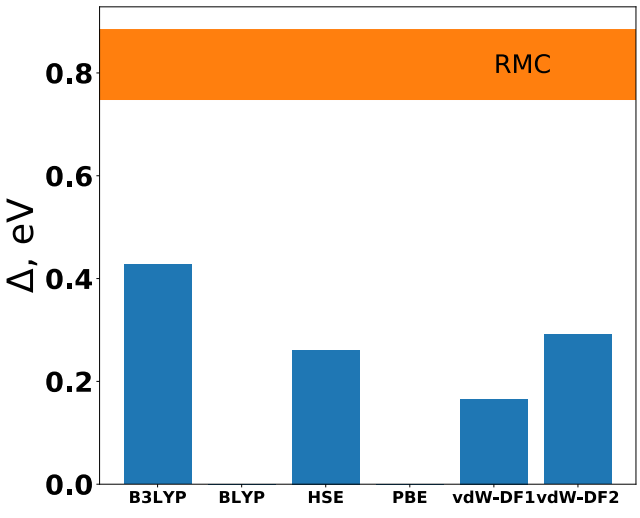

FIG. 3: The fundamental gap using various XC approximations for the ideal C2/c-24 structure at the nominal pressure of 500GPa. The orange horizontal bar reports the value of the RQMC-GCTABC gap and its thickness corresponds to its statistical uncertainty.

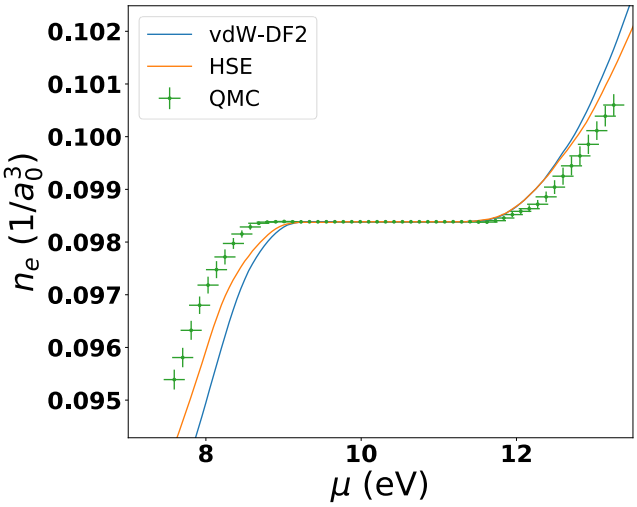

FIG. 4: Integrated density of states around the band edges from RQMC-GCTABC (green points with error bars), HSE (orange) and vdW-DF2 (blue) XC approximations for the ideal C2/c-24 crystal at 300GPa.

C. QMC benchmark of XC functionals

In this section we report some benchmark of various XC approximations for the band gap. We will focus on the ideal crystal in the C2/c-24 structure at two values of pressure: 300GPa and 500GPa. Figs. (2-3) show the values of the gap using various functionals compared with the gap from RQMC-GCTABC. HSE and vdW-DF2 are of comparable accuracy and perform better than the other approximations when the gap is large (300GPa). At 500GPa all XC approximations provides too small gaps. Again HSE and vdW-DF2 accuracy is comparable which is somehow surprising[14, 15]. In order to get more information about energy bands around the gap edges we compare in Fig. (4) the integrated density of states from vdW-DF2, HSE and RQMC-GCTABC. Besides a slightly larger incompressible region, the deviation of the QMC profile from the plateau is smoother than using either DFT approximations. The number of available

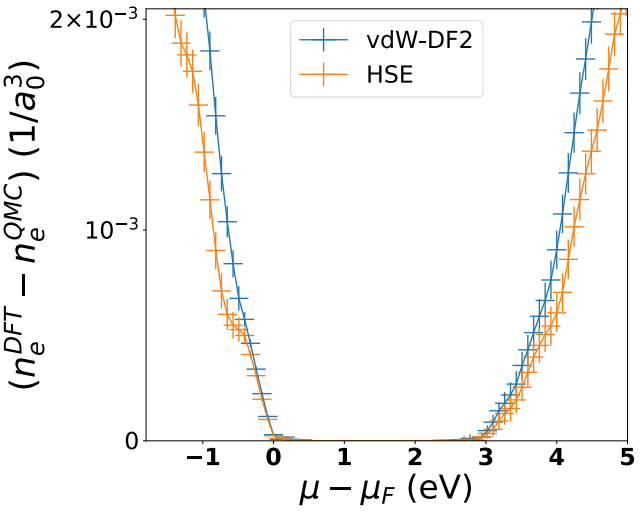

FIG. 5: The difference of the integrated density of states between DFT and QMC for the ideal C2/c-24 crystal at 300GPa. The scissor correction on the horizontal axis from the gap value has been applied to the DFT profiles before subtracting the QMC profile. For each profile, μ_F has been assigned at the maximum of the valence band.

states around the band edges is related to the intensity of optical properties. Unfortunately we cannot easily obtain optical properties with QMC methods. Therefore, it is important to try assessing the accuracy of the XC approximations. As is the common practice, for each DFT approximation we first correct the value of the DFT gap to match the QMC one, and then subtract from the corrected integrated density of states the QMC DOS. This gives an indirect measure of the expected accuracy of optical profiles of various XC approximations. The difference is shown in Fig. (5) for HSE and vdW-DF2 functionals. The two functionals perform in similar way but the HSE one has slightly smaller deviations from the QMC profile and it should provide better optical properties. In the following we have used HSE in computing optical spectra.

D. Kubo-Greenwood conductivity for thermal crystals

In order to compute optical properties, we employed the Kubo-Greenwood (KG) formalism implemented in the KGEC code [16]. Let us assume the product approximation to the exact electron-nuclear wave function $|\alpha n\rangle \simeq |\Psi_\alpha^{\mathbf{R}}\rangle |\chi_{\alpha n}\rangle$, where $\{\Psi_\alpha^{\mathbf{R}}, E_\alpha^{\mathbf{R}}\}$ is the solution of electronic problem depending parametrically on the nuclear configuration $\mathbf{R}$ and $|\chi_{\alpha n}\rangle$ is the nuclear wavefunction on the Born-Oppenheimer energy surface $E_\alpha^{\mathbf{R}}$. Assuming the electrons are initially in the ground state, one can write the real part of KG conductivity (and similarly for the imaginary part) as a thermal average over nuclear states

$$\sigma_1(\omega, T) \propto \frac{1}{Z} \sum_n e^{-\frac{E_{0n}}{k_B T}} \sigma_1(\omega, n) \tag{19}$$

where E_{0n} are the eigenvalues of the nuclear motion in the BO ground electronic states, $Z = \sum_n e^{-E_{0n}/k_B T}$ is the partition function and k_B is Boltzmann constant. In the single electron representation of the Kubo-Greenwood theory, $\sigma_1(\omega, n)$ takes the form

$$\sigma_1(\omega, n) = \frac{1}{\omega} \sum_\alpha^{occ.} \sum_{\beta,m}^{unocc.} \langle \chi_{\alpha n} | P_{\alpha\beta}^{\mathbf{R}} | \chi_{\beta m} \rangle \times \langle \chi_{\beta m} | P_{\beta\alpha}^{\mathbf{R}} | \chi_{\alpha n} \rangle \, \delta(\epsilon_{\alpha n} - \epsilon_{\beta m} - \hbar\omega) \tag{20}$$

where α indicate Kohn-Sham initial states in the valence band $|\phi_\alpha^{\mathbf{R}}\rangle$, β and m indicate, respectively, final electronic and nuclear states in the conduction band, $P_{\alpha\beta}^{\mathbf{R}} = \langle \phi_\alpha^{\mathbf{R}} | \nabla | \phi_\beta^{\mathbf{R}} \rangle$ is the matrix element of the single electron momentum operator at fixed nuclear configuration $\mathbf{R}$, and $\epsilon_{\alpha n}$ are the joint electron-nuclear eigenvalues. The conventional quasiclassical procedure introduced by Williams [17] and Lax [18] (WL) substitutes the final nuclear states with a continuum. In practice it replaces the eigenvalues $\epsilon_{\alpha n}$ in Eq. (20) by the eigenvalues evaluated at fixed nuclear configuration $\epsilon_\alpha^{\mathbf{R}}$, a procedure that can be justified as discussed in Refs [18, 19].

$$\sigma_1^{WL}(\omega, n) = \frac{1}{\omega} \sum_\alpha^{occ.} \sum_\beta^{unocc.} \langle \chi_{\alpha n} | \, |P_{\alpha\beta}^{\mathbf{R}}|^2 \delta(\Delta\epsilon_{\alpha,\beta}^{\mathbf{R}} - \hbar\omega) \, | \chi_{\alpha n} \rangle \tag{21}$$

Using second order perturbation theory, it can be shown that this expression considers in an effective way the phonon-assisted indirect transitions[19, 20]. However, for light nuclei as in the case of hydrogen, replacing the nuclear spectrum by a classical continuum might not be accurate enough.

An alternative method is to consider only direct transitions between pairs of electronic states of thermally averaged bands. This procedure will include temperature renormalization of the bands but, assuming momentum conservation, does not include indirect transitions. In practice we replace $\epsilon_{\alpha n}$ in eq. (20) by its thermal average $\langle \epsilon_\alpha \rangle$. In the following we denote this approximation as Quantum Averaging (QA). At low temperatures in the quantum regime, the QA for the conductivity,

$$\sigma_1^{QA}(\omega, n) = \frac{1}{\omega} \sum_\alpha^{occ.} \sum_\beta^{unocc.} \langle \chi_{\alpha n} | \, |P_{\alpha\beta}^{\mathbf{R}}|^2 \delta(\langle \Delta\epsilon_{\alpha,\beta}^{\mathbf{R}} \rangle - \hbar\omega) \, | \chi_{\alpha n} \rangle \, . \tag{22}$$

should be a better approximation. Note that this expression is exact when the nuclei are in their ground state and the nuclear ground state wave function does not change significantly when one electron is excited.

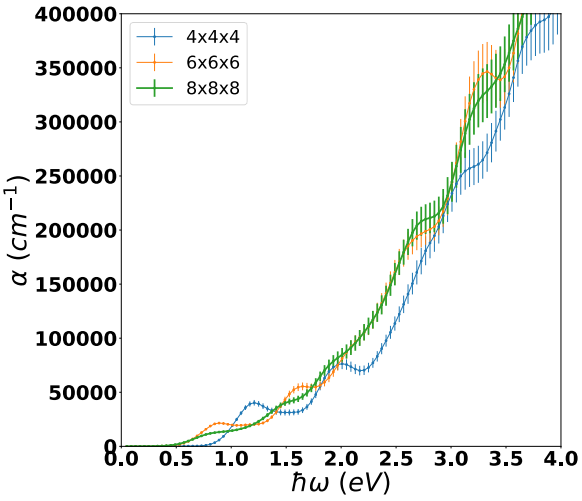

FIG. 6: K-grid convergence for C2/c-24 structure at 200K and 300 GPa

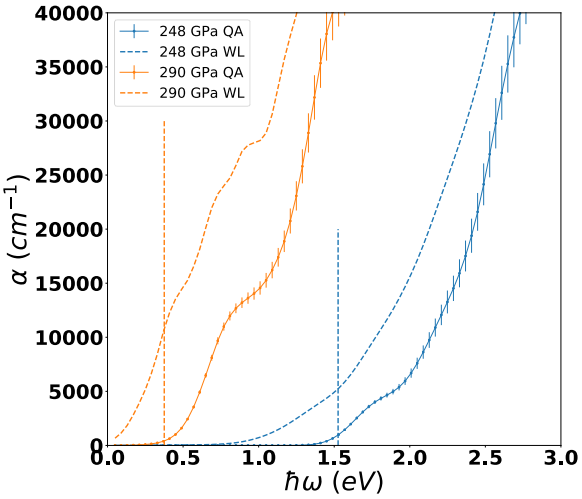

FIG. 7: Optical absorption for C2/c-24 quantum crystals at 200K using the semiclassical (WL) and quantum (QA) averaging procedures. The dashed vertical lines indicate the values of indirect band-gaps computed using the HSE functional.

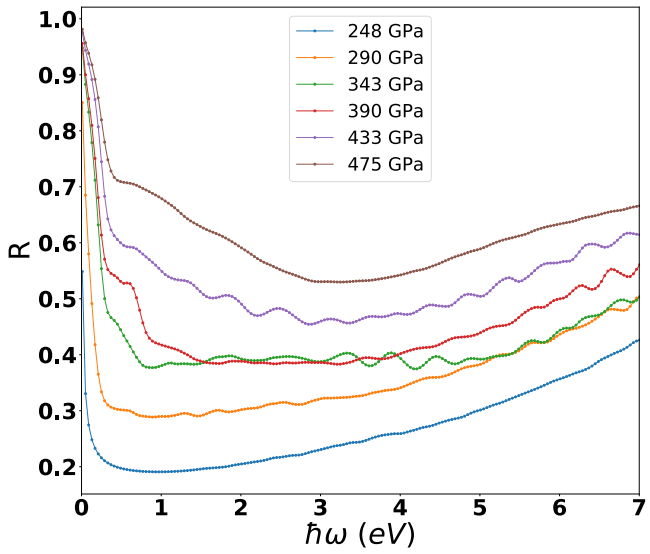

FIG. 8: Reflectivity for C2/c-24 quantum crystals at 200K from the QA procedure.

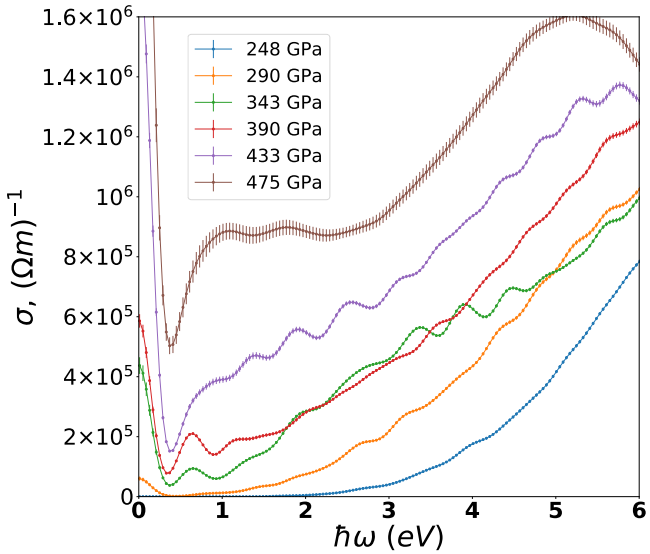

FIG. 9: The conductivity for C2/c-24 quantum crystals at 200K using the QA procedure. The smearing was 0.2 eV

E. Optical properties

We consider 40 statistically independent nuclear configurations from the CEIMC trajectory to perform the thermal averaging. For each configuration, we employ HSE-DFT electronic structure with $8 \times 8 \times 8$ k-grid and $2 \times 2 \times 2$ q-grid to sample the Fock operator. The q-grid convergence was tested on one configuration for C2/c-24 structure at 200K and 400 GPa. Figure 6 illustrates the k-grid convergence. We see that going from $6 \times 6 \times 6$ grid to $8 \times 8 \times 8$ does not modify the onset of absorption (i.e. band gap is converged), but for the larger grid, oscillations are smaller, allowing a better comparison to experiment. We note that smearing can be used to improve k-grid convergence, i.e. larger k-grid usually implies that one can use smaller smearing. However, due to dispersion of the eigenvalues, which is of the order of 0.2 eV, we are limited on the resolution of band structure and cannot take a smaller smearing.

In Fig. (7) we compare the WL and QA results for the absorption coefficients of the C2/c-24 structure at two pressures. We observe that the WL procedure does not allow one to extract the value of the indirect gap using a Tauc analysis [8], as in the case of silicon [20]. However, the QA procedure provides an interesting alternative. It deviates from zero at about the value of the indirect

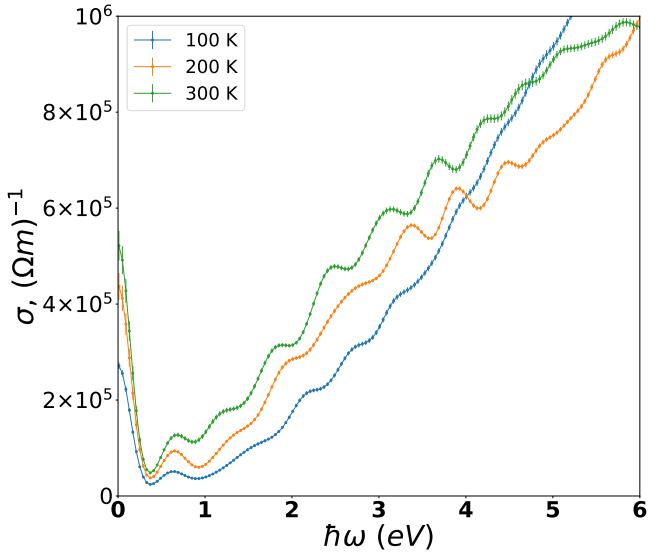

FIG. 10: The optical conductivity of the C2/c-24 quantum crystals at 343GPa and at three temperatures using the QA procedure.

gap, indicating that considering phonon mediated bands includes some indirect transitions as well, similar to the WL procedure.

In Figs. (8- 9) we show the reflectivity and the optical conductivity of the C2/c-24 structure at various pressures. The peak at $\omega = 0$ comes from the intraband transitions with a gaussian widening of 0.2eV. We see a progressive increase in optical conductivity with pressure. These results are in agreement with ref. [21]. Finally in Fig. (10) we show the conductivity of the C2/c-24 structure at 343GPa at three different temperatures. We see an increase of conductivity with temperature, hence a semiconducting behaviour, in agreement with the observation of a small but still open gap at this pressure.

F. Quasi-momentum of the electronic wave function of quantum crystals

Consider the Nth eigenstate $\Psi_N(\mathbf{r},\mathbf{R})$ of the electrons for ground state wave function $\chi_0(\mathbf{R})$ of the quantum protons and expand it around around $\psi^{\mathbf{R_0}}_{\mathbf{k}n}(\mathbf{r})$, the eigenstates of the ideal crystal with Bloch vector $\mathbf{k}$ and band index n

$$\Psi_N(\mathbf{r},\mathbf{R}) = \sum_{\mathbf{k}n} \psi^{\mathbf{R_0}}_{\mathbf{k}n}(\mathbf{r})\alpha^N_{\mathbf{k}n}(\mathbf{R}) \tag{23}$$

Further expanding the coefficients $\alpha^N_{\mathbf{k}m}(\mathbf{R})$ around the ideal crystal positions $\mathbf{R}_0$, terms linear in $\mathbf{R}-\mathbf{R}_0$ will necessarily involve phonon transitions inside electronic matrix elements. Therefore, $\alpha^N_{\mathbf{k}m}(\mathbf{R_0})$ determines the Bloch vector we can attribute to $\Psi_N(\mathbf{r},\mathbf{R})$. Assuming that $\Psi_N(\mathbf{r},\mathbf{R})$ is adiabatically connected to the ideal crystal state $\psi^{\mathbf{R_0}}_{\mathbf{k}n}(\mathbf{r})$, the crystal quasi-momentum $\mathbf{k}$ of the ideal structure is conserved for purely electronic matrix elements. Going beyond linear order in the deviations of the proton's position will involve virtual phonon-exchange processes resulting in violations of the quasi-momentum conservation.

Consistent with the Born-Oppenheimer approximation, we have assumed adiabaticity between our GCTABC electronic excitations and those of the ideal crystal ones. Thus, we can determine the crystal quasi-momentum of our excitations by taking into account the refolding of the Brillouin zone of the elementary cell to the supercell of an ideal crystal. Assuming virtual phonon exchanges to be negligible, the direct ("optical") gap is determined as the the minimum energy gap which conserves the crystal quasi-momentum in the Brillouin zone of the ideal crystal. Conservation of the twist vector in the supercell is a neccessary but not sufficient condition.

Analysing tailored overlap matrix elements between the Born-Openheimer and the ideal cyrstal wave functions, averaged over protonic configurations, we have verified within a DFT-HSE wave function that assuming adiabaticity and neglecting virtual phonon exchanges is well justified for C2/c-24 and Cmca-12 structures and pressures considered here. Whereas for $\Psi_N(\mathbf{r},\mathbf{R})$, this information can be simply obtained by projecting to ideal crystal states, Eq. (23), this analysis is less direct within the Born-Oppenheimer approximation, since the parametric dependence on $\mathbf{R}$ can only describe $\Psi_N(\mathbf{r},\mathbf{R})$ up to an arbitrary phase $\xi(\mathbf{R})$. A detailed description goes beyond the scope of this Supplementary Material and will be presented elsewhere.

[1] R. M. Martin, L. Reining, and D. M. Ceperley, *Interacting Electrons: Theory and Computational Approaches* (Cambridge University Press, 2016).
[2] In fact we consider a semi-grand-canonical ensemble where only electrons are exchanged while the number of nuclei is fixed.
[3] J. P. Perdew, Int. J. Quantum Chem. **28**, 497 (1985).
[4] C. Lin, F. H. Zong, and D. M. Ceperley, Phys. Rev. E **64**, 016702 (2001), arXiv:0101339 [cond-mat].
[5] M. Holzmann, B. Bernu, V. Olevano, R. M. Martin, and D. M. Ceperley, Phys. Rev. B **79**, 2 (2009).
[6] Y. Yang, V. Gorelov, C. Pierleoni, D. M. Ceperley, and M. Holzmann, arXiv:1910.07531.
[7] P. Loubeyre, F. Occelli, and R. LeToullec, Nature **416**, 613 (2002).
[8] J. Tauc, R. Grigorovici, and A. Vancu, Physica Status Solidi (b) **15**, 627 (1966).
[9] A. F. Goncharov, J. S. Tse, H. Wang, J. Yang, V. V. Struzhkin, R. T. Howie, and E. Gregoryanz, Phys. Rev. B **87**, 024101 (2013).

[10] R. T. Howie, T. Scheler, C. L. Guillaume, and E. Gregoryanz, Phys. Rev. B **86**, 3 (2012).
[11] P. Loubeyre, F. Occelli, and P. Dumas, (2019), arXiv:1906.05634.
[12] G. Grosso and G. Parravicini, *Solid State Physics*, 2nd ed. (Academic Press, 2014).
[13] F. Wooten, *Optical properties of Solids* (Academic Press, NY, 1972).
[14] R. C. Clay, J. McMinis, J. M. McMahon, C. Pierleoni, D. M. Ceperley, and M. A. Morales, Phys. Rev. B **89**, 184106 (2014).
[15] M. A. Morales, R. Clay, C. Pierleoni, and D. M. Ceperley, Entropy **16**, 287 (2014).
[16] L. Calderín, V. V. Karasiev, and S. B. Trickey, Comput. Phys. Commun. **221**, 118 (2017), 1707.08437.
[17] F. Williams, Phys. Rev. **82**, 281 (1951).
[18] M. Lax, The Journal of Chemical Physics **20**, 1752 (1952).
[19] C. E. Patrick and F. Giustino, Journal of Physics Condensed Matter **26**, 365503 (2014).
[20] M. Zacharias and F. Giustino, Phys. Rev. B **94**, 75125 (2016).
[21] G. Rillo, M. A. Morales, D. M. Ceperley, and C. Pierleoni, J. Chem Phys. **148**, 102314 (2018).