Distinguishing Basis Set Superposition Error (BSSE) from Basis Set Incompleteness Error (BSIE)

C. David Sherrill
School of Chemistry and Biochemistry
Georgia Institute of Technology
February 2017

Basis set superposition error (BSSE) is often blamed for the overbinding of van der Waals complexes. It is said that each monomer in the dimer can "steal" basis functions from the other monomer, lowering the energy of the dimer in a way that is not possible for the isolated monomers, thus leading to overbinding of the complex when one computes interaction energies,

$$E_{\mathrm{AB}}^{\mathrm{int}} = E_{\mathrm{AB}} - E_{\mathrm{A}} - E_{\mathrm{B}}. \tag{1}$$

Note that interaction energies are defined such that a stabilizing interaction energy is negative. Several ways to correct for overbinding due to BSSE have been proposed, including the well-known Boys-Bernardi counterpoise correction,[1] which asserts that the overbinding of the dimer is given by the amount that monomer A is stabilized by basis functions of B in the dimer, plus the amount that monomer B is stabilized by basis functions A in the dimer, or

$$E_{BSSE} = \left(E_{\mathrm{A}}(\mathrm{AB}) - E_{\mathrm{A}}(\mathrm{A})\right) + \left(E_{\mathrm{B}}(\mathrm{AB}) - E_{\mathrm{B}}(\mathrm{B})\right), \tag{2}$$

where the symbols in parenthesis denote the basis set used. The AB basis is the union of the basis functions on A and B. The computation of $E_{\mathrm{A}}(AB)$ thus involves basis functions on the atoms of B, even though no atoms or electrons from monomer B are present in this computation. Thus, these empty basis functions from monomer B are denoted "ghost functions." The larger AB basis set will lead to lower energies, so $E_{\mathrm{A}}(AB) < E_{\mathrm{A}}(A)$ and $E_{\mathrm{B}}(AB) < E_{\mathrm{B}}(B)$. This means that the BSSE is negative, or that the error leads to interaction energies that are overstabilized. If we subtract the BSSE from $E_{\mathrm{AB}}^{\mathrm{int}}$, we get the counterpoise corrected energy,

$$\begin{aligned} \mathcal{E}_{\mathrm{AB}}^{\mathrm{int}} &= E_{\mathrm{AB}}(\mathrm{AB}) - E_{\mathrm{A}}(\mathrm{A}) - E_{\mathrm{B}}(\mathrm{B}) - \left(E_{\mathrm{A}}(\mathrm{AB}) - E_{\mathrm{A}}(\mathrm{A})\right) - \left(E_{\mathrm{B}}(\mathrm{AB}) - E_{\mathrm{B}}(\mathrm{B})\right) \quad (3) \\ &= E_{\mathrm{AB}}(\mathrm{AB}) - E_{\mathrm{A}}(\mathrm{AB}) - E_{\mathrm{B}}(\mathrm{AB}). \quad (4) \end{aligned}$$

Here we denote a BSSE-free quantity by $\mathcal{E}$. In the first line, we have recognized that in the original (uncorrected) interaction energy computation, each species A, B, and AB is computed in its own basis set. The last line shows that the corrected interaction energy is essentially the same as the original definition, except that now each monomer is computed in the dimer basis set. This effectively gives each monomer as much basis flexibility as the dimer, removing the BSSE (there are debates in the literature about whether the Boys-Bernardi counterpoise correction exactly or only approximately removes BSSE).

In the complete basis set (CBS) limit, all species have all the basis functions they require, and thus there is no longer an imbalance in the basis set treatment of the dimer vs. the monomers. Hence, in the CBS limit, the BSSE goes to zero. If we denote energies in the CBS limit by $\mathfrak{E}$, we get

$$\mathfrak{E}^{\text{int}}_{\text{AB}} = \mathfrak{E}_{\text{AB}} - \mathfrak{E}_{\text{A}} - \mathfrak{E}_{\text{B}}. \tag{5}$$

Even after counterpoise correction, generally counterpoise-corrected values $\mathcal{E}_{\text{int}}$ do not match estimates of the CBS limit values $\mathfrak{E}_{\text{AB}}$. This is generally ascribed to "basis set incompleteness error" (BSIE). That is, it is generally understood that the error in $E^{\text{int}}_{\text{AB}}$ vs. $\mathfrak{E}^{\text{int}}_{\text{AB}}$ arises from *two* sources, BSSE and BSIE. In principle, subtracting both these errors from the approximate $E^{\text{int}}_{\text{AB}}$ should yield the CBS limit value $\mathcal{E}_{\text{int}}$:

$$\mathfrak{E}_{\text{AB}} = E^{\text{int}}_{\text{AB}} - (\text{BSSE}) - (\text{BSIE}), \tag{6}$$

$$= \mathcal{E}^{\text{int}}_{\text{AB}} - (\text{BSIE}). \tag{7}$$

Unfortunately, unless we simply accept the Boys-Bernardi counterpoise correction (or one of the alternative proposed BSSE corrections) as being exactly correct, it is difficult or impossible to clearly define the BSIE, and the definitions for BSSE and BSIE seem to get entangled with each other, with no precise definition for either. Without being able to solve this problem, these notes attempt to clarify the issue somewhat.

Consider the relationship between the energies computed in some basis set (e.g., 6-31G*), denoted E, and the CBS limiting values $\mathfrak{E}$. Each individual energy will have some error, which we could call a basis set error denoted γ:

$$\mathfrak{E} = E - \gamma. \tag{8}$$

The equation simply asserts that if we subtract the basis set error from an energy computed in some basis set, we recover the CBS value of that energy. Note that these individual basis set errors γ are not to be confused with the quantity (BSIE) above, which is the non-BSSE contribution to the error in the *interaction energy.* Note also that necessarily $\mathfrak{E} < E$, so that the basis set errors γ are positive by definition.

With this definition, we can relate the interaction energy in some basis, $E^{\text{int}}_{\text{AB}}$, with the CBS value $\mathfrak{E}^{\text{int}}_{\text{AB}}$, as

$$\mathfrak{E}^{\text{int}}_{\text{AB}} = \mathfrak{E}_{\text{AB}} - \mathfrak{E}_{\text{A}} - \mathfrak{E}_{\text{B}}, \tag{9}$$

$$= (E_{AB} - \gamma_{\text{AB}}) - (E_{\text{A}} - \gamma_{\text{A}}) - (E_{\text{B}} - \gamma_{\text{B}}), \tag{10}$$

$$= E^{\text{int}}_{\text{AB}} - (\gamma_{\text{AB}} - \gamma_{\text{A}} - \gamma_{\text{B}}), \tag{11}$$

$$= E^{\text{int}}_{\text{AB}} - \Delta\gamma_{\text{AB}}. \tag{12}$$

Thus, we see that the error in $E^{\text{int}}_{\text{AB}}$ is due to a mismatch in the basis set error of AB vs the sum of the basis set errors of A and B.

Generally speaking, the basis set requirements for the dimer AB could be more or less demanding than the basis set requirements for A and B. Thus, $\Delta\gamma_{\mathrm{AB}}$ could be positive or negative. For the $(\mathrm{Ar})_2$ dimer, for example, the basis set requirements for the dimer are actually greater than for the monomers, such that $\gamma_{\mathrm{AB}} > 2\gamma_{\mathrm{A}}$, or $\Delta\gamma_{\mathrm{AB}} > 0$. In this case, then, $\mathfrak{E}^{\mathrm{int}}_{\mathrm{AB}} < E^{\mathrm{int}}_{\mathrm{AB}}$, and the finite basis set computations (without counterpoise correction) *underbind* the dimer rather than overbind it. Because the BSSE defined by Boys and Bernardi is necessarily negative, when we subtract this negative error to get counterpoise-corrected interaction energies $\mathcal{E}^{\mathrm{int}}_{\mathrm{AB}}$, our answers get *worse* for this case compared to the CBS limit (underbound interaction energies get even more underbound compared to the CBS limit).

The case of $(\mathrm{Ar})_2$ notwithstanding, it quite frequently occurs that the borrowing of basis functions in B by monomer A, and vice versa, outweighs any greater basis set requirements of the dimer, so that $\gamma_{\mathrm{AB}} < \gamma_{\mathrm{A}} + \gamma_{\mathrm{B}}$, or $\Delta\gamma_{\mathrm{AB}} < 0$. This would be associated with the classic overbinding of the dimer due to BSSE (resulting in $E^{\mathrm{int}}_{\mathrm{AB}} < \mathfrak{E}^{\mathrm{int}}_{\mathrm{AB}}$).

Based on experience as well as the literature, we wish to maintain that (a) there is a real effect of basis set mismatch between dimers and monomers that can adversely affect computations of interaction energies, and that we will term BSSE, and that (b) even after we correct for this error, we still have an error in our interaction energies vs. the CBS limit, which we will call BSIE. This means that the overall error in our interaction energies vs. the CBS limit, $\Delta\gamma_{\mathrm{AB}} = E^{\mathrm{int}}_{\mathrm{AB}} - \mathfrak{E}^{\mathrm{int}}_{\mathrm{AB}}$, is a sum of two terms, the BSSE and BSIE:

$$\Delta\gamma_{\mathrm{AB}} = (\mathrm{BSSE}) + (\mathrm{BSIE}). \tag{13}$$

Repeating what was said above, if we have good estimates of the CBS limits of the quantities involved, we can compute the left-hand side of this equation, but there is no unique way to separate it into (BSSE) vs. (BSIE) unless we make some assumption about one term or the other (e.g., assuming that the Boys-Bernardi counterpoise correction gives the BSSE exactly).

Nevertheless, we return to the basic idea that the dimer AB is treated in a "better" effective basis set than the monomers, due to the opportunity for basis set borrowing by each monomer in the dimer. This will lower the basis set error for the dimer AB, γ_{AB}, relative to what it would have been otherwise (which we might call hypothetically γ'_{AB}). Thus,

$$\gamma_{\mathrm{AB}} = \gamma'_{\mathrm{AB}} - \alpha_{\mathrm{AB}}, \tag{14}$$

where α_{AB} is the amount γ'_{AB} has been lowered by basis set borrowing in the dimer. (Again, because the basis set requirements for the dimer could be greater than those of the monomers, even though $\alpha_{\mathrm{AB}} > 0$, we do not necessarily know that $\gamma_{\mathrm{AB}} < \gamma_{\mathrm{A}} + \gamma_{\mathrm{B}}$).

Although we wish to associate the concept of BSSE with a relative energy, rather than an absolute energy, nevertheless in our current bookkeeping scheme, it is clear that α_{AB} is none other than the negative of the BSSE. Thus we can split the BSSE and BSIE this way:

$$\begin{aligned} \Delta\gamma_{\mathrm{AB}} &= \gamma_{\mathrm{AB}} - \gamma_{\mathrm{A}} - \gamma_{\mathrm{B}}, && (15) \\ &= (\gamma'_{\mathrm{AB}} - \alpha_{\mathrm{AB}}) - \gamma_{\mathrm{A}} - \gamma_{\mathrm{B}}, && (16) \\ &= (\mathrm{BSIE}) + (\mathrm{BSSE}), && (17) \end{aligned}$$

where we have made the associations

$$
\begin{aligned}
\text{(BSSE)} &= -\alpha_{\mathrm{AB}}, && (18)\\
\text{(BSIE)} &= \gamma'_{\mathrm{AB}} - \gamma_{\mathrm{A}} - \gamma_{\mathrm{B}}. && (19)
\end{aligned}
$$

Note that the BSIE is an imbalance between the basis set errors of the dimer vs the sum of the basis set errors of the monomers, but we have corrected the dimer basis set error with a hypothetical one computed without any basis set borrowing allowed.

We can quickly check that the above equations are all consistent. For example, the BSSE-free interaction energy could be obtained if we removed the basis set imbalance between the dimer and the monomers by raising the dimer energy by α_{AB}:

$$
\begin{aligned}
\mathcal{E}^{\mathrm{int}}_{\mathrm{AB}} &= (E_{\mathrm{AB}} + \alpha_{\mathrm{AB}}) - E_{\mathrm{A}} - E_{\mathrm{B}}, && (20)\\
&= E^{\mathrm{int}}_{\mathrm{AB}} + \alpha_{\mathrm{AB}}, && (21)\\
&= E^{\mathrm{int}}_{\mathrm{AB}} - \text{(BSSE)}. && (22)
\end{aligned}
$$

Likewise,

$$
\begin{aligned}
\mathfrak{E}^{\mathrm{int}}_{\mathrm{AB}} &= \mathfrak{E}_{\mathrm{AB}} - \mathfrak{E}_{\mathrm{A}} - \mathfrak{E}_{\mathrm{B}}, && (23)\\
&= (E_{\mathrm{AB}} - \gamma_{\mathrm{AB}}) - (E_{\mathrm{A}} - \gamma_{\mathrm{A}}) - (E_{\mathrm{B}} - \gamma_{\mathrm{B}}), && (24)\\
&= (E_{\mathrm{AB}} - (\gamma'_{\mathrm{AB}} - \alpha_{\mathrm{AB}})) - (E_{\mathrm{A}} - \gamma_{\mathrm{A}}) - (E_{\mathrm{B}} - \gamma_{\mathrm{B}}), && (25)\\
&= E^{\mathrm{int}}_{\mathrm{AB}} + \alpha_{\mathrm{AB}} - (\gamma'_{\mathrm{AB}} - \gamma_{\mathrm{A}} - \gamma_{\mathrm{B}}), && (26)\\
&= E^{\mathrm{int}}_{\mathrm{AB}} - \text{(BSSE)} - \text{(BSIE)}. && (27)
\end{aligned}
$$

If a precise definition were made (or asserted) for BSSE or BSIE, the other is then also precisely defined with the equations above.

References

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