

# G4(MP2)-XK: A Variant of the G4(MP2)-6X Composite Method with Expanded Applicability for Main-Group Elements up to Radon

Bun Chan,<sup>\*,†</sup> Amir Karton,<sup>‡</sup> and Krishnan Raghavachari<sup>§</sup>

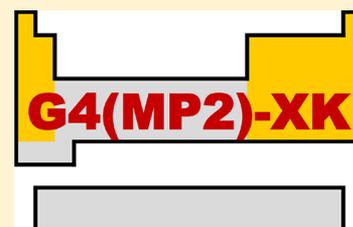
<sup>†</sup>Graduate School of Engineering, Nagasaki University, Bunkyo-machi 1-14, Nagasaki 852-8521, Japan

<sup>‡</sup>School of Molecular Sciences, The University of Western Australia, Perth, WA 6009, Australia

<sup>§</sup>Department of Chemistry, Indiana University, Bloomington, Indiana 47405, United States

## Supporting Information

**ABSTRACT:** In the present study, we have devised the G4(MP2)-XK composite method that covers species with up to fifth-row main-group elements (i.e., up to Rn). This new protocol is based on the previously published G4(MP2)-6X method, which has a general accuracy of  $\sim 5$  kJ mol<sup>-1</sup> for a diverse range of first- and second-row systems. The main difference between G4(MP2)-6X and G4(MP2)-XK is that the Pople-type basis sets in the former are replaced by Karlsruhe-type basis sets, with adjustments to the standard Karlsruhe basis sets to mimic the ones that they replace. Generally, G4(MP2)-XK is comparable in accuracy to G4(MP2)-6X. It is somewhat computationally more efficient than G4(MP2)-6X for the larger species that we have examined (e.g., a pentaglycine peptide). Importantly, the accuracy of G4(MP2)-XK for heavier elements is similar to that for first- and second-row species, even though it contains parameters that are fitted only to systems of the first two rows. This is indicative of the transferability of G4(MP2)-XK, and it paves the way for further expansion of its scope in future studies.



## INTRODUCTION

Composite ab initio procedures constitute one of the most widely used class of computational quantum chemistry methods for obtaining reliable thermochemical and kinetic quantities.<sup>1–5</sup> Among what is now becoming a “mini-zoo” of composite procedures, one of the most widely used subset of methods are the  $G_n(\text{MP2})$ -type protocols.<sup>1</sup> They are formulated in ways that are computationally highly efficient, while providing fairly good general accuracy. One of these is the G4(MP2)-6X method<sup>6</sup> that we have formulated several years ago, which has been subsequently applied in a wide range of theoretical investigations.<sup>7–13</sup>

The major (electron-correlated) quantum chemistry components of G4(MP2)-6X are based on Pople-style basis sets.<sup>14</sup> In this regard, there has been limited development of this class of basis sets for many years. To this day, they are only generally applicable up to third-row elements (K–Kr). Further expansion of G4(MP2)-6X to cover a wider range of systems requires availability of appropriate additional or alternative basis sets. In comparison to Pople-style basis sets, Karlsruhe-type<sup>15</sup> and Dunning-type<sup>16</sup> basis sets (and associated pseudopotentials where applicable) are defined for almost all elements. In our preliminary explorations, we have found that, among these two alternatives, Karlsruhe-type basis sets represent a more easily adaptable platform for formulating an alternative G4(MP2)-6X protocol, with future potential for an expanded scope.

As mentioned above, the G4 and G4(MP2) methods are applicable to elements up to the third row of the periodic table.<sup>2,17</sup> The present study formulates a G4(MP2)-6X-type method for a larger range of elements. We evaluate alternative

(Karlsruhe-type) basis sets for G4(MP2)-6X, and we make appropriate adjustments to emulate the performance of the original G4(MP2)-6X, in terms of both accuracy as well as computational efficiency. Our new protocol can be a platform for future formulations with an even wider scope. We also note that the use of pseudopotentials in a  $G_n(\text{MP2})$ -type method has been attempted,<sup>18,19</sup> but the focus of those studies is mainly on the reduction of computational cost.

## COMPUTATIONAL DETAILS

Standard quantum chemistry calculations were performed with the Gaussian 16 program.<sup>20</sup> Geometries were either obtained from previous studies<sup>6,8,12,21,22</sup> or optimized with the BMK functional<sup>23</sup> using various Karlsruhe-type basis sets, as described in the Results and Discussion section. Where applicable, zero-point vibrational energies (ZPVEs) and thermal corrections to enthalpy at 298 K ( $\Delta H_{298}$ ), derived from scaled BMK frequencies, were incorporated into the total energies. Frequency scale factors were determined using literature procedures as needed.<sup>24,25</sup> Single-point energies were obtained at the HF, MP2, and CCSD(T) levels for the composite procedures. In accordance with previous practice,<sup>26–29</sup> fitting of parameters was accomplished by minimizing the average of the mean absolute deviation (MAD) from the benchmark values and the standard deviation (SD) of the deviations for the training set. Relative energies in the text are reported in units of kJ mol<sup>-1</sup>.

Received: May 8, 2019

Published: July 9, 2019

## RESULTS AND DISCUSSION

**The Original G4(MP2)-6X Procedure.** Before we present the results of the present study, let us first provide a brief overview of the G4(MP2)-6X procedure<sup>6</sup> as a basis for later discussion. The G4(MP2)-6X protocol employs the BMK/6-31+G(2df,p) model for geometry optimization and vibrational frequency calculations. A series of single-point energies are obtained with the BMK-optimized structure and they are combined, together with an empirical “higher-level correction” (HLC) term, to yield the total vibrationless energy:

$$E_{G4(MP2)-6X} = E_{HF/CBS} + \Delta E_{SCS-MP2} + \Delta E_{scal-CCSD} + \Delta E_{scal-CCSD(T)} + HLC + E_{SO}$$

$E_{HF/CBS}$  is the Hartree–Fock energy at the complete-basis-set (CBS) limit, obtained using the GFHF3 and GFHF4 basis sets<sup>30</sup> (as implemented in Gaussian) and the extrapolation formula  $E_{CBS} = [E_{n+1} - E_n \exp(-1.63)]/[1 - \exp(-1.63)]$ . The GFHF3 and GFHF4 basis sets are modified variants of aug-cc-pVTZ and aug-cc-pVQZ, respectively. Thus,  $n = 3$  and  $4$  in the extrapolation formula. A major difference between GFHF3/4 basis sets and the corresponding aug-cc-pVnZ is that only a minimal  $sp$  set of diffuse functions is included in each GFHF3/4 set.

The  $\Delta E_{SCS-MP2}$  term is a correction for correlation effects at the MP2 level with the (triple- $\zeta$ ) G3MP2LargeXP basis set<sup>30</sup> (Gaussian keyword GTMP2LargeXP). It is defined by  $\Delta E_{SCS-MP2} = c_3 E_{C,OS}/G3MP2Large + c_4 E_{C,SS}/G3MP2LargeXP$ , where  $E_{C,OS}$  and  $E_{C,SS}$  are, respectively, the opposite- and same-spin components of the MP2 correlation energy, and  $c_3 = 1.249$  and  $c_4 = 0.486$ . The G3MP2LargeXP is a variant of the Pople basis set 6-311+G(3df,2p) with some extra polarization functions.

Additional corrections for higher-order correlation effects, i.e.,  $\Delta E_{scal-CCSD}$  and  $\Delta E_{scal-CCSD(T)}$ , are obtained with the smaller 6-31G(d) basis set (Gaussian keyword GTbas1). These terms are given by  $\Delta E_{scal-CCSD} = c_5 E_{C,CCSD}/6-31G(d) - c_1 E_{C,OS}/6-31G(d) - c_2 E_{C,SS}/6-31G(d)$ , and  $\Delta E_{scal-CCSD(T)} = c_6 E_{C,(T)}/6-31G(d)$ , where  $E_{C,CCSD}$  and  $E_{C,(T)}$  are the CCSD and perturbative triples contributions to the CCSD(T) correlation energies, and  $c_1 = 1.327$  and  $c_2 = 0.403$ ,  $c_5 = 1.077$ , and  $c_6 = 0.824$ .

In all correlation computations, the frozen-core approximation is used. The definition of core is identical to that for the G4(MP2) method. Generally, the largest noble-gas core is frozen but with a few exceptions. Specifically, the outer  $sp$ -core of second-row (Na and Mg) and later alkali and alkaline-earth metals, and the outer  $d$ -core of third-row and later  $p$ -block elements are kept in the valence. This definition of frozen core can be invoked within the Gaussian program by the FrzG4 option in correlation calculations.

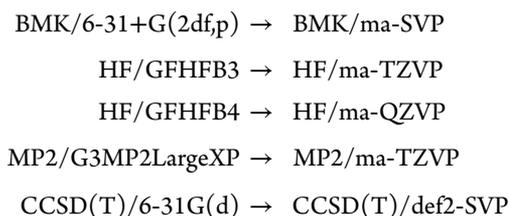
The HLC term is dependent on the number of valence electrons in the species according to conventional largest-noble-gas-core definition rather than FrzG4 used for correlation calculations. It is given by

$$HLC = \begin{cases} -An_\beta & \text{for closed-shell molecules} \\ -A'n_\beta - B(n_\alpha - n_\beta) & \text{for open-shell molecules} \\ -Cn_\beta - D(n_\alpha - n_\beta) & \text{for atomic species} \\ -En_\beta & \text{for “single-electron-pair” species, such as Li}_2 \end{cases}$$

with  $n_\alpha$  and  $n_\beta$  being the number of valence  $\alpha$  and  $\beta$  electrons, and  $A = 7.173$ ,  $A' = 7.264$ ,  $B = 3.677$ ,  $C = 7.239$ ,  $D = 2.404$ , and  $E = 1.021$  mhartree.

A term for spin–orbit correction ( $E_{SO}$ ) is included when it is already available from experiments or from existing high-level theoretical computations. This component is typically included for atomic species and for a few molecules. Scaled BMK vibrational frequencies are used to obtain zero-point vibrational energies (scale factor = 0.9770), thermal corrections to enthalpies at 298 K (0.9627) and entropies at 298 K (0.9695). These are then used to obtain the total enthalpies at 0 and 298 K, and total free energies at 298 K.

**Modifying G4(MP2)-6X with Standard Karlsruhe Basis Sets.** Let us now explore the possibility of revising G4(MP2)-6X with Karlsruhe basis sets (def2-SVP, def2-TZVP, and def2-QZVP) that are already available in the literature, which cover all elements up to astatine. They do not contain diffuse functions that are essential for many applications in chemistry. However, Truhlar and co-workers have provided a protocol for formulating minimally augmented Karlsruhe basis sets.<sup>31</sup> This approach involves adding a set of  $s$  and  $p$  diffuse functions to a nonaugmented basis set, with the exponents of these additional functions determined by dividing the smallest exponents of that angular momentum in that Karlsruhe set by a factor of 3. We can use these minimally augmented Karlsruhe sets (denoted ma-SVP, ma-TZVP, and ma-QZVP) to formulate a simple Karlsruhe variant for G4(MP2)-6X:



We denote this method G4(MP2)-6X-K1, with the “K” indicating the use of Karlsruhe basis sets. Based on the performance of this model (vide infra), we will also propose an improved variant, which we will refer to as G4(MP2)-6X-K2 later in this Article.

The parameters in the original G4(MP2)-6X protocol are optimized for the E2 set<sup>6</sup> of diverse chemical properties. This dataset contains ~500 data points of accurate experimental and theoretical values. It covers fundamental chemical properties (e.g., atomization energies, ionization energies, and electron affinities), quantities of relevance to chemical reactions (reaction energies and barriers), and noncovalent interactions (hydrogen bonding and van der Waals interactions). We employ the same set of data to optimize the parameters in G4(MP2)-6X-K(1/2).

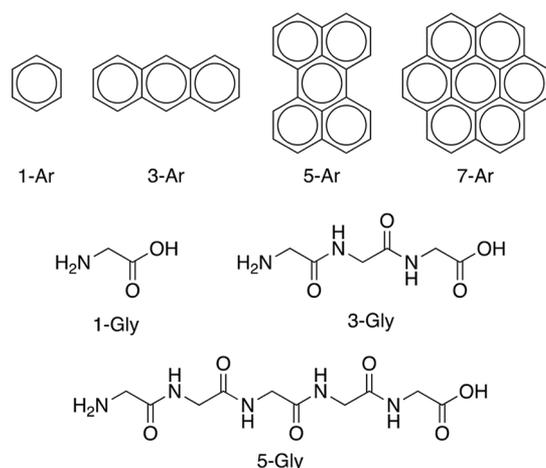
The mean absolute deviations (MADs) from benchmark values for the resulting methods for the E2 set and its subsets are shown in Table 1, along with those for the original G4(MP2)-6X for comparison. Generally, the G4(MP2)-6X-K1 protocol performs comparably with G4(MP2)-6X but is slightly less accurate. For the entire E2 set, the MADs are 3.5 kJ mol<sup>-1</sup> for [G4(MP2)-6X] and 4.1 kJ mol<sup>-1</sup> for [G4(MP2)-6X-K1]. We find this observation to be quite general across the subsets.

Let us turn our attention to the computational cost of the new protocol in comparison with the original. We use a series of aromatic molecules and glycine oligopeptides of different

**Table 1.** Mean Absolute Deviations from Benchmark Values for G4(MP2)-6X and Its Variants Based on Karlsruhe-Type (“K”) Basis Sets for the E2 Set and Its Subsets of Thermochemical Properties

	Mean Absolute Deviation (kJ mol <sup>-1</sup> )		
	G4(MP2)-6X	G4(MP2)-6X-K1	G4(MP2)-6X-K2 [= G4(MP2)-XK]
E2	3.5	4.1	3.8
<b>Atomization Energies and Heats of Formation</b>			
W4/08	3.8	4.6	4.6
G2/97'	3.0	3.7	3.1
G3/99'	3.4	3.5	3.4
<b>Other Fundamental Properties</b>			
IE	4.3	5.3	4.6
EA	5.3	5.7	5.0
PA	2.8	2.5	2.4
<b>Radical Addition (ADD) and Abstraction (ABS) Reaction Energies</b>			
ADD	2.5	2.1	2.5
ABS	1.7	2.2	1.9
<b>Barriers for Atom Transfer (DBH24) and Pericyclic (PR8) Reactions</b>			
DBH24	3.0	3.7	3.7
PR8	4.4	3.9	4.5
<b>Hydrogen Bond (HB16) and Weak Interaction (WI9/04) Energies</b>			
HB16	1.8	2.2	2.2
WI9/04	1.3	1.1	2.9

sizes to probe the scaling behavior of the various methods. These species are shown in Figure 1, and the relative times for



**Figure 1.** Molecules used for the assessment of relative times required for their computation using various G4(MP2)-6X-type methods.

their computations are given in Table 2. Note that the total time depends on many factors, including the number of cycles in the iterative component calculations. These relative timings do represent what one can expect from single-point energy calculations computations performed in the Gaussian program using default algorithms.

We can see that G4(MP2)-6X-K1 is more costly than G4(MP2)-6X. This is somewhat more apparent for the glycine oligopeptides, for which the calculations on 5-Gly with the new variant is ~50% more costly than that for G4(MP2)-6X. In absolute terms, the most costly calculation is the G4(MP2)-6X-K1 computation of 7-Ar, and it corresponds to 5.4 days of computing (on an octa-core processor). This is by no means an insignificant amount of resource but is not prohibitive with

**Table 2.** Relative Times Used for the Computation of Molecules in Figure 1 for G4(MP2)-6X and Its Variants

	Relative Computation Time		
	G4(MP2)-6X	G4(MP2)-6X-K1	G4(MP2)-6X-K2 [= G4(MP2)-XK]
1-Ar	2	6	6
3-Ar	41	43	43
5-Ar	209	215	208
7-Ar	614	790	604
1-Gly	1	1	2
3-Gly	33	39	37
5-Gly	445	632	344

the ubiquitous availability of standard multicore consumer hardware. When we take the results in Tables 1 and 2 into account, it seems that, while G4(MP2)-6X-K1 is somewhat less accurate and computationally more costly than G4(MP2)-6X, the difference is small enough, such that it appears to be a suitable platform for further improvement. Indeed, in the next section, we show that both accuracy and computational cost can be improved with minor modifications to the basis sets.

**Modifying Karlsruhe Basis Sets for G4(MP2)-6X.** Let us now inspect the differences between the basis sets used in the original G4(MP2)-6X and those for G4(MP2)-6X-K1, in order to determine whether there is an opportunity for straightforward refinement of the new protocol. The def2-SVP basis set used for the CCSD(T) calculation in G4(MP2)-6X-K1 is comparable in size to 6-31G(d) employed in G4(MP2)-6X, but it contains a set of *p* polarization functions for hydrogen and helium, whereas 6-31G(d) does not. This would account for the significant difference in cost for 5-Gly (C<sub>10</sub>H<sub>17</sub>N<sub>5</sub>O<sub>6</sub>, vs C<sub>20</sub>H<sub>12</sub> for 5-Ar) between the two protocols, for which the CCSD(T) computations dominate the total computational cost (see the Supporting Information). Therefore, a straightforward modification that would reduce the cost would be the removal of *p* functions from H and He. We denote this basis set as def2-SVSP, with the “SP” suffix signifying a “smaller polarization” set.

The definition for G3MP2LargeXP is somewhat more complex. According to the notation for Pople-type basis sets, it is essentially 311+G(*p*) for H and He but with separate *s* and *p* diffuse functions, instead of a combined *sp* function: 6-311+G(3df) for first-row elements, ~6-311+G(4d2f) for second-row elements, and ~6-311+G(3d2f) for third-row elements. Additional core–valence functions are also included for some elements for correlation treatment using the FrzG4 frozen-core convention. In comparison, the ma-TZVP basis set used in G4(MP2)-6X-K1 is smaller. It is smaller in terms of the valence functions, the polarization functions, and, in some cases, core–valence functions. We note that the polarization functions in the MP2 calculations for the various *G<sub>n</sub>*-type methods have played an important role in providing an adequate and balanced description for all elements covered by the protocol.

To improve ma-TZVP, we use the simple approach of selectively substituting its components with the ones in def2-QZVP, guided by comparison with G3MP2LargeXP. We term the new basis set ma-TZVXP. It consists of the following:

- H and He: ma-TZVP
- Li, Be: *sp* from ma-TZVP and *df* from def2-QZVP
- B–Ne: *spf* from ma-TZVP and *d* from def2-QZVP
- Na, Mg: ma-TZVP and *f* from def2-QZVP

- Al–Ar: *sp* from ma-TZVP and *df* from def2-QZVP
- K, Ca: ma-TZVP and *f* from def2-QZVP
- Ga–Kr: *spd* from ma-TZVP with the *5d* decontracted to *311d* (with no readjustment of exponents and coefficients) and *f* from def2-QZVP

The basis sets used for obtaining HF/CBS in G4(MP2)-6X-K1, namely, ma-TZVP and ma-QZVP, are not designed for basis-set extrapolation. Nonetheless, upon inspection of their composition, we believe that they would be adequate for such a purpose, because of their fairly systematic nature. Lastly, for geometry optimization within the G4(MP2)-6X protocol, the basis set employed is 6-31+G(2df,p), which has a larger set of polarization functions than that in ma-SVP used for G4(MP2)-6X-K1. To mimic the composition of 6-31+G(2df,p), we make the following adjustments to ma-SVP, with the resulting basis set termed ma-SVXP:

- H and He: def2-SVP
- Li, Be, Na, Mg, K, Ca: *sp* from ma-SVP and *df* from def2-QZVP
- B–Ne, Al–Ar, Ga–Kr: *sp* from ma-SVP and *df* from def2-TZVP

By making the above changes in basis set and redetermining the parameters  $c_{1-6}$  and A, A', B–E by fitting to the E2 set, we arrive at the G4(MP2)-6X-K2 procedure. The E2-fitted parameters are  $c_1 = 1.131$ ,  $c_2 = 0.512$ ,  $c_3 = 1.041$ ,  $c_4 = 0.704$ ,  $c_5 = 1.048$ ,  $c_6 = 0.526$ , A = 9.369, A' = 9.449, B = 3.832, C = 9.594, D = 1.874, and E = 2.491. As we can see from Table 1, G4(MP2)-6X-K2 is generally more accurate than G4(MP2)-6X-K1. The MAD values for the complete E2 set for G4(MP2)-6X, G4(MP2)-6X-K1, and G4(MP2)-6X-K2 are 3.5, 4.1, and 3.8 kJ mol<sup>-1</sup>, respectively. For most of the subsets of E2, the MADs for G4(MP2)-6X-K2 are fairly close to those for G4(MP2)-6X. If we now look at Table 2, we can see that G4(MP2)-6X-K2 is notably less costly than G4(MP2)-6X-K1 for the large systems examined. In comparison with G4(MP2)-6X, G4(MP2)-6X-K2 is somewhat less costly for the large species, in particular for 5-Gly. From this point onward, we will focus on G4(MP2)-6X-K2 and, for the sake of simplicity, we will refer to this protocol as G4(MP2)-XK.

**Independent Testing of the New G4(MP2)-XK Method on First- and Second-Row Systems.** We will now expand our assessment of the G4(MP2)-XK protocol to systems that are independent to those in the E2 set that we employ for determining the HLC and scale parameters. In this section, we will focus on a few additional first- and second-row main-group compounds. The test sets include the BDE261 set of a diverse range of bond dissociation energies,<sup>8</sup> the C24ISO set of isomerization energies with the formula C<sub>24</sub>H<sub>12</sub>,<sup>12</sup> MB08 set of reaction energies for artificially generated molecules,<sup>32</sup> and the ORBH36 set of reaction barriers.<sup>22</sup> Among these compilations, the MB08 and ORBH36 sets contain species with considerable multireference characters [with post-CCSD(T) effects up to ~10 kJ mol<sup>-1</sup>]. They represent challenging cases for quantum chemistry methods.

The MAD values for these various test sets for G4(MP2)-XK are shown in Table 3, along with those for G4(MP2)-6X for comparison. Generally, the performances of the two methods are reasonably close to each other. For the BDE261 and MB08 sets, the MADs for G4(MP2)-XK are somewhat smaller than those for G4(MP2)-6X, whereas the opposite is true for C24ISO and ORBH36. The largest margin between the MADs for the two methods can be seen for the ORBH36

**Table 3. Mean Absolute Deviations for the BDE261 Set of Bond Dissociation Energies, C24ISO Set of C<sub>24</sub>H<sub>12</sub> Isomerization Energies, ORBH36 Set of Reaction Barriers, G3/05' Set of Thermochemical Properties with Species Having Elements Up to the 3rd Row, and HP48 Set of Thermochemical Properties for Heavy (3rd Row to 5th Row) *p*-Block Species**

	Mean Absolute Deviation (kJ mol <sup>-1</sup> )	
	G4(MP2)-6X	G4(MP2)-XK
1st- and 2nd-Row Systems		
BDE261	4.8	2.9
C24ISO	0.9	1.4
MB08	22.0	18.9
ORBH36	6.2	10.5
Systems with Elements up to the 5th Row		
G3/05'	5.4	5.5
HP48	n/a	3.9

set [10.5 kJ mol<sup>-1</sup> for G4(MP2)-XK and 5.1 kJ mol<sup>-1</sup> for G4(MP2)-6X]. Interestingly, if we do not apply the fitted parameters for both protocols, the MAD values become 7.1 kJ mol<sup>-1</sup> for G4(MP2)-XK and 6.5 kJ mol<sup>-1</sup> for G4(MP2)-6X. Of course, we can adjust the parameters in G4(MP2)-XK to improve its performance for the ORBH36 set, but we deem such an approach somewhat arbitrary. We thus retain the E2-optimized parameters for G4(MP2)-XK, but caution that its performance for some challenging systems, similar to that for G4(MP2)-6X upon which it is based, can be less than ideal.

**Extension to Heavier Main-Group Elements.** A key motivation for the development of the G4(MP2)-XK method, as we presented earlier, is to provide a platform for expanding the scope of G4(MP2)-6X to cover more chemical elements. We have briefly investigated the viability of applying G4(MP2)-XK to additional elements. We focus on third-row and heavier main-group elements; we do not intend to use G4(MP2)-XK for transition metals. This is because the treatment of transition metals has been shown to be problematic for G4(MP2),<sup>17,33</sup> upon which G4(MP2)-6X and G4(MP2)-XK are based. Importantly, for transition-metal systems that are often highly multireference, CCSD(T) itself, which is the highest-level method employed in these G4(MP2)-type methods, can be inadequate.<sup>34</sup> We do not envisage that a straightforward adaptation of G4(MP2)-XK would lead to a reliable general methodology for transition-metal chemistry.

We use two sets of reference data that include heavy main-group species to further examine the performance of G4(MP2)-XK. One of these is the additional species in the G3/05 set<sup>35</sup> introduced to supplement the G3/99 set.<sup>36</sup> We will refer to this subset of G3/05 as G3/05'. It represents a good collection of experimental data, and it includes a considerable number of third-row systems. For fourth- and fifth-row systems for which reliable experimental data are limited, high-level quantum chemistry computations provide a means for obtaining benchmark values. We employ a modest set of small third-row to fifth-row *p*-block species for which reference values have been previously obtained at the CCSD(T) level extrapolated to the complete-basis-set (CBS) limit.<sup>37,38</sup> Specifically, this set consists of electron affinities for Ga–Br, In–I, and Tl–At, and bond dissociation energies for Ga<sub>2</sub>, In<sub>2</sub>, Tl<sub>2</sub>, As<sub>2</sub>, Sb<sub>2</sub>, Bi<sub>2</sub>, Br<sub>2</sub>, I<sub>2</sub>, At<sub>2</sub>, GaH, InH, TlH, BrH, IH, AtH, KrH<sup>+</sup> (to Kr + H<sup>+</sup>), XeH<sup>+</sup> (to Xe<sup>+</sup> + H), RnH<sup>+</sup> (to

Rn<sup>+</sup> + H), GaCl, InCl, TlCl, SeSi, TeSi, PoSi, AsN, SbN, BiN, GeO, SnO, PbO, BrF, IF, and AtF. In the present study, we will abbreviate this collection of thermochemical properties for heavy *p*-block species as the HP48 set.

As noted earlier, we have defined the basis sets used in G4(MP2)-XK for third-row elements by emulating the compositions of the corresponding ones in G4(MP2)-6X. The specific approach is to combine the components of the def2-SVP, def2-TZVP and def2-QZVP basis sets, as well as adding maug-type diffuse functions. We use the changes to the standard Karlsruhe basis sets for third-row as a guide to adjust those for fourth- and fifth-row *p*-block elements. The ma-SVXP and ma-TZVXP basis sets for In–Xe and Tl–Rn are given as follows:

- ma-SVXP: *sp* from ma-SVP and *df* from def2-TZVP
- ma-TZVXP: *spf* and core-*d* functions from ma-TZVP with the *6d* decontracted to *411d* and polarization-*d* functions from def2-QZVP

For the G3/05' set for which both the G4(MP2)-6X and G4(MP2)-XK methods are applicable, the MAD values for the two methods are comparable [5.4 and 5.5 kJ mol<sup>-1</sup>, respectively, for G4(MP2)-6X and G4(MP2)-XK; see Table 3]. The MAD for the HP48 set for G4(MP2)-XK is 3.9 kJ mol<sup>-1</sup>. Importantly, these MAD values are not very different from those for systems of lighter elements (see Tables 1 and 3), and our adaptations of G4(MP2)-6X and G4(MP2)-XK to cover additional elements do not involve redetermination of the HLC and scale parameters. Thus, the good performance of these methods for the G3/05' and HP48 sets indicates decent transferability to elements that are independent of those in the training set. It is also noteworthy to reiterate that, while the G4(MP2)-XK method employs basis sets that are not standard, their formulations are fairly straightforward. This creates a pathway for further expansion of the scope of the method.

**Description of G4(MP2)-XK.** Before we conclude our discussion, we provide a full description of the G4(MP2)-XK method for the convenience of the readers. The modified Karlsruhe basis sets are included as part of the Supporting Information, and we also provide example input files for performing G4(MP2)-XK computations in Gaussian, as well as a Perl script for obtaining the total energies.

(1) The geometry used a G4(MP2)-XK computation is obtained at the BMK/ma-SVXP level. The ma-SVXP basis set is a modified def2-SVP basis set. It includes a minimal set of diffuse functions (“ma-”) and a larger set of polarization functions.

(2) Zero-point vibrational energy (ZPVE), thermal correction for enthalpy at 298 K ( $\Delta H_{298}$ ), and entropy at 298 K ( $S_{298}$ ) are obtained using scaled BMK/ma-SVXP frequencies. The scale factors are 0.9766 (ZPVE), 0.9791 ( $\Delta H_{298}$ ), and 0.9647 ( $S_{298}$ ).

(3) The G4(MP2)-XK electronic energy is defined by the formulas

$$E_{\text{G4(MP2)-XK}} = E_{\text{HF/CBS}} + \Delta E_{\text{SCS-MP2}} + \Delta E_{\text{scal-CCSD}} \\ + \Delta E_{\text{scal-CCSD(T)}} + \text{HLC} + E_{\text{SO}} \\ E_{\text{HF/CBS}} = \frac{[E_{\text{HF/ma-QZVP}} - E_{\text{HF/ma-TZVP}} \exp(-1.63)]}{[1 - \exp(-1.63)]}$$

The ma-TZVP and ma-QZVP basis sets are, respectively, def2-TZVP and def2-QZVP supplemented with ma- functions.

$$\Delta E_{\text{SCS-MP2}} = c_3 E_{\text{C,OS/ma-TZVXP}} + c_4 E_{\text{C,SS/ma-TZVXP}}$$

The ma-TZVXP basis set is derived from def2-TZVP. It contains ma- functions and a larger set of polarization functions.  $E_{\text{C,OS}}$  and  $E_{\text{C,SS}}$  are, respectively, the opposite- and same-spin components of the MP2 correlation energy, and  $c_3 = 1.041$  and  $c_4 = 0.704$ .

$$\Delta E_{\text{scal-CCSD}} = c_5 E_{\text{C,CCSD/def2-SVSP}} - c_1 E_{\text{C,OS/def2-SVSP}} - c_2 E_{\text{C,SS/def2-SVSP}}$$

$E_{\text{C,CCSD}}$  is the CCSD correlation energy and def2-SVSP is a modified def2-SVP basis set with the *p* polarization functions on H and He removed, and  $c_5 = 1.048$ ,  $c_1 = 1.131$ ,  $c_2 = 0.512$ .

$$\Delta E_{\text{scal-CCSD(T)}} = c_6 E_{\text{C,(T)/def2-SVSP}}$$

$E_{\text{C,(T)}}$  is the perturbative triples contribution to the CCSD(T) correlation energy,  $c_6 = 0.526$ .

(4) In the MP2 and CCSD(T) calculations, the largest noble-gas core is generally frozen. However, the outer *s* and *p*-core of second-row (Na and Mg) and later alkali and alkaline-earth metals, and the outer *d*-core of third-row and later *p*-block elements are kept in the valence. This corresponds to the FrzG4 option in the Gaussian program.

(5) The HLC term is dependent on the number of valence  $\alpha$  ( $n_\alpha$ ) and  $\beta$  ( $n_\beta$ ) electrons in the species according to conventional largest-noble-gas-core definition rather than the FrzG4 definition. It is given by

$$\text{HLC} = \begin{cases} -An_\beta & \text{for closed-shell molecules} \\ -A'n_\beta - B(n_\alpha - n_\beta) & \text{for open-shell molecules} \\ -Cn_\beta - D(n_\alpha - n_\beta) & \text{for atomic species} \\ -En_\beta & \text{for "single-electron-pair" species, such} \\ & \text{as Li}_2 \end{cases}$$

with  $A = 9.369$ ,  $A' = 9.449$ ,  $B = 3.832$ ,  $C = 9.594$ ,  $D = 1.874$ , and  $E = 2.491$  mhartree.

(6) A spin-orbit correction term ( $E_{\text{SO}}$ ) is included when it is already available from experiment or from existing high-level theoretical computations. It is typically included for atomic species.

## CONCLUDING REMARKS

In the present study, we have devised the G4(MP2)-XK composite method as a platform for broadening the scope of G4(MP2)-6X, which is formulated for species containing up to second-row elements (i.e., H–Ar). Our approach involves replacing the Pople-type basis sets used in G4(MP2)-6X, which are applicable up to third-row elements, with Karlsruhe-type basis sets that are defined for elements up to the fifth row (i.e., H–Rn). We then use the composition of the Pople-type basis sets as a guide to modify the standard Karlsruhe-type basis sets, with the aim to emulate the accuracy and computational efficiency of G4(MP2)-6X with G4(MP2)-XK.

We have assessed G4(MP2)-XK with a wide range of main-group thermochemical quantities. The results show that, generally, G4(MP2)-XK is comparable in accuracy to G4(MP2)-6X. For some of the more challenging systems, both methods can lead to large deviations, and there can be a larger difference in the performances of the two protocols. In these cases, it is not straightforward to anticipate whether G4(MP2)-

XK would be more accurate than G4(MP2)-6X, or vice versa. In terms of computational efficiency, we find G4(MP2)-XK to have somewhat better scaling behavior than G4(MP2)-6X. Therefore, the new method is more suitable for the treatment of larger species.

For test sets that contain fourth- and fifth-row species for which only G4(MP2)-XK is applicable, we find that the accuracy is generally similar to that for first- and second-row species. Notably, G4(MP2)-XK contains parameters fitted to species with only first- and second-row elements [in accordance with the formulation of G4(MP2)-6X], and the good performance of heavier elements is achieved without reparametrization. This is indicative of the transferability of G4(MP2)-XK, and it paves the way for an expanded scope in future studies.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.9b00449.

Results of preliminary comparison between methods based on Karlsruhe-type and Dunning-type basis sets (PDF)

Basis sets for G4(MP2)-XK in the format of the Gaussian program, an example input Gaussian input file for computing the G4(MP2)-XK total electronic energy, enthalpy at 0 K, enthalpy at 298 K, and free energy at 298 K, and a computer script (written in the Perl language) for obtaining these quantities from the corresponding Gaussian output file (ZIP)

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: bun.chan@nagasaki-u.ac.jp.

### ORCID

Bun Chan: 0000-0002-0082-5497

Amir Karton: 0000-0002-7981-508X

Krishnan Raghavachari: 0000-0003-3275-1426

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

B.C. gratefully acknowledges research funding from the Japan Society for the Promotion of Science (JSPS) (Program for Advancing Strategic International Networks to Accelerate the Circulation of Talented Researchers) and generous grants of computer time from the RIKEN Advanced Center for Computing and Communication (ACCC) (Project No. Q19266). A.K. gratefully acknowledges an Australian Research Council (ARC) Future Fellowship (Project No. FT170100373). K.R. acknowledges support from U.S. National Science Foundation (Grant No. CHE-1665427) at Indiana University.

## ■ REFERENCES

(1) DeYonker, N.; Cundari, T. R.; Wilson, A. K. The Correlation Consistent Composite Approach (ccCA): Efficient and Pan-Periodic Kinetics and Thermodynamics. In *Advances in the Theory of Atomic and Molecular Systems*; Piecuch, P., Maruani, J., Delgado-Barrio, G., Wilson, S., Eds.; Progress in Theoretical Chemistry and Physics, Vol. 19; Springer: Dordrecht, The Netherlands; 2009; pp 197–224.

(2) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *Gn Theory*. *WIREs Comput. Mol. Sci.* **2011**, *1*, 810–825.

(3) Karton, A. A Computational Chemist's Guide to Accurate Thermochemistry for Organic Molecules. *WIREs Comput. Mol. Sci.* **2016**, *6*, 292–310.

(4) Chan, B. How to Computationally Calculate Thermochemical Properties Objectively, Accurately, and as Economically as Possible. *Pure Appl. Chem.* **2017**, *89*, 699–713.

(5) Martin, J. M. L. Computational thermochemistry: a brief overview of quantum mechanical approaches. *Annu. Rep. Comput. Chem.* **2005**, *1*, 31–43.

(6) Chan, B.; Deng, J.; Radom, L. G4(MP2)-6X: a Cost-Effective Improvement to G4(MP2). *J. Chem. Theory Comput.* **2011**, *7*, 112–120.

(7) Chan, B.; Morris, M.; Radom, L. O–H Bond Dissociation Energies. *Aust. J. Chem.* **2011**, *64*, 394–402.

(8) Chan, B.; Radom, L. BDE261: a Comprehensive Set of High-Level Theoretical Bond Dissociation Enthalpies. *J. Phys. Chem. A* **2012**, *116*, 4975–4986.

(9) Chan, B.; Radom, L. Hierarchy of Relative Bond Dissociation Enthalpies and Their Use to Efficiently Compute Accurate Absolute Bond Dissociation Enthalpies for C–H, C–C, and C–F Bonds. *J. Phys. Chem. A* **2013**, *117*, 3666–3675.

(10) Morris, M.; Chan, B.; Radom, L. Effect of Protonation State and Interposed Connector Groups on Bond Dissociation Enthalpies of Alcohols and Related Systems. *J. Phys. Chem. A* **2014**, *118*, 2810–2819.

(11) Chan, B.; Karton, A.; Easton, C. J.; Radom, L.  $\alpha$ -Hydrogen Abstraction by  $\bullet$ OH and  $\bullet$ SH Radicals from Amino Acids and Their Peptide Derivatives. *J. Chem. Theory Comput.* **2016**, *12*, 1606–1613.

(12) Chan, B. Unification of the W1X and G4(MP2)-6X Composite Protocols. *J. Chem. Theory Comput.* **2017**, *13*, 2642–2649.

(13) Chan, B.; Easton, C. J.; Radom, L. Effect of Hydrogen Bonding and Partial Deprotonation on the Oxidation of Peptides. *J. Phys. Chem. A* **2018**, *122*, 1741–1746.

(14) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54*, 724–728.

(15) Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H To Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.

(16) Dunning, T. H., Jr. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron Through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007.

(17) Mayhall, N. J.; Raghavachari, K.; Redfern, P. C.; Curtiss, L. A. Investigation of Gaussian4 Theory for Transition Metal Thermochemistry. *J. Phys. Chem. A* **2009**, *113*, 5170–5175.

(18) Silva, C. de S.; Pereira, D. H.; Custodio, R. G4CEP: A G4 Theory Modification by Including Pseudopotential for Molecules Containing First-, Second- and Third-Row Representative Elements. *J. Chem. Phys.* **2016**, *144*, 204118–204126.

(19) Leal, R. C.; Custodio, R. G3(MP2)//B3-SBK: A Revision of a Composite Theory for Calculations of Thermochemical Properties Including Some Non-Transition Elements Beyond the Fourth Period. *Comput. Theor. Chem.* **2019**, *1149*, 1–7.

(20) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.;

Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision C.01*; Gaussian, Inc.: Wallingford, CT, 2009.

(21) Chan, B.; Radom, L. Obtaining Good Performance With Triple- $\zeta$ -Type Basis Sets in Double-Hybrid Density Functional Theory Procedures. *J. Chem. Theory Comput.* **2011**, *7*, 2852–2863.

(22) Chan, B.; Simmie, J. M. Barriometry – An Enhanced Database of Accurate Barrier Heights for Gas-Phase Reactions. *Phys. Chem. Chem. Phys.* **2018**, *20*, 10732–10740.

(23) Boese, A. D.; Martin, J. M. L. Development of Density Functionals for Thermochemical Kinetics. *J. Chem. Phys.* **2004**, *121*, 3405–3416.

(24) Chan, B.; Radom, L. Frequency Scale Factors for Some Double-Hybrid Density Functional Theory Procedures: Accurate Thermochemical Components for High-Level Composite Protocols. *J. Chem. Theory Comput.* **2016**, *12*, 3774–3780.

(25) Chan, B. Use of Low-Cost Quantum Chemistry Procedures for Geometry Optimization and Vibrational Frequency Calculations: Determination of Frequency Scale Factors and Application to Reactions of Large Systems. *J. Chem. Theory Comput.* **2017**, *13*, 6052–6060.

(26) Chan, B.; Radom, L. W1X-1 and W1X-2: W1-Quality Accuracy with an Order of Magnitude Reduction in Computational Cost. *J. Chem. Theory Comput.* **2012**, *8*, 4259–4269.

(27) Chan, B.; Radom, L. W3X: A Cost-Effective Post-CCSD(T) Composite Procedure. *J. Chem. Theory Comput.* **2013**, *9*, 4769–4778.

(28) Chan, B.; Radom, L. Accurate Quadruple- $\zeta$  Basis-Set Approximation for Double-Hybrid Density Functional Theory with an Order of Magnitude Reduction in Computational Cost. *Theor. Chem. Acc.* **2014**, *133*, 277–286.

(29) Chan, B.; Radom, L. W2X and W3X-L: Cost-Effective Approximations to W2 and W4 with  $\text{kJ mol}^{-1}$  Accuracy. *J. Chem. Theory Comput.* **2015**, *11*, 2109–2119.

(30) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gaussian-4 Theory Using Reduced Order Perturbation Theory. *J. Chem. Phys.* **2007**, *127*, 124105–8.

(31) Zheng, J.; Xu, X.; Truhlar, D. G. Minimally Augmented Karlsruhe Basis Sets. *Theor. Chem. Acc.* **2011**, *128*, 295–305.

(32) Korth, M.; Grimme, S. “Mindless” DFT Benchmarking. *J. Chem. Theory Comput.* **2009**, *5*, 993–1003.

(33) Chan, B.; Karton, A.; Raghavachari, K.; Radom, L. Heats of Formation for  $\text{CrO}$ ,  $\text{CrO}_2$ , and  $\text{CrO}_3$ : An Extreme Challenge for Black-Box Composite Procedures. *J. Chem. Theory Comput.* **2012**, *8*, 3159–3166.

(34) Jiang, W.; DeYonker, N. J.; Wilson, A. K. Multireference Character for 3d Transition-Metal-Containing Molecules. *J. Chem. Theory Comput.* **2012**, *8*, 460–468.

(35) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Assessment of Gaussian-3 and Density Functional Theories on the G3/05 Test Set of Experimental Geometries. *J. Chem. Phys.* **2005**, *123*, 124107–12.

(36) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. Assessment of Gaussian-3 and Density Functional Theories for a Larger Test Set. *J. Chem. Phys.* **2000**, *112*, 7374.

(37) Peterson, K. A. Systematically Convergent Basis Sets with Relativistic Pseudopotentials. I. Correlationconsistent Basis Sets for the Post- $d$  Group 13–15 Elements. *J. Chem. Phys.* **2003**, *119*, 11099–11112.

(38) Peterson, K. A.; Figgen, D.; Goll, E.; Stoll, H.; Dolg, M. Systematically Convergent Basis Sets with Relativistic Pseudopotentials. II. Small-Core Pseudopotentials and Correlation Consistent Basis Sets for the Post- $d$  Group 16–18 Elements. *J. Chem. Phys.* **2003**, *119*, 11113–11123.