# Correlation-consistent Gaussian basis sets for solids made simple

Hong-Zhou Ye[1, a)] and Timothy C. Berkelbach[1, 2, b)]

[1)] *Department of Chemistry, Columbia University, New York, New York 10027, USA*
[2)] *Center for Computational Quantum Physics, Flatiron Institute, New York, New York 10010, USA*

The rapidly growing interest in simulating condensed-phase materials using quantum chemistry methods calls for a library of high-quality Gaussian basis sets suitable for periodic calculations. Unfortunately, most standard Gaussian basis sets commonly used in molecular simulation show significant linear dependencies when used in close-packed solids, leading to severe numerical issues that hamper the convergence to the complete basis set (CBS) limit, especially in correlated calculations. In this work, we revisit Dunning's strategy for construction of correlation-consistent basis sets and examine the relationship between accuracy and numerical stability in periodic settings. We find that limiting the number of primitive functions avoids the appearance of problematic small exponents while still providing smooth convergence to the CBS limit. As an example, we generate double-, triple-, and quadruple-zeta correlation-consistent Gaussian basis sets for periodic calculations with Goedecker-Teter-Hutter (GTH) pseudopotentials. Our basis sets cover the main-group elements from the first three rows of the periodic table. Especially for atoms on the left side of the periodic table, our basis sets are less diffuse than those used in molecular calculations. We verify the fast and reliable convergence to the CBS limit in both Hartree-Fock and post-Hartree-Fock (MP2) calculations, using a diverse test set of 19 semiconductors and insulators.

TOC graphic
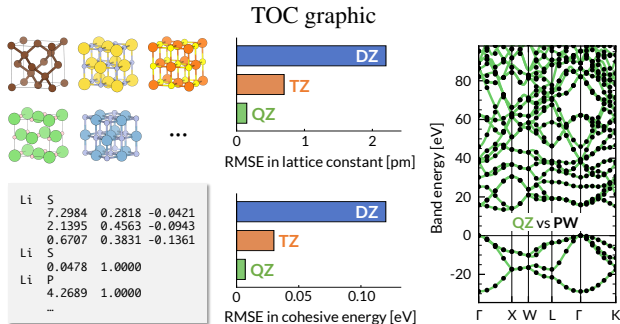
I. INTRODUCTION

Recent years have witnessed a rapid growth of interest in understanding condensed-phase materials using quantum chemistry methods,[1–13] especially those beyond density functional theory[14,15] (DFT) with local and semi-local exchange-correlation functionals.[16,17] The non-local exchange and the many-body electron correlation in the quantum chemistry methods promise many advantages, such as systematic improvability[7] and the ability to describe dispersion interactions[18–20] and strong electron correlations,[21,22] but are also computationally demanding, especially when using a plane-wave (PW) basis set due to the slow convergence with the number of virtual bands.[1,23–30]

Atom-centered Gaussian basis sets are most popular in molecular quantum chemistry,[31] where the correlation-consistent basis sets[32] allow systematic convergence to the complete basis set (CBS) limit in correlated calculations.[33–36] However, the properties that define a good basis set for molecules are not the same as those for periodic solids. For

example, the standard Gaussian basis sets,[37–39] often optimized on free atoms, contain relatively diffuse functions that are needed to correctly describe the wavefunction in distant regions of space. These diffuse functions cause significant linear dependencies when used in periodic calculations,[40–43] leading to numerical instabilities in the self-consistent field (SCF) calculations.[44] While this SCF convergence issue can sometimes be solved by discarding diffuse primitives in the basis set[44] or by canonical orthogonalization,[45] these modifications hinder reproducibility, cause discontinuities in potential energy surfaces, and degrade the quality of virtual orbitals, which affects subsequent correlated calculations. More importantly, the linear dependency problem is worse for larger basis sets, preventing convergence of a periodic calculation to the CBS limit.

One way to mitigate the linear dependency issue is re-optimizing the Gaussian exponents ζ_i and contraction coefficients c_i of existing basis sets based on a cost function such as[41,46–49]

$$\Omega(\{\zeta_i, c_i\}; \gamma) = E(\{\zeta_i, c_i\}) + \gamma \log \text{cond} \mathbf{S}(\{\zeta_i, c_i\}), \quad (1)$$

the minimization of which trades some of the energy E for a lower condition number of the basis set overlap matrix $\mathbf{S}$ to the extent controlled by the developer-selected parameter $\gamma > 0$. Such a cost function can be minimized on a paradigmatic system that exhibits the linear dependency of concern[41,47] or on each system under study,[46,48,49] and recent works have demonstrated the success of such approaches for producing Gaussian basis sets with better behavior.[46–49] However, aside from the extra cost associated with frequent basis set reoptimization, such approaches obviously hinder—or forfeit—transferability and reproducibility.

In this work, we take a different approach to constructing Gaussian basis sets for periodic systems, which are designed to be universal and transferable, by revisiting Dunning's strategy for generating correlation-consistent Gaussian basis sets.[32,50] The key modification needed for extended systems is found to be restricting the size of the valence basis to

[a)]Electronic mail: hzyechem@gmail.com
[b)]Electronic mail: tim.berkelbach@gmail.com

reach a balance between the accuracy and the numerical stability of a basis set. Our strategy is general and applies to both all-electron and pseudopotential-based calculations; as a specific example, we use the Goedecker-Teter-Hutter (GTH) family of pseudopotentials[51,52] optimized for Hartree-Fock[53] (HF) calculations and generate correlation-consistent Gaussian basis sets up to the quadruple-zeta (QZ) level for the main-group elements from the first three rows of the periodic table. The resulting GTH-cc-pVXZ (X = D, T, and Q) basis set series show fast convergence to the CBS limit (verified using a PW basis with the same pseudopotential) on the bulk properties of 19 semiconductors calculated at both mean-field (HF) and correlated (second-order Møller-Plesset perturbation theory,[54] MP2) levels.

This article is organized as follows. In Section II, we review Dunning's original scheme for generating correlation-consistent basis sets and provide a high-level description of our adaptation for periodic systems, using the GTH-cc-pVXZ basis set of carbon as an illustrative example. In Section III, we describe the computational details of the basis optimization and the numerical tests. In Section IV, we evaluate the quality of the GTH-cc-pVXZ basis sets on a variety of bulk properties by comparing to results obtained with a PW basis. In Section V, we conclude this work by pointing out future directions.

II. METHODOLOGY

A. Correlation consistent basis sets

We begin with a brief review of Dunning's original approach to constructing the cc-pVXZ basis set series for the main-group elements,[32] which our strategy closely follows. A cc-pVXZ basis set consists of a valence basis and a set of polarization functions. The valence basis has s and p primitive orbitals (only s for hydrogen and helium) whose exponents are determined by minimizing the HF ground state energy of a free atom. These optimized primitive orbitals are then contracted with coefficients obtained from spherically averaging the atomic HF orbitals. The most diffuse one, two, etc. primitive orbitals are freed from the contraction for DZ, TZ, etc. to better describe the electron density in the bonding region of a molecule. The size of the valence basis (i.e., the number of primitive s and p orbitals) is typically chosen by the desired accuracy of the atomic HF energy.

The polarization functions are primitive orbitals of d angular momentum or higher (p or higher for hydrogen and helium) whose exponents are determined by minimizing the correlation energy of a free atom. The rule of correlation consistency—an empirical observation first made by Dunning[32] and later confirmed by other[50,55,56]—states that the increase in the magnitude of the correlation energy $|\Delta E_c|$ obtained by adding the nth polarization function of angular momentum l is roughly equal to that of adding the $(n-1)$th polarization function of angular momentum $l+1$. For this reason, the polarization functions are added in groups, $1d$ for DZ, $2d1f$ for TZ, $3d2f1g$ for QZ, etc., and the correlation energies
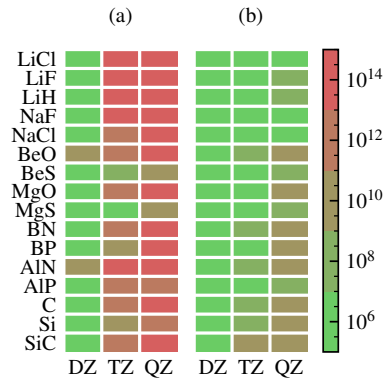

FIG. 1. Condition number of (a) the GTH-XZVP (taken from the CP2K software package[57]) and (b) the GTH-cc-pVXZ (this work) basis sets evaluated on 16 three-dimensional bulk materials at their respective experimental lattice parameters using a $5 \times 5 \times 5$ k-point mesh (the maximum condition number from all k-points is plotted). For materials containing s-block elements, the condition numbers from the small-core and the large-core GTH-cc-pVXZ basis sets are comparable and the former is shown here.

obtained with the cc-pVXZ series can often be extrapolated to the CBS limit using simple functional forms.[33,34]

B. The linear dependency problem

The primary problem that must be solved for generating Gaussian basis sets for solids is the potential high linear dependency of the basis sets,[40–43] which is particularly severe for three-dimensional solids. This is illustrated in Fig. 1(a) for the original GTH-XZVP basis sets[58] on 16 bulk solids composed of the main-group elements from the first three rows. The GTH-XZVP basis sets, which were first reported in Ref. 58 and are now distributed with the CP2K package,[57] were constructed by combining a valence basis optimized on free atoms at the DFT level and polarization functions of d angular momentum taken from the corresponding cc-pVXZ basis sets. In practice, a condition number higher than 10^{10} is found to be problematic in the manners discussed in Section I. As a result, most solids listed in Fig. 1(a) can only be studied at the DZ level when using the original GTH basis sets. The situation is similar for other all-electron[32,39,50] or pseudopotential-based basis sets,[59,60] including the cc-pVXZ series,[32,50] as shown in Fig. S1. We emphasize that this issue is mostly isolated to hard crystalline solids and is less severe for molecular solids or liquids.

C. Balancing accuracy and numerical stability

In our approach to basis set design for solids, we control the linear dependency by limiting the size of the valence basis, being careful not to introduce large basis set incompleteness errors. In this section, we use the carbon element with a GTH pseudopotential as an example to discuss how a balance
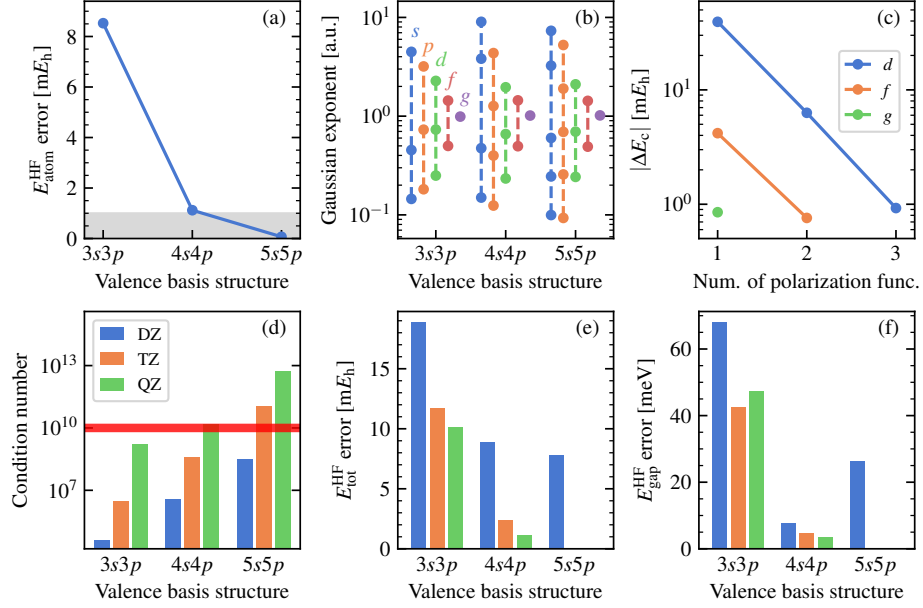
FIG. 2. Correlation consistent basis sets generated for carbon in the GTH pseudopotential using three different valence bases. (a) Convergence of the atomic HF energy. The grey shaded area indicates an error below 1 mE_h. (b) Gaussian exponents of the QZ basis sets. Different colors label shells of different angular momentum as indicated by the text in the corresponding color. (c) Increment of the atomic CCSD correlation energy with the number of polarization functions in each angular momentum channel for the $4s4p$ valence basis. The $3s3p$ and $5s5p$ valence bases give virtually the same plot (not shown). (d) Condition number of the basis overlap matrix evaluated for bulk diamond at the experimental geometry (the maximum condition number from a $5 \times 5 \times 5$ k-point mesh is plotted). The red horizontal line highlights a condition number of 10^{10}. (e) Error of the per-cell HF energy of bulk diamond evaluated using the Gaussian basis sets against a PW benchmark calculation. (f) Same as (e) for the HF band gap of diamond.

between accuracy and numerical stability can be reached. We postpone a discussion of computational details to Section III.

Figure 2(a) shows the error in the atomic HF energy of carbon with three optimized valence bases of increasing size (i.e., number of primitives): $3s3p$, $4s4p$, and $5s5p$. A relatively small basis of $4s4p$ achieves an error of about 1 mE_h, and that of $5s5p$ is already below 0.1 mE_h. For each of the three valence bases, we generate correlation-consistent DZ, TZ, and QZ basis sets, by optimizing the polarization functions based on a correlated calculation. The optimized exponents of the QZ primitives are shown in Fig. 2(b). Focusing on the s primitives, we note that the *largest* exponent splits into two from $3s$ to $4s$, but the *smallest* exponent splits into two from $4s$ to $5s$. Only the latter is consistent with the conventional "split-valence" picture. We will see that this split-valence structure yields unacceptably high condition numbers in solids, but is not necessary for accurate predictions. Despite the difference in the underlying valence basis, the optimized exponents for the polarization functions share a similar structure [Fig. 2(b)] and all exhibit perfect correlation consistency as shown explicitly for the $4s4p$ case in Fig. 2(c).

To test the applicability and the performance of the nine correlation-consistent bases obtained above in periodic calculations, we consider bulk diamond with its experimental lattice constant. The condition numbers plotted in Fig. 2(d) exhibit a quick and monotonic increase with both the zeta-level and the size of the valence basis. As a result, we ob-

served convergence issues in the SCF calculations using the $5s5p$-derived TZ and QZ basis sets, consistent with the high condition numbers of the two basis sets (greater than 10^{10} as indicated by the red line), which is due to the presence of exponents near 0.1 or below for this system.

For the remaining seven basis sets (DZ to QZ for $3s3p$ and $4s4p$ and DZ for $5s5p$), we used HF and MP2 to calculate various structural and energetic properties of diamond, which were then compared with benchmark results obtained using a PW basis. For most of these properties, the performance of the correlation-consistent basis sets at the same zeta-level shows a weak dependence on the underlying valence basis (Figs. S2 and S3). However, the $4s4p$ family is a clear winner on the more sensitive properties, including the HF total energy and the HF band gap as shown in Fig. 2(e-f); large residual errors persist in the $3s3p$-derived basis sets, even at the QZ level, and little improvement is gained in the $5s5p$-derived basis sets (at least at the DZ level which is the only one we can test due to their high linear dependencies).

The results in this section reveal the strong effect of the choice of a valence basis on the accuracy and the numerical stability of the resulting correlation-consistent basis sets. In particular, we see that limiting the size of the valence basis can preclude the appearance of problematic diffuse functions without significantly compromising the accuracy of calculated properties. In the next section, we extend the strategy used here to obtain correlation-consistent basis sets for all main-

group elements from the first three rows of the periodic table.

D. The GTH-cc-pV*X*Z basis sets

In the following subsections, we describe the detailed construction of our GTH-cc-pV*X*Z basis sets (X = D, T, and Q) for the first three rows of the periodic table. For selected elements (*vide infra*), we also construct basis sets augmented by diffusion functions (GTH-aug-cc-pV*X*Z). All of our basis sets are available for download in an online repository[61] and full details of their primitive and contracted structure is given in Table S1 and Figs. S4 and S5.

1. The valence basis

The scheme for choosing the optimal valence basis for the carbon atom with a GTH pseudopotential, presented in Section II C, can be made general. Based on atomic calculations, we generate candidate correlation-consistent basis sets using valence bases of multiple sizes, which are then tested on a few reference materials. The final basis set is then chosen as the one that remains numerically stable while predicting bulk properties that are converged with the size of the valence basis (or as converged as possible before large linear dependencies arise). In contrast to Dunning's original approach, we use these primitives for all zeta levels, with the final valence basis differing only in the number of uncontracted functions (however, see II D 3 for a modification of this procedure for group VI to VIII elements). We emphasize that the reference periodic system serves only as a *guide* for choosing an appropriate valence basis and is not used in the optimization of any parameters, unlike in previous works based on Eq. (1).[41,46–49] As we will see in the numerical results (Section IV), our scheme maintains the important atomic electronic structure of a basis set, which is crucial for its transferability and high accuracy.

The structure of the valence basis determined this way for all elements from the first three rows of the periodic table is summarized in Table I. The reference systems are chosen to be simple semiconductors and insulators formed by these elements. The bulk properties being monitored include the equilibrium lattice constant and bulk modulus evaluated at both the HF and the MP2 levels and the HF band gap at equilibrium geometry, all evaluated with the Brillouin zone sampled by a $3 \times 3 \times 3$ k-point mesh. The s-block elements, Li, Be, Na, and Mg, can be simulated using either a large-core pseudopotential or a small-core pseudopotential (which differ according to the treatment of core/semi-core electrons), and we present optimized basis sets for them separately. Furthermore, for these s-block elements, the exponents of the valence p orbitals cannot be determined in the usual way because these orbitals are unoccupied in the atomic HF ground state. In the literature, these exponents are often determined by minimizing the HF energy of the corresponding $s \rightarrow p$ valence excited state.[56] We found that this approach leads to valence p orbitals that are too diffuse and cause significant linear dependencies in the oxides and fluorides of these elements using the QZ ba-
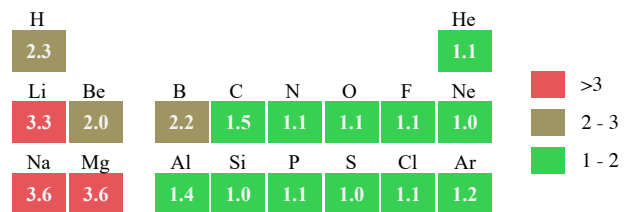

FIG. 3. Compactness of the valence basis developed in this work, defined as the ratio of the smallest exponent in our basis to that in the all-electron cc-pV5Z basis set for the same element. For s-block elements, the compactness of the valence basis optimized for the small-core and the large-core pseudopotentials is comparable, and we only show the results for the former.

sis sets. Therefore, we optimize the valence p orbitals at the correlated level in the same way as in the determination of the polarization functions, described more in the next section. We verified that the DZ and TZ basis sets obtained from both schemes give very similar numerical results for all properties tested in Section IV.

The size of the valence basis is found to correlate with the hardness of the underlying pseudopotentials as reported in previous work[47] but is otherwise smaller than the valence bases in the original GTH-TZVP and QZVP basis sets. For this reason, the condition numbers of the new basis sets are significantly reduced compared to the original GTH-*X*ZVP series as shown in Fig. 1(b). This can also be seen from the *compactness* of our valence basis (Fig. 3), defined as the ratio of the smallest exponent in our valence basis to that in the corresponding all-electron cc-pV5Z basis set (geometric mean is taken in case of multiple angular momentum channels). From this perspective, the constraint on the size of the valence basis is strongest for the s-block metals (compactness > 3), which explains the relatively large error of the atomic HF energy for these elements, but gradually relaxed for elements of higher group numbers (compactness $\approx$ 1). As we will see by thorough numerical tests in Section IV, this way of constraining the size of the valence basis does not degrade the performance of the full correlation-consistent basis sets in bulk calculations. The compactness of the valence basis in Fig. 3 also provides a practical guide to constructing valence bases of similar quality for other nuclear potentials (including the full Coulomb potential for all-electron calculations).

2. The polarization functions

We applied Dunning's scheme for determining the polarization functions up to QZ (i.e., $3d2f1g$) without modifications to the elements from group III to VIII. However, we used coupled-cluster theory with single and double excitations (CCSD)[62] instead of the more common configuration interaction with single and double excitations (CISD),[32,50,56] for the calculation of the correlation energy. The increment of the correlation energy with the number of polarization functions follows the rule of correlation consistency for all these elements. Plots similar to Fig. 2(c) for other atoms are shown in

TABLE I. The active electrons (not covered by the pseudopotential), the valence basis structure (i.e., number of primitives), the errors of the atomic HF energy (in mE_h, evaluated as the energy difference from a sufficiently large valence basis), and the reference systems on which the basis linear dependency and accuracy are monitored for the GTH-cc-pVXZ basis sets of all elements studied in this work. For the s-block elements, the valence bases are listed for both the large-core and the small-core pseudopotentials. The valence p orbitals of the s-block elements are obtained in a different manner (see the discussion at the end of Section II D 1) and denoted by a "+" sign in the table.

Element	Active electrons	Valence basis	HF error	Ref. sys.
H	$[1s^1]$	$4s$	0.19	LiH
He	$[1s^2]$	$5s$	0.05	solid He
Li	$[2s^1]$	$2s + 2p$	1.24	LiH, LiF, LiCl
	$[1s^2 2s^1]$	$4s + 4p$	2.39	LiH, LiF, LiCl
Be	$[2s^2]$	$3s + 3p$	0.12	BeO, BeS
	$[1s^2 2s^2]$	$5s + 4p$	1.53	BeO, BeS
B	$[2s^2 2p^1]$	$3s3p$	8.15	BN, BP
C	$[2s^2 2p^2]$	$4s4p$	1.13	diamond, SiC
N	$[2s^2 2p^3]$	$5s5p$	0.19	BN, AlN
O	$[2s^2 2p^4]$	$5s5p$	0.41	BeO, MgO
F	$[2s^2 2p^5]$	$5s5p$	0.72	LiF, NaF
Ne	$[2s^2 2p^6]$	$6s6p$	0.16	solid Ne
Na	$[3s^1]$	$2s + 2p$	0.93	NaF, NaCl
	$[2s^2 2p^6 3s^1]$	$5s5p + 1p$	2.44	NaF, NaCl
Mg	$[3s^2]$	$2s + 1p$	4.32	MgO, MgS
	$[2s^2 2p^6 3s^2]$	$4s4p + 1p$	17.27	MgO, MgS
Al	$[3s^2 3p^1]$	$3s2p$	0.62	AlN, AlP
Si	$[3s^2 3p^2]$	$4s4p$	0.05	Si, SiC
P	$[3s^2 3p^3]$	$4s4p$	0.13	AlN, AlP
S	$[3s^2 3p^4]$	$5s4p$	0.18	BeS, MgS
Cl	$[3s^2 3p^5]$	$5s4p$	0.24	LiCl, NaCl
Ar	$[3s^2 3p^6]$	$4s4p$	0.76	solid Ar

Fig. S6.

The s-block elements (including hydrogen) need some special treatments. The hydrogen atom only has one electron and hence no correlation energy. We follow ref 56 and minimize the correlation energy of a H_2 molecule with the experimental bond length of 0.7414 Å,[63] which gives polarization functions showing good correlation consistency [Fig. S6(a)].

For s-block elements with small-core pseudopotentials, the $[1s^2]$ core electrons (for Li and Be) or the $[2s^2 2p^6]$ semi-core electrons (for Na and Mg) are frozen in the CCSD calculations in order to determine the polarization functions that only account for the valence electron correlation, following ref 56. With this choice, the same number of electrons are correlated when using either the large-core or the small-core pseudopotentials, leading to similar exponents for the optimized polarization functions for both pseudopotentials. For Li and Na with only one correlated valence electron, we apply a similar treatment as we did for the hydrogen atom and minimize the correlation energy of a Li_2 molecule and a Na_2 molecule with the respective experimental bond lengths of 2.673 Å and 2.303 Å.[63] However, we found significant linear dependencies for both elements in periodic calculations upon adding the second polarization function in each angular momentum channel. We thus choose a $1d$, $1d1f$, and $1d1f1g$ structure

for the *valence*-correlated DZ, TZ, and QZ polarization functions of these elements. For Be and Mg, the regular structure of polarization functions is kept, but deviation from ideal correlation consistency is observed, wherein the third d function recovers much less correlation energy than the second f function [Figs. S6(e,f,o,p)]. Similar observations have also been reported for these elements in the all-electron cases.[56]

Despite the use of the CCSD correlation energy and a pseudopotential, the exponents of the polarization functions in our GTH-cc-pVXZ basis sets in general agree very well with those in the all-electron cc-pVXZ basis sets (Figs. S4 and S5). The few exceptions come from the third-row elements such as Al and Si, where our CCSD-optimized $3d$ set (for GTH-cc-pVQZ) shows a larger exponent splitting from the $2d$ set (for GTH-cc-pVTZ) than the corresponding $3d$ set in the all-electron cc-pVQZ basis sets. In these cases, we verified explicitly that very similar exponents are obtained by minimizing the CISD correlation energy instead. Thus, the observed difference between our GTH-cc-pVXZ series and the cc-pVXZ series for these elements is due to the pseudopotential.

3. Contraction of valence orbitals

In an all-electron cc-pVXZ basis set, the primitive orbitals in the valence basis describe both the core and the valence electrons of an atom, with those for the core being contracted in the way discussed in Section II A to reduce the computational cost. We follow this rule formally in constructing the GTH-cc-pVXZ basis sets for most elements, where the most diffuse one, two, and three primitive orbitals in each angular momentum channel of the valence basis are released from the contraction for X = D, T, and Q. (In case of, e.g., three primitive orbitals in an angular momentum channel, TZ and QZ will have the same valence basis, which is fully uncontracted in that angular momentum channel.) However, for group VI to VIII elements where the size of the valence basis is only weakly constrained (Fig. 3), the procedure above results in suboptimal performance, especially at the DZ level. For these elements, we instead augment the fully contracted valence basis with $1s1p$, $2s2p$, and $3s3p$ primitives (s-only for helium), determined separately from minimizing the atomic correlation energy (in the presence of the polarization functions), to make the DZ, TZ, and QZ basis sets.

The contraction coefficients determined from atomic HF orbitals might have limited transferability especially for group III to VIII elements and s-block elements with large-core pseudopotentials, because they have no core electrons. For this reason, we also generate valence-uncontracted DZ and TZ basis sets where the valence basis is made to match higher zeta-levels. For example, a GTH-cc-pV(T)DZ basis set has the polarization functions taken from GTH-cc-pVDZ and the valence basis from GTH-cc-pVTZ. We will see the importance of such valence uncontraction in Section IV in the calculation of virtual bands.

4. Extensions

Like the all-electron cc-pV*XZ* series, our basis sets can be straightforwardly extended by core-valence correlating functions,[64,65] tight *d* functions,[66] etc. In this work, we explore one such extension, namely the augmentation with diffuse functions,[67] which may be appropriate for the simulation of molecular crystals[3,6] or surface phenomena.[11,68,69] Particularly, we find that the noble gas solids, which are used as the reference materials for the evaluation of our noble gas basis sets, benefit substantially from augmentation with diffuse functions. We thus augment all Gaussian basis sets for the three noble gas elements used in this work by adding one diffuse function to each angular momentum channel. Because of the low density of noble gas solids, the condition numbers of their overlap matrices are 10^7 or less, even after augmentation. The exponent of this augmentation function is chosen to be proportional to the exponent of the most diffuse function in the non-augmented basis: $\alpha_{aug} = x\alpha_{min}$, where x is the analogous ratio of exponents in the all-electron aug-cc-pV*XZ* basis set for the same element.[67] We name these basis sets "GTH-aug-cc-pV*XZ*".

III. COMPUTATIONAL DETAILS

The protocol for generating correlation-consistent basis sets described in Section II was followed, and all calculations are performed using the PySCF software package.[70,71] The spin-restricted (or spin-restricted open-shell) HF and the spin-unrestricted CCSD are used to optimize the valence basis and the polarization functions, respectively. For periodic calculations, the recently developed range-separated Gaussian density fitting[72,73] (RSGDF) is used to handle the electron repulsion integrals. The density fitting auxiliary basis is an even-tempered Gaussian basis with a progression factor $\beta = 2.0$ (generated automatically by PySCF). Finite-size errors associated with the divergence of the HF exchange integral at $G = 0$ are handled using a Madelung constant, as described in ref 74–76. The basis set parameters are optimized using the Nelder-Mead algorithm[77] (as implemented in the SciPy library[78]), which we found to give the same exponents as the Broyden-Fletcher-Goldfarb-Shanno[79] or the conjugate gradient[80] algorithms (as used in previous work[46,48]) in most cases, but to be more robust against local minima in more challenging situations.

We assess the quality of the GTH-cc-pV*XZ* basis sets along with their valence-uncontracted counterparts at both HF and MP2 levels on a test set of 19 three-dimensional bulk systems listed in Table S2 (the 16 materials shown in Fig. 1 plus the solids of helium, neon, and argon). Results from the original GTH-DZVP basis sets, which show no linear dependency issues on these materials, will also be reported. For the three noble gas elements, we augment the GTH-DZVP basis with extra diffuse functions in the same manner as described in Section II D 4. For systems containing *s*-block elements that have small-core and large-core pseudopotentials, calculations in the GTH-cc-pV*XZ* family use the corresponding small-core

and large-core basis sets that we developed, while those in the GTH-DZVP family use the small-core basis sets for both pseudopotentials, because large-core GTH-DZVP basis sets do not exist.

The selected bulk properties include the cell energy (E_{cell}), the cohesive energy (E_{coh}), the band gap (E_{gap}) and band structure (only at the HF level), and the equilibrium lattice constant (a_0) and bulk modulus (B_0). For E_{tot}, E_{coh}, and E_{gap}, single-point calculations at experimental geometries are performed with the Brillouin zone sampled using a $5 \times 5 \times 5$ k-point mesh (evenly spaced and Γ-point included) without further extrapolation. The cohesive energy is counterpoise corrected for basis set superposition error. The band structure is obtained by performing individual single-point calculations using a $3 \times 3 \times 3$ k-point mesh shifted along a chosen k-point path. The a_0 and B_0 are obtained by scanning the lattice constant around the HF minimum and fitting the total energy curve to the Birch-Murnaghan equation of state.[81,82] A $3 \times 3 \times 3$ k-point mesh is used in all these calculations, except the results in Table II, which were calculated using a $5 \times 5 \times 5$ k-point mesh to facilitate comparison with literature values.

The errors in the above properties due to the Gaussian basis set incompleteness are determined by comparison to calculations with a PW basis made large enough to essentially achieve the CBS limit; the HF total energy, the HF band energy, and the MP2 energy are all converged to an accuracy better than 0.1 meV/cell. We note that the k-point meshes used in all of our calculations are sufficiently large to eliminate finite-size effects in *all* of the basis set errors and in *most* (but not all) of the predicted properties. In our PW calculations, the HF exchange is calculated with the adaptively compressed exchange (ACE) operator.[83] To converge the MP2 correlation energy to the CBS limit, we extrapolate according to the asymptotic behavior[25]

$$E_{corr}^{MP2}(n_{vir}) = An_{vir}^{-1} + E_{corr}^{MP2}(\infty),\qquad(2)$$

where n_{vir} is the number of virtual bands per k-point. The proper range of n_{vir} for a safe extrapolation using Eq. (2) is determined by monitoring the convergence of the properties calculated using the estimated $E_c^{MP2}(\infty)$. For the bulk systems composed of the elements from group III to V, well-converged estimates of a_0^{MP2} and B_0^{MP2} can be obtained by using $n_{vir} = 350 \sim 400$ with uncertainties of about 0.1 pm and 1 GPa, respectively. Unfortunately, a much larger n_{vir}, which is beyond the reach of our computational resources, is needed for the other bulk systems that contain the *s*-block or the noble gas elements and for all atomic calculations. Therefore, at the MP2 level, we are unable to evaluate basis set errors of all properties of these materials and the cohesive energy of all materials. However, we will study LiH with the small-core pseudopotential for Li as one such example of a difficult case. For this material, we will compare our results to MP2 calculations reported in literature and discuss how the Gaussian basis sets developed in this work can be leveraged to significantly improve the convergence of correlated calculations using a PW basis.
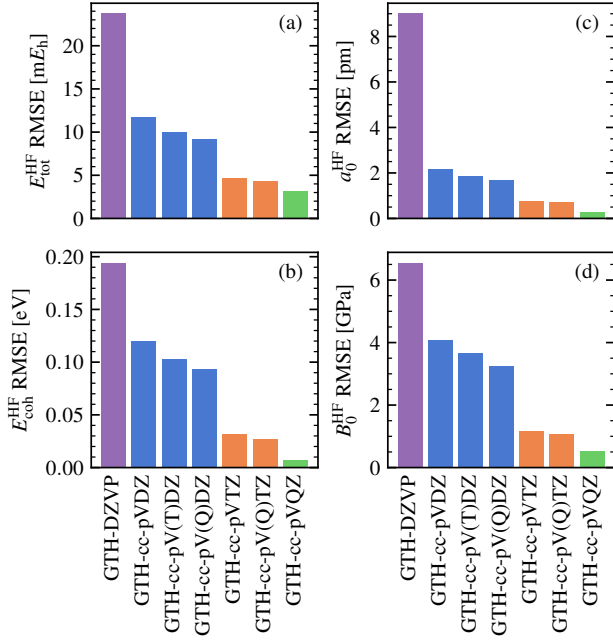
FIG. 4. The root-mean-square error of (a) the cell energy, (b) the cohesive energy, (c) the equilibrium lattice constant, and (d) the equilibrium bulk modulus calculated at the HF level using different Gaussian basis sets for the 19 bulk materials.

IV. RESULTS AND DISCUSSION

A. Occupied bands: HF ground-state properties

We first study the basis set performance for HF ground-state properties, which reflects the quality of the occupied bands. The root-mean-square error (RMSE) of the total energy $E_{\text{tot}}^{\text{HF}}$, the cohesive energy $E_{\text{coh}}^{\text{HF}}$, the equilibrium lattice constant a_0^{HF}, and the equilibrium bulk modulus B_0^{HF} are presented in Fig. 4 for different Gaussian basis sets, where the error is computed with respect to our PW results. The errors for each material are shown in Figs. S7–S10.

The most obvious trend in Fig. 4 is the monotonic decrease of the error of all four properties by following the hierarchy: GTH-DZVP, GTH-cc-pVDZ, GTH-cc-pV(T)DZ, GTH-cc-pV(Q)DZ, GTH-cc-pVTZ, GTH-cc-pV(Q)TZ, GTH-cc-pVQZ. This ranking is consistent with the flexibility of the basis sets except for GTH-DZVP, which differs from GTH-cc-pVDZ only in the size of the valence basis and the basis parameters. The larger error of the GTH-DZVP basis mainly comes from the two beryllium compounds where the contraction coefficients for the valence orbitals of Be are poor and the solid neon where the valence basis is too small (Figs. S7–S10). We verified that a modified GTH-DZVP basis for Be with the contraction coefficients re-computed using our code gives very similar accuracy as our GTH-cc-pVDZ basis. Similar reparametrization for the GTH-TZVP and GTH-QZVP basis sets can be performed but is not useful in practice due to the high linear dependencies of these basis sets.

Among our correlation-consistent basis sets, increasing the zeta-level is significantly more effective at reducing the errors than de-contracting the valence basis, which suggests reasonable transferability of the valence contraction coefficients determined from the atomic HF orbitals, at least for describing the occupied bands (see also the discussion on the HF band structure in Section IV B). The relatively large RMSE in the HF total energy [Fig. 4(a)] (about 4 mE_{h} even for the GTH-cc-pVQZ basis) is dominated by the two small-core magnesium compounds (Fig. S7), which inherit the large HF energy error of the magnesium atom, as shown in Table I. Nonetheless, the error in the total energy does not affect computed properties, suggesting a robust and systematic error cancellation in these basis sets.

B. Low-energy virtual bands: HF band gap and band structure

The RMSEs of the HF band gap calculated using different Gaussian basis sets are summarized in Fig. 5(a) (the errors for each material are shown in Fig. S11), which exhibits an overall trend similar to that discussed in Section IV A for the HF ground-state properties. However, a major distinction in the band gap calculations is the significant reduction of error by de-contracting the valence basis of GTH-cc-pVDZ, which indicates limited transferability of the valence contraction coefficients for describing the virtual bands. Nonetheless, even without the valence de-contraction, the smallest GTH-cc-pVDZ basis already achieves a RMSE below 0.1 eV, which is sufficient for most band gap calculations. We emphasize that the additional diffuse functions determined in the way described in Section II D 4 are crucial for obtaining accurate band gaps for the noble gas solids (Fig. S11), reducing the error from a few eVs to less than 0.1 eV in the most extreme case.

As an example of the performance for the band structure, in Fig. 5(b-e) we compare the the valence and low-energy conduction bands of diamond calculated using our Gaussian basis sets to those usings PWs. Similar plots for all other materials are displayed in Figs. S14–S41. The smallest GTH-cc-pVDZ basis already gives an accurate description of the valence and the first few conduction bands as shown in Fig. 5(b), which is consistent with the good performance observed for the HF ground-state properties (Fig. 4) and the band gap [Fig. 5(a)]. The fixed contraction coefficients in the valence basis of GTH-cc-pVDZ are responsible for the deviations from the PW band structure immediately beyond the first few virtual bands (e.g., at the Γ and L points, about 30 eV above the valence band maximum). Using the valence-uncontracted GTH-cc-pV(T)DZ basis fixes this problem and shows quantitative agreement with the PW bands up to about 40 eV, as shown in Fig. 5(c). The quantitative agreement between the Gaussian and the PW band structures extends to even higher energies by using the GTH-cc-pVTZ (~ 70 eV) and the GTH-cc-pVQZ (~ 90 eV) basis sets, as shown in Fig. 5(d) and (e), respectively.

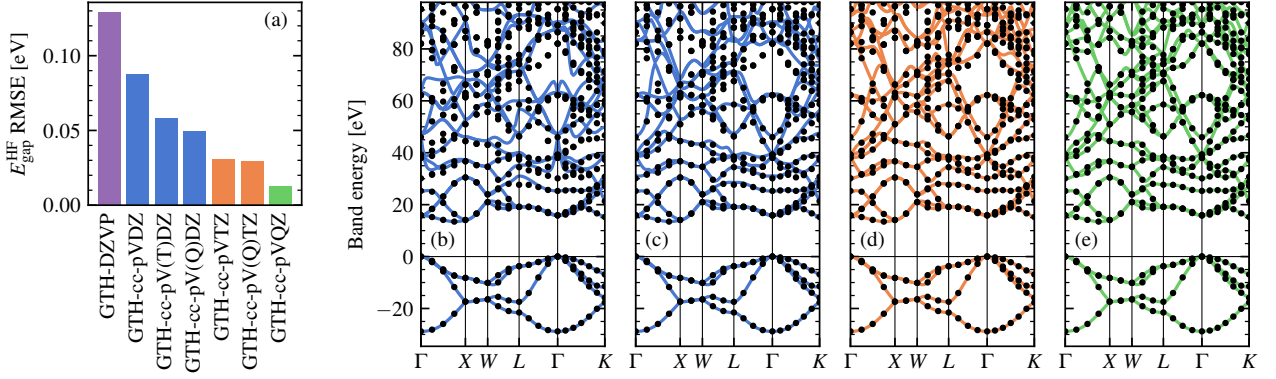We note that polarization functions of angular momentum

FIG. 5. (a) The root-mean-square error of the HF band gap calculated using different Gaussian basis sets for the 19 bulk materials. (b-e) HF band structure for diamond calculated using different Gaussian basis sets: (b) GTH-cc-pVDZ, (c) GTH-cc-pV(T)DZ, (d) GTH-cc-pVTZ, and (e) GTH-cc-pVQZ. The PW bands are shown as black dots.

f or higher, which are commonly absent from Gaussian basis sets meant for use in DFT calculations, can be important for the low-energy band structure. For example, the authors of ref 45 highlighted a missing state in the band structure of MgO (the fifth conduction band at the Γ point) unless a very large QZ basis set is used (167 basis functions per MgO unit). By contrast, our calculations on the same system (Fig. S29) suggest that the observed missing state is primarily an f-state localized on the oxygen atom and can already be captured accurately using our GTH-cc-pVTZ basis set with only 59 basis functions per MgO unit.

C. Convergence to the full virtual space limit: MP2 ground-state properties

The discussion in the previous two sections have focused on the occupied and the low-lying virtual bands. In this section, we study the basis set quality in correlated calculations at the MP2 level, which in principle requires an infinite number of virtual bands in order to reach the CBS limit. In a PW basis, the CBS limit is approached in a *dense* manner by increasing the number of virtual bands n_{vir} being correlated from low to high energy. The Gaussian virtual bands follow the dense manifold of the PW bands in the low-energy regime (as discussed in Section IV B), but become *sparse* at higher energy, effectively skipping some states. Ideally, the correlation energy obtained using either basis should show the asymptotic n_{vir}^{-1} convergence (2) for sufficiently large n_{vir}, but this may or may not be achievable with the available computational resources.

For the seven materials that do not contain s-block or noble gas elements, i.e., BN, BP, AlN, AlP, C, Si, and SiC, reliable extrapolations using Eq. (2) can be performed to obtain accurate estimates of $E_{\mathrm{corr}}^{\mathrm{MP2}}(\infty)$ in the PW basis (see Section III), from which we compute reference values for the equilibrium lattice constant a_0^{MP2} and the equilibrium bulk modulus B_0^{MP2}. The RMSEs of these two properties calculated using different Gaussian basis sets are shown in Fig. 6(a-b). The errors for each material are shown in Figs. S12 and S13. We also compute these two properties using the PW basis without extrapolation for a series of n_{vir} and plot the RMSEs in Fig. 6(c-d). The first three points in Fig. 6(c-d) with $n_{\mathrm{vir}} = 20$, 50, and 100 are chosen to match roughly the number of virtual bands in the GTH-cc-pVXZ basis set for X = D, T, and Q, respectively.

For both properties, the Gaussian basis exhibits the familiar hierarchy observed in the HF band gap calculations (Section IV B), where increasing the zeta-level significantly improves the accuracy, and de-contracting the valence basis is also effective at the DZ level [Fig. 6(a-b)]. The difference between the GTH-DZVP basis and the GTH-cc-pVDZ basis is somewhat smaller than in the HF calculations mainly because the problematic beryllium compounds are not included in the statistics here.

For correlated calculations with a Gaussian basis, basis set errors enter through both the HF energy and the correlation energy. In contrast, for correlated calculations with a PW basis, basis set errors enter through the correlation energy only, because the HF energy is essentially converged with respect to the number of PWs. Indeed, the errors observed in Fig. 6(a-b) for the DZ and TZ bases are dominated by errors in the HF energy, and the results can be significantly improved by combining the MP2 correlation energies calculated in a given basis set with the more accurate HF energies obtained from the GTH-cc-pVQZ basis, as shown by the thinner white bars in Fig. 6(a-b). In a similar spirit, one could perform a HF calculation in a large Gaussian basis and then perform an MP2 calculation with some number of frozen virtual orbitals.[10,84] Similar corrections are impossible with the original GTH basis set series due to the high linear dependencies at the QZ (or even TZ) level [Fig. 1(a)].

For the MP2 calculation of a_0 in a PW basis [Fig. 6(c)], the error of the smallest calculations with only 20 virtual bands is much smaller than the error of the calculations with the GTH-cc-pVDZ basis without the HF correction, but is similar to the error after the HF correction. The situation is different for B_0 [Fig. 6(d)], where the error of the PW calculations with 20 virtual bands is notably larger than that of the calculations with the GTH-cc-pVDZ basis, even without the HF correction. As n_{vir} increases, the errors of both properties decay
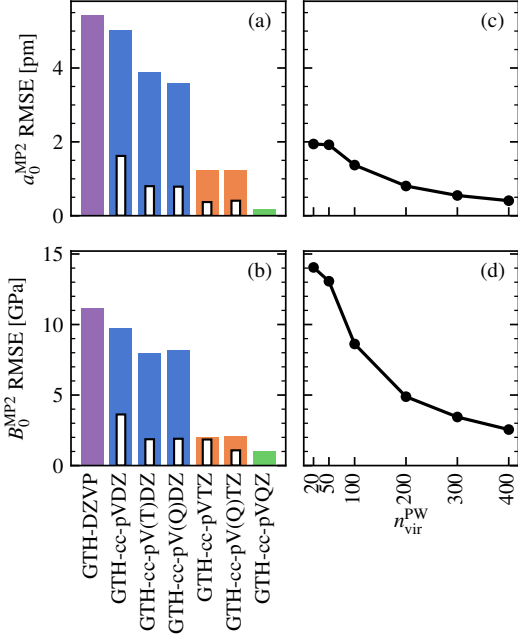
FIG. 6. Root-mean-square errors of (a,c) the MP2 equilibrium lattice constant and (b,d) the MP2 equilibrium bulk modulus calculated using (a-b) different Gaussian basis sets and (c-d) the PW basis with increasing number of virtual bands, n_{vir}^{PW}, where the first three points (n_{vir}^{PW} = 20, 50, and 100) are chosen to match the size of the virtual space of the GTH-cc-pVXZ basis for X = D, T, and Q, respectively. In each case, the errors are evaluated against CBS-extrapolated PW results for the seven bulk materials not containing the s-block or the noble gas elements: BN, BP, AlN, AlP, C, Si, and SiC. For the DZ and TZ Gaussian basis sets (except for GTH-DZVP), the HF-corrected results (see main text for explanation) are shown as the thinner bar with black edge.
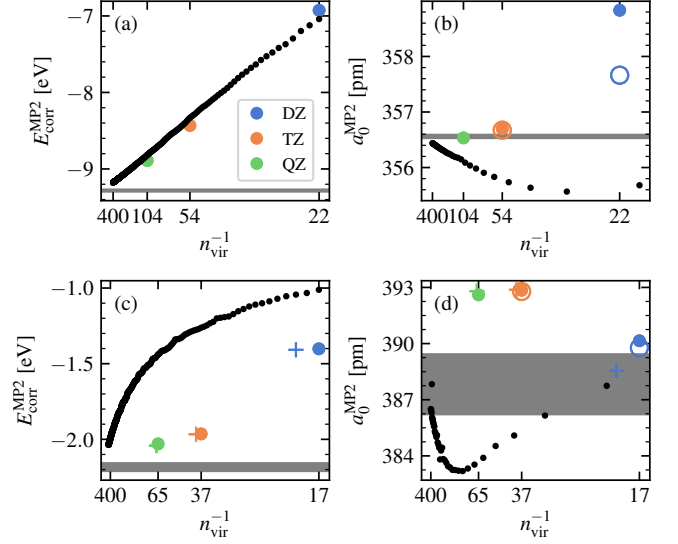

FIG. 7. Convergence of the MP2 correlation energy (left column) and the MP2 equilibrium lattice constant (right column) with the number of virtual bands included in the calculations for two materials: diamond (a-b) and LiH (c-d). The small-core pseudopotential is used for Li. In each case, the small black dots are results obtained using a PW basis, while the blue, orange, and green filled circles are results obtained using the GTH-cc-pVDZ, TZ, and QZ basis sets, respectively. For DZ and TZ, the open circles are a_0^{MP2} calculated by combining the MP2 correlation energy with the HF energy computed using the QZ basis. Gray shaded areas indicate the extrapolated PW results and their uncertainty. For LiH, the "+" symbols are results obtained by using the PW-resolved Gaussian virtual bands derived from the GTH-cc-pVXZ basis sets for the virtual space.

describing electron correlation in these systems.

The poor performance of PWs is exacerbated for elements with hard pseudopotentials, e.g., the s-block elements with core or semi-core electrons. We illustrate this behavior for the ionic crystal LiH, using the small-core pseudopotential for Li. As shown in Fig. 7(c), the convergence of the PW MP2 correlation energy is much slower than for diamond, with the asymptotic n_{vir}^{-1} convergence (2) not achieved even when $n_{vir} \approx 400$. This slow convergence yields a large uncertainty in the extrapolated $E_{corr}^{MP2}(\infty)$ (grey shaded area), which is similar to the GTH-cc-pVQZ result (green circle) obtained with only 65 virtual bands. Extrapolation of the Gaussian basis results suggests that the extrapolated PW result is likely an underestimate. The convergence in the PW basis is even slower for a_0 as shown in Fig. 7(d), which yields an extrapolated value that is about 4 pm below that obtained with our Gaussian basis sets. The Gaussian basis results converge much more quickly and their correctness is verified by comparison to literature values[23] as shown in Table II.

We end the discussion by showing that the convergence of PW-based correlated calculations can be significantly improved by leveraging a good Gaussian basis, such as the one developed in this work. Specifically, we compute the virtual bands in a PW-resolved Gaussian basis[27,28] generated by projecting out the converged PW occupied bands from our

very slowly in the PW basis, especially for small n_{vir}. The largest PW calculations with n_{vir} = 400 only achieve a RMSE comparable to the HF-corrected GTH-cc-pVTZ basis for a_0 and the HF-corrected GTH-cc-pVDZ basis for B_0, where the two Gaussian basis sets use only about 50 (TZ) and 20 (DZ) virtual bands in the correlated calculations.

The slow convergence of a_0 and B_0 with the number of PWs is caused by the small imbalance of the correlation energies evaluated using a fixed n_{vir} at the different cell volumes needed for the equation of state. This is illustrated in Fig. 7(a-b) for diamond. At a given lattice constant (here, the equilibrium lattice constant), the MP2 correlation energies evaluated in both the PW and the Gaussian basis sets exhibit the desired n_{vir}^{-1} convergence (2) [Fig. 7(a)]. But only the Gaussian basis sets converge quickly to the CBS limit for a_0 [Fig. 7(b)]. This behavior occurs because the PW basis is ignorant of the underlying atomic structure and thus exhibits an unphysical sensitivity to the cell volume. The situation is even worse in correlated calculations of molecular crystals[3,4] and free molecules or atoms,[24] due to the large amount of empty space between atoms or molecules. The Gaussian basis sets, along with the well-established BSSE correction,[85,86] are more suitable for

TABLE II. Comparison of the bulk properties of LiH (the small-core pseudopotential is used for Li) to those from Ref. 23 obtained using PWs and the projector augmented-wave (PAW) method.[87] The Gaussian basis calculations use an unshifted $5\times5\times5$ k-point mesh without extrapolation.

	a_0 [pm]		B_0 [GPa]		E_{coh} [eV]	
	HF	MP2	HF	MP2	HF	MP2
GTH-cc-pVDZ	410.4	394.3	32.5	39.4	1.81	2.34
GTH-cc-pVTZ	410.2	396.7	32.5	38.5	1.81	2.37
GTH-cc-pVQZ	410.1	396.5	32.6	38.6	1.81	2.38
PAW+PW	411.1	397.1	32	38	1.79	2.39

GTH-cc-pVXZ basis sets, followed by orthonormalization. As shown in Fig. 7(c-d) for LiH, the MP2 calculations that use the PW occupied bands plus the PW-resolved Gaussian virtual bands show significantly faster convergence than those that use the bare PW virtual bands. This example shows that our Gaussian basis sets are also useful for PW-based correlated calculations.

V. CONCLUSION

To conclude, we extended Dunning's strategy for constructing correlation-consistent Gaussian basis sets to periodic systems by controlling the size of the valence basis to reach a balance between accuracy and numerical stability. The generated GTH-cc-pVXZ basis sets are found to be well-conditioned for solid-state calculations and show fast convergence to the CBS limit at both mean-field and correlated levels of theory on a number of bulk properties. Our scheme can also be used straightforwardly to design all-electron basis sets for solids, which will differ only by the addition of primitives with large exponents that do not significantly contribute to linear dependencies.

Although our basis sets were tested using MP2, they will be valuable in work using more accurate ab initio correlated methods, such as coupled-cluster theory,[7–9,11,12,29] auxiliary field quantum Monte Carlo,[28,88–91] or quantum embedding approaches.[92–100] In particular, our basis sets remain to be tested on three-dimensional metals, where MP2 is inapplicable.[101] Reliable and standardized Gaussian basis sets for periodic systems also call for the development of optimized auxiliary basis sets for density fitting.[102–104] Finally, future work will proceed down the periodic table to obtain performant Gaussian basis sets for more elements. Of special interest are the d and f-block metals due to their appearance in a variety of functional materials,[105–107] whose accurate description demands correlated electronic structure theories.

ACKNOWLEDGEMENTS

SUPPORTING INFORMATION

See the supporting information for (i) plot of the condition number of the overlap matrix of commonly used Gaussian basis sets for 16 bulk materials, (ii) convergence of the HF and MP2 bulk properties for diamond with respect to the size of the valence basis of carbon, (iii) comparison of the exponents of the GTH-cc-pVXZ basis sets with the all-electron cc-pVXZ series for all elements studied in this work, (iv) increment of the atomic CCSD correlation energy as a function of the number of polarization functions for all elements studied in this work, (v) material-specific errors at both HF and MP2 levels for all properties reported in Figs. 4 to 6, (vi) band structure calculated using the GTH-cc-pVXZ basis sets compared to the PW CBS result for all materials studied in this work, (vii) detailed basis size information of the GTH-cc-pVXZ basis sets, and (viii) lattice structure and experimental lattice constants for all materials studied in this work.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

[1] M. Marsman, A. Grüneis, J. Paier, and G. Kresse, J. Chem. Phys. **130**, 184103 (2009).

[2] C. Müller and B. Paulus, Phys. Chem. Chem. Phys. **14**, 7605 (2012).

[3] M. Del Ben, J. Hutter, and J. VandeVondele, J. Chem. Theory Comput. **8**, 4177 (2012).

[4] M. Del Ben, J. Hutter, and J. VandeVondele, J. Chem. Theory Comput. **9**, 2654 (2013).

[5] G. H. Booth, A. Grüneis, G. Kresse, and A. Alavi, Nature **493**, 365 (2013).

[6] J. Yang, W. Hu, D. Usvyat, D. Matthews, M. Schütz, and G. K.-L. Chan, Science **345**, 640 (2014).

[7] J. McClain, Q. Sun, G. K.-L. Chan, and T. C. Berkelbach, J. Chem. Theory Comput. **13**, 1209 (2017).

[8] T. Gruber, K. Liao, T. Tsatsoulis, F. Hummel, and A. Grüneis, Phys. Rev. X **8**, 021043 (2018).

[9] I. Y. Zhang and A. Grüneis, Front. Mater. **6**, 123 (2019).

[10] X. Wang and T. C. Berkelbach, J. Chem. Theory Comput. **16**, 3095 (2020).

[11] B. T. G. Lau, G. Knizia, and T. C. Berkelbach, J. Phys. Chem. Lett. **12**, 1104 (2021).

[12] M. F. Lange and T. C. Berkelbach, J. Chem. Phys. **155**, 081101 (2021).

[13] X. Wang and T. C. Berkelbach, J. Chem. Theory Comput. **17**, 6387 (2021).

[14] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

[15] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[16] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[17] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou, and K. Burke, Phys. Rev. Lett. **100**, 136406 (2008).

[18] B. Jeziorski, R. Moszynski, and K. Szalewicz, Chem. Rev. **94**, 1887 (1994).

[19] M. O. Sinnokrot, E. F. Valeev, and C. D. Sherrill, **124**, 10887 (2002).

[20] C. D. Sherrill, T. Takatani, and E. G. Hohenstein, J. Phys. Chem. A **113**, 10146 (2009).

[21] B.-X. Zheng, C.-M. Chung, P. Corboz, G. Ehlers, M.-P. Qin, R. M. Noack, H. Shi, S. R. White, S. Zhang, and G. K.-L. Chan, Science **358**, 1155 (2017).

[22] Z. Li, S. Guo, Q. Sun, and G. K.-L. Chan, Nat. Chem. **11**, 1026 (2019).

[23] A. Grüneis, M. Marsman, and G. Kresse, J. Chem. Phys. **133**, 074107 (2010).

[24] A. Grüneis, G. H. Booth, M. Marsman, J. Spencer, A. Alavi, and G. Kresse, J. Chem. Theory Comput. **7**, 2780 (2011).

[25] J. J. Shepherd, A. Grüneis, G. H. Booth, G. Kresse, and A. Alavi, Phys. Rev. B **86**, 035111 (2012).

[26] H. Jiang and P. Blaha, Phys. Rev. B **93**, 115203 (2016).

[27] G. H. Booth, T. Tsatsoulis, G. K.-L. Chan, and A. Grüneis, J. Chem. Phys. **145**, 084111 (2016).

[28] M. A. Morales and F. D. Malone, J. Chem. Phys. **153**, 194111 (2020).

[29] J. M. Callahan, M. F. Lange, and T. C. Berkelbach, J. Chem. Phys. **154**, 211105 (2021).

[30] Z. Wei, F. Göltl, and P. Sautet, J. Chem. Theory Comput. **17**, 7862 (2021).

[31] J. G. Hill, Int. J. Quantum Chem. **113**, 21 (2013).

[32] T. H. Dunning, J. Chem. Phys. **90**, 1007 (1989).

[33] D. Feller, K. A. Peterson, and J. Grant Hill, J. Chem. Phys..

[34] F. Neese and E. F. Valeev, J. Chem. Theory Comput. **7**, 33 (2011).

[35] I. Y. Zhang, X. Ren, P. Rinke, V. Blum, and M. Scheffler, New J. Phys. **15**, 123033 (2013).

[36] A. J. C. Varandas, Phys. Chem. Chem. Phys. **23**, 9571 (2021).

[37] D. Feller, J. Comput. Chem. **17**, 1571 (1996).

[38] F. Weigend, F. Furche, and R. Ahlrichs, J. Chem. Phys. **119**, 12753 (2003).

[39] F. Weigend and R. Ahlrichs, Phys. Chem. Chem. Phys. **7**, 3297 (2005).

[40] B. Klahn and W. A. Bingel, Int. J. Quantum Chem. **11**, 943 (1977).

[41] J. VandeVondele and J. Hutter, J. Chem. Phys. **127**, 114105 (2007).

[42] M. Peintinger, D. Oliveira, and T. Bredow, J. Comput. Chem. **34**, 451 (2013).

[43] D. Vilela Oliveira, J. Laun, M. F. Peintinger, and T. Bredow, J. Comput. Chem. **40**, 2364 (2019).

[44] C. C. J. Roothaan, Rev. Mod. Phys. **23**, 69 (1951).

[45] J. Lee, X. Feng, L. A. Cunha, J. F. Gonthier, E. Epifanovsky, and M. Head-Gordon, J. Chem. Phys. **155**, 164102 (2021).

[46] L. E. Daga, B. Civalleri, and L. Maschio, J. Chem. Theory Comput. **16**, 2192 (2020).

[47] W.-L. Li, K. Chen, E. Rossomme, M. Head-Gordon, and T. Head-Gordon, J. Phys. Chem. Lett. **12**, 10304 (2021).

[48] Y. Zhou, E. Gull, and D. Zgid, J. Chem. Theory Comput. **17**, 5611 (2021).

[49] V. A. Neufeld, H.-Z. Ye, and T. C. Berkelbach, To be published.

[50] D. E. Woon and T. H. Dunning, J. Chem. Phys. **98**, 1358 (1993).

[51] S. Goedecker, M. Teter, and J. Hutter, Phys. Rev. B **54**, 1703 (1996).

[52] C. Hartwigsen, S. Goedecker, and J. Hutter, Phys. Rev. B **58**, 3641 (1998).

[53] J. Hutter, "New optimization of GTH pseudopotentials for PBE, SCAN, PBE0 functionals. GTH pseudopotentials for Hartree-Fock. NLCC pseudopotentials for PBE." https://github.com/juerghutter/GTH (2019).

[54] C. Møller and M. S. Plesset, Phys. Rev. **46**, 618 (1934).

[55] N. B. Balabanov and K. A. Peterson, J. Chem. Phys. **123**, 064107 (2005).

[56] B. P. Prascher, D. E. Woon, K. A. Peterson, T. H. Dunning, and A. K. Wilson, Theor. Chim. Acta **128**, 69 (2011).

[57] T. D. Kühne, M. Iannuzzi, M. Del Ben, V. V. Rybkin, P. Seewald, F. Stein, T. Laino, R. Z. Khaliullin, O. Schütt, F. Schiffmann, D. Golze, J. Wilhelm, S. Chulkov, M. H. Bani-Hashemian, V. Weber, U. Borštnik, M. Taillefumier, A. S. Jakobovits, A. Lazzaro, H. Pabst, T. Müller, R. Schade, M. Guidon, S. Andermatt, N. Holmberg, G. K. Schenter, A. Hehn, A. Bussy, F. Belleflamme, G. Tabacchi, A. Glöß, M. Lass, I. Bethune, C. J. Mundy, C. Plessl, M. Watkins, J. VandeVondele, M. Krack, and J. Hutter, J. Chem. Phys. **152**, 194103 (2020).

[58] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter, Comput. Phys. Commun. **167**, 103 (2005).

[59] M. C. Bennett, C. A. Melton, A. Annaberdiyev, G. Wang, L. Shulenburger, and L. Mitas, J. Chem. Phys. **147**, 224106 (2017).

[60] M. C. Bennett, G. Wang, A. Annaberdiyev, C. A. Melton, L. Shulenburger, and L. Mitas, J. Chem. Phys. **149**, 104108 (2018).

[61] H.-Z. Ye, "Correlation-consistent gaussian basis sets for periodic calculations." https://github.com/hongzhouye/ccgto (2021).

[62] G. D. P. III and R. J. Bartlett, J. Chem. Phys. **76**, 1910 (1982).

[63] K. P. Huber and G. Herzberg, *Molecular Spectra and Molecular Structure IV. Constants of Diatomic Molecules* (Springer, Boston, MA, 1979).

[64] D. E. Woon and T. H. Dunning, J. Chem. Phys. **103**, 4572 (1995).

[65] K. A. Peterson and T. H. Dunning, J. Chem. Phys. **117**, 10548 (2002).

[66] T. H. Dunning, K. A. Peterson, and A. K. Wilson, J. Chem. Phys. **114**, 9244 (2001).

[67] D. E. Woon and T. H. Dunning, J. Chem. Phys. **100**, 2975 (1994).

[68] L. Schimka, J. Harl, A. Stroppa, A. Grüneis, M. Marsman, F. Mittendorfer, and G. Kresse, Nat. Mater. **9**, 741 (2010).

[69] T. Tsatsoulis, F. Hummel, D. Usvyat, M. Schütz, G. H. Booth, S. S. Binnie, M. J. Gillan, D. Alfè, A. Michaelides, and A. Grüneis, J. Chem. Phys. **146**, 204108 (2017).

[70] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K.-L. Chan, Wiley Interdisc. Rev. Comput. Mol. Sci **8**, e1340 (2018).

[71] Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui, J. J. Eriksen, Y. Gao, S. Guo, J. Hermann, M. R. Hermes, K. Koh, P. Koval, S. Lehtola, Z. Li, J. Liu, N. Mardirossian, J. D. McClain, M. Motta, B. Mussard, H. Q. Pham, A. Pulkin, W. Purwanto, P. J. Robinson, E. Ronca, E. R. Sayfutyarova, M. Scheurer, H. F. Schurkus, J. E. T. Smith, C. Sun, S.-N. Sun, S. Upadhyay, L. K. Wagner, X. Wang, A. White, J. D. Whitfield, M. J. Williamson, S. Wouters, J. Yang, J. M. Yu, T. Zhu, T. C. Berkelbach, S. Sharma, A. Y. Sokolov, and G. K.-L. Chan, J. Chem. Phys. **153**, 024109 (2020).

[72] H.-Z. Ye and T. C. Berkelbach, J. Chem. Phys. **154**, 131104 (2021).

[73] H.-Z. Ye and T. C. Berkelbach, J. Chem. Phys. **155**, 124106 (2021).

[74] J. Paier, M. Marsman, K. Hummer, G. Kresse, I. C. Gerber, and J. G. Ángyán, J. Chem. Phys. **124**, 154709 (2006).

[75] P. Broqvist, A. Alkauskas, and A. Pasquarello, Phys. Rev. B **80**, 085114 (2009).

[76] R. Sundararaman and T. A. Arias, Phys. Rev. B **87**, 165122 (2013).

[77] J. A. Nelder and R. Mead, Comput. J. **7**, 308 (1965).

[78] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, Nat. Methods **17**, 261 (2020).

[79] W. Press and W. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*, bk. 4 (Cambridge University Press, 1992).

[80] M. R. Hestenes and E. Stiefel, J. Res. Natl. Bur. Stand. **49**, 409 (1952).

[81] F. D. Murnaghan, Proc. Natl. Acad. Sci. **30**, 244 (1944).

[82] F. Birch, Phys. Rev. **71**, 809 (1947).

[83] L. Lin, J. Chem. Theory Comput. **12**, 2242 (2016).

[84] M. F. Lange and T. C. Berkelbach, Mol. Phys. **118**, e1808726 (2020).

[85] S. Boys and F. Bernardi, Mol. Phys. **19**, 553 (1970).

[86] F. B. van Duijneveldt, J. G. C. M. van Duijneveldt-van de Rijdt, and J. H. van Lenthe, Chem. Rev. **94**, 1873 (1994).

[87] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).

[88] F. Ma, W. Purwanto, S. Zhang, and H. Krakauer, Phys. Rev. Lett. **114**, 226401 (2015).

[89] B. Rudshteyn, D. Coskun, J. L. Weber, E. J. Arthur, S. Zhang, D. R. Reichman, R. A. Friesner, and J. Shee, J. Chem. Theory Comput. **16**, 3041 (2020).

[90] F. D. Malone, S. Zhang, and M. A. Morales, J. Chem. Theory Comput. **16**, 4286 (2020).

[91] H. Shi and S. Zhang, J. Chem. Phys. **154**, 024107 (2021).

[92] I. W. Bulik, W. Chen, and G. E. Scuseria, J. Chem. Phys. **141**, 054113 (2014).

[93] A. A. Rusakov, S. Iskakov, L. N. Tran, and D. Zgid, J. Chem. Theory Comput. **15**, 229 (2019).

[94] S. Iskakov, C.-N. Yeh, E. Gull, and D. Zgid, Phys. Rev. B **102**, 085105 (2020).

[95] C.-N. Yeh, A. Shee, S. Iskakov, and D. Zgid, Phys. Rev. B **103**, 155158 (2021).

[96]H. Q. Pham, M. R. Hermes, and L. Gagliardi, J. Chem. Theory Comput. **16**, 130 (2020).

[97]Z.-H. Cui, T. Zhu, and G. K.-L. Chan, J. Chem. Theory Comput. **16**, 119 (2020).

[98]T. Zhu, Z.-H. Cui, and G. K.-L. Chan, J. Chem. Theory Comput. **16**, 141 (2020).

[99]H.-Z. Ye, H. K. Tran, and T. Van Voorhis, J. Chem. Theory Comput. **16**, 5035 (2020).

[100]T. Zhu and G. K.-L. Chan, Phys. Rev. X **11**, 021006 (2021).

[101]J. J. Shepherd and A. Grüneis, Phys. Rev. Lett. **110**, 226401 (2013).

[102]F. Weigend, M. Häser, H. Patzelt, and R. Ahlrichs, Chem. Phys. Lett. **294**, 143 (1998).

[103]F. Weigend, Phys. Chem. Chem. Phys. **4**, 4285 (2002).

[104]F. Weigend, J. Comput. Chem. **29**, 167 (2008).

[105]B. Keimer, S. A. Kivelson, M. R. Norman, and S. U. . J. Zaanen, Nature **518**, 179 (2015).

[106]Z. Du, S. Yang, S. Li, J. Lou, S. Zhang, S. Wang, B. Li, Y. Gong, L. Song, X. Zou, and P. M. Ajayan, Nature **577**, 492 (2020).

[107]G. Giuffredi, T. Asset, Y. Liu, P. Atanassov, and F. Di Fonzo, ACS Mater. Au **1**, 6 (2021).