Photoelectron properties of DNA and RNA bases from many-body perturbation theory

Xiaofeng Qian, Paolo Umari, and Nicola Marzari

Photoelectron properties of DNA and RNA bases from many-body perturbation theory

Xiaofeng Qian[1], Paolo Umari[2], and Nicola Marzari[1,3]

[1]*Department of Materials Science and Engineering,*
Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
[2]*Theory at Elettra Group, CNR-IOM Democritos, Basovizza-Trieste, Italy and*
[3]*Department of Materials, University of Oxford, Oxford OX1 3PH, UK*
(Dated: June 27, 2011)

The photoelectron properties of DNA and RNA bases are studied using many-body perturbation theory within the GW approximation, together with a recently developed Lanczos-chain approach. Calculated vertical ionization potentials, electron affinities, and total density of states are in good agreement with experimental values and photoemission spectra. Convergence benchmark demonstrates the importance of using an optimal polarizability basis in the GW calculations. A detailed analysis of the role of exchange and correlation in both many-body and density-functional theory calculations shows that while self-energy corrections are strongly orbital-dependent, they nevertheless remain almost constant for states that share the same bonding character. Finally, we report on the inverse lifetimes of DNA and RNA bases, that are found to depend linearly on quasi-particle energies for all deep valence states. In general, our G_0W_0-Lanczos approach provides an efficient yet accurate and fully-converged description of quasiparticle properties of five DNA and RNA bases.

I. INTRODUCTION

Understanding the photoelectron properties of DNA and RNA bases and strands is of central importance to the study of DNA damage following exposure to ultraviolet light or ionizing radiation[1], and to the development of fast DNA sequencing techniques[2] and DNA and RNA-based molecular electronics and sensors[3,4]. Extensive experimental efforts[5–11] have been made since the 1970's to measure the photoelectron properties of DNA and RNA bases. Meanwhile, theoretical calculations on their ionization potentials and electron affinities have been carried out using density-functional theory (DFT) and high-level quantum chemistry methods[8,10,12–15]. However, the results from DFT calculations are highly dependent on exchange-correlation functionals, and quantum chemistry methods, though more accurate, require considerably more computational effort. In contrast, many-body perturbation theory within Hedin's GW approximation[16,17] presents a unique framework that allows access to both quasi-particle (QP) energies and lifetimes on the same footing. This method has been successfully applied to quasi-1D, 2D and 3D semiconductors, insulators, and metals[18–23], and very recently to molecular systems[24–33].

In this work, we present the *entire* QP spectrum of DNA and RNA bases using many-body perturbation theory within Hedin's GW approximation, obtained with a recently developed approach that is particularly effective in reaching numerical convergence[27,28]. In the GW approximation, the self-energy operator is expressed as a convolution of the QP Green's function G with the screened Coulomb interaction W. Therefore, at increasing system sizes (as is the case for the present work) two computational challenges arise: (i) first, a large basis set has to be adopted to represent operators such as polarizability, and (ii) the calculation of the irreducible dynamical polarizability and that of the self-energy require sums over single-particle conduction states that converge very slowly. We overcome these two obstacles through (i) the use of optimal basis sets for representing the polarization operators[27] and (ii) the use of a Lanczos-chain algorithm[28,29] to avoid explicit sums over empty single-particle states. In addition, the G_0W_0 approximation is adopted, in which the dynamical polarizability is calculated within the random-phase approximation and the QP Green's function is replaced by its unperturbed single-particle counterpart. This approach is implemented in the open-source QUANTUM-ESPRESSO distribution[34]. It is applied here to achieve fully-converged QP spectra and inverse lifetimes of the five isolated DNA and RNA bases and to investigate the important but distinct roles of exchange and correlation in the G_0W_0 self-energy corrections.

Photoelectron properties of DNA and RNA bases using many-body GW have not been reported until a very recent study by Faber *et al.*[33]. The work by Faber *et al.* presents a many-body GW study on QP energies (including ionization potentials and electron affinities) of DNA and RNA bases at several levels of self-consistency within the GW approximation. Their calculations are based on the conventional implementation of the GW method using a localized basis-set and a direct sum-over-states approach, and it demonstrated that self-consistent GW calculations indeed further improve the results of G_0W_0 (one-shot GW) calculations. Although localized basis-set can significantly improve the computational efficiency and a direct sum-over-states approach can be easily implemented, both of them could have several potential drawbacks, and can introduce large errors in QP

energies. One may solve the former issue by systematically increasing the size of basis-sets; however, there is no simple solution to the convergence problem introduced by the direct sum-over-state approach. Second, dipole-bound conduction states will not be obtained from localized basis-sets due to their highly diffuse character in the vacuum region. In fact, only electron affinities of covalent-bound conduction states were reported in the Faber's work. Therefore, it would be desirable to calculate GW QP energies in plane-wave basis without suffering from the above issues, which is one of the subjects of this research.

The paper is organized as follows. Computational details of our calculations are given in Sec. II. Real-space representations of optimal polarizability basis are displayed in Sec. III. We then report convergence benchmark in Sec. IV. In Sec. V, we present QP energies and inverse lifetimes as well as the entire QP spectra for all five DNA and RNA bases, including guanine (G), adenine (A), cytosine (C), thymine (T), and uracil (U). Vertical ionization potentials (VIPs) and vertical electron affinities (VEAs) are compared to experimental data and other theoretical results. Two types of VEAs are reported using plane-wave basis, including valence-bound (VB, also called covalent-bound) VEA and dipole-bound (DB) VEAs. In Sec. VI, we reveal the role of exchange and correlation in GW self-energy corrections to the DFT Kohn-Sham eigenvalues. Finally, we summarize our work in Sec. VII.
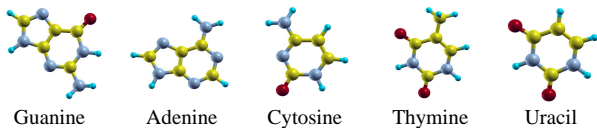
II. COMPUTATIONAL DETAILS

FIG. 1: (color online). Ground-state structures of five DNA and RNA bases including G9K-guanine, adenine, C1-cytosine, thymine, and uracil.

Ground-state DFT calculations are performed in a cubic supercell of 18.0^3 Å^3, using the Perdew-Burke-Ernzerhof's (PBE) exchange-correlation functional, Troullier-Martins's norm-conserving pseudopotentials, and a plane-wave basis set with a cutoff of 544 eV. Structures are optimized with a residual force threshold of 0.026 eV/Å. A truncated Coulomb potential with radius cutoff of 7.4 Å is employed to remove artificial interactions from periodic images. The vacuum level is corrected by an exponential fitting of E_{HOMO} with respect to the supercell volume. The polarizability basis sets have been obtained using a parameter E^* of 136.1 eV and a threshold q^* of 0.1 a.u., giving an accuracy of 0.05 eV for the calculated QP energies (E^* and q^* will be explained in the next section). The final accu-

racy including the errors from the analytic continuation is about 0.05 to 0.1 eV. The structures of five DNA and RNA bases are shown in Fig. 1. Here the effect of gas-phase tautomeric forms[15] of guanine and cytosine on QP properties are beyond the scope of this work, and we only focus on the G9K form of guanine and the C1 form of cytosine[15].

III. OPTIMAL POLARIZABILITY BASIS

The key quantity in many-body GW calculations is the irreducible dynamic polarizability $\hat{P}_0$ in the random-phase approximation:

$$\hat{P}_0(\omega) = -i \sum_{v,c} \frac{|\psi_v \psi_c\rangle \langle \psi_c \psi_v|}{\omega + i\eta - (\varepsilon_c - \varepsilon_v)}, \qquad (1)$$

where η is an infinitesimal positive real number. $|\psi_v \psi_c\rangle$ denotes the direct product of a valence state ψ_v and a conduction state ψ_c in real space and ψ_v and ψ_c are considered to be real. A strategy was proposed in Refs. 27 and 28 for obtaining a compact basis set, referred to as *optimal polarizability basis*, to represent $\hat{P}_0$ at all frequencies. First, we consider the frequency average of $\hat{P}_0(\omega)$ which corresponds to the element at time $t = 0$ of its Fourier transform $\tilde{\hat{P}}_0(t)$, without considering the constant $(-i)$:

$$\tilde{\hat{P}}_0(t = 0) = \sum_{v,c} |\psi_v \psi_c\rangle \langle \psi_c \psi_v|. \qquad (2)$$

We not that $\tilde{\hat{P}}_0(t = 0)$ is positive definite. Then, the *optimal polarizability basis*, $\{\Phi_\mu\}$, is built from the most important eigenvectors of $\tilde{\hat{P}}_0(t = 0)$, corresponding to the largest eigenvalues q_μ above a given threshold q^*:

$$\tilde{\hat{P}}_0(t = 0) |\Phi_\mu\rangle = q_\mu |\Phi_\mu\rangle. \qquad (3)$$

It must be noted that this does not require any explicit calculation of empty (i.e., conduction) states as we can use the closure relation:

$$\hat{P}_c = 1 - \hat{P}_v, \qquad (4)$$

together with an iterative diagonalization scheme. However, the latter procedure would build polarizability basis sets which are larger than what is necessary for a good convergence of the quasi-particle energy levels. This stems from treating all the one-particle excitations on the same footing, independent of their energy. A practical solution would be to limit the sum in Eq. (2) on the conduction states below a given energy cutoff E^*:

$$\tilde{\hat{P}}_0' = \sum_{v,c}^{\epsilon_c < E^*} |\psi_v \psi_c\rangle \langle \psi_c \psi_v|, \qquad (5)$$

However, limiting the sum over the empty states laying in the lower part of the conduction manifold does not allow to use the closure relation alluded to above.

Thus, in order to keep avoiding the calculation of empty states we replace them in Eq. (5) with a set of plane waves $\{\mathbf{G}\}$ with their kinetic energies lower than E^*, which are first projected onto the conduction manifold using Eq. (4) and then orthonormalized. We indicate these *augmented plane-waves* as $\{\tilde{\mathbf{G}}\}$ and arrive at the following modified operator,

$$\tilde{\tilde{P}}_0'' = \sum_{v,\tilde{\mathbf{G}}} |\psi_v\tilde{\mathbf{G}}\rangle\langle\tilde{\mathbf{G}}\psi_v|, \qquad (6)$$

which is also positive definite. An optimal polarizability basis $\{\Phi_\mu\}$ is finally obtained by replacing $\tilde{P}_0(t=0)$ in Eq. (3) with $\tilde{\tilde{P}}_0''$.

It should be stressed that the above approximation is used only for obtaining a set of optimal basis vectors for representing the polarization operators and not for the actual calculation of the irreducible dynamic polarizability at finite frequency in Eq. (1); the latter is performed using a Lanczos-chain algorithm[28]. Moreover, due to the completeness of the eigenvectors of $\tilde{P}''$, for any value of E^* the GW results will converge to the same values by lowering the threshold q^*, and eventually reach the same results as those obtained by directly using a dense basis of plane-waves. However, compared to the pure plane-waves which are completely delocalized in real space, the optimal polarizability basis is particularly convenient for isolated systems since the most important eigenvectors of $\tilde{\tilde{P}}_0''$ will be mostly localized in the regions with higher electron density. Thus, converged results can be obtained using much smaller optimal-polarizability basis sets than plane-waves basis sets.

these provide well converged results. It is clearly seen that the eigenvalues of the optimal polarizability basis decay exponentially and change by almost four orders of magnitude from the first to the last basis. In Fig. 3 we show the real-space representations of a few selected elements. The first five, corresponding to the five largest eigenvalues, are strongly localized around the chemical bonds of the molecule. The second row contains five elements which are more delocalized, and those in the last row are completely delocalized. This indicates that even though localized optimal basis like those shown in the first two rows can be easily captured by localized basis-sets, the delocalized ones with smaller eigenvalues q_μ (like those in the last row) are more difficult to capture if diffuse functions are not employed.

FIG. 3: (color online). Real-space representation of optimal-polarizability basis elements for cytosine, labeled with their eigenvalue indexes. Due to the delocalized nature of the optimal basis in the third row, the images in the third row were generated with a smaller isovalue and shown at a larger scale than those in the first two rows.
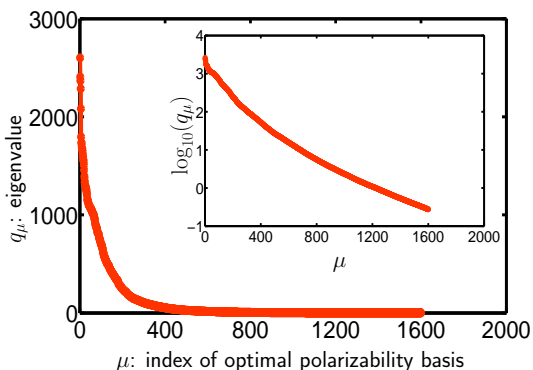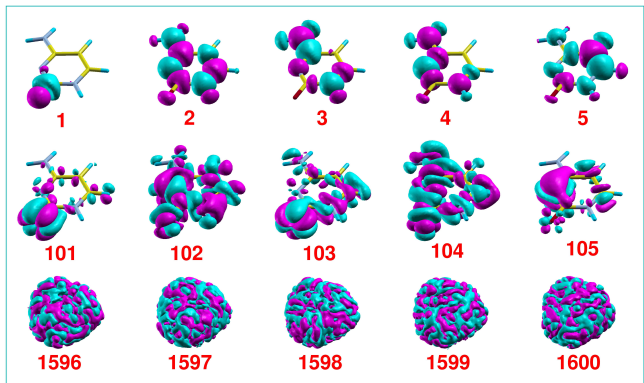

FIG. 2: Eigenvalue distribution of the optimal polarizability basis for cytosine. The inset plot shows the eigenvalues in a log scale.

Now, we want to have a closer look at the optimal polarizability basis. The eigenvalue distribution of $\tilde{P}_0''$ for cytosine is displayed in Fig. 2. We only show the largest 1600 eigenvalues with $E^* = 136.1$ eV in the plot, since

IV. CONVERGENCE BENCHMARK

The number of optimal-polarizability basis elements N_P and the energy cutoff of the augmented plane-waves E^* are two critical parameters used in our G_0W_0 calculations in order to achieve both efficiency and accuracy. Therefore, we performed a series of calculations to benchmark the convergence with respect to these two parameters. In Fig. 4, we present the convergence behavior of VIPs and VB-VEAs of five DNA and RNA bases for the highest-occupied molecular orbital (HOMO) and the lowest-unoccupied molecular orbital (LUMO), respectively: $\text{VIP} \equiv -\text{Re}(\varepsilon_{\text{HOMO}}^{\text{QP}})$ and $\text{VEA} \equiv -\text{Re}(\varepsilon_{\text{LUMO}}^{\text{QP}})$. We find that for both VIPs and VEAs convergence within 0.1 eV is achieved with ~ 600 optimal basis elements for $E^* = 95.2$ eV and with ~ 750 optimal basis elements for $E^* = 136.1$ eV. Indeed, similar trends were reported in Ref. 27. VIPs and VEAs reported in the following sections are calculated using the most strict parameters ($N_P = 2400$ and $E^* = 136.1$ eV).
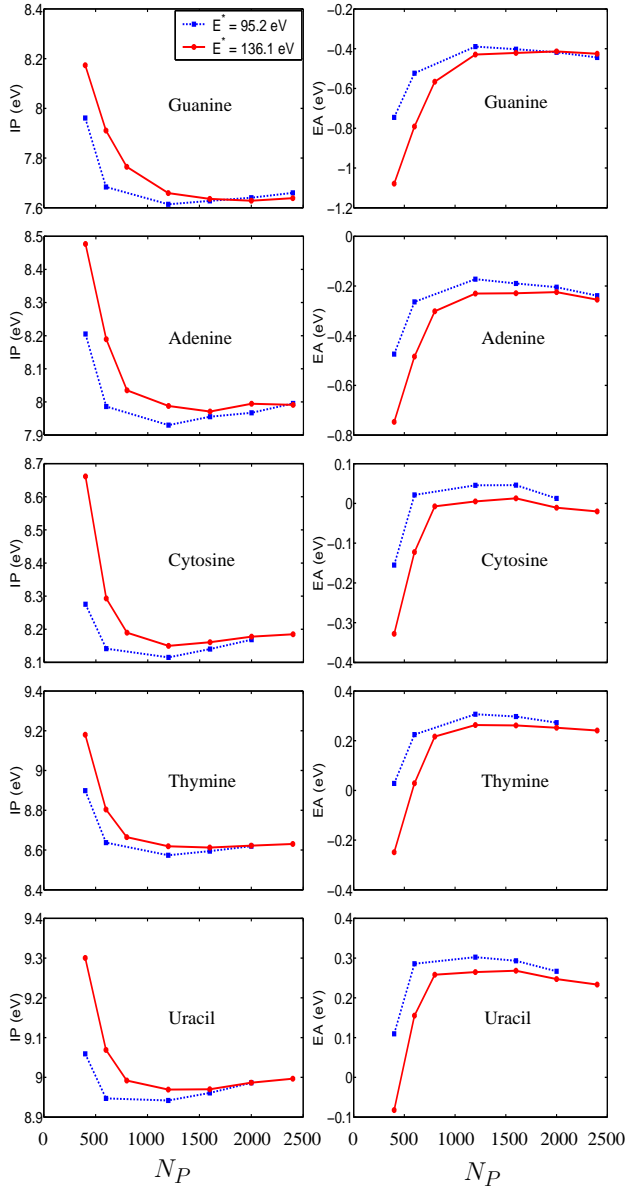
FIG. 4: (color online). Convergence benchmark of VIP and VEA of five DNA and RNA bases with respect to number of optimal-polarizability basis elements N_P, and augmented plane-wave cutoff E^*. Results using $E^* = 95.2$ and 136.1 eV are plotted in dashed-blue lines and solid-red lines, respectively.

The above benchmark indicates that, if basis-sets and conduction states in DFT calculations are not properly tested, one could easily obtain non-converged results from G_0W_0 calculations, resulting in higher VIPs and lower VEAs for all five bases. We also note that the choice of N_P and E^* remains the same for all the DNA and RNA bases, indicating portability for these parameters.

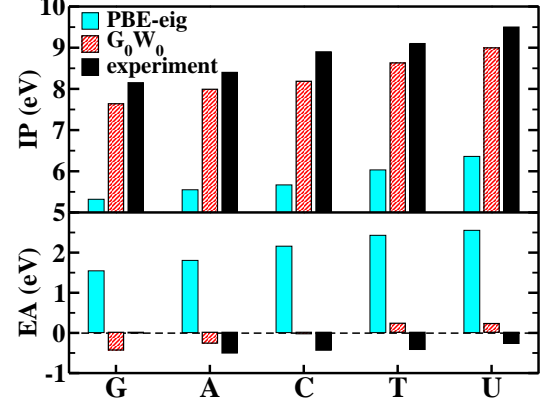## V. IONIZATION POTENTIALS AND ELECTRON AFFINITIES

FIG. 5: (color online). VIP and VB-VEA of five DNA and RNA bases from our DFT and G_0W_0 calculations. Here we adopt the mean values of various experimental data listed in Table I. The experimental ranges are $8.0 \sim 8.3$, $8.3 \sim 8.5$, $8.8 \sim 8.9$, $9.0 \sim 9.2$, and $9.4 \sim 9.6$ eV for G, A, C, T, and U, respectively.

VIPs and VB-VEAs from our G_0W_0 calculations and experimental data are shown in Fig. 5 for all five bases, together with the DFT-PBE eigenvalues for the HOMO and LUMO levels. Only the mean values of experimental VIPs and VEAs are plotted in Fig. 5. G_0W_0 dramatically improves VIPs and VEAs compared to DFT-PBE eigenvalues, providing VIPs of 7.64, 7.99, 8.18, 8.63 and 8.99 eV and VEAs of -0.43, -0.25, -0.02, 0.24, and 0.23 eV for G, A, C, T, and U, respectively. The experimental VIPs are compiled in Table I, and span a range of $8.0 \sim 8.3$, $8.3 \sim 8.5$, $8.8 \sim 8.9$, $9.0 \sim 9.2$, and $9.4 \sim 9.6$ eV for G, A, C, T, and U. Compared to the mean values of experimental VIPs, the mean absolute error of the calculated VIPs for all five bases is 0.52 eV. Furthermore, experimental VB-VEAs are negative for all five bases, indicating that excited π^* states are unstable upon electron attachment. This leads to challenging measurements of VEAs and a wide range of measured values[12,14] listed in Table I: $-0.56 \sim -0.45$, $-0.55 \sim -0.32$, $-0.53 \sim -0.29$, and $-0.30 \sim -0.22$, for A, C, T, and U. Compared to the mean values of experimental VEAs, the mean absolute errors of the calculated VEAs for four bases is 0.45 eV. Interestingly, the VEA of guanine has never been measured successfully, possibly due to a large negative value. This is clearly reflected in our calculated G_0W_0 VEA of -0.43 eV, which is the most negative one among all five bases. Even though the G_0W_0 VEAs of thymine and uracil are slightly positive, the trend for all the calculated VEAs agrees well with experiments. In addition, the DFT-PBE HOMO-LUMO gaps for the five bases are about 45% of the G_0W_0 gaps. This is in agreement with previous observations that DFT with the local density approximation (LDA) or the generalized gradient

approximation (GGA) of exchange-correlation functionals usually underestimates by 30-50% the true QP energy gap[35,36].

We further compare several low-lying G_0W_0 VIPs and their excitation characters with experimental and other theoretical results and assignments. First, as shown in Table I, both G_0W_0 VIPs and their orbital assignments agree well with experiments and other theoretical works for all the five bases, where the corresponding excitation character is either π or n (lone-pair). Second, our G_0W_0 VIPs, especially those corresponding to the five HOMO levels, are in good agreement with Faber's G_0W_0 values calculated in localized basis sets. However, larger deviations are clearly observed in some of the lone-pair valence states. Their G_0W_0 VIPs are higher than our values by 0.30, 0.40, 0.38, and 0.37 eV for HOMO-1 (the first lone-pair state) of cytosine, HOMO-1 (the first lone-pair state) of thymine, and HOMO-1 and HOMO-3 (the first and second lone-pair states) of uracil, respectively. We plot in Fig. 6 the convergence behavior of VIPs with respect to the dimension of the polarizability basis for these lone-pair states to check whether convergence issues are present. But it is apparent that VIPs from our G_0W_0 calculations are fully converged. Another significant difference is found in the valence-bound VEAs for all five LUMO levels. Moreover, Faber's G_0W_0 VB-VEAs are lower than the present results by 0.61, 0.39, 0.43, 0.38, and 0.34 eV for G, A, C, T, and U, respectively. It is interesting to notice that similar trends of increased VIPs and decreased VEAs are observed in the previous convergence benchmark of Fig. 4, when a small optimal polarizability basis was employed. However, since we do not find significant difference in the G_0W_0 VIPs for other QP states, the source of the above deviations is not clear. Furthermore, as listed in Table I, the work by Faber *et al.* demonstrates the importance of self-consistency of QP energies in GW calculations with QP wavefunctions unchanged. This self-consistent GW method increases the G_0W_0 VIPs of the HOMO levels by 0.32, 0.32, 0.52, 0.41, and 0.44 eV and decreases the G_0W_0 VEAs of the LUMO levels by 0.54, 0.50, 0.46, 0.53, and 0.53 eV for G, A, C, T, and U, respectively. Results from advanced quantum chemistry methods are also listed in Table I, including complete active space with second-order perturbation theory (CASPT2)[13,14], coupled-cluster with singles, doubles, and perturbative triple excitations (CCSD(T))[13,14], and equation of motion ionization potential coupled-cluster (EOM-IP-CCSD)[15]. VIPs from CASPT2, CCSD(T), and EOM-IP-CCSD for the HOMO levels are very similar, and close to the experimental mean values within 0.07, 0.07, and 0.05 eV, respectively. VEAs from CASPT2 and CCSD(T) for the LUMO levels are also close to each other; however, they are less close to the mean experimental values (within 0.30 and 0.33 eV, respectively). Among all the theoretical approaches, self-consistent GW and quantum chemistry methods provide the VIPs and VEAs with the smaller errors with respect to the experimental data.
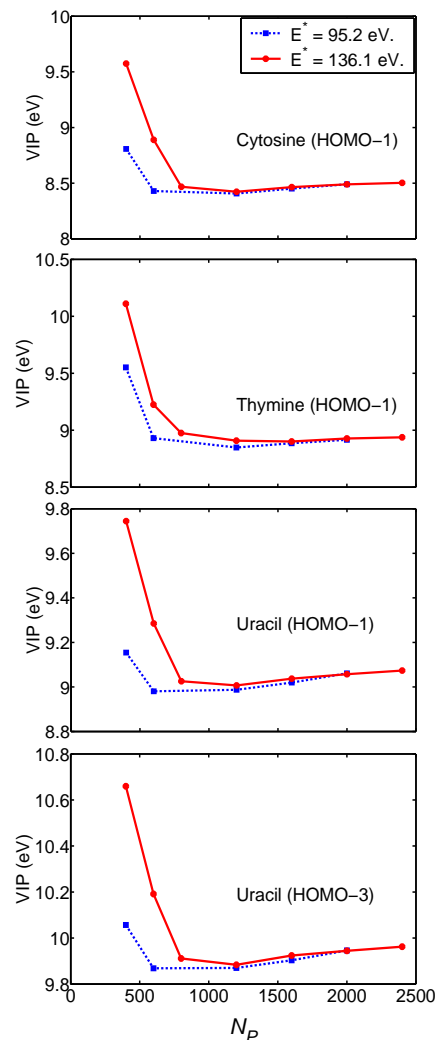

FIG. 6: (color online). Convergence behavior of vertical ionization potentials of several n states in DNA and RNA bases with respect to number of optimal polarizability basis, N_P, and augmented plane-wave cutoff, E^*. These states are cytosine's HOMO-1 state, thymine's HOMO-1 state, and uracil's HOMO-1 and HOMO-3 states. Results using $E^* = 95.2$ and 136.1 eV are plotted in dashed-blue lines and solid-red lines, respectively.

Beside the VB-VEAs, there also exist dipole-bound (DB) VEAs, which correspond to having the additional electron weakly bound to the DNA and RNA bases by local electrostatic dipoles[14]. Both types of QP states are shown in Fig. 7. It is clear that all five VB states are localized π^* states, while DB states present large lobes, highly extended outside the molecules. These lobes are mainly located in the vicinity of the N-H bond, and with a non-negligible dipole moment along their bond axis. The energy difference between the VB-VEAs and their nearest DB-VEAs, $\Delta_{\mathrm{VEA}} \equiv \mathrm{VEA(VB)} - \mathrm{VEA(DB)}$, are -0.23, 0.06, 0.21, 0.48, and 0.52 eV for G, A, C, T, and U, respectively. This suggests that at the G_0W_0 level VB states in the latter four bases are energetically more

TABLE I: Vertical ionization potentials and vertical electron affinities for several low-lying Kohn-Sham eigenstates close to the HOMO and LUMO levels obtained from negative DFT-PBE eigenvalues and G_0W_0(PBE) in comparison with other GW and quantum chemistry calculations and experimental data. CASPT2: complete active space with second-order perturbation theory; CCSD(T): coupled-cluster with singles, doubles, and perturbative triple excitations; EOM: equation of motion ionization potential coupled-cluster. Only the valence-bound vertical electron affinities are shown in this table. The experimental mean values are taken as the reference in the calculations of mean absolute error(MAE) for both LUMO and HOMO levels.

		DFT-PBE[a]	G_0W_0(PBE)[a]	G_0W_0(LDA)[b]	GW(LDA)[b]	CASPT2[c,d]/CCSD(T)[c,d]	EOM[e]	Experiment[f,g,h,i,j]
G	[LUMO] 1.12 (π)	-0.43	-1.04	-1.58	-1.14[c]/		
	[HOMO] 5.32 (π)	7.64	7.49	7.81	8.09[d]/8.09[d]	8.15	8.0∼8.3[f]/8.30[i]/8.26[j]
	5.88 (n)	8.67	8.78	9.82	9.56[d]/	9.86	9.90[i]/9.81[j]
	6.37 (n)	9.38			9.61[d]/	10.13	
	7.04 (π)	9.43			10.05[d]/	10.29	
	6.94 (π)	9.48			10.24[d]/	10.58	10.45 (n)[i]/10.36[j]
	7.76 (π)	10.37			10.90[d]/	11.38	11.15[i]/11.14[j]
	7.64 (n)	10.57					
A	[LUMO] 1.81 (π)	-0.25	-0.64	-1.14	-0.91[c]/		-0.56∼-0.45[g]
	[HOMO] 5.55 (π)	7.99	7.90	8.22	8.37[d]/8.40[d]	8.37	8.3∼8.5[f]/8.47[h]
	5.89 (n)	8.80	8.75	9.47	9.05[d]/	9.37	9.45[h]
	6.65 (π)	9.06			9.54[d]/	9.60	9.54[h]
	6.74 (n)	9.71			9.96[d]/	10.42	10.45[h]
	7.22 (π)	9.78			10.38[d]/	10.58	10.51[h]
	7.58 (n)	10.65			11.06[d]/	11.47	11.35[h]
C	[LUMO] 2.16 (π)	-0.02	-0.45	-0.91	-0.69[c]/-0.79[c]		-0.55∼-0.32[g]
	[HOMO] 5.67 (π)	8.18	8.21	8.73	8.73[d]/8.76[d]	8.78	8.8∼9.0[f]/8.89[h]
	5.63 (n)	8.50	8.80	9.89	9.42[d]/	9.65	9.45[i]/9.55[h]
	6.28 (π)	8.94	8.92	9.52	9.49[d]/	9.55	9.89[h]
	6.38 (n)	9.39	9.38	10.22	9.88[d]/	10.06	11.20[h]
	8.44 (π)	11.08			11.84[d]/	12.28	11.64[h]
	9.27 (π)	11.98			12.71[d]/	13.27	12.93 (σ, π)[h]
T	[LUMO] 2.43 (π)	0.24	-0.14	-0.67	-0.60[c]/-0.65[c]		-0.53∼-0.29[g]
	[HOMO] 6.03 (π)	8.63	8.64	9.05	9.07[d]/9.04[d]	9.13	9.0∼9.2[f]/9.19[h]
	6.12 (n)	8.94	9.34	10.41	9.81[d]/	10.13	9.95∼10.05[f]/10.14[h]
	6.80 (π)	9.52			10.27[d]/	10.52	10.39∼10.44[f]/10.45[h]
	6.93 (n)	9.77			10.49[d]/	11.04	10.80∼10.88[f]/10.89[h]
	8.79 (π)	11.53			12.37[d]/	12.67	12.10∼12.30[f]/12.27[h]
U	[LUMO] 2.55 (π)	0.23	-0.11	-0.64	-0.61[c]/-0.64[c]		-0.30∼-0.22[g]
	[HOMO] 6.36 (π)	8.99	9.03	9.47	9.42[d]/9.43[d]		9.4∼9.6[f]
	6.14 (n)	9.07	9.45	10.54	9.83[d]/		10.02∼10.13[f]
	7.00 (π)	9.68	9.88	10.66	10.41[d]/		10.51∼10.56[f]
	6.92 (n)	9.96	10.33	11.48	10.86[d]/		10.90∼11.16[f]
	9.17 (π)	11.90			12.59[d]/		12.50∼12.70[f]
MAE	[LUMO] 2.64 (π)	0.45	0.14	0.44	0.30/0.33		
	[HOMO] 3.02 (π)	0.52	0.56	0.15	0.07/0.07	0.05	

[a] This work.
[b] Reference 33.
[c] Reference 14.
[d] Reference 13.
[e] Reference 15.
[f] Collected in Reference 13.
[g] Collected in Reference 14.
[h] Reference 8.
[i] Reference 6.
[j] Reference 10.

FIG. 7: (color online). Valence-bound and dipole-bound VEAs and their corresponding QP states, calculated at the DFT-PBE level, in the five DNA and RNA bases. Values listed below are VEAs in the unit of eV.

stable than the DB ones.

Experimental valence photoemission spectra extends into deep valence states[8,37,38], allowing us to further evaluate our G_0W_0 results at a broader energy range. The DFT-PBE and G_0W_0 densities of states (DOS) for all five bases, neglecting any oscillator strength effect, are compared to valence photoemission spectra in Fig. 8(a). For better comparison, both curves are shifted in order to match the first experimental VIP. It is clearly shown that for all five bases the G_0W_0 DOS agrees much better with experiment than the DFT-PBE DOS, thanks to the correct relative position of the various peaks. Moreover, the G_0W_0 self-energy not only leads to large corrections to DFT eigenvalues, but also provides an estimation of QP intrinsic lifetimes due to inelastic electron-electron scattering, as reflected in the imaginary part of QP energies, with $1/\tau_n = 2|\text{Im}(\varepsilon_n^{\text{QP}})|$. The calculated QP inverse lifetimes at the G_0W_0 level are plotted in Fig. 8(b) against the corresponding QP valence energies. Although G_0W_0 permits only a rough estimate of QP lifetimes (the exact ones are expected to be zero in the range $[2\text{Re}(\varepsilon_{\text{HOMO}}^{\text{QP}}), \text{Re}(\varepsilon_{\text{HOMO}}^{\text{QP}})])$, we note that the QP inverse lifetimes decrease almost linearly with respect to QP energies for the deep valence states in all five cases. However, it is still unknown to what extent the G_0W_0 estimation of inverse lifetime would be modified by fully self-consistent GW calculations.

VI. ROLE OF EXCHANGE AND CORRELATION IN GROUND-STATE DFT AND GW CALCULATIONS

In order to understand the role of exchange and correlation in the self-energy corrections to the DFT-PBE results, we first express each Kohn-Sham eigenvalue $\varepsilon_n^{\text{KS}}$ of eigenstate ψ_n for the n-th state as the sum of a single-particle energy ε_n^{S} and an exchange-correlation energy $\varepsilon_n^{\text{XC}}$: $\varepsilon_n^{\text{KS}} = \varepsilon_n^{\text{S}} + \varepsilon_n^{\text{XC}}$, where ε_n^{S} contains the energy contributions from the kinetic energy operator, the exter-
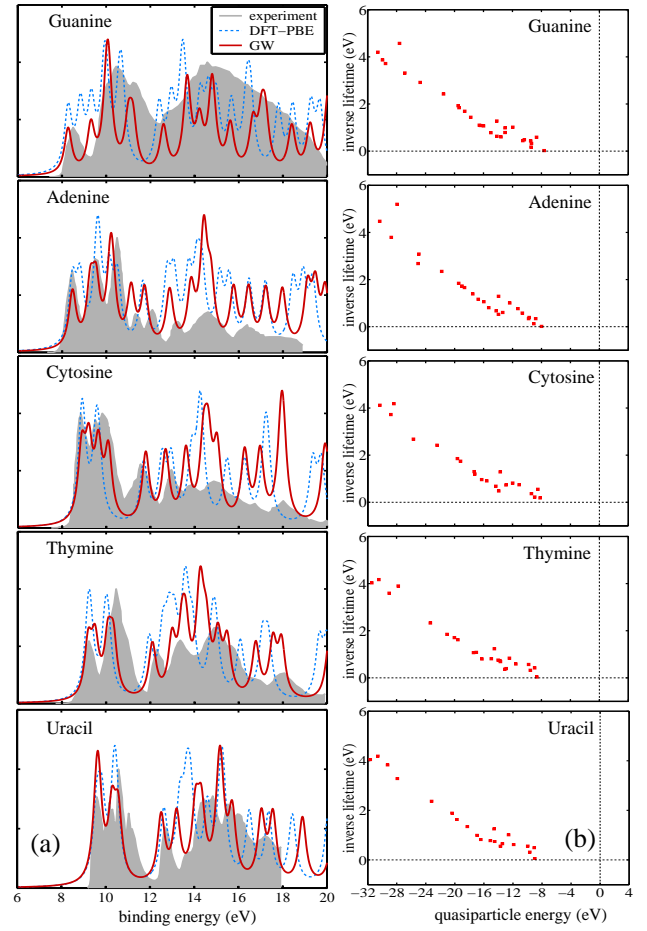

FIG. 8: (color online). (a) Experimental valence photoemission spectrum (shaded gray area), DFT-PBE DOS (blue dashed lines), and G_0W_0 DOS (red solid lines). Both DOS curves are shifted to match the first VIP of experimental data. Experimental PES spectra of G, A, C, T, and U are extracted from Ref. 37, 8, 8, 8, and 38, respectively. The theoretical DOS have been obtained through a Lorentzian broadening defined by a width of 0.4 eV. (b) G_0W_0 QP energies and inverse lifetime for valence states (unit: eV).

nal ionic potential, and the Hartree term. Furthermore, the G_0W_0 QP energy can be written in terms of the exchange self-energy Σ_n^{X} and of the correlation self-energy Σ_n^{C}: $\varepsilon_n^{G_0W_0} = \varepsilon_n^{\text{S}} + \Sigma_n^{\text{X}} + \Sigma_n^{\text{C}}$. Exchange and correlation effects can then be systematically investigated by analyzing $\Sigma_n^{\text{X}}, \Sigma_n^{\text{C}}, \Sigma_n^{\text{XC}}, \varepsilon_n^{\text{XC}}$, and Δ_n^{XC}, with $\Sigma_n^{\text{XC}} \equiv \Sigma_n^{\text{X}} + \Sigma_n^{\text{C}}$ and $\Delta_n^{\text{XC}} \equiv \Sigma_n^{\text{XC}} - \varepsilon_n^{\text{XC}}$. We consider the adenine molecule and plot the above quantities with respect to the G_0W_0 QP energy $\varepsilon_n^{G_0W_0}$. As shown in Fig. 9(a) and (b), the G_0W_0 exchange energy Σ_n^{X} increases from -28.0 to -17.8 eV for the 25 valence states and from -9.2 to -0.2 eV for the 10 conduction states, while the G_0W_0 correlation energy Σ_n^{C} decreases from 7.1 eV down to 0.2 eV for the valence states and from -0.3 to -3.1 eV for the conduction states. This clearly shows that Σ_n^{X} is always negative, stabilizing both electron and hole excitations; however, Σ_n^{C} is positive for valence states and negative for
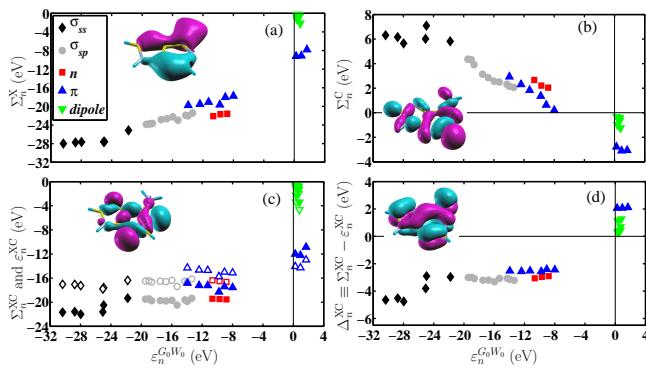
FIG. 9: (color online). The role of exchange and correlation in the G_0W_0 self-energy corrections to Kohn-Sham eigenvalues of 25 valence states and 10 conduction states in adenine. (a) G_0W_0 exchange energy Σ_n^{X}, (b) G_0W_0 correlation energy Σ_n^{C}, (c) the sum of G_0W_0 exchange and correlation energy Σ_n^{XC} (filled symbols) and DFT XC energy $\varepsilon_n^{\mathrm{XC}}$ (unfilled symbols), and (d) the difference between G_0W_0 and DFT exchange-correlation energy, $\Delta_n^{\mathrm{XC}} \equiv \Sigma_n^{\mathrm{XC}} - \varepsilon_n^{\mathrm{XC}}$. Four types of molecular orbitals are illustrated in (a)-(d), corresponding to σ_{ss}, σ_{sp}, n, and π characters.

conduction states, indicating that the effect of correlation is that of destabilizing hole excitations and of stabilizing electron excitations. Although Σ_n^{X} and Σ_n^{C} have opposite trends for hole excitations, exchange interactions eventually dominate due to their larger magnitude, leading to the negative Σ_n^{XC} of Fig. 9(c). Interestingly, the G_0W_0 Σ_n^{XC} is lower than the DFT-PBE $\varepsilon_n^{\mathrm{XC}}$ for the valence states, but higher than $\varepsilon_n^{\mathrm{XC}}$ for the conduction states. Consequently, the difference Δ_n^{XC} between Σ_n^{XC} and $\varepsilon_n^{\mathrm{XC}}$, shown in shown in Fig. 9(d), is negative for the valence manifold and positive for the conduction manifold, resulting in an increased HOMO-LUMO gap. The same behavior is observed for the other four bases as well.

As shown in Fig. 9, we can recognize five major orbital types among the valence and conduction orbitals of the isolated adenine molecule: σ_{ss}, σ_{sp}, n, π, and dipole-bound states. The lowest six states correspond to σ orbitals due to s-s hybridization, which have larger G_0W_0 exchange, correlation, and total self-energy corrections than the other states. The following ten states at higher energy levels exhibit σ_{sp} character, and their Σ_n^{X} and Σ_n^{C} show a linear but opposite dependence with respect to the G_0W_0 QP energy $\varepsilon_n^{G_0W_0}$. Thus, their sum Σ_n^{XC} is shown to be almost constant, ranging from -20.5 to -19.4 eV. Since the same trend is present in $\varepsilon_n^{\mathrm{XC}}$, the final difference Δ_n^{XC} between G_0W_0 and DFT results stays almost constant, between -3.3 and -3.0 eV. The

next three n and six π valence states and three π^* conduction states have a similar behavior, despite different magnitudes in their self-energy corrections. In particular, the six π valence states are lowered by about -2.5 eV, while the three π^* conduction states are lifted by 2.1 eV, leading to an increase of 4.6 eV for the HOMO-LUMO gap. The above observations provide an important evidence that the G_0W_0 self-energy corrections are highly orbital-dependent and on average $\Sigma^{\mathrm{X}}(\sigma_{ss}) < \Sigma^{\mathrm{X}}(\sigma_{sp}) < \Sigma^{\mathrm{X}}(n) < \Sigma^{\mathrm{X}}(\pi)$, $\Sigma^{\mathrm{C}}(\sigma_{ss}) > \Sigma^{\mathrm{C}}(\sigma_{sp}) > \Sigma^{\mathrm{C}}(n) > \Sigma^{\mathrm{C}}(\pi)$, and $\Delta^{\mathrm{XC}}(\sigma_{ss}) < \Delta^{\mathrm{XC}}(\sigma_{sp}) \approx \Delta^{\mathrm{XC}}(n) < \Delta^{\mathrm{XC}}(\pi)$. Consequently, the commonly-used "scissor operator" to correct bandgaps by rigidly lowering the valence levels and increasing the conduction levels by the same amount will never be adequate for describing the entire QP spectrum.

VII. SUMMARY

In summary, VIPs, VEAs and DOS of five DNA and RNA bases obtained from a fully-converged many-body G_0W_0 approach are found to be in very good agreement with experiments and other theoretical works. Two types of vertical electron affinities are found, corresponding to localized valence-bound excitations and delocalized dipole-bound excitations. Our calculations further reveal that QP inverse lifetimes depend linearly on QP energies for the deep valence states. They, however, come from the zero-th order G_0W_0 estimation, and may be significantly affected in self-consistent GW calculations. Interestingly, the G_0W_0 self-energy corrections are highly orbital-dependent, but remain relatively constant for the states with similar bonding character. Moreover, G_0W_0 VIPs of lone-pair states deviate from the experimental ones more than those for π states. Whether this difference comes from the different self-interaction errors in Kohn-Sham eigenstates will require further studies using self-interaction corrected functionals[39–41]; work is in progress along this direction.

[1] A. O. Colson and M. D. Sevilla, J. Phys. Chem. **99**, 3867 (1995).

[2] M. Zwolak and M. Di Ventra, Nano Lett. **5**, 421 (2005).

[3] D. Porath, A. Bezryadin, S. de Vries, and C. Dekker, Na-

ture **403**, 635 (2000).

[4] K. Kawai, H. Kodera, Y. Osakada, and T. Majima, Nat. Chem. **1**, 156 (2009).

[5] N. S. Hush and A. S. Cheung, Chem. Phys. Lett. **34**, 11

(1975).

[6] D. Dougherty, E. S. Younathan, R. Voll, S. Abdulnur, and S. P. McGlynn, J. Electron Spectrosc. Relat. Phenom. **13**, 379 (1978).

[7] K. W. Choi, J. H. Lee, and S. K. Kim, J. Am. Chem. Soc. **127**, 15674 (2005).

[8] A. B. Trofimov, J. Schirmer, V. B. Kobychev, A. W. Potts, D. M. P. Holland, and L. Karlsson, J. Phys. B-At. Mol. Opt. Phys. **39**, 305 (2006).

[9] M. Schwell, H. W. Jochims, H. Baumgartel, and S. Leach, Chem. Phys. **353**, 145 (2008).

[10] I. L. Zaytseva, A. B. Trofimov, J. Schirmer, O. Plekan, V. Feyer, R. Richter, M. Coreno, and K. C. Prince, J. Phys. Chem. A **113**, 15142 (2009).

[11] O. Kostko, K. Bravaya, A. Krylov, and M. Ahmed, Phys. Chem. Chem. Phys. **12**, 2860 (2010).

[12] N. Russo, M. Toscano, and A. Grand, J. Comput. Chem. **21**, 1243 (2000).

[13] D. Roca-Sanjuan, M. Rubio, M. Merchan, and L. Serrano-Andres, J. Chem. Phys. **125**, 084302 (2006).

[14] D. Roca-Sanjuan, M. Merchan, L. Serrano-Andres, and M. Rubio, J. Chem. Phys. **129**, 095104 (2008).

[15] K. B. Bravaya, O. Kostko, S. Dolgikh, A. Landau, M. Ahmed, and A. I. Krylov, J. Phys. Chem. A **114**, 12305 (2010).

[16] L. Hedin, Phys. Rev. **139**, A796 (1965).

[17] L. Hedin and S. Lundqvist, in *Solid State Physics, Advances in Research and Application*, edited by F. Seitz, D. Turnbull, and H. Ehrenreich (Academic Press, New York, 1969), vol. 23, pp. 1–181.

[18] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).

[19] H. N. Rojas, R. W. Godby, and R. J. Needs, Phys. Rev. Lett. **74**, 1827 (1995).

[20] M. M. Rieger, L. Steinbeck, I. D. White, H. N. Rojas, and R. W. Godby, Comput. Phys. Commun. **117**, 211 (1999).

[21] I. Campillo, J. M. Pitarke, A. Rubio, E. Zarate, and P. M. Echenique, Phys. Rev. Lett. **83**, 2230 (1999).

[22] C. D. Spataru, M. A. Cazalilla, A. Rubio, L. X. Benedict, P. M. Echenique, and S. G. Louie, Phys. Rev. Lett. **87**, 246405 (2001).

[23] G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. **74**, 601 (2002).

[24] N. Dori, M. Menon, L. Kilian, M. Sokolowski, L. Kronik, and E. Umbach, Phys. Rev. B **73**, 195208 (2006).

[25] M. L. Tiago, P. R. C. Kent, R. Q. Hood, and F. A. Reboredo, J. Chem. Phys. **129**, 084311 (2008).

[26] M. Palummo, C. Hogan, F. Sottile, P. Bagala, and A. Rubio, J. Chem. Phys. **131**, 084102 (2009).

[27] P. Umari, G. Stenuit, and S. Baroni, Phys. Rev. B **79**, 201104 (2009).

[28] P. Umari, G. Stenuit, and S. Baroni, Phys. Rev. B **81**, 115104 (2010).

[29] G. Stenuit, C. Castellarin-Cudia, O. Plekan, V. Feyer, K. C. Prince, A. Goldoni, and P. Umari, Phys. Chem. Chem. Phys. **12**, 10812 (2010).

[30] P. Umari, X. Qian, N. Marzari, G. Stenuit, L. Giacomazzi, and S. Baroni, Phys. Status Solidi (b) **248**, 527 (2011).

[31] C. Rostgaard, K. W. Jacobsen, and K. S. Thygesen, Phys. Rev. B **81**, 085103 (2010).

[32] X. Blase, C. Attaccalite, and V. Olevano, Phys. Rev. B **83**, 115103 (2011).

[33] C. Faber, C. Attaccalite, V. Olevano, E. Runge, and X. Blase, Phys. Rev. B **83**, 115123 (2011).

[34] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, et al., J. Phys.-Condes. Matter **21**, 395502 (2009), URL http://www.quantum-espresso.org.

[35] R. W. Godby, M. Schluter, and L. J. Sham, Phys. Rev. Lett. **56**, 2415 (1986).

[36] M. Gruning, A. Marini, and A. Rubio, Phys. Rev. B **74**, 161103 (2006).

[37] J. Lin, C. Yu, S. Peng, I. Akiyama, K. Li, L. K. Lee, and P. R. Lebreton, J. Phys. Chem. **84**, 1006 (1980).

[38] T. J. Odonnell, J. D. Petke, L. L. Shipman, and P. R. Lebreton, J. Phys. Chem. **84**, 1975 (1980).

[39] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[40] I. Dabo, A. Ferretti, N. Poilvert, Y. L. Li, N. Marzari, and M. Cococcioni, Phys. Rev. B **82**, 115121 (2010).

[41] T. Korzdorfer, J. Chem. Phys. **134**, 094111 (2011).