Theor Chem Acc (2002) 107:180–186
DOI 10.1007/s00214-001-0310-1

Theoretical
Chemistry Accounts

Letter

An overlap criterion for selection of core orbitals

Amy J. Austin,[1] **Michael J. Frisch,**[2] **John A. Montgomery Jr**[2], **George A. Petersson**[1]

[1] Hall-Atwater Laboratories of Chemistry, Wesleyan University, Middletown, Connecticut 06459-0180, USA
[2] Gaussian Inc., 140 Washington Ave., North Haven, Connecticut 06473, USA

Received: 31 August 2001 / Accepted: 11 October 2001 / Published online: 9 January 2002

Abstract. An overlap criterion is defined that connects the identification of core orbitals in a molecular system, which can be problematic, to that in isolated atoms, which is well defined. This approach has been tested on a variety of troublesome systems that have been identified in the literature, including molecules containing third-row main-group elements, and is shown to remove errors of up to 100 kcal/mol arising from an inconsistent treatment of core orbitals at different locations on a potential-energy surface. For some systems and choices of core orbitals, errors as large as 19 kcal/mol can be introduced even when consistent sets of orbitals are frozen, and the new method is shown to identify these cases of substantial core–valence mixing. Finally, even when there is limited core–valence mixing, the frozen-core approximation can introduce errors of more than 5 kcal/mol, which is much larger than the presumed accuracy of models such as G2 and CBS-QB3. The source of these errors includes interatomic core–core and core–valence dispersion forces.

Key words: Core – Frozen core – Valence – Dispersion forces – Orbital energies

1 Introduction

The partition of atomic orbitals and molecular orbitals (MOs) into valence sets that form chemical bonds and core sets that remain relatively unchanged is one of the most basic concepts of chemistry. The core–valence partition is employed almost universally in qualitative descriptions of chemical bonding and also provides one of the more common approximations helping to make quantitative calculations practical. This article will focus on resolving problems encountered in the latter application [1, 2].

Recent studies reveal that sizeable errors are obtained when methods that correlate only the valence electrons are applied to certain molecules containing third-row elements [1, 2, 3, 4]. In their examination of gallium fluorides, Bauschlicher et al. [1] reported that G2 atomization energies "disagree dramatically with experiment; the difference between theory and experiment is 40 kcal/mol for GaF growing to more than 100 kcal/mol for GaF_3. This is hardly the 2 kcal/mol error typically associated with the G2 approach." In this article we will examine carefully the problems contributing to these effects and present a general solution that can be implemented in systematic model chemistries.

2 Basic considerations

The usual approach for partitioning MOs into valence and core sets involves an energy criterion. The lowest-energy MOs are selected for the core. The number of such core orbitals is set equal to the sum of the core sizes for the constituent atoms. The definition of core orbitals is straightforward for the elements from hydrogen to argon. Partially filled shells and other atomic orbitals with the same principal quantum number constitute the valence set leaving a rare-gas core of orbitals with lower principal quantum numbers. Specifically, hydrogen and helium have no core, lithium through neon have a $1s^2$ helium core, and sodium to argon have a $1s^22s^22p^6$ neon core (Fig. 1) [5]. The core–valence partition is more subjective for the third row, since the 4*s* orbitals of potassium and calcium are below the 3*d* orbitals. However, once the 3*d* shell is filled, this shell is generally considered part of the core for gallium to krypton, even though $1s^22s^22p^63s^23p^63d^{10}$ does not correspond to a noble gas core. The analogous partition is repeated for rubidium to xenon, giving a $1s^22s^22p^63s^23p^63d^{10}4s^24p^6 4d^{10}$ core. Rather than honor the principal quantum number, some authors prefer honoring the actual rare-gas configurations, using a $1s^22s^22p^63s^23p^6$ argon core for gallium to krypton, and a $1s^22s^22p^63s^23p^63d^{10}4s^24p^6$ krypton core for indium to xenon. Cesium, barium, and

---

Correspondence to: G. A. Petersson
e-mail: gpetersson@wesleyan.edu

lanthanum begin filling the 6*s* and 4*f* shells and provide us with examples of relatively high energy 5*s* and 5*p* core orbitals. Regardless of which choice we make for the cores of the third- and fourth-row main-group elements, it is evident from Fig. 1 that there is no energy gap separating core orbitals from valence orbitals [5].

The energetic mixing of core and valence orbitals is illustrated in Fig. 2 [5]. If we employ the principal

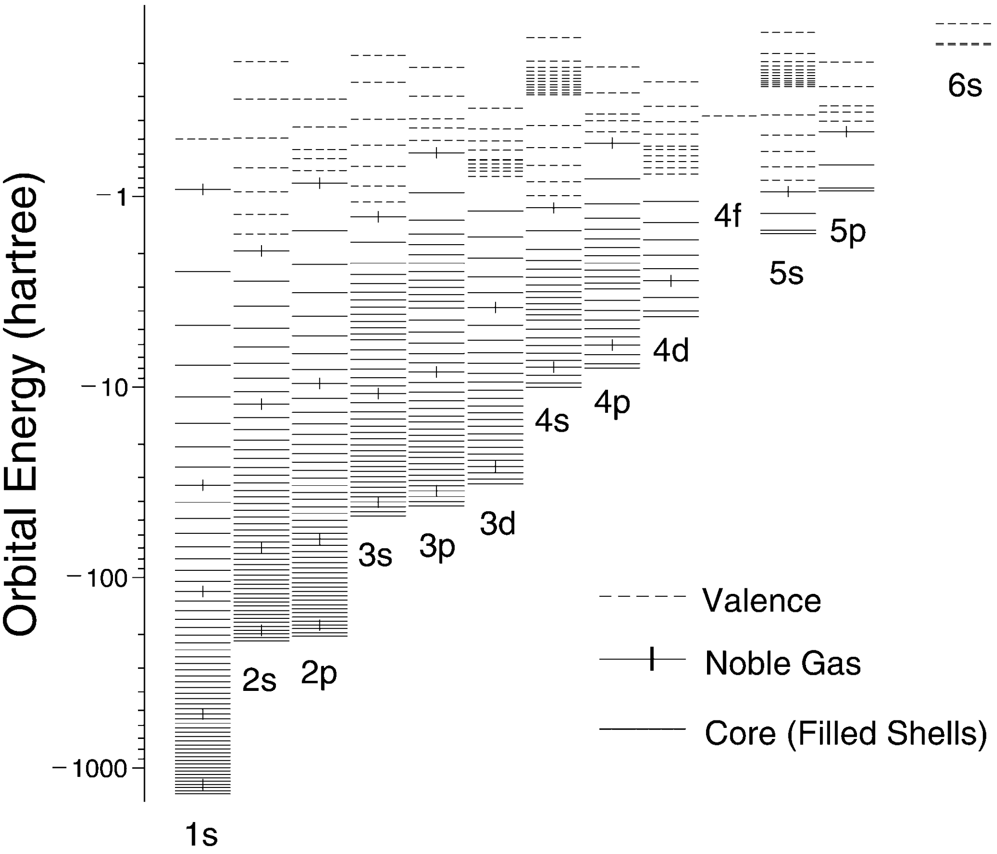

Fig. 1. Occupied atomic orbital energies for the elements from hydrogen to lanthanum

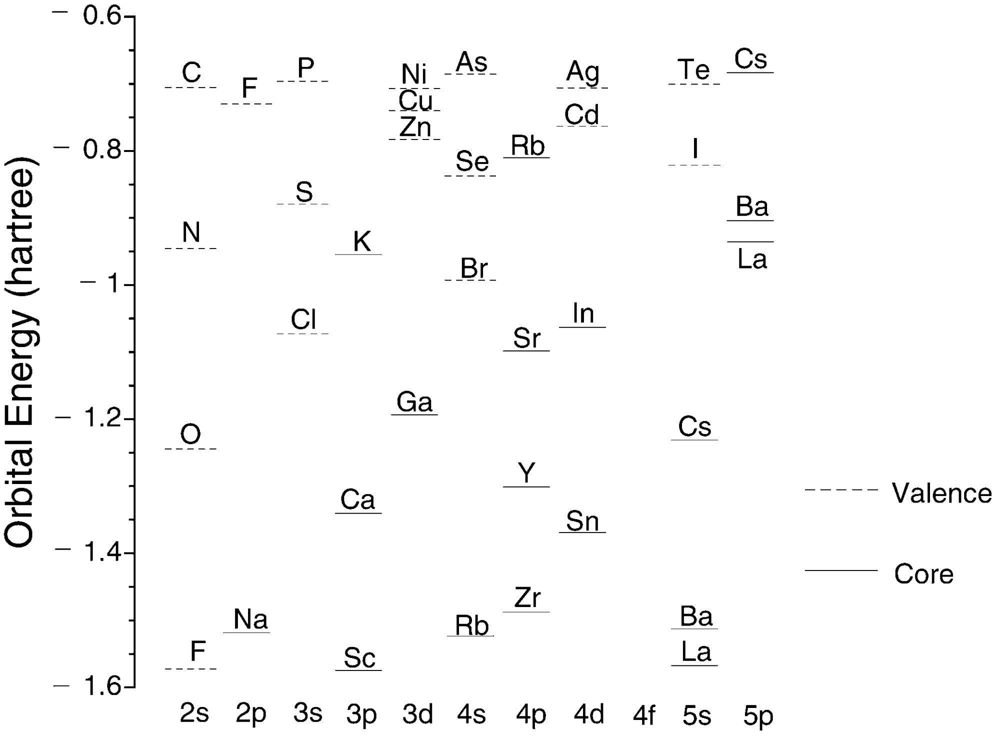

Fig. 2. The lowest-energy atomic valence orbital (i.e., partially filled shell) is the fluorine 2*s* orbital. The highest-energy atomic core orbital (i.e., filled shell) to lanthanum is the cesium 5*p* orbital. All atomic orbitals within this range are shown

quantum number for the core–valence partition of the third- and fourth-row main-group elements, the lowest energy valence orbital is the fluorine 2*s* orbital, and the highest-energy core orbital (to lanthanum) is the cesium 5*p* orbital [5]. All atomic orbitals within this range are included in Fig. 2. These are the orbitals that create problems for the orbital-energy-based core–valence partitions of MOs for systems containing third-row elements [1, 2, 3, 4]. It is clear from Fig. 2 that an energy criterion for the core–valence partitioning of the gallium fluorides will include the fluorine 2*s* atomic valence orbital in the molecular core while placing one of the gallium 3*d* atomic core orbitals in the molecular valence space. Inclusion of the gallium 3*d* orbitals in the correlation treatment has been shown to rectify such inversions in most examples [1, 2]; however, this correction would not be suitable for sodium bromide, since the sodium 2*s* and 2*p* atomic core orbitals are higher in energy than the bromine 3*d* atomic valence orbitals. Core–valence inversions of this type are easily remedied by selecting the molecular core orbitals on the basis of their overlap with the atomic core orbitals. Other problematic cases arise when core and valence MOs are physically mixed [1, 2, 3, 4]. The gallium oxides exhibit such mixing between one of the gallium 3*d* atomic core orbitals and the oxygen 2*s* atomic valence orbital. In these cases, inclusion of the gallium 3*d* orbitals in the molecular valence is quite essential [2]. Implementing an overlap criterion for the selection of core orbitals permits the consistent inclusion of 3*d* orbitals in the correlation space for molecules containing third-row main-group elements and also provides a simple test for the type of core–valence mixing found in the gallium oxides.

3 Methods

Our goal is to identify which canonical orbitals in a molecular system are most like the known core orbitals of the isolated atoms, and to do this with a prescription which is applicable to any atomic orbital basis and with any of the usual one-electron Hamiltonians (i.e., all-electron, using effective core potentials on some atoms, or using a scalar relativistic approximation for the kinetic energy and potential-energy operators). It would be difficult to tabulate one set of atomic orbitals for reference that could be used reliably with many different molecular basis sets and with many different choices of one-electron Hamiltonian. Consequently, an efficient and reliable method of solving for atomic orbitals in a given basis and using a given one-electron Hamiltonian is required. This need not be an exact atomic solution, but must be sufficiently accurate that the resulting core orbitals are representative. Since the solution of the self-consistent-field problem for open-shell atomic states can present problems, especially for transition metals, a noniterative atomic model is preferable. One obvious choice would be to diagonalize the bare one-electron Hamiltonian (i.e., to neglect all electron–electron interactions). Unfortunately, this choice is not sufficiently accurate (vide infra). A much more accurate model Hamiltonian, particularly for atomic calculations, can be produced using the Harris functional [6], a first-order approximation to density functional theory. The details of the Harris approximation to the molecular total energy are not relevant to the present use of the approximation; what is important is that the orbitals in the Harris approximation are produced by a single diagonalization of a Fock matrix which includes Coulomb and exchange–correlation potentials computed from a superposition of electron densities for the isolated, neutral atoms. In the general case

$$\mathbf{F}_{\mu\nu} = \mathbf{T}_{\mu\nu} + \mathbf{V}_{\mu\nu} + \int \phi^0(\mathbf{r})\chi_\mu(\mathbf{r})\chi_\nu(\mathbf{r})\mathrm{d}\mathbf{r} + \int \frac{\delta f^{\mathrm{xc}}[\rho^0(\mathbf{r})]}{\delta\rho}\chi_\mu(\mathbf{r})\chi_\nu(\mathbf{r})\mathrm{d}\mathbf{r} \ , \tag{1}$$

$$\rho^0(\mathbf{r}) = \sum_{\mathrm{A}} \rho^0_{\mathrm{A}}(r) \ , \tag{2}$$

$$\phi^0(\mathbf{r}) = \sum_{\mathrm{A}} \phi^0_{\mathrm{A}}(r) \ . \tag{3}$$

Here, $\rho^0_{\mathrm{A}}(r)$ is the electron density from neutral atom A and $\boldsymbol{\phi^0}_{\mathbf{A}}$ is the electrostatic potential of this distribution. $\mathbf{T}_{\mu\nu}$ and $\mathbf{V}_{\mu\nu}$ are the normal kinetic energy and potential-energy matrices for the system. Since the atoms are spherically symmetric, the radial values of $\rho^0_{\mathrm{A}}(r)$ and ϕ^0_{A} are tabulated for each atom and then interpolated during the numerical quadrature of the Coulomb and exchange–correlation integrals. Any of the standard density functionals can be used in this approximation. In our tests, the resulting orbitals, especially the core orbitals, are quite insensitive to the specific choice of functional. Consequently, we use a simple local density approximation with the VWN5 parameterization.

In each molecular calculation, we form and diagonalize the Fock matrix in the Harris approximation for each isolated atom $[\rho^0(\mathbf{r}) = \rho^0_{\mathrm{A}}(\mathbf{r})$, etc.] using only the basis functions on that atom, and then select the lowest-energy eigenvectors for each atom to represent the core orbital space for that atom. Collecting eigenvectors from all the atoms produces a set of vectors that represents the desired frozen-core orbital subspace for the molecule. A projector onto this space can be formed by orthogonalizing the atomic core orbitals and is invariant to the choice of orthogonalization procedure. Using symmetric orthogonalization, the projector $\mathbf{P}^{\mathrm{core}}$ can be formed from the N_{c} core orbitals having coefficients $O_{\mu\mathrm{I}}$, $\mu = 1, N_{\mathrm{b}}$; $\mathbf{I} = 1, N_{\mathrm{c}}$ (N_{b} being the number of atomic orbital basis functions in the molecule) via

$$\bar{\mathbf{O}} = \mathbf{O}(\mathbf{O}^{\mathrm{T}}\mathbf{S}\mathbf{O})^{\frac{1}{2}} \ , \tag{4}$$

$$\mathbf{P}^{\mathrm{core}} = \bar{\mathbf{O}}\bar{\mathbf{O}}^{\mathrm{T}}\mathbf{S} \ . \tag{5}$$

However, for classification of canonical MOs, it is more convenient to form

$$\bar{\mathbf{C}}_{pi} = \bar{\mathbf{O}}^{\mathrm{T}}_{p\mu}\mathbf{S}_{\mu\nu}\mathbf{C}_{\nu i} \ , \tag{6}$$

since $\bar{\mathbf{C}}_{pi}$ gives the contribution of orthogonalized core orbital *p* to projected MO *i*. Hence, $\sum_p \mathbf{C}^2_{pi}$ is the norm of MO i after projection onto the core orbital space.

Those norms should be very close to 1 or 0 after the projection, depending on whether the orbital is core or valence, respectively. Intermediate values indicate mixing of core and valence and their significance will be discussed in the context of our example calculations.

4 Results and discussion

The overlap-based core–valence partition was tested against several published cases for which the energy criterion selection is known to fail, including GaF, $\mathrm{GaF_2}$, $\mathrm{GaF_3}$, GaO, GaOH, HGaO, $\mathrm{GeF_2}$, NaBr, KF, and NaF [1, 2, 3, 4]. The three definitions of frozen core that were applied using both energy and overlap criteria are illustrated in Table 1. To be consistent with the majority of the examples in the recent literature, our first core definition employs the variant of the principal quantum number definition (principal Q#) used in the G2 model [7], which treats the 3*s* and 3*p* orbitals of potassium and calcium as valence. In order to evaluate the effects of also including the 3*d* orbitals in the correlation space for third-row main-group element containing molecules, our second definition freezes the noble gas core. Finally, our third definition freezes only the inner noble gas core, with the intention of including the principal core correlation effects.

A comparison of the core orbitals (in the principal quantum number definition) either from diagonalization of the Harris Fock matrix or from diagonalization of the bare one-electron Hamiltonian, with the Hartree–Fock orbitals, is given in Table 2. Since the 6-311 + G (2df, p) basis set used in our molecular calculations has limited flexibility in the core, we also carried out this comparison using an uncontracted basis containing the same primitive Gaussians. This agreement between the core orbitals from the Harris model and from the Hartree–Fock model is excellent, better than 99.7% in all cases, and is equally good for the contracted and uncontracted basis sets. In contrast, the core orbitals from diagonalizing the one-electron Hamiltonian are very poor for the contracted basis and become significantly worse for the uncontracted basis. For example, the 3*d* orbitals of gallium from the one-electron Hamiltonian have only a 68% overlap with those from Hartree–Fock! Clearly, the Harris model provides atomic core orbitals which are in close quantitative agreement with Hartree–Fock. The alternativee noniterative model based on the one-electron Hamiltonian does not provide useful simulations of atomic orbitals.

The total energies of the molecules and atoms at the second-order Møller–Plesset (MP2) level of theory with all electrons correlated and the energy changes arising from use of each of the three frozen-core definitions along with either the energy or overlap criterion for selection of the core orbitals are presented in Table 3. The 6-311 + G (2df, p) basis and MP2(full)/6-31G* optimized geometries were used throughout. The resulting MP2(full) atomization energies along with the errors resulting from application of the frozen-core alternatives are presented in Table 4. Note that we are not advocating the nonrelativistic MP2(full)/6-311 + G (2df, p) model as an adequate approximation to experiment but simply as a benchmark for the corresponding frozen-core calculations. The core–valence mixings for the three core definitions are shown in Table 5. Components of the MP2 correlation energy contribution to the dissociation energy of NaF are given in Table 6. All the calculations were performed with a development version of the Gaussian programs [8].

4.1 The inversion of core and valence orbitals

When the principal quantum number definition is employed, the use of an energy criterion for partitioning the MOs of the gallium fluorides results in deviations from full correlation atomization energy calculations ranging from 41.60 kcal/mol for GaF to 100.99 kcal/mol for GaF_3, as evidenced in Table 4 and previously noted by Bauschlicher et al. [1]. Because the gallium 3*d* orbitals are higher in energy than the fluorine 2*s* orbital (Fig. 2), the 15 lowest-energy occupied MOs of GaF include the fluorine $1s^2$, the gallium $1s^22s^22p^63s^23p^6$ (MOs 1–10), the fluorine $2s^2$ (MO 11), and four of the gallium 3*d* orbitals (MOs 12–15). The highest-energy gallium 3*d* orbital (MO 16) is thus forced into the valence;

Table 1. Core–valence definitions for selected atoms

Atom	Principal Q#		Noble gas core		Inner noble gas core	
	Core	Valence	Core	Valence	Core	Valence
O, F	$1s^2$	$2s^22p^6$	$1s^2$	$2s^22p$	None	$1s^22s^22p$
Na	$1s^22s^22p^6$	$3s^1$	$1s^22s^22p^6$	$3s^1$	$1s^2$	$2s^22p^63s^1$
K[a]	$1s^22s^22p^6$	$3s^23p^64s^1$	$1s^22s^22p^63s^23p^6$	$4s^1$	$1s^22s^22p^6$	$3s^23p^64s^1$
Ga–Kr	$1s^22s^22p^63s^23p^63d^{10}$	$4s^24p$	$1s^22s^22p^63s^23p^6$	$3d^{10}4s^24p$	$1s^22s^22p^6$	$3s^23p^63d^{10}4s^24p$

[a] In accord with standard G2 theory, potassium has the $1s^22s^22p^6$ in the atomic core, while the $3s^23p^64s^1$ are valence orbitals [7]

Table 2. Percentage error in model core orbitals for atoms

Atom	Harris functional 6-311 + G (2df, p)		Bare one-electron Hamiltonian 6-311 + G (2df, p)	
	Contracted	Uncontracted	Contracted	Uncontracted
O	0.011	0.016	0.183	0.279
F	0.005	0.010	0.148	0.230
Na	0.053	0.048	27.437	27.803
Cl	0.034	0.034	7.378	11.113
K[a]	0.026	0.026	6.038	8.953
K[b]	0.015	0.015	54.986	68.124
Ga	0.264	0.261	31.813	56.772
Ge	0.151	0.150	30.624	53.347
Br	0.038	0.039	27.140	44.679

[a] In accord with standard G2 theory, potassium has the $1s^22s^22p^6$ in the atomic core, while the $3s^23p^64s^1$ are valence orbitals [7]

[b] Potassium has $1s^22s^22p^63s^23p^6$ in the atomic core, while the $4s^1$ orbital is in the valence set

Table 3. Energy differences for MP2 (FC)/6-311 + G (2df, p) compared to MP2 (full)/6-311 + G (2df, p)

Molecule	ΔE (hartrees) = E(frozen) − E(full)						E(full) (hartrees)
	Principal Q#		Noble gas core		Inner noble gas core		
	Energy	Overlap	Energy	Overlap	Energy	Overlap	
GaF	0.51300	0.45033	0.27655	0.27655	0.08638	0.08638	−2023.49166
GaF_2	0.58384	0.47083	0.29768	0.29768	0.08657	0.08657	−2123.24192
GaF_3	0.64975	0.49537	0.31941	0.31941	0.08675	0.08675	−2223.07147
GaO	0.48531	0.47070	0.27596	0.27596	0.08643	0.08643	−1998.75628
GaOH	0.50904	0.46024	0.27628	0.27628	0.08642	0.08642	−1999.46021
HGaO	0.47619	0.47619	0.27513	0.27513	0.08653	0.08653	−1999.40219
GeF_2	0.46644	0.46644	0.29793	0.29793	0.08838	0.08838	−2275.40609
NaBr	0.55436	0.55436	0.42090	0.38756	0.10754	0.10754	−2735.02225
KF	0.13642	0.13642	0.40655	0.35562	0.12402	0.11500	−699.29172
NaF	0.15911	0.15911	0.15911	0.15911	0.01521	0.01521	−261.78896
Ga	0.42563	0.42564	0.25452	0.25452	0.08619	0.08619	−1923.63746
F	0.02106	0.02106	0.02106	0.02106	0.00000	0.00000	−99.61962
O	0.02021	0.02021	0.02021	0.02021	0.00000	0.00000	−74.97025
H	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	−0.49981
Ge	0.41417	0.41417	0.25305	0.25305	0.08796	0.08796	−2075.74047
Na	0.13621	0.13621	0.13621	0.13621	0.01515	0.01515	−161.98219
Br	0.41615	0.41615	0.25156	0.25156	0.09221	0.09221	−2572.89691
K	0.11475	0.11475	0.33386	0.33386	0.11475	0.11475	−599.48290

Table 4. Errors in atomization energies for MP2(FC)/6-311 + G (2df, p) compared to MP2 (full)/6-311 + G (2df, p)

Molecule	ΔE (kcal/mol) = E(full) − E(frozen)						E(full) (kcal/mol)
	Principal Q#		Noble gas core		Inner noble gas core		
	Energy	Overlap	Energy	Overlap	Energy	Overlap	
GaF	41.60	2.28	0.61	0.61	0.12	0.12	147.20
GaF_2	72.84	1.93	0.65	0.65	0.24	0.24	229.20
GaF_3	100.99	4.11	1.07	1.07	0.35	0.35	360.90
GaO	24.76	15.60	0.77	0.77	0.15	0.15	93.23
GaOH	39.66	9.03	0.97	0.97	0.15	0.15	221.32
HGaO	19.04	19.04	0.25	0.25	0.21	0.21	184.91
GeF_2	6.36	6.36	1.73	1.73	0.26	0.26	267.55
NaBr	1.25	1.25	20.79	−0.13	0.11	0.11	89.83
KF	0.38	0.38	32.40	0.44	5.82	0.16	118.73
NaF	1.15	1.15	1.15	1.15	0.04	0.04	117.44

Table 5. Largest percentage of core–valence mixing for any occupied orbital

Molecule	Principal Q#	Noble gas core	Inner noble gas core
GaF	4.42, 3.80	0.012	0.006
GaF_2	9.17	0.012	0.006
GaF_3	6.83	0.009	0.006
GaO	23.8	0.012	0.006
GaOH	15.5	0.012	0.006
HGaO	24.4	0.012	0.006
GeF_2	6.97	0.014	0.005
NaBr	0.068	0.068	0.037
KF	0.026	3.46	0.026
NaF	2.15	2.15	0.002

however, all five 3*d* orbitals are frozen in the gallium atom, leading to inconsistencies. The use of an overlap criterion for selecting core orbitals removes this inconsistency. Projection of the fluorine 2*s* MO onto the space spanned by the atomic orbitals reveals that this orbital is valencelike (norm^2 after projection = 0.0380, Table 5), and MO 11 is permuted to the valence of the gallium fluoride. The 3*d* orbital of gallium (MO 16) is identified as core (norm^2 after projection = 0.9558, Table 5) and is not included in the correlation space. The result is a reduction of the error from 41.60 to 2.28 kcal/mol. Likewise, two and three gallium 3*d* orbitals are included in the correlation space for GaF_2 and GaF_3, respectively. Deviations from full correlation calculations are reduced to 1.93 kcal/mol for GaF_2 and 4.11 kcal/mol for GaF_3.

It has been suggested [1, 2] that the problems with the gallium fluorides can be resolved by including the 3*d* orbitals of the third-row main-group elements in the valence space. We concur that this core definition gives more reliable thermochemistry as demonstrated later; however, this remedy creates a new inversion problem if

Table 6. Components of MP2/6-311 + G (2df, p) correlation energy contribution to the dissociation energy of NaF

Correlation orbital set	Changes in components of correlation energy (kcal/mol)				Total
	F (val, val)	Na (core, val)	Na (core, core)	Na (core) F (val)	
FC	48.70	0.00	0.00	0.00	48.70
FC + Na (2*p*)	48.70	−1.91	−1.88	4.37	49.27
FC + Na (2*s*, 2*p*)	48.70	−2.09	−2.05	5.00	49.57
Full	48.95[a]	−2.09	−2.06	5.06[b]	49.86

[a] Includes F (core, core) and F (core, val)
[b] Includes Na (core) F (core)

the energy criterion is used to select molecular core orbitals. Under the energy criterion and a noble gas core definition, the 14 lowest-energy MOs of sodium bromide are frozen. These orbitals are the bromine $1s^2$ (MO 1), the sodium $1s^2$ (MO 2), the bromine $2s^2 2p^6 3s^2 3p^6$ and four of the 3*d* orbitals (MOs 3–14). The sodium $2s^2$ and $2p^6$ orbitals are included in the valence shell of sodium bromide; however, these orbitals are treated as core in the sodium atom. The deviation induced from substituting the sodium 2*s* and 2*p* orbitals for four of the bromine 3*d* orbitals in the correlation space is 20.79 kcal/mol compared to 0.13 kcal/mol when an overlap criterion is used and consistent orbitals are frozen.

These energetic inversions are mirrored in the selection of core MOs for potassium fluoride. On the application of an inner noble gas core definition, an energy criterion freezes the five lowest-energy MOs of potassium fluoride. These include the potassium $1s^2$ (MO 1), the fluorine $1s^2$ (MO 2), the potassium $2s^2$ (MO 3), and two of the potassium 2*p* orbitals (MOs 4 and 5). The remaining potassium 2*p* orbital is included in the correlation; however, all three 2*p* orbitals are treated as core in the potassium atom. As demonstrated in Table 4, this inconsistency gives rise to a deviation of 5.82 kcal/mol compared to 0.16 kcal/mol when the overlap criterion properly identifies all three potassium 3*p* orbitals as valence.

4.2 The intrinsic effects of the 3d orbitals

Even with consistent selection of core and valence orbitals, a definition which includes too large a core space can lead to substantial errors in calculated energy differences. The gallium oxides provide an instructive example of such errors. As observed in the cases of the gallium fluorides, orbital inversion was found to be present between one of the gallium 3*d* orbitals and the 2*s* orbital of oxygen for both GaO and GaOH. When the principal quantum number definition is applied using an energy criterion, one of the gallium 3*d* orbitals is substituted for the oxygen 2*s* orbital in the correlation space of GaO and GaOH. The errors in these atomization energies are 24.76 and 39.66 kcal/mol, respectively. As with the gallium fluorides, the inversion effects are corrected by employing an overlap criterion, with which the errors in atomization energy fall to 15.60 and 9.03 kcal/mol, respectively. HGaO, which has no orbital inversions, exhibits an error of 19.04 kcal/mol. Clearly, in these cases sizeable errors remain even after errors due to inversion have been remedied. These large errors arise owing to significant mixing between one of the gallium 3*d* orbitals and the oxygen 2*s* orbital of the gallium oxides, MOs 11 and 16 having substantial contributions from both. When the principal quantum number definition is applied, Table 5 indicates that the largest core–valence mixings for GaO, GaOH, and HGaO are 23.8, 15.5, and 24.4%, respectively. In contrast to the gallium fluorides, the gallium 3*d* orbitals of the gallium oxides do not remain essentially atomic-like in the molecule. Rather, the gallium 3*d* orbitals participate in the bonding. The chemistry inherent in the 3*d*–2*s* hybrid orbitals of the gallium oxides cannot be physically removed, and necessitates their inclusion in the correlation treatment. When a noble gas core definition is applied, errors in the atomization energy range from 0.25 to 0.97 kcal/mol.

This example shows that core–valence mixings of the order of 10% indicate significant participation in bonding by orbitals chosen to be in the core, and that such a choice of core size is associated with unacceptably large errors.

Germanium difluoride and sodium fluoride provide examples of cases in which effects of several kilocalories per mole are present when the principal quantum number definition is employed in the absence of either inversion or the dramatic mixing exhibited by the gallium fluorides. The results of Table 4 suggest that the sodium 2*s* and 2*p* electrons have very little effect on the binding of species such as NaF. A more detailed analysis in Table 6 demonstrates that this is quite misleading. Sodium fluoride is essentially an ion pair, making a partition of the second-order correlation energy into intraatomic and interatomic pair energies unambiguous. The sodium 3*s* electron is transferred to the fluorine, thereby markedly increasing the fluorine valence-shell correlation energy. This transfer also eliminates the sodium (core, 3*s*) correlation energy. The correlation energy within the sodium core is reduced by a similar amount. These latter two effects are very nearly equal, but opposite in sign to the dispersion attraction (i.e., the sum of the interatomic pair energies) between the sodium (2*s*, 2*p*) core and the fluorine (2*s*, 2*p*) valence electrons. The dispersion attraction in NaF is considerably larger than that for Ne_2 (1.52 kcal/mol) or NeF^- (2.80 kcal/mol) at the same bond length. The loosely bound valence electrons of F^- are more easily polarized, and the Na^+ ion pulls the F^- electrons closer.

The dispersion and sodium intraatomic effects are each large compared to the accuracy sought in models such as CBS-QB3 [9] and G2 [10], and there is no obvious reason to expect them to cancel in general. Germanium difluoride provides such an example. The dispersion effect is responsible for the sizeable core contribution to the bond energy of GeF_2. If we exclude the two bonding pairs, the remaining Ge (3*d*)–F (2*s*, 2*p*) lone-pair correlation energies account for 6.43 kcal/mol of the GeF_2 binding energy, comparable to the 6.36 kcal/mol error found in Table 4 for this species. The covalent bonds formed by valence electrons will generally bring the cores within their "van der Waals radii", making dispersion forces comparable to those of van der Waals complexes and providing a significant component of bond energies for elements from sodium on. This important effect of core electron energies seems to have been overlooked in previous studies. It is clear from Table 6 that the dispersion effects inherent to NaF and GeF_2 are included only by the application of the inner noble gas core definition.

5 Conclusions

Two fundamental issues pertinent to the use of the frozen-core approximation in systematic model chemistries are the definition of atomic core orbitals and the identification of canonical molecular core orbitals that correspond to the chosen definition of atomic core orbitals.

We have shown that no definition of the atomic core orbitals permits a consistent identification of molecular core orbitals on the basis of orbital energies, except for the trivial case of defining all occupied orbitals as valence. In contrast, the overlap criterion presented here provides a consistent identification of core MOs for any atomic core definition based on filled shells. Our implementation using the Harris functional is applicable to any atom-centered basis set, to both nonrelativistic and relativistic Hamiltonians, and is compatible with the use of effective core potentials to replace some of the atomic cores.

Having solved the problem of consistent identification, we examined the implications of the most common definitions for atomic cores for several recently reported problematic species. We identified two sources of systematic error in thermochemistry that can result from an overly optimistic definition of atomic core orbitals: inclusion of orbitals in the core that form filled shells yet distort or mix significantly with valence orbitals upon bonding; neglect of dispersion interactions between the filled shell of one atom and the electrons of neighboring atoms.

Several examples of orbital mixing are known to give rise to errors in atomization energy of 10 to 20 kcal/mol; however, the dispersion interactions between filled shells in covalent systems that can contribute 5 kcal/mol or more to binding energies have not been noted previously. These interactions must, therefore, be considered if the accuracy goals of methods such as CBS-QB3 or G2 are to be achieved, particularly for systems containing third-row main-group elements.

The core–valence mixings computed during the application of the overlap criterion for the identification of core orbitals provide a useful diagnostic for the possible presence of either of these sources for systematic error. In particular, mixings larger than 10% suggest very serious mixing between designated core and valence orbitals, and the associated errors are greater than 10 kcal/mol for some of the cases considered here. Mixings larger than 1% suggest that the designated core orbitals have undergone noticeable static polarization by the molecular environment and consequently are also likely to be amenable to dynamic polarization, thereby providing a significant dispersion contribution to the thermochemistry of the molecule.

References

1. Bauschlicher CW, Melius CF, Allendorf MD (1999) J Chem Phys 110: 1897
2. Duke BJ, Radom L (1998) J Chem Phys 109: 3352
3. Yamaguchi Y, Schaefer HF (1996) J Chem Phys 104: 9841
4. Richards CA, Yamaguchi Y, Kim SJ, Schaefer HF (1996) J Chem Phys 104: 8516
5. Fischer C (1977) The Hartree–Fock method for atoms. Wiley, New York
6. Harris J (1985) Phys Rev B 31: 1770
7. Blaudeau JF, McGrath MP, Curtiss LA, Radom L (1997) J Chem Phys 107: 5016
8. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery JA Jr, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Mennucci B, Cossi M, Adamo C, Jaramillo J, Cammi R, Pomelli C, Ochterski J, Petersson GA, Ayala PY, Morokuma K, Salvador P, Dannenberg JJ, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Andres JL, Gonzalez C, Head-Gordon M, Replogle ES, Pople JA (2001) Gaussian 01, development version (revision A.01). Gaussian, Pittsburgh, Pa
9. Montgomery JA Jr, Frisch MJ, Ochterski JW, Petersson GA (1999) J Chem Phys 110: 2822
10. Curtiss LA, Raghavachari K, Trucks GW, Pople JA (1991) J Chem Phys 94: 7221