

# Gaussian Basis Sets for All-Electron Excited-State Calculations of Large Molecules

Rémi Pasquier,<sup>\*,†</sup> Maximilian Graml,<sup>\*,†</sup> and Jan Wilhelm<sup>\*</sup>Cite This: *J. Chem. Theory Comput.* 2026, 22, 540–557

Read Online

ACCESS |



Metrics &amp; More

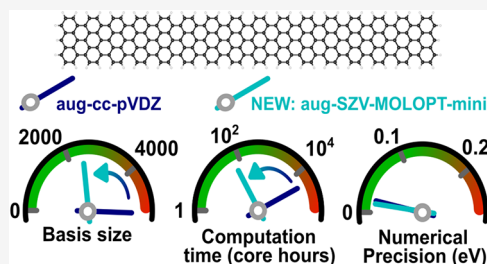


Article Recommendations



Supporting Information

**ABSTRACT:** We introduce a family of all-electron Gaussian basis sets, augmented MOLOPT, optimized for excited-state calculations on large molecules. We generate these basis sets by augmenting existing STO-3G, STO-6G, and MOLOPT basis sets optimized for ground state energy calculations. The augmented MOLOPT basis sets achieve fast convergence of GW gaps and Bethe–Salpeter excitation energies, while maintaining low condition numbers of the overlap matrix to ensure numerical stability. For GW HOMO–LUMO gaps, the double- $\zeta$  augmented MOLOPT basis yields a mean absolute deviation of 60 meV to the complete basis set limit. The basis set convergence for excitation energies from time-dependent density functional theory and the Bethe–Salpeter equation is similar. We use our smallest generated augmented MOLOPT basis (aug-SZV-MOLOPT-ae-mini) to demonstrate GW calculations on nanographenes with 9224 atoms requiring only 34300 core hours of computational resources.



## 1. INTRODUCTION

First-principles electronic structure calculations<sup>1</sup> are now widely employed across various fields, including the computation of electronic band structures of crystals and the investigation of reaction mechanisms in chemistry. A fundamental initial step in nearly all such calculations is specifying the atomic geometry, that is, the positions of the atomic nuclei. While this task is relatively straightforward for small molecules with a few atoms in the unit cell, it quickly becomes complex as the number of atoms increases. For instance, determining the atomic geometry of a liquid–solid interface can be a challenge. Recently, there has been a transformative shift in how atomic geometries are determined, driven by advances in machine learning. Techniques such as machine-learned interatomic potentials<sup>2</sup> and direct structure prediction approaches, like those used in protein folding,<sup>3</sup> are rapidly becoming standard tools in the field.

As a result, increasingly complex atomic structures, comprising 100,000 atoms or more,<sup>4</sup> are now available as starting points for first-principles calculations. A particularly interesting branch of these calculations is the study of electronically excited states.<sup>5–7</sup> Understanding these excitations is important for interpreting optical experiments, from conventional optical absorption spectroscopy to ultrafast phenomena induced by femtosecond laser pulses.<sup>8</sup> On the theoretical side, this poses a major challenge as first-principles methods for excited-state calculations are significantly more computationally demanding than those for ground-state properties.<sup>5,6</sup> The most widely used approaches for excited-state calculations include time-dependent density functional theory (TDDFT),<sup>9,10</sup> the GW approximation for quasiparticle energies, i.e., electron removal and addition energies,<sup>11–13</sup> and the GW plus Bethe–Salpeter equation (GW-BSE) for optical excitations.<sup>5,7</sup> All of

these methods have in common that the computational cost can quickly grow with the number of atoms in the molecule or unit cell, depending on the specific algorithm.

One approach to restrict this growth in computation time is the usage of low-scaling algorithms which often employ spatial locality. As an example, we consider the irreducible density response  $\chi^0(\mathbf{r}, \mathbf{r}')$ , which describes how the electron density at position  $\mathbf{r}$  changes in response to an external potential applied at position  $\mathbf{r}'$ .  $\chi^0(\mathbf{r}, \mathbf{r}')$  neglects the Coulomb interaction of the induced electron density and reflects the system's intrinsic nonlocal polarizability. In semiconductors and dielectrics,  $\chi^0(\mathbf{r}, \mathbf{r}')$  decays exponentially with increasing distance between  $\mathbf{r}$  and  $\mathbf{r}'$ , i.e.,  $\chi^0(\mathbf{r}, \mathbf{r}') \rightarrow 0$  as  $|\mathbf{r} - \mathbf{r}'| \rightarrow \infty$ , which is a sign of the spatial locality of the electronic structure, the "nearsightedness".<sup>14,15</sup>

Spatial locality can be exploited in the GW space-time algorithm,<sup>16,17</sup> which scales as  $O(N^3)$  with the number of atoms  $N$ , compared to the  $O(N^4)$  scaling of conventional frequency-based GW algorithms.<sup>13</sup> The space-time algorithm uses a real-space grid representation and switches between real-space and plane-wave bases. To retain the favorable scaling, key steps must be performed in real space; using plane waves throughout would increase scaling again to  $O(N^4)$ . Fast Fourier transforms (FFTs) enable efficient switching between representations but require

Received: August 19, 2025

Revised: December 2, 2025

Accepted: December 3, 2025

Published: December 22, 2025



an equidistant real-space grid. Such grids are not ideal for electronic structure calculations, as they fail to exploit the characteristic shapes of atomic orbitals, such as *s*-, *p*-, or *d*-type functions. We do need to mention recent advancements in the generation of nonequidistant real-space grids<sup>18–22</sup> which have reduced the number of grid points dramatically. Still, a compact atomic-orbital basis set is required for the generation of these real-space grids. Reformulating the GW space-time method in a localized atomic-orbital basis can significantly reduce matrix sizes, enabling efficient calculations for two-dimensional crystals<sup>23</sup> as well as large and complex systems.<sup>21,24–27</sup>

The computational effort of excited-state methods like GW, TDDFT and GW-BSE depends sensitively on the size of the atomic-orbital basis set. As an example, the computational cost of space-time GW<sup>24,26,27</sup> increases with the fourth power in the number of basis functions per atom. It is thus highly desirable to employ optimal atomic-orbital basis sets that provide converged excitation energies with a small number of basis functions. Atomic-orbital basis sets come in various forms, including numeric atom-centered orbitals (NAOs),<sup>28,29</sup> Slater-type orbitals (STOs),<sup>30</sup> and Gaussian-type orbitals (GTOs).<sup>31</sup> NAOs offer high flexibility and accuracy with compact sets, while STOs closely resemble atomic orbitals and also allow for small basis sizes. GTOs achieve radial flexibility via superposition of several Gaussians (“contractions”) and enable efficient evaluation of Coulomb integrals through analytical expressions.<sup>31–33</sup> This efficiency has made GTOs a standard in quantum chemistry software and motivates their use in this work.

Most GTO basis sets have been optimized for the computation of the ground state energy. However, when such basis sets are used in excited-state methods like GW-BSE, they yield a slow basis-set convergence. To cure this issue, one can add further Gaussian functions to ground-state optimized basis sets, as it is done in the aug-cc-pVXZ basis set family, X = D, T, Q, 5.<sup>34</sup> Here, the additional Gaussians are optimized to match the LUMO wave function via optimization of the total energy of charged atoms. In this way, the additional Gaussians describe electronically excited states, which often involve the excitation from occupied orbitals to LUMO. The aug-cc-pVXZ basis sets yield good accuracy for excitation energies of small molecules, but their application to large systems is severely limited by numerical issues mainly related to the inclusion of very diffuse Gaussian functions, i.e., those with very small exponents that decay slowly. As a result, the condition number of the overlap matrix is large,<sup>35</sup> leading to numerical instability and convergence problems in the self-consistent-field iterations of large molecules. Consequently, the use of aug-cc-pVXZ is typically prohibitive for large molecules. This presents a gap in the current methodology: while Gaussian basis sets optimized for excited-state calculations of small molecules exist, Gaussian basis sets optimized for excited-state calculations of large, complex molecular systems are lacking.

In contrast to excited states, Gaussian basis sets tailored for computing the DFT ground-state energy in large molecular systems already exist. A prominent class of such basis sets are the MOLOPT-type basis sets, which were specifically designed to balance accuracy of DFT ground-state energy calculations with numerical stability. Not only the accuracy of ground-state energies of molecules<sup>35–38</sup> has been optimized, but also the condition number of the overlap matrix has been minimized, which is critical for ensuring numerically stable DFT calculations in extended and complex systems.

Despite their success in ground-state calculations, the currently available MOLOPT basis sets<sup>35–38</sup> do not offer a sufficiently accurate description of electronically excited states. In this work, we address this limitation by augmenting existing all-electron MOLOPT basis sets<sup>37,38</sup> with additional diffuse Gaussian functions, which are optimized to reproduce excitation energies from GW-BSE calculations performed with a large reference basis set (aug-cc-pV5Z). As a result, we introduce the all-electron (ae) aug-MOLOPT-ae basis set family containing aug-SZV-MOLOPT-ae, aug-DZVP-MOLOPT-ae, and aug-TZVP-MOLOPT-ae. Our basis sets cover the elements of periods I, II, and III (H to Cl) and are specifically designed for accurate and efficient excited-state calculations in large molecular systems.

The article is organized as follows: In Section 2.1, we give an overview of Kohn–Sham DFT in a Gaussian basis, where numerical instabilities due to the inverse overlap matrix can arise, as discussed in detail in Section 2.2. A theoretical perspective on basis set convergence for quasiparticle and excitation energies is provided in Section 2.3. We further provide the procedure for generating the augmented MOLOPT basis sets in Section 2.4. Benchmark results on HOMO–LUMO gaps from PBE0 and GW as well as excitation energies from BSE and TDDFT for the augmented MOLOPT basis sets are given in Sections 2.5 and 2.6, respectively. The procedure to generate the associated auxiliary RI basis sets is presented in Section 3.1. In Section 3.2, we evaluate how auxiliary RI basis size and Coulomb cutoff affect the accuracy in low-scaling GW calculations. Large-scale applicability of the augmented MOLOPT basis sets is demonstrated in Section 4, where we perform GW calculations on nanographenes with over 9000 atoms. We describe the molecular test set for our benchmark and the computational details in App. A. In App. B, we describe a memory-saving scheme for the computation of the self-energy using a repeated calculation of three-center integrals. Finally, in App. C, we carry out a test calculation of our new basis sets on a representative system (9,10-Dihydroanthracene), in order to quantitatively assess the quality of these basis sets with respect to their size.

## 2. ORBITAL BASIS SETS

**2.1. Expansion of Kohn–Sham Orbitals in a Gaussian Basis Set.** Many excited-state calculations of large molecules start from Kohn–Sham (KS) density functional theory (or Hartree–Fock theory),<sup>39,40</sup> where one needs to solve the KS equations

$$\left[ -\frac{\nabla^2}{2} + V_{\text{eff}}(\mathbf{r}) \right] \psi_n(\mathbf{r}) = \varepsilon_n \psi_n(\mathbf{r}) \quad (1)$$

where  $\psi_n(\mathbf{r})$  are the KS orbitals and  $\varepsilon_n$  are the KS eigenvalues. The effective potential  $V_{\text{eff}}(\mathbf{r})$  includes the external potential originating from the Coulomb potential of the nuclei, the Hartree potential, and the exchange–correlation potential.

To solve the KS equations (1) numerically, each KS orbital is expanded as a linear combination of predefined basis functions. When a Gaussian basis set is used, each orbital  $\psi_n(\mathbf{r})$  is written as a sum over basis functions  $\phi_\mu(\mathbf{r})$  of the molecule (or unit cell<sup>29</sup>) with expansion coefficients  $C_{\mu n}$

$$\psi_n(\mathbf{r}) = \sum_{\mu=1}^{N_{\text{bf}}} C_{\mu n} \phi_\mu(\mathbf{r}) \quad (2)$$

$N_{\text{bf}}$  is the number of basis functions. The coefficients  $C_{\mu n}$  are unknown and determined by solving the KS equations (1). More specifically, the KS equations (1) are reformulated by inserting the basis expansion (2) into the Kohn–Sham equations (1), then multiplied with a basis function  $\phi_\nu(\mathbf{r})$  and integrated over the whole space to obtain the matrix equations

$$\sum_\nu h_{\mu\nu} C_{\nu n} = \sum_\nu S_{\mu\nu} C_{\nu n} \epsilon_n \quad (3)$$

known as the Roothaan–Hall equations.<sup>40</sup> Here,  $h_{\mu\nu}$  is the Kohn–Sham matrix and the overlap matrix is defined as

$$S_{\mu\nu} = \int d\mathbf{r} \phi_\mu(\mathbf{r}) \phi_\nu(\mathbf{r}) \quad (4)$$

Equation 3 is a generalized eigenvalue problem because the basis functions  $\{\phi_\mu\}$  are nonorthogonal for molecules with more than a single atom, i.e.,  $\mathbf{S} \neq \text{Id}$ . To solve eq 3 for  $C_{\nu n}$  and  $\epsilon_n$ , one usually transforms eq 3 into a standard eigenvalue problem

$$\sum_\nu \tilde{h}_{\mu\nu} \tilde{C}_{\nu n} = \tilde{C}_{\mu n} \epsilon_n \quad (5)$$

by using the following transformations:

$$\tilde{\mathbf{h}} = \mathbf{S}^{-1/2} \mathbf{h} \mathbf{S}^{-1/2}, \mathbf{C} = \mathbf{S}^{-1/2} \tilde{\mathbf{C}} \quad (6)$$

The procedure is to first compute  $\tilde{\mathbf{h}} = \mathbf{S}^{-1/2} \mathbf{h} \mathbf{S}^{-1/2}$ , followed by the diagonalization of  $\tilde{\mathbf{h}}$  to obtain  $\tilde{\mathbf{C}}$  via eq 5, and  $\mathbf{C} = \mathbf{S}^{-1/2} \tilde{\mathbf{C}}$  to obtain the expansion coefficients  $C_{\mu n}$ .

A Gaussian-type basis function  $\phi_\mu(\mathbf{r})$  used in eq 2 is centered at an atom  $A$  and is a linear combination of Gaussian functions multiplied with a spherical harmonic  $Y_{l,\mu}^m$ .<sup>40,41</sup>

$$\phi_\mu(\mathbf{r}) = Y_{l_\mu}^{m_\mu}(\theta_A, \varphi_A) r_A^{l_\mu} \sum_{i=1}^{N_{\text{prim}}} \alpha_{\mu,i} \exp(-\beta_{\mu,i} r_A^2) \quad (7)$$

where  $\mathbf{r}_A = \mathbf{r} - \mathbf{R}_A$  is the displacement vector to nucleus  $A$  located at position  $\mathbf{R}_A$ ,  $(\theta_A, \varphi_A)$  are the polar angles of  $\mathbf{r}_A$  and  $r_A = |\mathbf{r}_A|$ .  $l_\mu$  is the angular momentum quantum number and  $m_\mu$  the magnetic quantum number of  $\phi_\mu$ .  $\alpha_{\mu,i}$  are the contraction coefficients,  $\beta_{\mu,i}$  the Gaussian exponents and  $N_{\text{prim}}$  the number of Gaussian primitives. The parameters  $l_\mu$ ,  $m_\mu$ ,  $\{\alpha_{\mu,i}\}_{i=1}^{N_{\text{prim}}}$  and  $\{\beta_{\mu,i}\}_{i=1}^{N_{\text{prim}}}$  entering eq 7 are determined and fixed before the KS-DFT calculation. A Gaussian basis set  $B^A$  of atom  $A$  is defined as a finite set

$$B^A = \{\phi_\mu(\mathbf{r})\}_{\mu=1}^{N_{\text{bf}}^A} \quad (8)$$

containing  $N_{\text{bf}}^A$  Gaussian-type basis functions  $\phi_\mu(\mathbf{r})$  from eq 7, all centered at atom  $A$ .

**2.2. Numerical Instability Computing  $\mathbf{S}^{-1/2}$  and the Condition Number of  $\mathbf{S}$ .** While Gaussian basis sets offer powerful flexibility by tuning contraction coefficients  $\alpha$  and exponents  $\beta$ , eq 7, the inclusion of diffuse Gaussian functions, i.e., those with small exponents, can introduce serious numerical challenges. These diffuse functions are often essential for accurately capturing excited-state quantities, because empty KS orbitals are usually more diffuse than occupied KS orbitals. Diffuse Gaussians decay slowly and exhibit significant spatial overlap even across distant atoms. As a result, the overlap matrix  $\mathbf{S}$  becomes increasingly ill-conditioned, with eigenvalues that span several orders of magnitude. This poor conditioning leads to numerical instability when computing  $\mathbf{S}^{-1/2}$  for transforming the generalized KS eigenvalue problem (3) into a standard one (6). Such instabilities are particularly challenging in large

molecules which can lead to convergence issues of the self-consistent-field (SCF) cycle. In the following, we analyze this numerical instability in detail.

To illustrate the numerical instability of computing  $\mathbf{S}^{-1/2}$  introduced by diffuse Gaussian basis functions, we consider the minimal example of the hydrogen molecule,  $\text{H}_2$ . Each atom has a single  $s$ -type Gaussian basis function with identical exponent  $\beta$ , i.e., the two basis functions of the molecule read

$$\phi_{1,2}(\mathbf{r}) = (2\beta/\pi)^{3/4} \exp(-\beta[(x \pm d/2)^2 + y^2 + z^2]) \quad (9)$$

where  $d$  is the distance between both atoms. Both basis functions  $\phi_{1,2}(\mathbf{r})$  are normalized, such that the diagonal elements of the overlap matrix are equal to one,  $S_{11} = \int d\mathbf{r} \phi_1^2(\mathbf{r}) = S_{22} = \int d\mathbf{r} \phi_2^2(\mathbf{r}) = 1$ . We further have  $S_{12} = S_{21} = \int d\mathbf{r} \phi_1(\mathbf{r}) \phi_2(\mathbf{r}) = e^{-\beta d^2/2}$ . The eigenvalues of the  $2 \times 2$  overlap matrix  $\mathbf{S}$  are then

$$s_{1,2} = 1 \pm e^{-\beta d^2/2} \quad (10)$$

For a very diffuse Gaussian with  $\beta = 10^{-3}/a_0^2$  and bond distance  $d = 1.4 a_0$ , we have  $s_1 = 1.9990$  and  $s_2 = 9.8 \cdot 10^{-4}$ . Note that  $\mathbf{S}$  of any molecule is positive semidefinite, so  $s_i \geq 0$ . The condition number of  $\mathbf{S}$  then is  $\kappa(\mathbf{S}) = 2041$ , computed from

$$\kappa(\mathbf{S}) = \frac{\max s_i}{\min s_i} \quad (11)$$

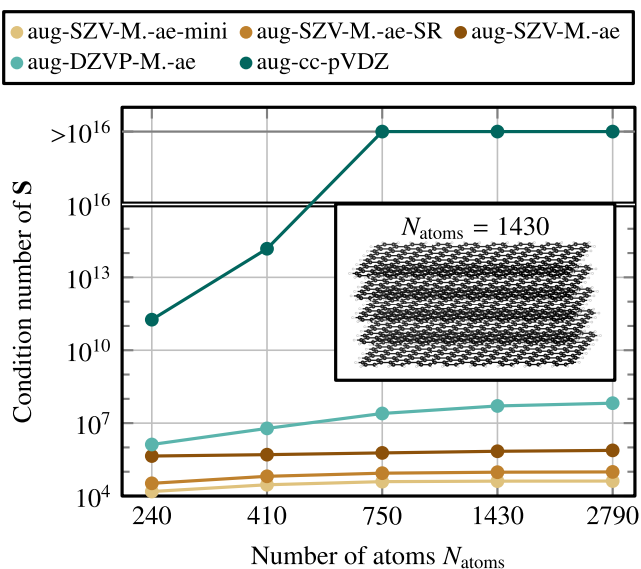
For computing  $\mathbf{S}^{-1/2}$  required in eq 6, one needs to compute  $1/\sqrt{s_i}$ , which gets increasingly large for decreasing  $s_i$ . This introduces numerical instability, which is quantified by the condition number  $\kappa(\mathbf{S})$ . The numerical example of  $\kappa(\mathbf{S}) = 2041$  for  $\text{H}_2$  with two diffuse Gaussians demonstrates how even a small molecule with only two diffuse basis functions can lead to poor conditioning.

Numerical instabilities arise if  $\kappa(\mathbf{S})$  hits the inverse machine precision. For double precision arithmetic, machine precision is  $2^{-52} \approx 10^{-16}$  and thus numerical instabilities arise if  $\kappa(\mathbf{S}) \gtrsim 10^{16}$ . Several numerical tricks have been used to circumvent these instabilities related to large  $\kappa(\mathbf{S})$ , for example the removal of small eigenvalues of the overlap matrix.<sup>35</sup> In our experience, we have observed that this technique can help to certain extent, but for too large condition numbers, the SCF cycle fails to converge nevertheless. One of the possible reasons for this behavior is that the eigendecomposition of the overlap matrix can become highly unstable for overcomplete basis sets,<sup>42</sup> leading to an unreliable regularization and thus SCF calculation.

$\kappa(\mathbf{S})$  increases rapidly with additional diffuse functions per atom and for systems with more atoms. We illustrate the effect of the molecular size on the condition number (11) in Figure 1: When adding more atoms to the system,  $\kappa(\mathbf{S})$  increases by orders of magnitude for the commonly used Gaussian aug-cc-pVDZ basis set (dark green curve),<sup>43,44</sup> reaching values above the inverse machine precision,  $\kappa(\mathbf{S}) > 10^{16}$ . In contrast, for all four augmented MOLOPT basis sets presented later in this work, the condition number remains well below this threshold, even in the limit of infinite system size (bulk graphite).

**2.3. Basis Set Convergence for Excitation Energies of Charged and Neutral Excitations.** As discussed in the last Section 2.2, an essential requirement for basis sets used in excited-state electronic structure methods is numerical stability, ensured for example by keeping the condition number of the overlap matrix  $\mathbf{S}$  sufficiently low. Equally important is that the basis sets enable fast convergence of excited-state energies with





**Figure 1.** Condition number  $\kappa(S)$  of the overlap matrix for a finite graphite-like cutout computed from eq 11 using five different Gaussian basis sets. The cutout with 1430 atoms is shown in the inset; we vary its horizontal length, and the corresponding number of atoms is plotted on the horizontal axis. For the aug-cc-pVDZ basis set,<sup>43,44</sup>  $\kappa(S)$  exceeds the inverse machine precision ( $\sim 10^{16}$ ) for cutouts with 750 atoms and more, leading convergence issues in the SCF. We also show the condition number of the four augmented MOLOPT basis sets (aug-SZV-MOLOPT-ae-mini, aug-SZV-MOLOPT-ae-SR, aug-SZV-MOLOPT-ae, aug-DZVP-MOLOPT-ae) developed in this work, where the condition number remains well below this threshold.

respect to the basis set size. In this section, we analyze the convergence behavior of two types of excitations: (i) charged excitations, corresponding to quasiparticle (QP) energies as obtained from GW, and (ii) charge-neutral excitations, such as those calculated from BSE or TDDFT.<sup>13</sup>

**2.3.1. Charged Excitations: GW Quasiparticle Energies.** It is well established that absolute GW quasiparticle energies  $\epsilon_n$  converge slowly with increasing basis set size  $N_{\text{bf}}$ <sup>45,46</sup> and thus require basis set extrapolation.<sup>47,48</sup> However, energy differences between QP states, such as the HOMO–LUMO gap, converge much faster.<sup>26,46</sup> We rationalize this behavior in this section, starting by dividing the QP energies into ionization potentials (IPs) and electron affinities (EAs):<sup>13</sup>

$$\begin{aligned}\epsilon_n &= E_N^0 - E_{N-1}^n \text{ if } \epsilon_n \leq \epsilon_F \text{ (IPs)} \\ \epsilon_n &= E_{N+1}^n - E_N^0 \text{ if } \epsilon_n > \epsilon_F \text{ (EAs)}\end{aligned}\quad (12)$$

Here,  $E_{N\pm 1}^n$  is the  $n$ th excited state energy of the  $N \pm 1$  electron system and  $\epsilon_F$  is the Fermi energy.

According to the second Hohenberg–Kohn theorem,<sup>49</sup> the ground state energy  $E_N^0$  of KS-DFT converges from above toward the complete-basis-set (CBS) limit as the basis size increases, i.e.,

$$E_N^0 \equiv \lim_{N_{\text{bf}} \rightarrow \infty} E_N^0(N_{\text{bf}}) \leq E_N^0(N_{\text{bf}}) \quad (13)$$

We make this assumption also for the excited-state energies  $E_N^n$  and  $E_{N\pm 1}^n$ , following the Hylleraas–Undheim–MacDonald theorem:<sup>30,51</sup>

$$E_N^n \equiv \lim_{N_{\text{bf}} \rightarrow \infty} E_N^n(N_{\text{bf}}) \leq E_N^n(N_{\text{bf}}) \quad (14)$$

We assume the deviation from the complete basis set limit is extensive in the number of electrons, i.e., it scales linearly with the number of electrons, and it is independent of excitation index  $n$ . This behavior can be expressed as

$$E_N^0(N_{\text{bf}}) = E_N^0 + N \cdot f(N_{\text{bf}}) \quad (15)$$

$$E_N^n(N_{\text{bf}}) = E_N^n + N \cdot f(N_{\text{bf}}) \quad (16)$$

$$E_{N\pm 1}^n(N_{\text{bf}}) = E_{N\pm 1}^n + (N \pm 1) \cdot f(N_{\text{bf}}) \quad (17)$$

with

$$f(N_{\text{bf}}) \geq 0 \text{ and } \lim_{N_{\text{bf}} \rightarrow \infty} f(N_{\text{bf}}) = 0 \quad (18)$$

Combining eqs 12, 15, 17, and 18 implies that the absolute QP energy levels (i.e., IPs and EAs) converge slowly from above with increasing basis set size:

$$\epsilon_n(N_{\text{bf}}) = \epsilon_n + f(N_{\text{bf}}) \geq \epsilon_n \equiv \lim_{N_{\text{bf}} \rightarrow \infty} \epsilon_n(N_{\text{bf}}) \quad (19)$$

In contrast, when using the simplified model (19) for differences of QP energy levels,  $\epsilon_n(N_{\text{bf}}) - \epsilon_m(N_{\text{bf}})$ , they are independent of  $f(N_{\text{bf}})$ :

$$\epsilon_n(N_{\text{bf}}) - \epsilon_m(N_{\text{bf}}) = \epsilon_n - \epsilon_m \equiv \lim_{N_{\text{bf}} \rightarrow \infty} [\epsilon_n(N_{\text{bf}}) - \epsilon_m(N_{\text{bf}})] \quad (20)$$

This analysis suggests that the convergence of QP energy differences is significantly faster than that of absolute QP energies, provided that the basis set error satisfies the convergence forms given in eq 15–17, at least approximately. Finally, we assume that GW QP energies are good approximations to the QP energies (12) such that this analysis carries over to GW QP energies.

**2.3.2. Charge-Neutral Excitations: TDDFT and BSE Excitation Energies.** For charge-neutral excitations one considers the energy difference  $\Delta E_n$  between the ground state and the electronically excited state  $n$ , both with  $N$  electrons:

$$\Delta E_n = E_N^n - E_N^0 \quad (21)$$

Again, assuming the convergence form (16), we see that the function  $f(N_{\text{bf}})$  cancels out for the basis set convergence of excitation energies,

$$\Delta E_n(N_{\text{bf}}) = \Delta E_n \equiv \lim_{N_{\text{bf}} \rightarrow \infty} \Delta E_n(N_{\text{bf}}) \quad (22)$$

where again  $\Delta E_n$  is the excitation energy in the CBS limit.

Summarizing, we can expect relatively fast convergence of BSE excitation energies and GW QP energy differences like the GW HOMO–LUMO gap with the basis set size, given that the basis sets are well optimized, while absolute values of GW QP energy levels are hard to converge with the basis set size.

**2.4. Basis Set Generation Recipe.** A Gaussian basis set (8) for an element  $A$  is constructed by specifying the total number of basis functions  $N_{\text{bf}}^A$ , selecting the number of functions  $\phi_\mu(\mathbf{r})$  for each angular momentum quantum number  $l$ . Each  $\phi_\mu(\mathbf{r})$  consists of a linear combination (7) of primitive Gaussians characterized by exponential decay parameters  $\beta_{\mu,i}$  and contraction coefficients  $\alpha_{\mu,i}$ . One motivation for this contraction scheme is to better approximate Slater-type orbitals, which decay exponentially as  $\exp(-\zeta r)$  and represent the shape of atomic orbitals in the hydrogen-like model.<sup>31</sup> The parameters of each contracted function  $\phi_\mu(\mathbf{r})$  are then optimized to reproduce one or more atomic or molecular properties. These may include,

**Table 1.** Composition of Gaussian Basis Sets: Dunning's Augmented Correlation-consistent Basis Sets Include aug-cc-pVDZ, aug-cc-pVTZ and aug-cc-pVQZ<sup>a</sup>

|         | basis              | composition        | $N_{bf}$ | root basis (rb)              | rb composition     | augmentation       | min exponent $\beta$ (au) |
|---------|--------------------|--------------------|----------|------------------------------|--------------------|--------------------|---------------------------|
| H, He   | aug-cc-pVDZ        | 3s, 2p             | 9        | cc-pVDZ <sup>43,44</sup>     | 2s, 1p             | 1s, 1p             | 0.030 (H), 0.072 (He)     |
| Li–Ne   | aug-cc-pVDZ        | 4s, 3p, 2d         | 23       | cc-pVDZ <sup>43,53</sup>     | 3s, 2p, 1d         | 1s, 1p, 1d         | 0.006 (Li)–0.106 (Ne)     |
| Na–Cl   | aug-cc-pVDZ        | 5s, 4p, 2d         | 27       | cc-pVDZ <sup>53,54</sup>     | 4s, 3p, 1d         | 1s, 1p, 1d         | 0.006 (Na)–0.047 (Cl)     |
| H, He   | aug-cc-pVTZ        | 4s, 3p, 2d         | 23       | cc-pVTZ <sup>43,44</sup>     | 3s, 2p, 1d         | 1s, 1p, 1d         | 0.025 (H), 0.052 (He)     |
| Li–Ne   | aug-cc-pVTZ        | 5s, 4p, 3d, 2f     | 46       | cc-pVTZ <sup>43,53</sup>     | 4s, 3p, 2d, 1f     | 1s, 1p, 1d, 1f     | 0.008 (Li)–0.092 (Ne)     |
| Na–Cl   | aug-cc-pVTZ        | 6s, 5p, 3d, 2f     | 50       | cc-pVTZ <sup>53,54</sup>     | 5s, 4p, 2d, 1f     | 1s, 1p, 1d, 1f     | 0.007 (Na)–0.042 (Cl)     |
| H, He   | aug-cc-pVQZ        | 5s, 4p, 3d, 2f     | 46       | cc-pVQZ <sup>43,44</sup>     | 4s, 3p, 2d, 1f     | 1s, 1p, 1d, 1f     | 0.024 (H), 0.048 (He)     |
| Li–Ne   | aug-cc-pVQZ        | 6s, 5p, 4d, 3f, 2g | 80       | cc-pVQZ <sup>43,53</sup>     | 5s, 4p, 3d, 2f, 1g | 1s, 1p, 1d, 1f, 1g | 0.006 (Li)–0.082 (Ne)     |
| Na–Cl   | aug-cc-pVQZ        | 7s, 6p, 4d, 3f, 2g | 84       | cc-pVQZ <sup>53,54</sup>     | 6s, 5p, 3d, 2f, 1g | 1s, 1p, 1d, 1f, 1g | 0.005 (Na)–0.038 (Cl)     |
| H, He   | aug-SZV-M.-ae-mini | 3s, 1p             | 6        | STO-3G <sup>31</sup>         | 1s                 | 2s, 1p             | 0.065 (H), 0.090 (He)     |
| Li–Ne   | aug-SZV-M.-ae-mini | 3s, 2p             | 9        | STO-3G <sup>31,55</sup>      | 2s, 1p             | 1s, 1p             | 0.048 (Li)–0.200 (Ne)     |
| Na–Cl   | aug-SZV-M.-ae-mini | 4s, 3p             | 13       | STO-3G <sup>55</sup>         | 3s, 2p             | 1s, 1p             | 0.050 (Na)–0.080 (Cl)     |
| H       | aug-SZV-M.-ae-SR   | 3s, 1p             | 6        | STO-3G <sup>31</sup>         | 1s                 | 2s, 1p             | 0.065 (H)                 |
| C, N, O | aug-SZV-M.-ae-SR   | 3s, 2p, 1d         | 14       | STO-3G <sup>31</sup>         | 2s, 1p             | 1s, 1p, 1d         | 0.115 (C)–0.162 (O)       |
| H, He   | aug-SZV-M.-ae      | 3s, 1p             | 6        | STO-6G <sup>31</sup>         | 1s                 | 2s, 1p             | 0.065 (H), 0.090 (He)     |
| Li–Ne   | aug-SZV-M.-ae      | 3s, 2p, 1d         | 14       | STO-6G <sup>31,55</sup>      | 2s, 1p             | 1s, 1p, 1d         | 0.031 (Li)–0.200 (Ne)     |
| Na–Cl   | aug-SZV-M.-ae      | 4s, 3p, 1d         | 18       | STO-6G <sup>55</sup>         | 3s, 2p             | 1s, 1p, 1d         | 0.050 (Na)–0.077 (Cl)     |
| H, He   | aug-DZVP-M.-ae     | 3s, 2p             | 9        | SVP-M.-ae <sup>37,38</sup>   | 2s, 1p             | 1s, 1p             | 0.035 (H), 0.060 (He)     |
| Li–Ne   | aug-DZVP-M.-ae     | 4s, 3p, 2d, 1f     | 30       | SVP-M.-ae <sup>37,38</sup>   | 3s, 2p, 1d         | 1s, 1p, 1d, 1f     | 0.025 (Li)–0.100 (Ne)     |
| Na–Cl   | aug-DZVP-M.-ae     | 5s, 4p, 2d, 1f     | 34       | SVP-M.-ae <sup>37,38</sup>   | 4s, 3p, 1d         | 1s, 1p, 1d, 1f     | 0.045 (Na)–0.077 (Cl)     |
| H, He   | aug-TZVP-M.-ae     | 4s, 3p, 2d         | 23       | TZVPP-M.-ae <sup>37,38</sup> | 3s, 2p, 1d         | 1s, 1p, 1d         | 0.030 (H), 0.050 (He)     |
| Li–Ne   | aug-TZVP-M.-ae     | 6s, 4p, 3d, 2f, 1g | 56       | TZVPP-M.-ae <sup>37,38</sup> | 5s, 2p, 2d, 1f     | 1s, 1p, 1d, 1f, 1g | 0.025 (Li)–0.100 (Ne)     |
| Na–Cl   | aug-TZVP-M.-ae     | 6s, 6p, 4d, 2f, 1g | 67       | TZVPP-M.-ae <sup>37,38</sup> | 5s, 5p, 3d, 1f     | 1s, 1p, 1d, 1f, 1g | 0.025 (Na)–0.130 (Cl)     |

<sup>a</sup>The basis sets developed in this work are aug-SZV-MOLOPT-ae-mini, aug-SZV-MOLOPT-ae-SR, aug-SZV-MOLOPT-ae, aug-DZVP-MOLOPT-ae and aug-TZVP-MOLOPT-ae.

for instance, the correlation energy of neutral atoms<sup>43</sup> or negatively charged atoms<sup>34</sup> or the ground state energy molecules from DFT with the PBE functional.<sup>35</sup>

By increasing the number of basis functions  $N_{bf}^A$  of the basis set, one can construct hierarchical families of basis sets with systematic improvements. The cc-pVXZ basis set family,  $X = D, T, Q, \dots$ <sup>43</sup> is designed to systematically improve the correlation energy of molecules obtained from post-Hartree–Fock methods like MP2,<sup>40</sup> the random phase approximation<sup>52</sup> or coupled cluster.<sup>40</sup> The aug-cc-pVXZ basis set family,  $X = D, T, Q, \dots$ <sup>34</sup> is constructed to yield systematic improvements of electron affinities of molecules computed from post-Hartree–Fock methods. In contrast, the XZVP-MOLOPT basis set family,  $X = S, D, T, Q$ <sup>35,37,38</sup> is designed to obtain basis-set converged groundstate DFT calculations of large molecules, crystals, liquids and material interfaces. For a more complete review of Gaussian basis sets, we refer to ref<sup>40</sup>

We recall the strategy behind the construction of Dunning's aug-cc-pVXZ basis sets<sup>34,56</sup> as summarized in Table 1, to motivate our approach of constructing basis sets. The aug-cc-pVXZ basis sets are designed to compute electron affinities, which requires accurate description of the lowest unoccupied molecular orbital (LUMO) in KS-DFT. The LUMO is typically much more diffuse than occupied orbitals as indicated by the decay length  $\zeta_n \approx \hbar / \sqrt{2m|\epsilon_n|}$  of a molecular orbital outside of the molecule;  $\epsilon_n$  is the orbital energy of a bound state and  $m$  the electron mass. For occupied orbitals, we usually have  $|\epsilon_n| < 5$  eV while  $|\epsilon_{LUMO}| \approx 0$  eV leads to a long decay length  $\zeta_{LUMO}$ , indicating a diffuse LUMO. Standard cc-pVXZ basis sets lack the necessary diffuse functions to describe such orbitals. To address this, aug-cc-pVXZ adds an uncontracted diffuse Gaussian  $\phi_\mu(\mathbf{r}) = Y_l^m(\theta_A, \varphi_A) r_A^l \alpha_l \exp(-\beta_l r_A^2)$  for each angular momentum  $l$  present in cc-pVXZ. The exponent  $\beta_l$  is optimized to match the

correlation energy of the corresponding anion at the complete-basis-set limit. These augmented basis sets have proven effective for excited-state properties, such as GW HOMO–LUMO gaps and TDDFT or GW+BSE excitation energies.<sup>46,57,58</sup>

While aug-cc-pVXZ basis sets yield good accuracy for excitation energies of small and medium-sized molecules, their application to large systems is severely limited by numerical issues. The inclusion of very diffuse Gaussians results in a large condition number of the overlap matrix  $\mathbf{S}$ , making the computation (6) of  $\mathbf{S}^{-1/2}$  numerically unstable, as discussed in Section 2.2. This instability often leads to convergence problems of the self-consistent-field cycle when treating large molecules. Consequently, the use of aug-cc-pVXZ is usually prohibitive for large molecules.

We develop a family of augmented all-electron (ae) MOLOPT basis sets specifically targeted for excited-state calculations of large molecules, following the analysis of numerical stability from Section 2.2 and basis set convergence from Section 2.3. We apply an augmentation strategy to the basis sets STO-3G,<sup>31</sup> STO-6G,<sup>31</sup> SVP-MOLOPT-ae,<sup>37,38</sup> and TZVPP-MOLOPT-ae,<sup>37,38</sup> resulting in the aug-SZV-MOLOPT-mini-ae, aug-SZV-MOLOPT-ae-SR, aug-SZV-MOLOPT-ae, aug-DZVP-MOLOPT-ae, and aug-TZVP-MOLOPT-ae basis sets, respectively, see Table 1. The "mini" bases correspond to smaller, more compact versions of the corresponding regular basis sets, and are therefore well-suited for intensive calculations in larger systems with a small cost in accuracy, whereas the "SR" (short-range) basis sets have been generated with less diffuse primitives and are therefore intended to reduce the computational cost in condensed phase systems at a similar accuracy. As such, we expect the following accuracy hierarchy to hold: aug-SZV-MOLOPT-ae-mini < aug-SZV-MOLOPT-ae-SR < aug-DZVP-MOLOPT-ae < aug-TZVP-

|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|---------------------------------|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| C aug-SZV-MOLOPT-ae-mini        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2                               |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1 0 0 3 1                       |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 71.617 0.154                    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 13.045 0.535                    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 3.531 0.445                     |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 0 1 4 2 2                     |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 0.115 0.000 0.008 0.000 -1.199  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 0.222 0.700 4.256 0.392 -2.375  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 0.683 0.400 -0.565 0.608 0.392  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2.941 -0.100 0.000 0.156 -0.022 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|                                 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**Figure 2.** aug-SZV-MOLOPT-ae-mini, aug-SZV-MOLOPT-ae-SR, aug-SZV-MOLOPT-ae and aug-DZVP-MOLOPT-ae basis sets developed in this work for carbon in CP2K basis set format.<sup>56,61</sup> Numbers marked in red have been optimized to match BSE excitation energies of the molecules contained in Thiel's set. The black numbers are the parameters taken from STO-3G<sup>31</sup> (for aug-SZV-MOLOPT-ae-SR), STO-6G<sup>31</sup> (for aug-SZV-MOLOPT-ae) and from SVP-MOLOPT-PBE-ae<sup>37,38</sup> (for aug-DZVP-MOLOPT-ae). The basis sets for other atoms and all corresponding auxiliary RI basis sets are listed in the Supporting Information S3, S4.

MOLOPT-ae. We add one additional angular momentum shell beyond  $l_{\max}$  of the root basis, and introduce one new diffuse Gaussian primitive with a smaller exponent per angular momentum to improve radial flexibility. One should therefore note that we are using the term *augmentation* in a broader way than the usual meaning that only involves diffuse functions.<sup>34</sup> The name was chosen to obtain a compact and practical name for the new family of basis sets and underline that these are built on and retain the stability of the MOLOPT basis set, but with better excited-state properties. We optimize the added basis functions to reproduce the lowest five  $G_0W_0$ -BSE@PBE0 excitation energies calculated with the aug-cc-pV5Z basis set for Thiel's set<sup>59</sup> (for elements H, C, N, O) and the available molecular set from ref 35 for the other elements up to chlorine. In the objective function, we also ensure that the condition number of the basis set stays limited, as in previous optimizations of Gaussian basis sets.<sup>35,36</sup> In practice, the optimization is performed using Powell's algorithm,<sup>60</sup> which is a local optimizer and may therefore converge to a local minimum. However, since the optimization in any case depends on the choice of the rather small molecular training set, reaching the global minimum is not a strict requirement to generate high-quality basis sets. Instead, we carefully validate the resulting basis functions on a much larger benchmark set containing 247 molecules, ensuring their general reliability.

As an example, the STO-3G basis set of carbon contains five basis functions, two  $s$ -functions ( $l = 0, m = 0$ ) and one  $p$ -function ( $l = 1, m = -1, 0, 1$ )

$$\begin{aligned}\phi_1(\mathbf{r}) &= \sum_{i=1}^3 \alpha_{i,1} \exp(-\beta_i r_C^2) \\ \phi_2(\mathbf{r}) &= \sum_{i=1}^3 \alpha_{i,2} \exp(-\gamma_i r_C^2) \\ \phi_{3,4,5}(\mathbf{r}) &= Y_1^{-1,0,1}(\theta_C, \varphi_C) r_C \sum_{i=1}^3 \alpha_{i,3} \exp(-\gamma_i r_C^2)\end{aligned}\quad (23)$$

where  $\alpha_{ij}$  can be found in the literature<sup>31,56</sup> and the exponents are

$$\begin{aligned}\beta_1 &= 71.62/a_0^2 & \beta_2 &= 13.05/a_0^2 & \beta_3 &= 3.53/a_0^2 \\ \gamma_1 &= 2.94/a_0^2 & \gamma_2 &= 0.68/a_0^2 & \gamma_3 &= 0.22/a_0^2\end{aligned}\quad (24)$$

$a_0 = 0.529 \text{ \AA}$  is the Bohr radius.

To obtain the aug-SZV-MOLOPT-ae-SR basis set for carbon, we use the STO-3G basis set (23) and add one  $s$ -function ( $l = 0, m = 0$ ), one  $p$ -function ( $l = 1, m = -1, 0, 1$ ), and one  $d$ -function ( $l = 2, m = -2, -1, 0, 1, 2$ ), each having the form

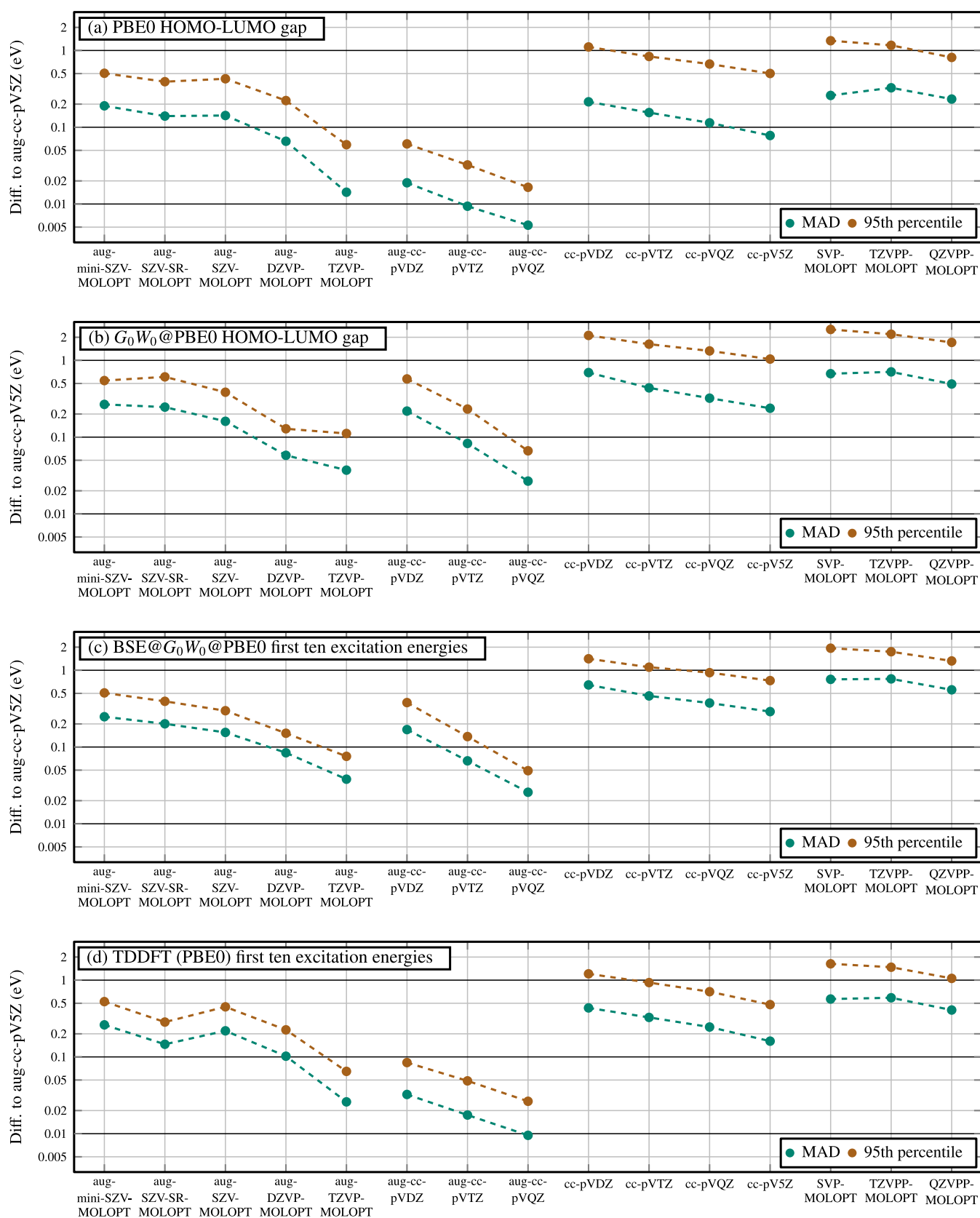
$$\begin{aligned}\phi_6(\mathbf{r}) &= \sum_{i=1}^4 \alpha_{i,4} \exp(-\gamma_i r_C^2) \\ \phi_{7,8,9}(\mathbf{r}) &= Y_1^{-1,0,1}(\theta_C, \varphi_C) r_C \sum_{i=1}^4 \alpha_{i,5} \exp(-\gamma_i r_C^2) \\ \phi_{10-14}(\mathbf{r}) &= Y_2^{-2,-1,0,1,2}(\theta_C, \varphi_C) r_C^2 \sum_{i=1}^4 \alpha_{i,6} \exp(-\gamma_i r_C^2)\end{aligned}\quad (25)$$

The free parameters are  $\gamma_4$  and  $\alpha_{ij}$ ,  $i = 1, 2, 3, 4$ ,  $j = 4, 5, 6$ , which we have optimized to match the lowest five BSE excitation energies in the complete-basis set limit of the molecules of Thiel's set. The result of the optimization is  $\gamma_4 = 0.115$  and the optimized  $\alpha_{ij}$  are listed in Figure 2.

All generated aug-SZV-MOLOPT-ae-mini, aug-SZV-MOLOPT-ae-SR, aug-SZV-MOLOPT-ae, aug-DZVP-MOLOPT-ae, and aug-TZVP-MOLOPT-ae basis sets are listed in the Supporting Information (Section S5); the size, augmentation procedure and minimum exponent of the basis sets are summarized in Table 1.

**2.5. PBE0 and GW HOMO–LUMO gaps.** We compute HOMO–LUMO gaps using PBE0 and  $G_0W_0$ @PBE0 for a subset of 247 molecules from the GW5000 benchmark set (see App. A for the description of the benchmark set and the computational details). Figure 3a,b shows the results for the aug-MOLOPT basis sets introduced in this work, together with the aug-cc-pVXZ, cc-pVXZ, and MOLOPT basis sets.

Figure 3a compares PBE0 HOMO–LUMO gaps across the four basis set families. We report the mean absolute deviation (MAD) with respect to the complete basis set (CBS) limit, taken here as aug-cc-pV5Z:



**Figure 3.** Basis set convergence of excited-state energies for a subset of 247 molecules from the GW5000 benchmark set. We report the mean absolute deviation (MAD) and 95th percentile error (95PE) relative to the aug-cc-pV5Z basis for the aug-MOLOPT basis sets developed in this work, aug-cc-pVXZ,<sup>34</sup> cc-pVXZ<sup>43</sup> and all-electron MOLOPT basis sets.<sup>37,38</sup> Panels show (a) PBE0 HOMO–LUMO gaps, (b)  $G_0W_0$ @PBE0 HOMO–LUMO gaps, (c) first ten excitation energies computed from BSE@ $G_0W_0$ @PBE0, and (d) from TDDFT (PBE0). All benchmark molecules contain between 10 and 20 atoms. For results on smaller molecules containing all elements H–Cl, see the Supporting Information S2, S3.



$$\text{MAD}^B = \frac{1}{N_{\text{mol}}} \sum_{i=1}^{N_{\text{mol}}} |\Delta_i^B - \Delta_i^{\text{aug-cc-pV5Z}}| \quad (26)$$

where  $N_{\text{mol}}=247$  is the number of molecules and  $\Delta_i^B$  is the PBE0 HOMO–LUMO gap of molecule  $i$  computed with basis set  $B$ . While MAD captures the average accuracy, we also report the 95th percentile error (95PE) to quantify the worst-case deviations of the worst 5% of molecules. The aug-MOLOPT basis sets show systematic improvement from aug-SZV-MOLOPT-ae to aug-TZVP-MOLOPT-ae, with both MAD and 95PE decreasing toward the CBS limit. The aug-TZVP-MOLOPT basis has a MAD of just 14 meV. The aug-cc-pVXZ basis sets generally show even smaller deviations at equivalent basis size—e.g., aug-cc-pVTZ is closer to the CBS than aug-TZVP-MOLOPT-ae. This is expected, as aug-cc-pVXZ are specifically optimized for electron affinities and include very diffuse functions well suited for describing the LUMO. In contrast, the cc-pVXZ and MOLOPT families exhibit significantly larger errors and slower convergence for the HOMO–LUMO gap, reflecting their optimization for ground-state energies rather than excited states. Overall, the aug-MOLOPT basis sets provide fast convergence of HOMO–LUMO gaps, while maintaining a well-conditioned overlap matrix (see Figure 1).

Figure 3b shows the basis set convergence of the four basis set families for  $G_0W_0$ @PBE0 HOMO–LUMO gaps, using aug-cc-pV5Z as the CBS reference. The aug-MOLOPT basis sets exhibit consistently small deviations from the CBS. Notably, the MAD of the small aug-SZV-MOLOPT-ae basis is 160 meV, better than the larger aug-cc-pVDZ basis, which yields a MAD of 220 meV. Likewise, aug-DZVP-MOLOPT-ae achieves a MAD of just 60 meV, below the 80 meV deviation of the larger aug-cc-pVTZ basis. This finding suggests that the aug-MOLOPT basis sets are the ideal choice for GW HOMO–LUMO gap calculations to ensure both fast basis set convergence of GW HOMO–LUMO gaps and numerical stability for large molecules. Again, the nonaugmented cc-pVXZ and MOLOPT basis sets exhibit larger and more slowly converging errors for  $G_0W_0$ @PBE0 HOMO–LUMO gaps. For example, the MAD of the large cc-pV5Z basis is 240 meV—only slightly lower than the minimal aug-SZV-MOLOPT-ae-mini basis which has a MAD of 270 meV. When excluding molecules with diffuse LUMOs (defined as LUMO eigenvalues above  $-2$  eV), the MAD for cc-pV5Z decreases by an order of magnitude to 30 meV (Figure S2 in Supporting Information). This indicates that these basis sets were optimized for ground-state properties, and lacking diffuse functions, they are inadequate for accurately describing GW gaps in systems with unoccupied states that significantly extend into the vacuum.

**2.6. GW+BSE and TDDFT Excitation Energies.** Figure 3c shows the basis set convergence of the four basis set families for the first ten BSE@ $G_0W_0$ @PBE0 excitation energies, where the deviation is again computed against aug-cc-pV5Z as the CBS reference. As with GW calculations, the aug-MOLOPT basis sets exhibit also consistently small deviations from the CBS in this case. The MAD of the compact aug-SZV-MOLOPT-ae basis is 160 meV, which is slightly below the 170 meV MAD of the larger aug-cc-pVDZ basis. The aug-DZVP-MOLOPT-ae basis exhibits a MAD of 80 meV, slightly worse than the 70 meV deviation of the larger aug-cc-pVTZ basis. The nonaugmented cc-pVXZ and MOLOPT basis sets show larger and more slowly converging errors; for example, the MAD of the large cc-pV5Z basis is 290

meV; almost double the error of aug-SZV-MOLOPT-ae. In this case, the aug-MOLOPT basis sets appear to be an excellent choice for BSE calculations to ensure fast basis set convergence of BSE excitation energies and numerical stability for large molecules.

Figure 3d shows the basis set convergence of the first ten excitation energies computed with TDDFT (PBE0). As before, the aug-MOLOPT basis sets exhibit systematic improvement with increasing basis size. However, in contrast to the BSE case, the aug-cc-pVXZ basis sets outperform the aug-MOLOPT family: for example, the MAD of aug-DZVP-MOLOPT-ae is 100 meV, whereas aug-cc-pVDZ achieves a significantly lower MAD of 17 meV. Comparing with BSE results in Figure 3c, the aug-cc-pVXZ basis sets converge more rapidly for TDDFT than for BSE, while the aug-MOLOPT sets show similarly fast convergence for both methods. We attribute this difference to the design philosophy behind the basis sets: the aug-MOLOPT sets were specifically optimized for BSE excitation energies (albeit on a different training set, the Thiel's set<sup>59</sup>), whereas the aug-cc-pVXZ family was not. Nevertheless, aug-cc-pVXZ basis sets feature ill-conditioned overlap matrices for large molecules, making the aug-MOLOPT basis sets numerically more robust for larger molecules. We report in App. C an example of a calculation of these orbital basis sets on the 9,10-Dihydroanthracene molecule, showing the variation of the error for all the test cases of Figure 3 with respect to the basis set size.

### 3. RI BASIS SETS

**3.1. Auxiliary RI Basis Set Generation.** The resolution-of-the-identity (RI) technique is widely used to reduce the computational cost of quantum chemical methods.<sup>62</sup> In RI, four-center integrals

$$(ialjb) = \int d\mathbf{r} d\mathbf{r}' \psi_i(\mathbf{r}) \psi_a(\mathbf{r}) \frac{1}{|\mathbf{r} - \mathbf{r}'|} \psi_j(\mathbf{r}') \psi_b(\mathbf{r}') \quad (27)$$

are expressed by products of two- and three-center integrals, which can enable substantial reduction of computational effort:

$$\begin{aligned} (ialjb)_{\text{RI}} &= \sum_{PQRT} (ialP)_m (\mathbf{M}^{-1})_{PQ} V_{QR} (\mathbf{M}^{-1})_{RT} (Tljb)_m \\ (ialP)_m &= (Plia)_m = \int d\mathbf{r} d\mathbf{r}' \phi_i(\mathbf{r}) \phi_a(\mathbf{r}) m(\mathbf{r}, \mathbf{r}') \phi_P(\mathbf{r}') \\ M_{PQ} &= \int d\mathbf{r} d\mathbf{r}' \phi_P(\mathbf{r}) m(\mathbf{r}, \mathbf{r}') \phi_Q(\mathbf{r}') \\ V_{PQ} &= \int d\mathbf{r} d\mathbf{r}' \phi_P(\mathbf{r}) \frac{1}{|\mathbf{r} - \mathbf{r}'|} \phi_Q(\mathbf{r}') \end{aligned} \quad (28)$$

Here, we introduced the auxiliary RI basis set  $\{\phi_P\}$ , which also consists of Gaussians.  $m(\mathbf{r}, \mathbf{r}')$  is the RI metric; convergence of the RI expansion (28) depends on  $m$ . It has been shown that the fastest convergence of the RI expansion is achieved using the Coulomb metric,  $m(\mathbf{r}, \mathbf{r}') = 1/|\mathbf{r} - \mathbf{r}'|$ .<sup>62</sup>

Early applications of RI include DFT<sup>63,64</sup> and MP2,<sup>65</sup> where it became a standard technique by now. In random phase approximation (RPA) calculations for the correlation energy, RI reduces the scaling from  $O(N^6)$  in the canonical Casida-based formulation to  $O(N^4)$ .<sup>66</sup> However, RI is not universally beneficial: in Hartree–Fock and hybrid functional calculations, RI typically improves performance only when large orbital basis sets are used.<sup>67</sup> For small orbital basis sets, conventional four-center formulations may remain more efficient. For the



computation of charged excitations based on GW, RI has also become a common tool, where it reduces the scaling from  $O(N^6)$  to  $O(N^4)$ ,<sup>68–70</sup> as well as for charge-neutral excitations based on the BSE, where the screened Coulomb interaction is computed using RI.

When using RI, an auxiliary RI basis set  $\{\varphi_p\}$  is required for the factorization (28) of four-center integrals into two- and three-center integrals. Although it is possible to generate auxiliary RI basis sets on the fly during the calculation,<sup>71,72</sup> this often results in large auxiliary RI basis sets. Recently, several schemes have been proposed to tackle this issue, such as the combined use of a contraction based on the singular value decomposition and a high-momentum truncation of the generated basis sets,<sup>73</sup> or a newly suggested approach with uncontracted, noneven-tempered sets that are truncated using the 2-body energy as a metric.<sup>74</sup> In this work, we instead generate auxiliary RI basis sets by matching the RI-MP2 correlation energy of isolated atoms to the corresponding MP2 reference energies.<sup>75</sup> For this purpose, as proposed in,<sup>75</sup> we generate auxiliary RI basis sets of different sizes by using the (relative)  $\Delta_I$  metric as an optimization parameter:

$$\Delta_I = -\frac{1}{4E_{\text{MP2}}} \sum_{ijab} \frac{|\langle ij||ab \rangle - \langle ij||ab \rangle_{\text{RI}}|^2}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \quad (29)$$

where the  $(ij)$  refer to occupied orbitals and  $(ab)$  to empty orbitals,  $E_{\text{MP2}}$  is the MP2 correlation energy and using a standard notation for the double bar integral defined as<sup>76</sup>

$$\langle ij||ab \rangle = (ia|jb) - (ib|ja) \quad (30)$$

$$\langle ij||ab \rangle_{\text{RI}} = (ia|jb)_{\text{RI}} - (ib|ja)_{\text{RI}} \quad (31)$$

Larger auxiliary sets lead to consistently lower values of  $\Delta_I$ , see Figure 4. Also, for smaller orbital basis sets, the required auxiliary

RI basis size to reach a given  $\Delta_I$  metric value is smaller. For the smallest aug-SZV-MOLOPT-ae-mini basis, an auxiliary RI basis set with only 25 basis functions is sufficient to reach a  $\Delta_I$  metric value below  $10^{-6}$ . This highlights the potential for efficient calculations using the aug-SZV-MOLOPT-ae-mini basis set.

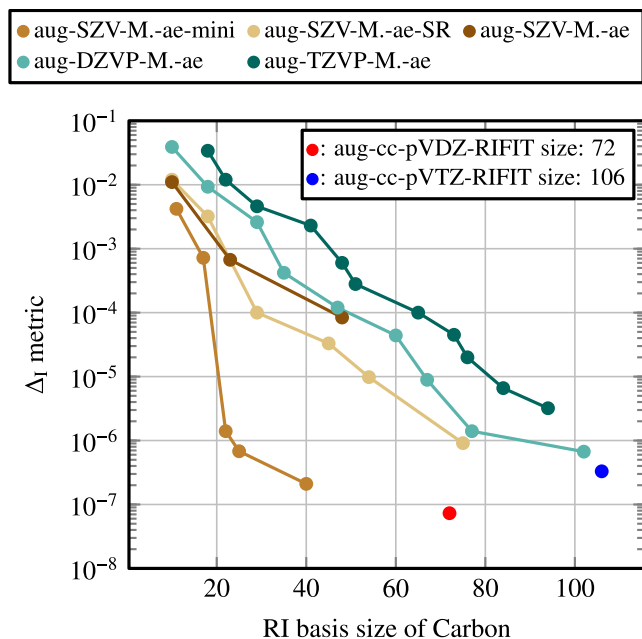
For comparison, we compute  $\Delta_I$  for the existing aug-cc-pVDZ and aug-cc-pVTZ with corresponding RI basis sets,<sup>77</sup> see Figure 4. These basis sets have very small  $\Delta_I$  metric values below  $10^{-6}$ , but are relatively large in size (72 and 106 functions for aug-cc-pVDZ-RIFIT and aug-cc-pVTZ-RIFIT, respectively). We also create smaller auxiliary RI basis sets with lower accuracy, which are still sufficient in applications as we demonstrate later for nanographenes (Section 4). All generated auxiliary RI basis sets are available in the Supporting Information (Section S6). The optimization was performed using the auxiliary RI basis set optimizer implemented in CP2K.<sup>78</sup>

**3.2. RI Basis Set Convergence: GW HOMO–LUMO Gaps from Low-Scaling  $O(N^3)$  GW.** For the GW and BSE basis set benchmark presented in Figure 3, we employed the largest available auxiliary RI basis sets (see Section 3.1 for generation details). To enable large-scale GW and BSE simulations, it is desirable to reduce the size of the auxiliary RI basis set while maintaining high numerical accuracy. Smaller auxiliary RI basis sets lead to lower computational cost and improved scalability, particularly in low-scaling GW algorithms.

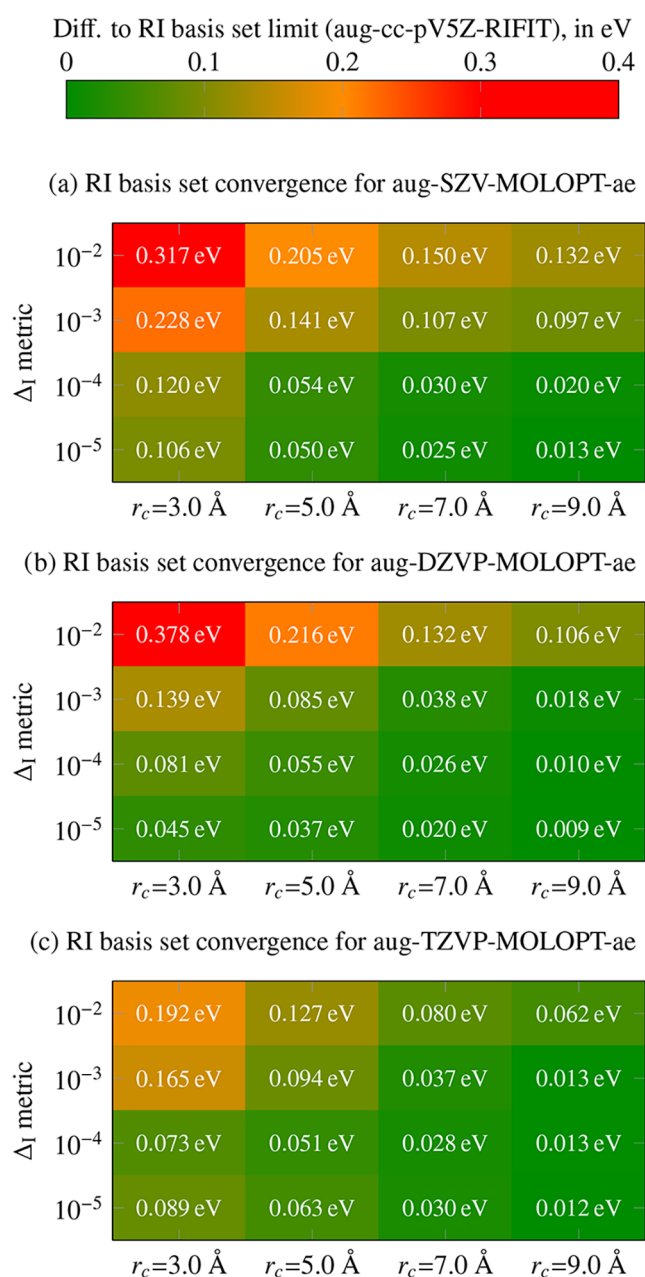
In this work, we employ the cubic-scaling GW implementation in CP2K for molecules,<sup>79</sup> which uses the truncated Coulomb metric<sup>80</sup> for the RI approximation. While the fastest convergence of RI-based methods is achieved when the cutoff radius of the Coulomb operator is infinite, this also increases the computational cost. Therefore, a balance must be found: the cutoff radius should be small enough to reduce computational requirements, yet large enough to ensure sufficiently fast convergence of the auxiliary RI basis set.

To evaluate this trade-off, we benchmark  $G_0W_0$ @PBE0 HOMO–LUMO gaps for the aug-SZV-MOLOPT-ae, aug-DZVP-MOLOPT-ae and aug-TZVP-MOLOPT-ae basis sets on the same subset of 247 molecules from the GW5000 benchmark set used in Figure 3. We consider four auxiliary RI basis sets of increasing size, corresponding to decreasing the  $\Delta_I$  metric threshold:  $10^{-2}$ ,  $10^{-3}$ ,  $10^{-4}$ , and  $10^{-5}$ . For each basis set, we evaluate four different cutoff values for the truncated Coulomb operator:  $r_c = 3, 5, 7$ , and  $9 \text{ \AA}$ .

Figure 5 summarizes the results. The color map shows the absolute deviation of the  $G_0W_0$  HOMO–LUMO gaps (averaged over all 247 molecules) from the converged reference obtained with the large aug-cc-pV5Z-RIFIT auxiliary RI basis set<sup>81</sup> and cutoff  $r_c = 9 \text{ \AA}$ . For the aug-SZV-MOLOPT basis set, at the loosest RI threshold ( $10^{-2}$ ) and smallest cutoff ( $r_c = 3 \text{ \AA}$ ), the average error exceeds 300 meV. However, increasing the cutoff to  $r_c = 9 \text{ \AA}$  reduces this error to  $\sim 130 \text{ meV}$ . For a larger auxiliary RI basis ( $\Delta_I$  threshold of  $10^{-4}$ ), numerical accuracy is substantially improved: for  $r_c = 7 \text{ \AA}$ , the deviation is 30 meV, and drops to 20 meV at  $r_c = 9 \text{ \AA}$ . The best overall agreement with aug-cc-pV5Z-RIFIT is obtained for an even larger auxiliary RI basis ( $\Delta_I$  threshold of  $10^{-5}$ ) with  $r_c = 9 \text{ \AA}$ , where the average absolute error is reduced to 13 meV. The results for the aug-DZVP-MOLOPT and aug-TZVP-MOLOPT basis sets show better convergence properties than the aug-SZV-MOLOPT benchmark tests, which can be easily explained by the larger size of these basis sets. However, the overall convergence trends are very similar between all the basis sets. These results demonstrate that accurate low-scaling GW calculations can be achieved using



**Figure 4.**  $\Delta_I$  metric for a carbon atom as a function of the auxiliary RI basis set size, using various augmented MOLOPT basis sets introduced in this work. The auxiliary RI basis sets are optimized for the carbon atom in a given basis set size to match the MP2 correlation energy. Reference auxiliary RI basis set sizes and  $\Delta_I$  for aug-cc-pVDZ-RIFIT and aug-cc-pVTZ-RIFIT are shown for comparison.



**Figure 5.** Convergence of low-scaling GW calculations<sup>79</sup> with respect to the cutoff radius of the truncated Coulomb metric and the auxiliary RI basis set size (here quantified by the  $\Delta_I$  metric threshold). As orbital basis set, we employ aug-SZV-MOLOPT-ae (top), aug-DZVP-MOLOPT-ae (middle) and aug-TZVP-MOLOPT-ae (bottom). The color map shows the mean absolute deviation of  $G_0W_0$ @PBE0 HOMO–LUMO gaps for the same subset of 247 molecules from the GW5000 benchmark set used in Figure 3, relative to a reference calculation using the aug-cc-pV5Z-RIFIT auxiliary RI basis set.<sup>81</sup> Each row corresponds to an auxiliary RI basis set generated with a given  $\Delta_I$  metric threshold (from  $10^{-2}$  to  $10^{-5}$ ). Smaller errors are achieved with tighter RI thresholds and larger Coulomb cutoffs. A practical compromise is reached with a  $\Delta_I$  metric threshold of  $10^{-4}$  and  $r_c \geq 7$  Å (error: 30 meV).

relatively compact auxiliary RI basis sets when paired with a sufficiently large Coulomb cutoff. For practical applications aiming at high numerical precision, we recommend a  $\Delta_I$  threshold of  $10^{-4}$  and a cutoff radius of at least 7 Å, giving excellent balance between efficiency and accuracy ( $\sim 30$  meV).

#### 4. LOW-SCALING $O(N^3)$ GW CALCULATIONS ON NANOGAPHENES

To demonstrate the suitability of the generated augmented MOLOPT basis sets for large-scale applications, we perform GW calculations on nanographenes of increasing size. Representative geometries are shown in Figure 6a. For these systems, we employ the PBE functional<sup>82</sup> as the DFT starting point and Hedin's shift<sup>34</sup> to avoid the higher cost of hybrid functionals during the SCF.

We use auxiliary RI basis sets with a  $\Delta_I$  value below  $1.5 \times 10^{-2}$ . The corresponding basis sizes are listed in Table 2.  $G_0W_0$ +H HOMO–LUMO gaps for the nanographenes are shown in Figure 6b. For  $L = 1$  (9,10-Dihydroanthracene), basis set convergence is challenging: the minimal aug-SZV-MOLOPT-ae-mini basis underestimates the gap by approximately 1 eV. This is partly due to the small auxiliary RI basis used (cf. Figure 5). For larger systems ( $L \geq 8$ ), convergence improves significantly: all five basis sets agree within 50 meV (inset of Figure 6b). This matches findings for two-dimensional materials,<sup>23</sup> where convergence within 100 meV was reached using the aug-SZV-MOLOPT basis.

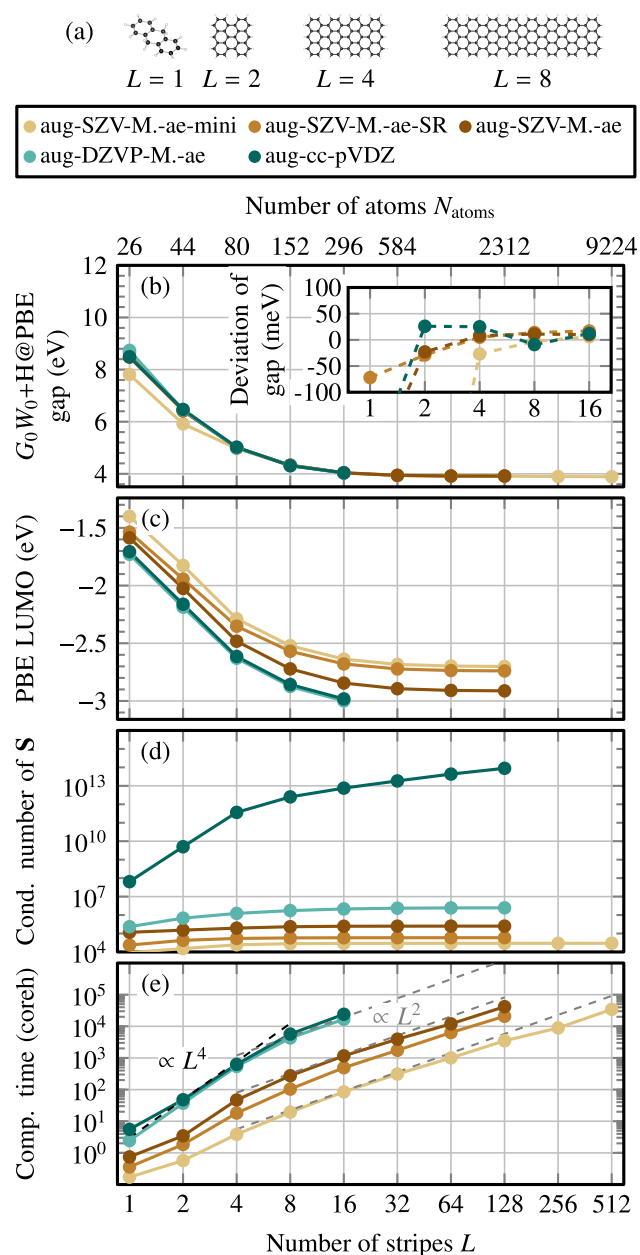
We attribute the improved basis set convergence for larger structures to three factors: (i) larger systems offer more basis functions, increasing flexibility; (ii) the PBE LUMO energy decreases with  $L$  (Figure 6c), making the LUMO less diffuse in vacuum and thus easier to represent (as discussed for the decay length  $\zeta_n \approx \hbar/\sqrt{2m|E_n|}$  in Section 2.4). For a benchmark of the numerical precision of basis sets for  $G_0W_0$  HOMO–LUMO gaps as a function of the DFT LUMO energy, see the Supporting Information, Figures S1 and S2 (and Figure S3 for differences of excitation energies); (iii) there may be cancellation of errors between an underconverged orbital and auxiliary RI basis set. Notably, error cancellation does not distort the size dependence: for  $L = 32, 64, 128$ , all aug-SZV-MOLOPT (SR, mini) basis sets yield size-converged gaps consistently within 33 meV. This indicates robust GW calculations for large systems. These results support two conclusions: (i) basis set convergence for nanostructures differs from that of small molecules and must be analyzed accordingly; (ii) further optimization of Gaussian basis sets for extended systems, in particular with pseudopotentials,<sup>83</sup> appears promising.

The condition number of the overlap matrix remains below  $10^7$  for all augmented MOLOPT basis sets (Figure 6d). In contrast, it exceeds  $10^{13}$  for aug-cc-pVDZ. The computational cost is roughly reduced by a factor of 280 when using aug-SZV-MOLOPT-ae-mini instead of aug-cc-pVDZ (Figure 6e). This aligns with expected GW scaling of  $N_{\text{bf}}^2 N_{\text{RI}}^2$ : According to Table 2,  $N_{\text{bf}}$  and  $N_{\text{RI}}$  decrease by factors of about 2.4 and 6.8, respectively, giving  $2.4^2 \times 6.8^2 \approx 270$ . Despite this enormous speedup, the gap difference between aug-SZV-MOLOPT-ae-mini and aug-cc-pVDZ is less than 10 meV for  $L = 16$ .

The small size of the aug-SZV-MOLOPT-ae-mini basis set enabled us to perform a GW calculation on a nanographene with 9224 atoms, requiring only 34,300 core hours.

#### 5. CONCLUSION

We introduced the augmented MOLOPT family of all-electron Gaussian basis sets optimized for accurate excited-state calculations of large molecules for the elements H to Cl. These basis sets achieve fast basis set convergence of GW quasiparticle energy differences and BSE excitation energies while ensuring low condition numbers of the overlap matrix  $S$ ,



**Figure 6.** GW calculations on nanographenes of increasing length (defined by the number of stripes  $L = 1, 2, 4, \dots, 512$ ). (a) Nanographene geometry for  $L = 1, 2, 4, \dots$ . Note that we put two hydrogen atoms at the center carbon atom at the zigzag edge to prevent for magnetic zigzag edge states. (b) Quasiparticle HOMO–LUMO gaps computed with  $G_0W_0$  + Hedin's shift @ PBE using different basis sets. Inset: deviation from the aug-DZVP-MOLOPT-ae basis set. (c) PBE LUMO eigenvalue, serving as a measure of LUMO diffuseness. (d) Condition number  $\kappa(S)$  of the overlap matrix, computed from eq 11. (e) Computation time (in core hours) of the  $G_0W_0$  calculations on Noctua2 (AMD Milan 7763) and Otus (AMD Turin 9655). The aug-MOLOPT basis sets exhibit low condition numbers and reduced computational cost, enabling stable and efficient calculations for nanographenes exceeding 9000 atoms. Details on the number of nodes used, wall time and memory consumption of the GW calculations are listed in Table 3.

thereby enabling numerically stable calculations. For  $G_0W_0$ @PBE0 gaps, aug-DZVP-MOLOPT-ae yields a mean absolute deviation (MAD) of 60 meV compared to the aug-cc-pV5Z complete basis set, outperforming the larger aug-cc-pVTZ basis

**Table 2.** Orbital Basis Set Size and Number of Auxiliary RI Basis Functions H and C Atom Used for the GW Calculations Shown in Fig. 6 Across Different Orbital Basis Sets. (cf. Figure 4)

| basis set          | $N_{\text{bf}}^{\text{H}}$ | $N_{\text{bf}}^{\text{C}}$ | $N_{\text{RI}}^{\text{H}}$ | $N_{\text{RI}}^{\text{C}}$ | $\Delta_{\text{I}}^{\text{H}}$ | $\Delta_{\text{I}}^{\text{C}}$ |
|--------------------|----------------------------|----------------------------|----------------------------|----------------------------|--------------------------------|--------------------------------|
| aug-SZV-M.-ae-mini | 6                          | 9                          | 2                          | 11                         | $1.5 \times 10^{-2}$           | $4.2 \times 10^{-3}$           |
| aug-SZV-M.-ae-SR   | 6                          | 14                         | 5                          | 18                         | $3.0 \times 10^{-3}$           | $3.2 \times 10^{-3}$           |
| aug-SZV-M.-ae      | 6                          | 14                         | 6                          | 23                         | $1.3 \times 10^{-4}$           | $6.7 \times 10^{-4}$           |
| aug-DZVP-M.-ae     | 9                          | 30                         | 9                          | 35                         | $4.4 \times 10^{-5}$           | $4.2 \times 10^{-4}$           |
| aug-cc-pVDZ        | 9                          | 23                         | 23                         | 72                         | $7.3 \times 10^{-8}$           | $9.7 \times 10^{-8}$           |

set (MAD: 80 meV) for organic molecules. Similar MAD are observed for BSE and TDDFT excitation energies. The augmented MOLOPT basis sets exhibit excellent numerical stability, with overlap matrix condition numbers below  $10^7$  even for 9000-atom nanographenes. We also generate very compact basis sets, aug-SZV-MOLOPT-ae-mini, which enable very efficient large-scale  $G_0W_0$  calculations, e.g., on a 9224-atom nanographene consuming only 34300 core hours. This demonstrates that the proposed augmented MOLOPT basis sets enable routine GW and BSE calculations on large-scale systems with several thousands of atoms, keeping good numerical accuracy and reducing the computational cost by 2 orders of magnitude compared to previously used aug-cc-pVXZ basis sets. All generated augmented MOLOPT basis sets are freely available in the Supporting Information.

## APP. A—COMPUTATIONAL DETAILS

### Description of Molecular Test Set

For benchmarking excited-state energies with our generated basis sets, we use the GW5000 dataset.<sup>87</sup> We exclude all molecules with less than ten atoms as small molecules tend to have very diffuse unoccupied states; and the purpose of the generated basis sets is to describe large molecules with less diffuse unoccupied states. To reduce the computational cost, we only use molecules with at most 20 atoms. We also remove all molecules larger than 15 atoms in which carbon atoms outnumber all other non-hydrogen elements by more than a factor of two. Our aim is to ensure a balanced benchmark set avoiding overrepresentation of unsubstituted or weakly substituted hydrocarbons. The precise criterion for removal is  $N_{\text{C}} > 2(N_{\text{tot}} - N_{\text{H}} - N_{\text{C}})$ , where  $N_{\text{C}}$  is the number of carbon atoms,  $N_{\text{H}}$  the number of hydrogen atoms and  $N_{\text{tot}}$  the total number of atoms in the molecule.

Applying these criteria gives 247 molecules in the GW5000 benchmark set, where the majority of the molecules contain C (98%), H (96%), N (76%), O (74%), while other elements are less often present: S (31%), Cl (23%), F (10%), P (2%), B (1%) and Si (1%). We also use a second molecular benchmark set that focuses on other elements (Li, Be, B, Na, Ca, Al, Si, P); we show the composition of this benchmark set and the calculations in the Supporting Information SI1 and SI2, respectively. We employ the CP2K package for all calculations.<sup>41,88</sup> CP2K employs a Gaussian basis set for representing KS orbitals [eq 2]. We use the Gaussian and augmented plane-waves scheme,<sup>89</sup> which enables all-electron calculations in CP2K. We use implementations in CP2K of conventional GW (Section 2.5) in imaginary-frequency formulation with analytic continuation,<sup>46</sup> BSE<sup>90</sup> and TDDFT<sup>91</sup> (Section 2.6), as well as low-scaling GW<sup>79</sup> (Sections 3.2 and 4) based on the space-time method<sup>16</sup> using minimax time-frequency grids.<sup>92–95</sup> We visualized atomic geometries using the VESTA program.<sup>96</sup>



## Numerical Aspects

We employ the PBE0 exchange–correlation functional<sup>97</sup> as starting point for our excited-state calculations. The usage of PBE0 as starting point for GW and Bethe–Salpeter avoids numerical instabilities due to multipole features of the self-energy close to the quasiparticle solution,<sup>98,99</sup> which can be present when starting from the PBE functional.<sup>82</sup> For the low-scaling GW calculations on nanographenes (Section 4), however, we use PBE for the SCF cycle to reduce computational cost. As discussed in ref 99,  $G_0W_0@PBE$  can suffer from numerical instabilities caused by poles in the self-energy  $\Sigma_n(\omega)$  close to the quasiparticle energy  $\omega \approx \epsilon_n^{G_0W_0}$ , where

$$\epsilon_n^{G_0W_0} = \epsilon_n^{\text{PBE}} + \text{Re}\Sigma_n(\epsilon_n^{G_0W_0}) - v_n^{\text{xc}} \quad (32)$$

Here,  $v_n^{\text{xc}}$  is the diagonal matrix element of the PBE exchange–correlation potential. These instabilities can be eliminated either by using eigenvalue self-consistent schemes (evGW)<sup>98,99</sup> or, more computationally efficient, by introducing a state-specific Hedin shift,<sup>100,101</sup>

$$\Delta H_n = \text{Re}\Sigma_n(\epsilon_n^{\text{PBE}}) - v_n^{\text{xc}} \quad (33)$$

leading to the modified quasiparticle equation

$$\epsilon_n^{G_0W_0+H} = \epsilon_n^{\text{PBE}} + \text{Re}\Sigma_n(\epsilon_n^{G_0W_0+H} - \Delta H_n) - v_n^{\text{xc}} \quad (34)$$

which we apply in Section 4 to obtain quasiparticle energies using the  $G_0W_0$  + Hedin's shift ( $G_0W_0+H$ ) method.

## APP. B—MEMORY-SAVING SCHEME FOR LOW-SCALING GW CALCULATIONS

In this appendix, we describe a memory saving scheme to reduce the random access memory (RAM) of low-scaling GW calculations<sup>27,79</sup> substantially. The RAM bottleneck of the GW algorithm<sup>27,79</sup> appears in the computation of the self-energy  $\Sigma$  in imaginary time  $i\tau$

$$\Sigma_{\lambda\sigma}(i\tau) = \sum_{\nu Q} \left[ \sum_{\mu} (\lambda\mu|Q) G_{\mu\nu}(i\tau) \right] \left[ \sum_P (\nu\sigma|P) W_{PQ}(i\tau) \right] \quad (35)$$

where  $\mu, \nu, \lambda, \sigma$  are atom-centered Gaussian basis functions for expanding molecular orbitals (MO) and  $P, Q$  are auxiliary RI basis functions for the screened Coulomb interaction  $W$ .  $G$  denotes the Green's function and

$$(\mu\nu|P) = \int d\mathbf{r} d\mathbf{r}' \phi_{\mu}(\mathbf{r}) \phi_{\nu}(\mathbf{r}') V_{r_c}(\mathbf{r}, \mathbf{r}') \phi_P(\mathbf{r}') \quad (36)$$

are three-center integrals (3cI) of the truncated Coulomb operator ( $r_c$ : truncation radius)

$$V_{r_c} = \begin{cases} 1/|\mathbf{r} - \mathbf{r}'| & \text{if } |\mathbf{r} - \mathbf{r}'| \leq r_c \\ 0 & \text{else} \end{cases} \quad (37)$$

The 3cIs  $(\mu\nu|P)$  are sparse, i.e., the numerical integral value  $(\mu\nu|P)$  is large if and only if the three Gaussian functions  $\phi_{\mu}(\mathbf{r})$ ,  $\phi_{\nu}(\mathbf{r})$  and  $\phi_P(\mathbf{r})$  are close together. The number of integrals  $(\mu\nu|P)$  that need to be stored in the calculation is  $N_{\text{bf}}^2 N_{\text{RI}} \alpha$ , where  $\alpha$  is the percentage of non-negligible  $(\mu\nu|P)$  elements kept in the calculation. Each integral requires storage of 8 B in double-precision arithmetic and thus the memory required to store all  $(\mu\nu|P)$  is

$$\mathcal{M} = N_{\text{bf}}^2 N_{\text{RI}} \alpha \cdot 8\text{B} \quad (38)$$

The challenge regarding memory comes in the intermediate tensor from eq 35

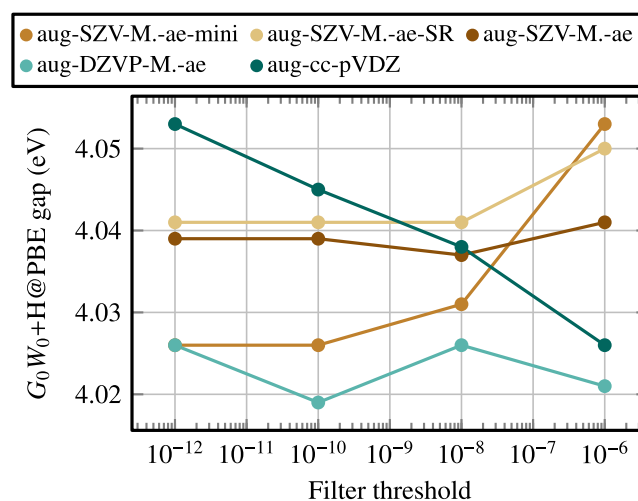
$$M_{\nu\sigma Q} := \sum_P (\nu\sigma|P) W_{PQ}(i\tau) \quad (39)$$

where the sparsity of  $M_{\nu\sigma Q}$  in the index pairs  $\nu$ - $Q$  and  $\sigma$ - $Q$  is lost because the screened Coulomb interaction  $W_{PQ}$  is long-ranged. Therefore, a larger fraction  $\beta \gg \alpha$  of  $M_{\nu\sigma Q}$  elements are non-negligible in the calculation.

We reduce the RAM consumption of eq 35 by a repeated calculation of 3cIs. Specifically, we rewrite eq 35 as sum over atomic contributions from atom  $A$  and atom  $B$

$$\Sigma_{\lambda\sigma}(i\tau) = \sum_{A,B} \sum_{\nu(\text{at } A)} \sum_{Q(\text{at } B)} \left[ \sum_{\mu} (\lambda\mu|Q) G_{\mu\nu}(i\tau) \right] \times \left[ \sum_P (\nu\sigma|P) W_{PQ}(i\tau) \right] \quad (40)$$

and we only keep the quantities on the right side of eq 40 in memory if  $\nu$  and  $Q$  belong to the atom-pair  $(A,B)$ . The result of the summation in eq 40 of the atom-pair  $(A,B)$  is added to  $\Sigma_{\lambda\sigma}(i\tau)$  and we release then all quantities from the right side of eq 40 belonging to atom pair  $(A,B)$  from the allocated memory. For the next atom pair  $(A',B')$ , we compute the 3cI  $(\mu\nu|P)$  from eq 36. This repeated calculation of 3cIs allows us to only keep a fraction of 3cIs and of intermediate tensors  $M_{\nu\sigma Q}$  (39) in memory, reducing the RAM consumption drastically.

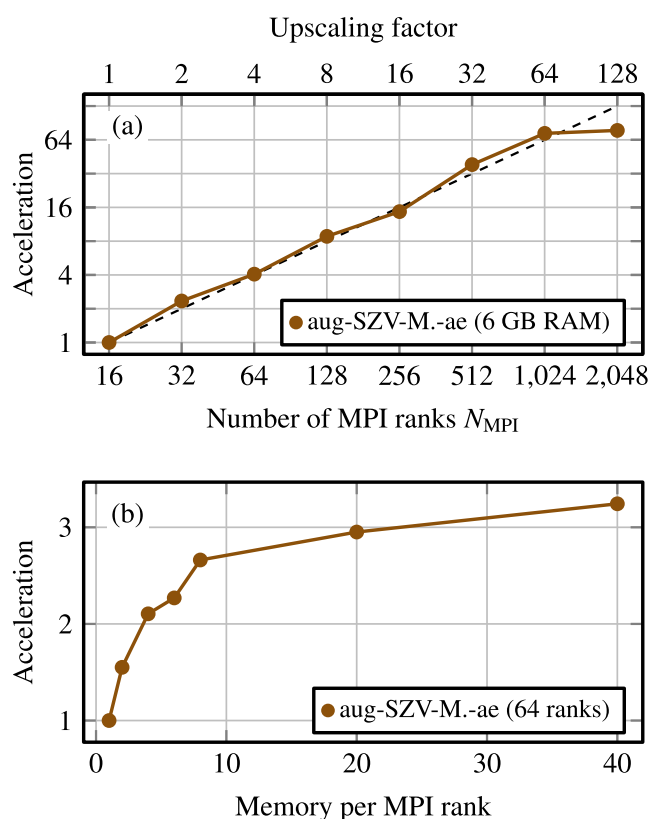


**Figure 7.** Convergence of the  $G_0W_0$  with Hedin's shift HOMO–LUMO gap with respect to the filter threshold parameter in the low-scaling GW implementation for the nanographene with length  $L = 16$ . Results are shown for five different basis sets. Deviations are decreasing for tighter thresholds.

The repeated calculation of 3cIs comes with the drawback that we need to compute the same integral  $(\mu\nu|P)$  several times. Here, we make use of the properties of the Gaussian basis that analytical integral expressions are available for  $(\mu\nu|P)$ , such that this additional computational load is small.<sup>102</sup> In fact, the computation of 3cIs for large systems exceeding 100 atoms only takes <0.1% of the total execution time in the present GW algorithm.<sup>27</sup>

We discuss now the numerical parameters of the GW algorithm in relation to the generated augmented MOLOPT





**Figure 8.** MPI and memory scaling for  $G_0W_0$  calculations on a nanographene with length  $L = 64$  using the aug-SZV-MOLOPT-ae basis set. (a) Parallel acceleration with increasing number of nodes (each node has 16 MPI ranks; each MPI rank has 8 OMP threads). The calculation scales efficiently up to over 1000 MPI ranks, with near-ideal scaling. (b) Acceleration when providing more available RAM per MPI rank to the GW calculation (at fixed 64 MPI ranks, i.e., 4 nodes). The runtime benefits from increased memory per MPI rank, with saturation reached around 8 GB per MPI rank.

basis sets. First, we discuss the filter threshold for sparse operations like computing the self-energy, eq 40, see Figure 7 for the nanographene with length  $L = 16$ . We observe that the GW HOMO–LUMO gap computed with smaller basis sets converges faster with the filter threshold; as an example, the GW HOMO–LUMO gap only changes by less than 1 meV in aug-SZV-MOLOPT-ae-mini when decreasing the filter threshold for atomic blocks from  $10^{-10}$  to  $10^{-12}$ . Instead, for the aug-cc-pVDZ basis set, the GW HOMO–LUMO gap changes by 8 meV. This finding suggests that with the developed compact augmented MOLOPT basis sets, larger filter thresholds can be chosen in the calculation, which contributes to further improve the computational efficiency and numerical stability.

Finally, we discuss the scaling of computation time with number of employed cores. In Figure 8a, we report the acceleration of the calculation for a nanographene of length  $L = 64$  (1160 atoms) with increasing number of MPI ranks. We observe almost perfect weak scaling from one node (16 MPI ranks) to 64 nodes (1024 MPI ranks). Another handle for the user to optimize the computation time, is the amount of RAM available to every MPI rank. While we have fixed the available RAM to 6 GB in Figure 8a, we report the acceleration with respect to the available RAM in Figure 8b. For 4 nodes (64 ranks), we vary the available memory between 1 GB per MPI rank and 40 GB per MPI rank (on large-memory nodes with

1024 GB per node). The calculation gets accelerated when increasing the available memory from 1 to 8 GB per MPI rank by a factor 2.8, because a smaller amount of three-center integrals need to be recomputed when more memory is available. Providing even more memory (20 or 40 GB per MPI rank), only leads to a minor additional acceleration.

For all GW calculations reported in Figure 6b, we provide the full details of basis sizes, memory requirements for storing three-center integrals, their sparsity, available RAM and the execution time in Table 3.

## APP. C—BASIS SET CONVERGENCE FOR 9,10-DIHYDROANTHRACENE

In this appendix, we report DFT and GW HOMO–LUMO gaps as well as GW+BSE and TDDFT excitation energies of 9,10-Dihydroanthracene ( $L = 1$  nanographene in Section 4) using various basis sets to assess the quality of the basis sets as a function of their basis set size. We plot in Figure 9 the error with respect to the aug-cc-pV5Z calculation of the excitation energies and band gaps for the 9,10-Dihydroanthracene molecule as a function of the number of basis functions.

The results show similar trends as observed in Figure 3. Both nonaugmented cc and MOLOPT basis sets show poor convergence with respect to the number of basis function, showing that these are not appropriate for the simulation of excited state energies. For the cc-pVQZ basis set, the error across all tests is around 150 meV on average, whereas it is around 450 meV on average for the comparable QZVPP-MOLOPT basis set, so that the cc basis sets perform better in this case (but still very poorly in comparison to the reference calculation, given the large size of the cc-pVQZ basis set).

The augmented basis sets show much better convergence across all four tests in Figure 9 w.r.t. the basis set size. For the aug-cc-pVTZ basis set, which is of comparable size as the nonaugmented cc-pVQZ and QZVPP-MOLOPT basis sets, the error is around 20 meV on average, and for the aug-TZVP-MOLOPT basis set it is around 10 meV. Note that for the PBE0 HOMO–LUMO gap (Figure 9a) and TDDFT excitation energies (Figure 9d), the aug-cc-pVXZ basis sets converge faster with the basis set size than the aug-MOLOPT basis sets; still aug-MOLOPT basis sets feature improved numerical stability for large-scale calculations due to the reduced condition number compared to aug-cc-pVXZ. For the GW HOMO–LUMO gap (Figure 9b) and GW-BSE excitation energies (Figure 9c), our aug-MOLOPT basis sets give faster convergence with the basis set size than aug-cc-pVXZ basis sets.

## ASSOCIATED CONTENT

### Data Availability Statement

Inputs and outputs of all calculations reported in this work are available in a Github repository<sup>85</sup> and in a Zenodo repository.<sup>86</sup> The augmented MOLOPT basis sets and corresponding auxiliary RI basis sets generated in this work are available in the Supporting Information S3, S4 and in the open-source package CP2K.<sup>41,61</sup> The GW, GW+BSE and TDDFT algorithms employed in this work are available in the open-source package CP2K.<sup>41,61</sup>

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.5c01386>.

Additional benchmark calculations to assess the numerical precision of our developed all-electron augmented

**Table 3. Basis Sizes, Memory Requirements for Storing Three-Center Integrals ( $\mu\nu|P$ ), Eq 38, Their Sparsity, Available RAM and the Execution Time for GW Calculations on Nanographenes from Figure 6<sup>a</sup>**

| $L$ | $N_C$ | $N_H$ | basis set          | $N_{bf}$ | $N_{RI}^H$ | $N_{RI}^C$ | $N_{RI}$ | Occ. of ( $\mu\nu P$ ) | RAM ( $\mu\nu P$ ) (GB) | $N_{nodes}$ | RAM of nodes (GB) | GW execution time (h) |
|-----|-------|-------|--------------------|----------|------------|------------|----------|------------------------|-------------------------|-------------|-------------------|-----------------------|
| 1   | 14    | 12    | aug-SZV-M.-ae-mini | 198      | 2          | 11         | 178      | 100.00%                | 0                       | 1           | 258               | 0.001                 |
| 1   | 14    | 12    | aug-SZV-M.-ae-SR   | 268      | 5          | 18         | 312      | 100.00%                | 0                       | 1           | 258               | 0.003                 |
| 1   | 14    | 12    | aug-SZV-M.-ae      | 268      | 6          | 23         | 394      | 100.00%                | 0                       | 1           | 258               | 0.006                 |
| 1   | 14    | 12    | aug-DZVP-M.-ae     | 528      | 9          | 35         | 598      | 100.00%                | 1                       | 1           | 258               | 0.019                 |
| 1   | 14    | 12    | aug-cc-pVDZ        | 430      | 23         | 72         | 1284     | 100.00%                | 1                       | 1           | 258               | 0.043                 |
| 2   | 28    | 16    | aug-SZV-M.-ae-mini | 348      | 2          | 11         | 340      | 100.00%                | 0                       | 1           | 258               | 0.004                 |
| 2   | 28    | 16    | aug-SZV-M.-ae-SR   | 488      | 5          | 18         | 584      | 100.00%                | 1                       | 1           | 258               | 0.014                 |
| 2   | 28    | 16    | aug-SZV-M.-ae      | 488      | 6          | 23         | 740      | 100.00%                | 1                       | 1           | 258               | 0.027                 |
| 2   | 28    | 16    | aug-DZVP-M.-ae     | 984      | 9          | 35         | 1124     | 100.00%                | 8                       | 1           | 258               | 0.289                 |
| 2   | 28    | 16    | aug-cc-pVDZ        | 788      | 23         | 72         | 2384     | 100.00%                | 11                      | 1           | 258               | 0.373                 |
| 4   | 56    | 24    | aug-SZV-M.-ae-mini | 648      | 2          | 11         | 664      | 70.31%                 | 1                       | 1           | 258               | 0.030                 |
| 4   | 56    | 24    | aug-SZV-M.-ae-SR   | 928      | 5          | 18         | 1128     | 77.07%                 | 5                       | 1           | 258               | 0.141                 |
| 4   | 56    | 24    | aug-SZV-M.-ae      | 928      | 6          | 23         | 1432     | 91.05%                 | 8                       | 1           | 258               | 0.366                 |
| 4   | 56    | 24    | aug-DZVP-M.-ae     | 1896     | 9          | 35         | 2176     | 99.82%                 | 62                      | 1           | 258               | 4.204                 |
| 4   | 56    | 24    | aug-cc-pVDZ        | 1504     | 23         | 72         | 4584     | 99.97%                 | 82                      | 1           | 258               | 4.907                 |
| 8   | 112   | 40    | aug-SZV-M.-ae-mini | 1248     | 2          | 11         | 1312     | 25.33%                 | 4                       | 1           | 258               | 0.151                 |
| 8   | 112   | 40    | aug-SZV-M.-ae-SR   | 1808     | 5          | 18         | 2216     | 28.62%                 | 16                      | 1           | 258               | 0.811                 |
| 8   | 112   | 40    | aug-SZV-M.-ae      | 1808     | 6          | 23         | 2816     | 38.33%                 | 28                      | 1           | 258               | 2.143                 |
| 8   | 112   | 40    | aug-DZVP-M.-ae     | 3720     | 9          | 35         | 4280     | 50.68%                 | 240                     | 1           | 258               | 33.487                |
| 8   | 112   | 40    | aug-cc-pVDZ        | 2936     | 23         | 72         | 8984     | 55.16%                 | 341                     | 1           | 258               | 44.360                |
| 16  | 224   | 72    | aug-SZV-M.-ae-mini | 2448     | 2          | 11         | 2608     | 7.43%                  | 9                       | 1           | 258               | 0.662                 |
| 16  | 224   | 72    | aug-SZV-M.-ae-SR   | 3568     | 5          | 18         | 4392     | 8.45%                  | 37                      | 1           | 258               | 3.843                 |
| 16  | 224   | 72    | aug-SZV-M.-ae      | 3568     | 6          | 23         | 5584     | 11.73%                 | 66                      | 1           | 258               | 9.099                 |
| 16  | 224   | 72    | aug-DZVP-M.-ae     | 7368     | 9          | 35         | 8488     | 16.15%                 | 595                     | 10          | 2577              | 13.034                |
| 16  | 224   | 72    | aug-cc-pVDZ        | 5800     | 23         | 72         | 17,784   | 18.01%                 | 861                     | 10          | 2577              | 18.726                |
| 32  | 448   | 136   | aug-SZV-M.-ae-mini | 4848     | 2          | 11         | 5200     | 2.00%                  | 19                      | 3           | 3060              | 0.808                 |
| 32  | 448   | 136   | aug-SZV-M.-ae-SR   | 7088     | 5          | 18         | 8744     | 2.28%                  | 80                      | 1           | 258               | 13.650                |
| 32  | 448   | 136   | aug-SZV-M.-ae      | 7088     | 6          | 23         | 11,120   | 3.21%                  | 143                     | 5           | 5100              | 6.116                 |
| 64  | 896   | 264   | aug-SZV-M.-ae-mini | 9648     | 2          | 11         | 103,84   | 0.52%                  | 40                      | 3           | 3060              | 2.611                 |
| 64  | 896   | 264   | aug-SZV-M.-ae-SR   | 14,128   | 5          | 18         | 17,448   | 0.59%                  | 164                     | 4           | 1031              | 12.398                |
| 64  | 896   | 264   | aug-SZV-M.-ae      | 14,128   | 6          | 23         | 22,192   | 0.84%                  | 297                     | 5           | 5100              | 18.621                |
| 128 | 1792  | 520   | aug-SZV-M.-ae-mini | 19,248   | 2          | 11         | 20,752   | 0.13%                  | 81                      | 3           | 3060              | 9.080                 |
| 128 | 1792  | 520   | aug-SZV-M.-ae-SR   | 28,208   | 5          | 18         | 34,856   | 0.15%                  | 334                     | 4           | 4080              | 40.488                |
| 128 | 1792  | 520   | aug-SZV-M.-ae      | 28,208   | 6          | 23         | 44,336   | 0.21%                  | 605                     | 5           | 5100              | 65.225                |
| 256 | 3584  | 1032  | aug-SZV-M.-ae-mini | 38,448   | 2          | 11         | 41,488   | 0.03%                  | 163                     | 10          | 16,287            | 4.636                 |
| 512 | 7168  | 2056  | aug-SZV-M.-ae-mini | 76,848   | 2          | 11         | 82,960   | 0.01%                  | 328                     | 20          | 32,573            | 8.928                 |

<sup>a</sup>The available memory for the storage of 3cIs is 6 GB per MPI rank throughout all the listed calculations. Execution time has been measured on the Noctua2 cluster at PC2 computing center in Paderborn, where each node is equipped with two AMD Milan 7763 processors, each providing 64 cores (128 cores per node). For  $L \geq 256$ , the computations have been executed on the Otus cluster at the PC2 computing center, where one node consists of two AMD Turin 9655 processors, each providing 96 cores (192 cores per node). In the basis set names, “M.” stands for “MOLOPT”.

## AUTHOR INFORMATION

### Corresponding Authors

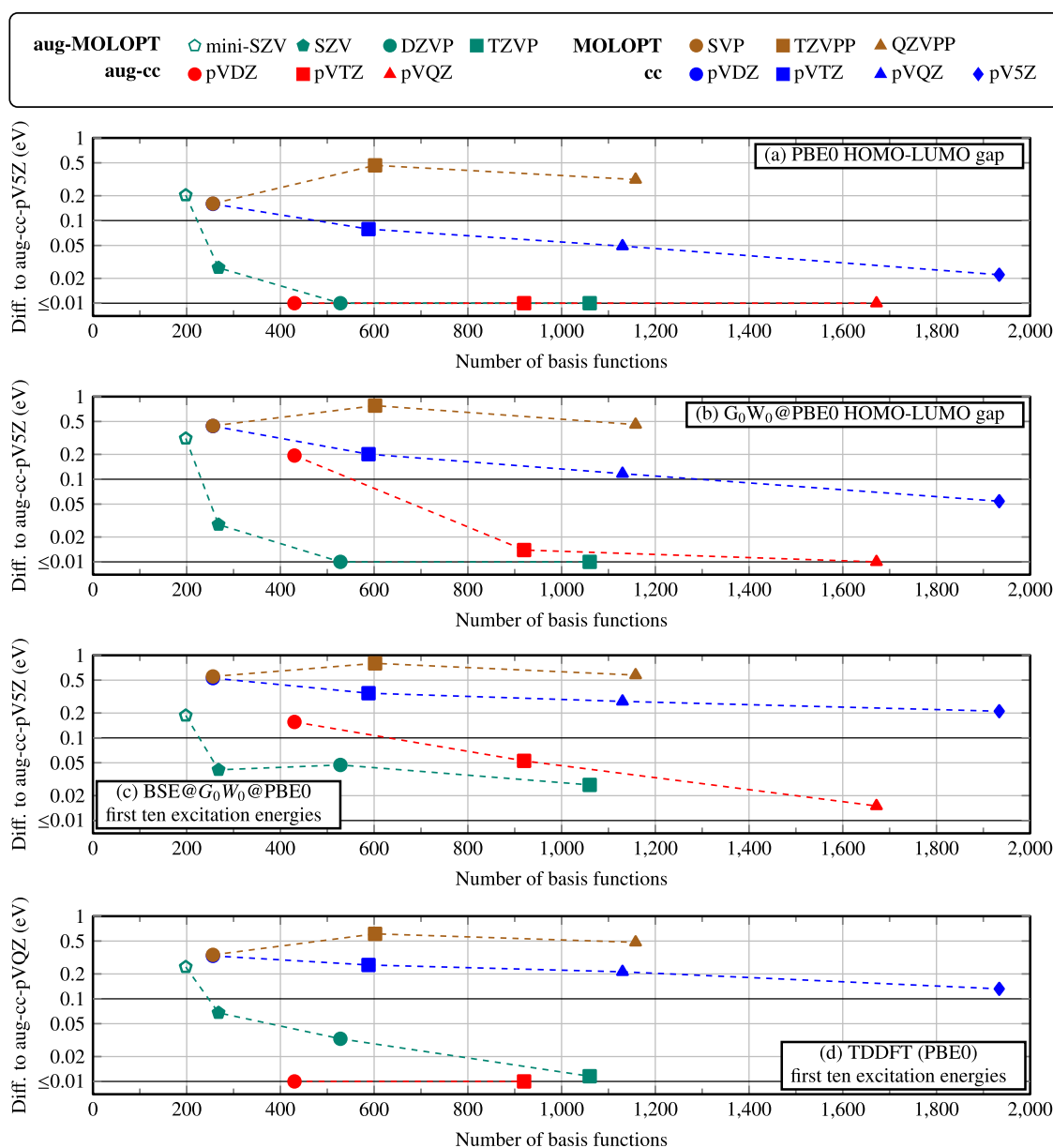
**Rémi Pasquier** – Institute of Theoretical Physics and Regensburg Center for Ultrafast Nanoscopy (RUN), University of Regensburg, 93053 Regensburg, Germany; [orcid.org/0000-0003-0058-4125](https://orcid.org/0000-0003-0058-4125); Email: [remi.pasquier@physik.uni-regensburg.de](mailto:remi.pasquier@physik.uni-regensburg.de)

**Maximilian Graml** – Institute of Theoretical Physics and Regensburg Center for Ultrafast Nanoscopy (RUN), University of Regensburg, 93053 Regensburg, Germany; [orcid.org/0000-0002-4279-8511](https://orcid.org/0000-0002-4279-8511); Email: [maximilian.graml@physik.uni-regensburg.de](mailto:maximilian.graml@physik.uni-regensburg.de)

**Jan Wilhelm** – Institute of Theoretical Physics and Regensburg Center for Ultrafast Nanoscopy (RUN), University of Regensburg, 93053 Regensburg, Germany; Email: [jan.wilhelm@physik.uni-regensburg.de](mailto:jan.wilhelm@physik.uni-regensburg.de)

Complete contact information is available at:

MOLOPT basis sets. We define another benchmark set containing 123 small molecules in Section S1 for benchmarking the augmented MOLOPT basis sets of Li, Be, B, Na, Ca, Al, Si, P, which are only rare in the GW5000 subset used in the main text. We report PBE0 and GW HOMO–LUMO gap and Bethe–Salpeter and TDDFT excitation energies in Section S2 computed with the augmented MOLOPT basis sets and compared to the complete basis set limit. We show additional results on the GW5000 subset with molecules with a LUMO energy below  $-2$  eV (Section S3). We also list excitation gaps obtained with BSE and TDDFT (Section S4). We provide all the newly generated orbital (Section S5) and auxiliary RI (Section S6) basis sets in the CP2K basis set file format (PDF)



**Figure 9.** Basis set convergence of excited-state energies for the 9,10-Dihydroanthracene molecule ( $L = 1$  nanographene geometry sketched in Figure 6a), as a function of the number of orbital basis functions, relative to the aug-cc-pV5Z basis for the aug-MOLOPT basis sets developed in this work, aug-cc-pVXZ,<sup>34</sup> cc-pVXZ<sup>43</sup> and all-electron MOLOPT basis sets.<sup>37,38</sup> Panels show (a) PBE0 HOMO–LUMO gaps, (b)  $G_0W_0@PBE0$  HOMO–LUMO gaps, (c) first ten excitation energies computed from  $BSE@G_0W_0@PBE0$ , and (d) from TDDFT (PBE0). For TDDFT, we use the results from the aug-cc-pVQZ calculations as reference data due to numerical instabilities with aug-cc-pV5Z.

<https://pubs.acs.org/10.1021/acs.jctc.5c01386>

## Author Contributions

<sup>†</sup>R.P. and M.G. contributed equally to this work.

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank Tilo Wettig for an important remark on recomputing sparse tensor elements, which led to the memory-saving scheme described in the App. B. The support of Robert Schade and Xin Wu from the NHR Center Paderborn Center for Parallel Computing (PC2) with optimizing the GW module in CP2K is gratefully acknowledged. We further thank Ritaj Tyagi for helpful comments on the basis sets and Momme Allalen, Gerald

Mathias, Ferdinand Evers, Dorothea Golze, Jürg Hutter, Patrick Rinke and Carlo Pignedoli for helpful discussions. We acknowledge funding by the Free State of Bavaria through the KONWIHR software initiative. DFG is acknowledged for funding via the Emmy Noether Programme (project number 503985532), CRC 1277 (project number 314695032, sub-project A03) and RTG 2905 (project number 502572516). The authors gratefully acknowledge the computing time made available to them on the high-performance computers Noctua<sup>284</sup> and Otus at the NHR Center Paderborn Center for Parallel Computing. This center is jointly supported by the Federal Ministry of Research, Technology and Space and the state governments participating in the National High-Performance Computing (NHR) joint funding program ([www.nhr-verein.de/en/our-partners](http://www.nhr-verein.de/en/our-partners)). The authors also gratefully ac-

knowledge the Gauss Centre for Supercomputing e.V. ([www.gauss-centre.eu](http://www.gauss-centre.eu)) for funding this project by providing computing time on the GCS Supercomputer SuperMUC-NG at Leibniz Supercomputing Centre ([www.lrz.de](http://www.lrz.de)).

## REFERENCES

- (1) Martin, R. M. *Electronic Structure: Basic Theory and Practical Methods*; Cambridge University Press, 2004.
- (2) Behler, J. Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.* **2016**, *145*, No. 170901.
- (3) Jumper, J.; Evans, R.; Pritzel, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589.
- (4) Deringer, V. L.; Bernstein, N.; Csányi, G.; Ben Mahmoud, C.; Ceriotti, M.; Wilson, M.; Drabold, D. A.; Elliott, S. R. Origins of structural and electronic transitions in disordered silicon. *Nature* **2021**, *589*, 59–64.
- (5) Onida, G.; Reining, L.; Rubio, A. Electronic excitations: density-functional versus many-body Green's-function approaches. *Rev. Mod. Phys.* **2002**, *74*, 601–659.
- (6) Dreuw, A.; Head-Gordon, M. Single-Reference ab Initio Methods for the Calculation of Excited States of Large Molecules. *Chem. Rev.* **2005**, *105*, 4009–4037.
- (7) Blase, X.; Duchemin, I.; Jacquemin, D.; Loos, P.-F. The Bethe-Salpeter Equation Formalism: From Physics to Chemistry. *J. Phys. Chem. Lett.* **2020**, *11*, 7371–7382.
- (8) Krausz, F.; Ivanov, M. Attosecond physics. *Rev. Mod. Phys.* **2009**, *81*, 163–234.
- (9) Runge, E.; Gross, E. K. U. Density-Functional Theory for Time-Dependent Systems. *Phys. Rev. Lett.* **1984**, *52*, 997–1000.
- (10) Ullrich, C. A. *Time-Dependent Density-Functional Theory: Concepts and Applications*; Oxford University Press, 2011.
- (11) Hedin, L. New method for calculating the one-particle Green's function with application to the electron-gas problem. *Phys. Rev.* **1965**, *139*, A796.
- (12) Reining, L. The GW approximation: content, successes and limitations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8*, No. e1344.
- (13) Golze, D.; Dvorak, M.; Rinke, P. The GW Compendium: A Practical Guide to Theoretical Photoemission Spectroscopy. *Front. Chem.* **2019**, *7*, No. 377.
- (14) Kohn, W. Density Functional and Density Matrix Method Scaling Linearly with the Number of Atoms. *Phys. Rev. Lett.* **1996**, *76*, 3168–3171.
- (15) Prodan, E.; Kohn, W. Nearsightedness of electronic matter. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 11635–11638.
- (16) Rojas, H. N.; Godby, R. W.; Needs, R. J. Space-Time Method for Ab Initio Calculations of Self-Energies and Dielectric Response Functions of Solids. *Phys. Rev. Lett.* **1995**, *74*, 1827–1830.
- (17) Liu, P.; Kaltak, M.; Klimeš, J.; Kresse, G. Cubic scaling GW: Towards fast quasiparticle calculations. *Phys. Rev. B* **2016**, *94*, No. 165109.
- (18) Lu, J.; Ying, L. Compression of the electron repulsion integral tensor in tensor hypercontraction format with cubic scaling cost. *J. Comput. Phys.* **2015**, *302*, 329–335.
- (19) Lu, J.; Thicke, K. Cubic scaling algorithms for RPA correlation using interpolative separable density fitting. *J. Comput. Phys.* **2017**, *351*, 187–202.
- (20) Duchemin, I.; Blase, X. Separable resolution-of-the-identity with all-electron Gaussian bases: Application to cubic-scaling RPA. *J. Chem. Phys.* **2019**, *150*, No. 174120.
- (21) Duchemin, I.; Blase, X. Cubic-Scaling All-Electron GW Calculations with a Separable Density-Fitting SpaceTime Approach. *J. Chem. Theory Comput.* **2021**, *17*, 2383–2393.
- (22) Delesma, F. A.; Leucke, M.; Golze, D.; Rinke, P. Benchmarking the accuracy of the separable resolution of the identity approach for correlated methods in the numeric atom-centered orbitals framework. *J. Chem. Phys.* **2024**, *160*, No. 024118.
- (23) Pasquier, R.; Camarasa-Gómez, M.; Hehn, A.-S.; Hernangómez-Pérez, D.; Wilhelm, J. Efficient GW band structure calculations using Gaussian basis functions and application to atomically thin transition-metal dichalcogenides. *Phys. Rev.* **2025**, *112*, No. 18411.
- (24) Wilhelm, J.; Golze, D.; Talirz, L.; Hutter, J.; Pignedoli, C. A. Toward GW Calculations on Thousands of Atoms. *J. Phys. Chem. Lett.* **2018**, *9*, 306–312.
- (25) Förster, A.; Visscher, L. Low-Order Scaling  $G_0W_0$  by Pair Atomic Density Fitting. *J. Chem. Theory Comput.* **2020**, *16*, 7381–7399.
- (26) Förster, A.; Visscher, L. Quasiparticle Self-Consistent GW-Bethe-Salpeter Equation Calculations for Large Chromophoric Systems. *J. Chem. Theory Comput.* **2022**, *18*, 6779–6793.
- (27) Graml, M.; Zollner, K.; Hernangómez-Pérez, D.; Faria Junior, P. E.; Wilhelm, J. Low-Scaling GW Algorithm Applied to Twisted Transition-Metal Dichalcogenide Heterobilayers. *J. Chem. Theory Comput.* **2024**, *20*, 2202–2208.
- (28) Delley, B. From molecules to solids with the DMol3 approach. *J. Chem. Phys.* **2000**, *113*, 7756–7764.
- (29) Blum, V.; Gehrke, R.; Hanke, F.; Havu, P.; Havu, V.; Ren, X.; Reuter, K.; Scheffler, M. Ab initio molecular simulations with numeric atom-centered orbitals. *Comput. Phys. Commun.* **2009**, *180*, 2175–2196.
- (30) te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Snijders, J. G.; Ziegler, T. Chemistry with ADF. *J. Comput. Chem.* **2001**, *22*, 931–967.
- (31) Hehre, W. J.; Stewart, R. F.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. I. Use of Gaussian Expansions of Slater-Type Atomic Orbitals. *J. Chem. Phys.* **1969**, *51*, 2657–2664.
- (32) Obara, S.; Saika, A. Efficient recursive computation of molecular integrals over Cartesian Gaussian functions. *J. Chem. Phys.* **1986**, *84*, 3963–3974.
- (33) Golze, D.; Benedikter, N.; Iannuzzi, M.; Wilhelm, J.; Hutter, J. Fast evaluation of solid harmonic Gaussian integrals for local resolution-of-the-identity methods and range-separated hybrid functionals. *J. Chem. Phys.* **2017**, *146*, No. 034105.
- (34) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796–6806.
- (35) VandeVondele, J.; Hutter, J. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J. Chem. Phys.* **2007**, *127*, No. 114105.
- (36) Ye, H.-Z.; Berkelbach, T. C. Correlation-Consistent Gaussian Basis Sets for Solids Made Simple. *J. Chem. Theory Comput.* **2022**, *18*, 1595–1606.
- (37) Hutter, J. Optimized MOLOPT basis sets for PBE DFT calculations, SVP-MOLOPT-PBE-ae and TZVPP-MOLOPT-PBE-ae - Basis sets based on the def2-SVP/def2-TZVPP/def2-QZVPP Ahlrichs sets. [https://github.com/cp2k/cp2k/blob/master/data/BASIS\\_MOLOPT\\_UZH](https://github.com/cp2k/cp2k/blob/master/data/BASIS_MOLOPT_UZH), accessed Dec 2, 2025.
- (38) Müller, T. From Benchmarking to Periodic Fock Exchange in the Auxiliary Density Matrix Method via k-Point Sampling with Gaussian Basis Sets. Ph.D. thesis, University of Zurich, Zürich, 2024.
- (39) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133–A1138.
- (40) Jensen, F. *Introduction to Computational Chemistry*; John Wiley & Sons, 2017.
- (41) Kühne, T. D.; Iannuzzi, M.; Del Ben, M.; et al. CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations. *J. Chem. Phys.* **2020**, *152*, No. 194103.
- (42) Lehtola, S. Curing basis set overcompleteness with pivoted Cholesky decompositions. *J. Chem. Phys.* **2019**, *151*, No. 241102.
- (43) Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- (44) Woon, D. E.; Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. IV. Calculation of static electrical response properties. *J. Chem. Phys.* **1994**, *100*, 2975–2988.
- (45) van Setten, M. J.; Caruso, F.; Sharifzadeh, S.; Ren, X.; Scheffler, M.; Liu, F.; Lischner, J.; Lin, L.; Deslippe, J. R.; Louie, S. G.; Yang, C.; Weigend, F.; Neaton, J. B.; Evers, F.; Rinke, P. GW100: Benchmarking



- $G_0W_0$  for Molecular Systems. *J. Chem. Theory Comput.* **2015**, *11*, 5665–5687.
- (46) Wilhelm, J.; Del Ben, M.; Hutter, J. GW in the Gaussian and Plane Waves Scheme with Application to Linear Acenes. *J. Chem. Theory Comput.* **2016**, *12*, 3623–3635.
- (47) Bruneval, F.; Maliyov, I.; Lapointe, C.; Marinica, M.-C. Extrapolating Unconverged GW Energies up to the Complete Basis Set Limit with Linear Regression. *J. Chem. Theory Comput.* **2020**, *16*, 4399–4407.
- (48) Baum, D.; Visscher, L.; Förster, A. Predicting complete basis set limit quasiparticle energies from triple- $\zeta$  calculations. arXiv:2511.22462. arXiv.org e-Print archive. <https://arxiv.org/abs/2511.22462>. 2025.
- (49) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864–B871.
- (50) Hylleraas, E. A.; Undheim, B. Numerische Berechnung der 2S-Terme von Ortho- und Par-Helium. *Z. Phys.* **1930**, *65*, 759–772.
- (51) MacDonald, J. K. L. Successive Approximations by the Rayleigh-Ritz Variation Method. *Phys. Rev.* **1933**, *43*, 830–833.
- (52) Ren, X.; Rinke, P.; Joas, C.; Scheffler, M. Random-phase approximation and its applications in computational chemistry and materials science. *J. Mater. Sci.* **2012**, *47*, 7447–7471.
- (53) Prascher, B. P.; Woon, D. E.; Peterson, K. A.; Dunning, T. H.; Wilson, A. K. Gaussian basis sets for use in correlated molecular calculations. VII. Valence, core-valence, and scalar relativistic basis sets for Li, Be, Na, and Mg. *Theor. Chem. Acc.* **2011**, *128*, 69–82.
- (54) Woon, D. E.; Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.* **1993**, *98*, 1358–1371.
- (55) Hehre, W. J.; Ditchfield, R.; Stewart, R. F.; Pople, J. A. Self-Consistent Molecular Orbital Methods. IV. Use of Gaussian Expansions of Slater-Type Orbitals. Extension to Second-Row Molecules. *J. Chem. Phys.* **1970**, *52*, 2769–2773.
- (56) Pritchard, B. P.; Altarawy, D.; Didier, B.; Gibbs, T. D.; Windus, T. L.; New Basis Set, A. Exchange: An Open, Up-to-date Resource for the Molecular Sciences Community. *J. Chem. Inf. Model.* **2019**, *59*, 4814–4820.
- (57) Jacquemin, D.; Duchemin, I.; Blase, X. 0–0 Energies Using Hybrid Schemes: Benchmarks of TD-DFT, CIS(D), ADC(2), CC2, and BSE/GW formalisms for 80 Real-Life Compounds. *J. Chem. Theory Comput.* **2015**, *11*, 5340–5359.
- (58) Knysh, I.; Lippert, F.; Blondel, A.; Duchemin, I.; Blase, X.; Loos, P.-F.; Jacquemin, D. Reference CC3 Excitation Energies for Organic Chromophores: Benchmarking TD-DFT, BSE/GW, and Wave Function Methods. *J. Chem. Theory Comput.* **2024**, *20*, 8152–8174.
- (59) Schreiber, M.; Silva-Junior, M. R.; Sauer, S. P. A.; Thiel, W. Benchmarks for electronically excited states: CASPT2, CC2, CCSD, and CC3. *J. Chem. Phys.* **2008**, *128*, No. 134110.
- (60) Powell, M. J. D. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Comput. J.* **1964**, *7*, 155–162.
- (61) CP2K <https://www.cp2k.org>, <https://github.com/cp2k/cp2k>, accessed Dec 2, 2025.
- (62) Vahtras, O.; Almlöf, J.; Feyereisen, M. Integral approximations for LCAO-SCF calculations. *Chem. Phys. Lett.* **1993**, *213*, 514.
- (63) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. Auxiliary basis sets to approximate Coulomb potentials. *Chem. Phys. Lett.* **1995**, *240*, 283–290.
- (64) Sierka, M.; Hogekamp, A.; Ahlrichs, R. Fast evaluation of the Coulomb potential for electron densities using multipole accelerated resolution of identity approximation. *J. Chem. Phys.* **2003**, *118*, 9136–9148.
- (65) Weigend, F.; Furche, F.; Ahlrichs, R. Gaussian basis sets of quadruple zeta valence quality for atoms H–Kr. *J. Chem. Phys.* **2003**, *119*, 12753–12762.
- (66) Eshuis, H.; Yarkony, J.; Furche, F. Fast computation of molecular random phase approximation correlation energies using resolution of the identity and imaginary frequency integration. *J. Chem. Phys.* **2010**, *132*, No. 234114.
- (67) Weigend, F. A fully direct RI-HF algorithm: Implementation, optimized auxiliary basis sets, demonstration of accuracy and efficiency. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285–4291.
- (68) Blase, X.; Attaccalite, C.; Olevano, V. First-principles GW calculations for fullerenes, porphyrins, phthalocyanine, and other molecules of interest for organic photovoltaic applications. *Phys. Rev. B* **2011**, *83*, No. 115103.
- (69) Ren, X.; Rinke, P.; Blum, V.; Wierfink, J.; Tkatchenko, A.; Sanfilippo, A.; Reuter, K.; Scheffler, M. Resolution-of-identity approach to Hartree-Fock, hybrid density functionals, RPA, MP2 and GW with numeric atom-centered orbital basis functions. *New J. Phys.* **2012**, *14*, No. 053020.
- (70) van Setten, M. J.; Weigend, F.; Evers, F. The GW-method for quantum chemistry applications: Theory and implementation. *J. Chem. Theory Comput.* **2013**, *9*, 232–246.
- (71) Stoychev, G. L.; Auer, A. A.; Neese, F. Automatic Generation of Auxiliary Basis Sets. *J. Chem. Theory Comput.* **2017**, *13*, 554–562.
- (72) Lehtola, S. Straightforward and Accurate Automatic Auxiliary Basis Set Generation for Molecular Calculations with Atomic Orbital Basis Sets. *J. Chem. Theory Comput.* **2021**, *17*, 6886–6900.
- (73) Lehtola, S. Automatic Generation of Accurate and Cost-Efficient Auxiliary Basis Sets. *J. Chem. Theory Comput.* **2023**, *19*, 6242–6254.
- (74) Surjuse, K. A.; Valeev, E. F. Physics-Driven Construction of Compact Primitive Gaussian Density Fitting Basis Sets. *J. Chem. Theory Comput.* **2025**, *21*, 10863–10878.
- (75) Weigend, F.; Häser, M.; Patzelt, H.; Ahlrichs, R. RI-MP2: optimized auxiliary basis sets and demonstration of efficiency. *Chem. Phys. Lett.* **1998**, *294*, 143–152.
- (76) Kraka, E.; Gauss, J.; Cremer, D. Determination and use of response densities. *J. Mol. Struct. THEOCHEM* **1991**, *234*, 95–126.
- (77) Weigend, F.; Köhn, A.; Hättig, C. Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *J. Chem. Phys.* **2002**, *116*, 3175–3183.
- (78) Del Ben, M.; Hutter, J.; VandeVondele, J. Electron Correlation in the Condensed Phase from a Resolution of Identity Approach Based on the Gaussian and Plane Waves Scheme. *J. Chem. Theory Comput.* **2013**, *9*, 2654–2671.
- (79) Wilhelm, J.; Seewald, P.; Golze, D. Low-Scaling GW with Benchmark Accuracy and Application to Phosphorene Nanosheets. *J. Chem. Theory Comput.* **2021**, *17*, 1662–1677.
- (80) Jung, Y.; Sodt, A.; Gill, P. M.; Head-Gordon, M. Auxiliary basis expansions for large-scale electronic structure calculations. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6692–6697.
- (81) Hättig, C. Optimization of auxiliary basis sets for RI-MP2 and RI-CC2 calculations: Core-valence and quintuple-zeta basis sets for H to Ar and QZVPP basis sets for Li to Kr. *Phys. Chem. Chem. Phys.* **2005**, *7*, 59–66.
- (82) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (83) Goedecker, S.; Teter, M.; Hutter, J. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B* **1996**, *54*, 1703–1710.
- (84) Bauer, C.; Kenter, T.; Lass, M.; Mazur, L.; Meyer, M.; Nitsche, H.; Riebler, H.; Schade, R.; Schwarz, M.; Winnwa, N.; Wiens, A.; Wu, X.; Plessl, C.; Simon, J. Noctua 2 Supercomputer. *J. Large-Scale Res. Facil.* **2024**, *8*, A187.
- (85) [https://github.com/RemiPasquier/Inputs\\_Outputs\\_Aug\\_Molopt\\_AE\\_benchmark.git](https://github.com/RemiPasquier/Inputs_Outputs_Aug_Molopt_AE_benchmark.git), accessed Dec 2, 2025.
- (86) <https://doi.org/10.5281/zenodo.17878850>, 2025.
- (87) Stuke, A.; Kunkel, C.; Golze, D.; Todorović, M.; Margraf, J. T.; Reuter, K.; Rinke, P.; Oberhofer, H. Atomic structures and orbital energies of 61,489 crystal-forming organic molecules. *Sci. Data* **2020**, *7*, No. 58.
- (88) Iannuzzi, M. et al. The CP2K Program Package Made Simple. arXiv:2508.15559. arXiv.org e-Print archive. <https://arxiv.org/abs/2508.15559>. 2025.
- (89) Lippert, G.; Hutter, J.; Parrinello, M. The Gaussian and augmented-plane-wave density functional method for ab initio

molecular dynamics simulations. *Theor. Chem. Acc.* **1999**, *103*, 124–140.

(90) Graml, M.; Wilhelm, J. Optical Excitations in Nanographenes from the Bethe-Salpeter Equation and Time-Dependent Density Functional Theory: Absorption Spectra and Spatial Descriptors. arXiv:2510.25658. arXiv.org e-Print archive. <https://arxiv.org/abs/2510.25658>. 2025.

(91) Hehn, A.-S.; Sertcan, B.; Belleflamme, F.; Chulkov, S. K.; Watkins, M. B.; Hutter, J. Excited-State Properties for Extended Systems: Efficient Hybrid Density Functional Methods. *J. Chem. Theory Comput.* **2022**, *18*, 4186–4202.

(92) Takatsuka, A.; Ten-no, S.; Hackbusch, W. Minimax approximation for the decomposition of energy denominators in Laplace-transformed Møller-Plesset perturbation theories. *J. Chem. Phys.* **2008**, *129*, No. 044112.

(93) Hackbusch, W. Computation of best  $L^\infty$  exponential sums for  $1/x$  by Remez' algorithm. *Comput. Visualization Sci.* **2019**, *20*, 1–11.

(94) Azizi, M.; Wilhelm, J.; Golze, D.; Giantomassi, M.; Panadés-Barrueta, R. L.; Delesma, F. A.; Bucchieri, A.; Gulans, A.; Rinke, P.; Draxl, C.; Gonze, X. Time-frequency component of the GreenX library: minimax grids for efficient RPA and GW calculations. *J. Open Source Software* **2023**, *8*, 5570.

(95) Azizi, M.; Wilhelm, J.; Golze, D.; Delesma, F. A.; Panadés-Barrueta, R. L.; Rinke, P.; Giantomassi, M.; Gonze, X. Validation of the GreenX library time-frequency component for efficient GW and RPA calculations. *Phys. Rev. B* **2024**, *109*, No. 245101.

(96) Momma, K.; Izumi, F. VESTA 3 for three-dimensional visualization of crystal, volumetric and morphology data. *J. Appl. Crystallogr.* **2011**, *44*, 1272–1276.

(97) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158–6170.

(98) Vèril, M.; Romaniello, P.; Berger, J. A.; Loos, P.-F. Unphysical Discontinuities in GW Methods. *J. Chem. Theory Comput.* **2018**, *14*, 5220–5228.

(99) Schambeck, M.; Golze, D.; Wilhelm, J. Solving multipole challenges in the GW100 benchmark enables precise low-scaling GW calculations. *Phys. Rev. B* **2024**, *110*, No. 125146.

(100) Hedin, L. On correlation effects in electron spectroscopies and the GW approximation. *J. Phys.: Condens. Matter* **1999**, *11*, R489.

(101) Li, J.; Jin, Y.; Rinke, P.; Yang, W.; Golze, D. Benchmark of GW Methods for Core-Level Binding Energies. *J. Chem. Theory Comput.* **2022**, *18*, 7570–7585.

(102) Valeev, E. F. A library for the evaluation of molecular integrals of operators over Gaussian functions. <https://github.com/evaleev/libint>. accessed Dec 2, 2025.



The image is a promotional graphic for CAS Insights. It features a collage of scientific and technological images, including a person in a lab coat, molecular structures, and data visualizations. The text 'CAS INSIGHTS™' is prominently displayed at the top. Below it, the main headline reads 'EXPLORE THE INNOVATIONS SHAPING TOMORROW'. A sub-headline states: 'Discover the latest scientific research and trends with CAS Insights. Subscribe for email updates on new articles, reports, and webinars at the intersection of science and innovation.' A yellow button with the text 'Subscribe today' is located below the sub-headline. At the bottom right, the CAS logo is shown, with the text 'A division of the American Chemical Society' underneath it.