

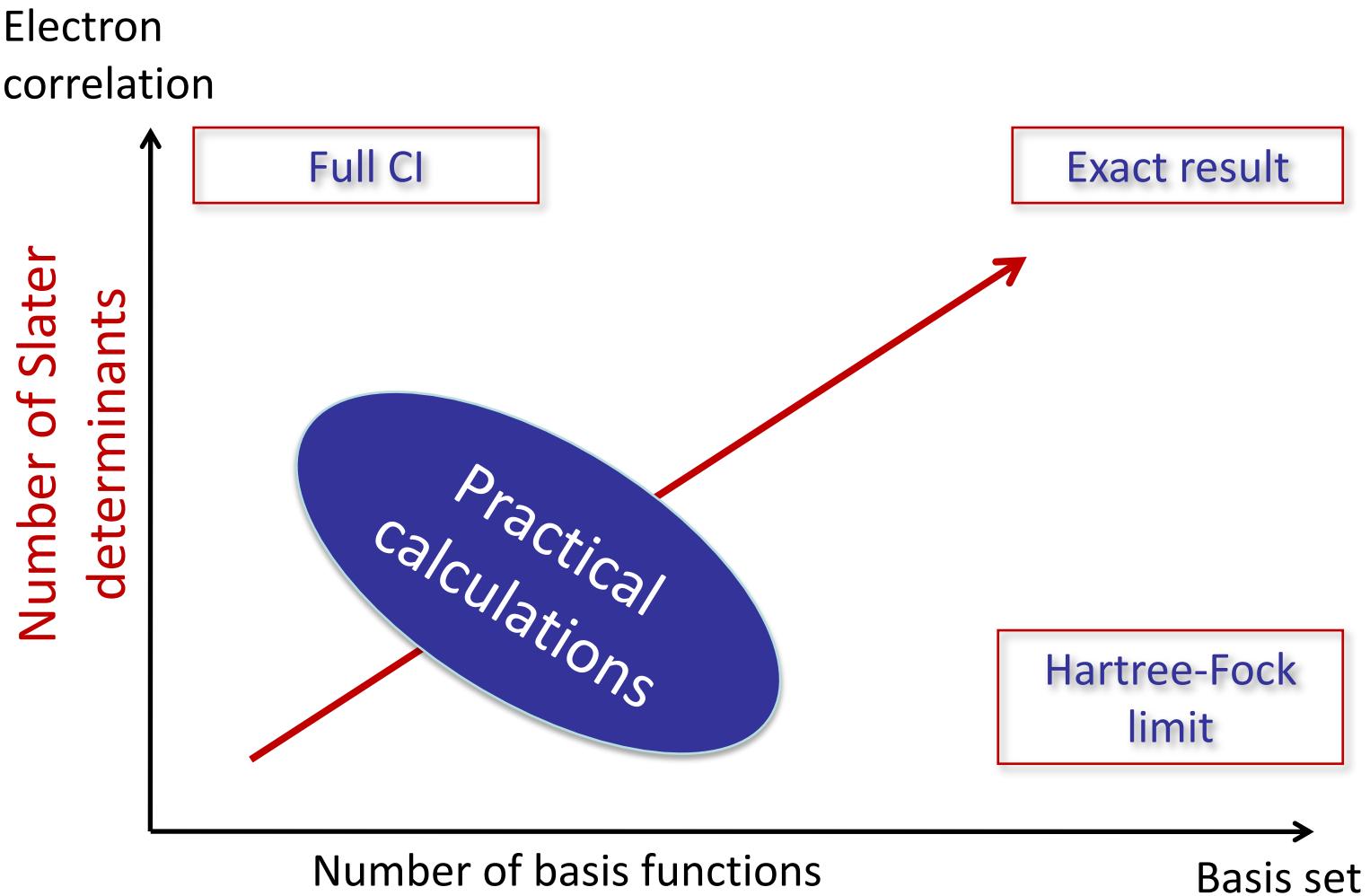
CHM 673

Lecture 15: Extrapolation techniques

Suggested reading:

Chapter 5.9, 5.10, 5.13 from Jensen

Theoretical model chemistries



Our goal now is to figure out how to reach the “exact” answer without crashing the computer
→ extrapolation techniques

Basis set extrapolation

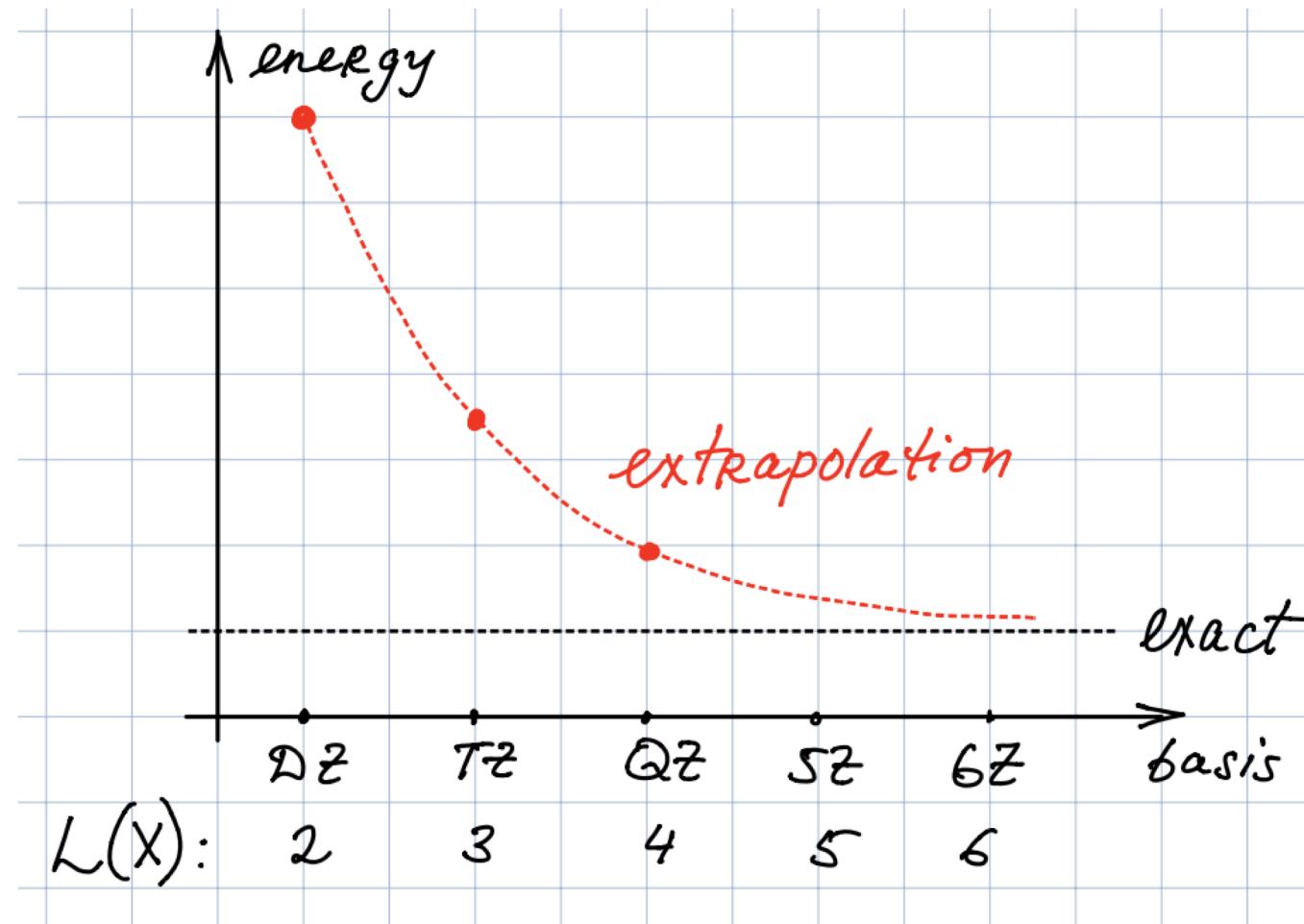
What happens if we perform a calculation on our molecule in a larger and larger basis?

→ Energy will go down and will eventually reach the “complete basis set” (CBS) limit

Can we find CBS limit without performing the calculation in the infinitely large (complete) basis set?

→ Typically YES

There are several explored basis set extrapolation schemes used in combination with correlation consistent basis sets cc-pVXZ and aug-cc-pVXZ



Basis set extrapolation

In **correlation consistent basis sets** cc-pVXZ and aug-cc-pVXZ, the highest angular momentum is equal to X:

$$\text{cc-pVDZ} \rightarrow X=2 \quad (\text{d})$$

$$\text{cc-pVTZ} \rightarrow X=3 \quad (\text{f})$$

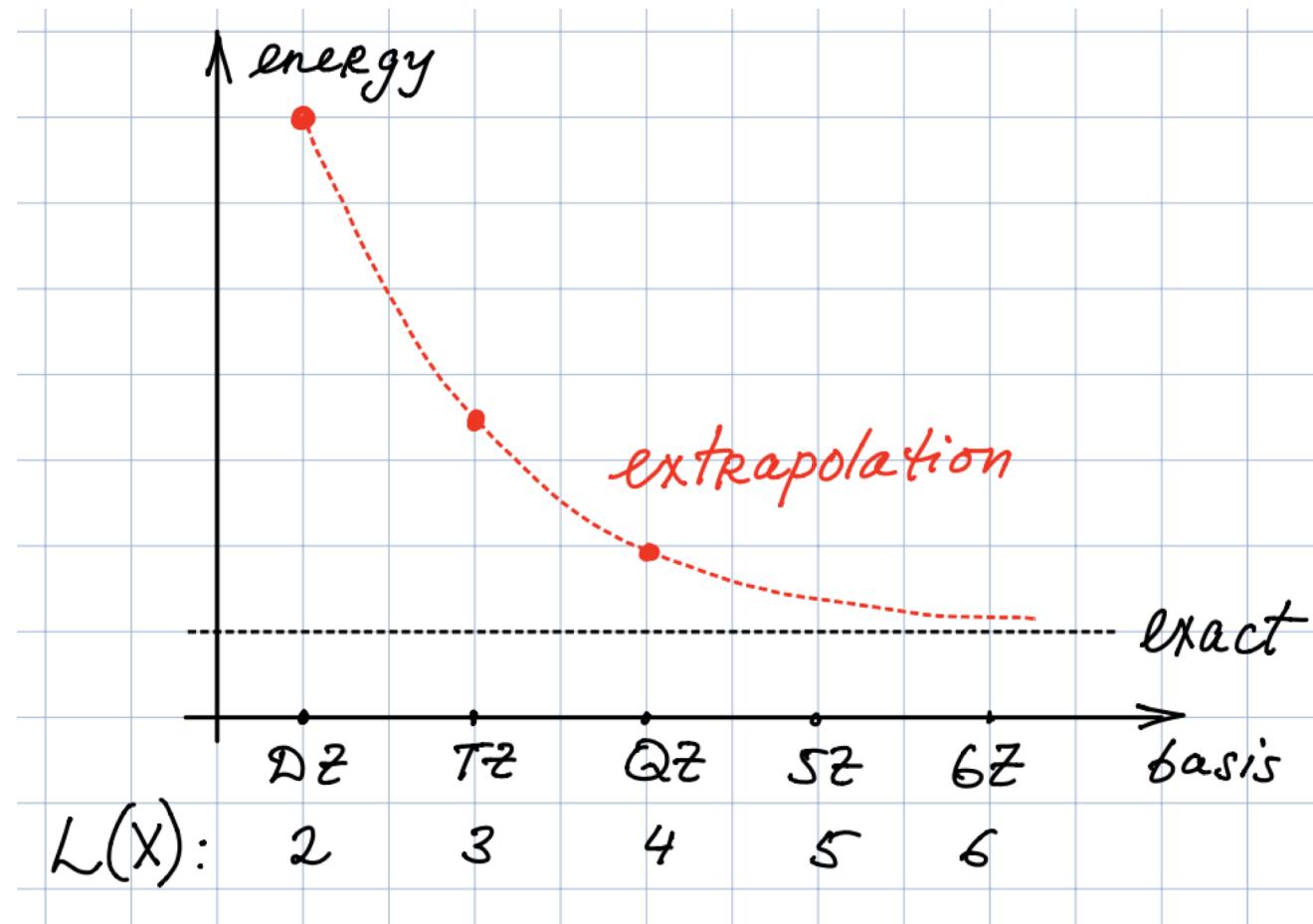
$$\text{cc-pVQZ} \rightarrow X=4 \quad (\text{g})$$

Hartree-Fock and DFT energies converge as:

$$E_X^{HF} = E_{CBS}^{HF} + A \exp(-B\sqrt{X})$$

Correlation energy (e.g. in MP2, CCSD(T)) converges slower:

$$\Delta E_X^{corr} = \Delta E_{CBS}^{corr} + C X^{-3}$$



Beware: DZ basis is often too small for good extrapolation. If possible, start extrapolation with TZ basis

Basis set extrapolation

Hartree-Fock and DFT energies converge as:

$$E_X^{HF} = E_{CBS}^{HF} + A \exp(-B\sqrt{X})$$

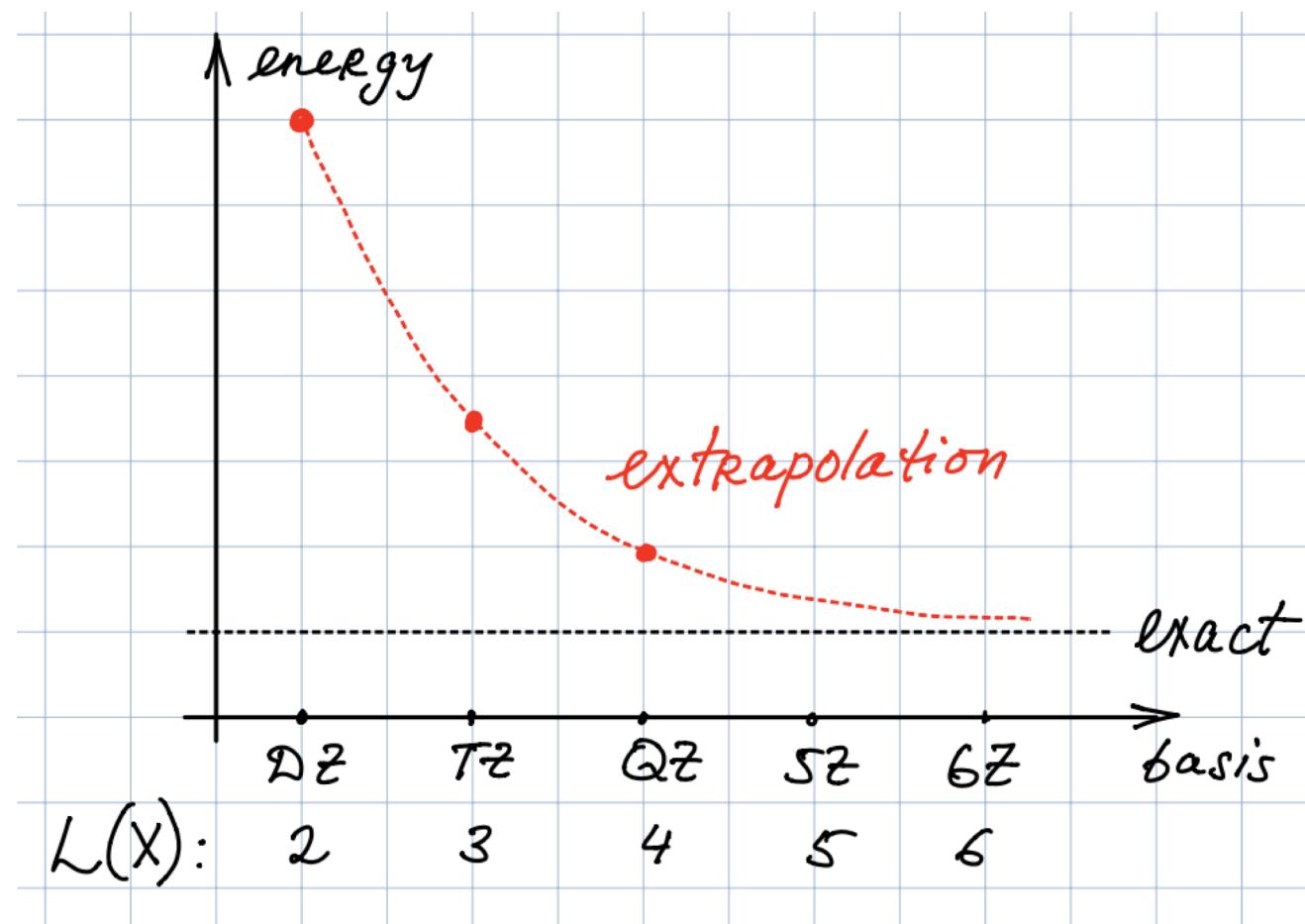
Correlation energy converges as:

$$\Delta E_X^{corr} = \Delta E_{CBS}^{corr} + C X^{-3}$$

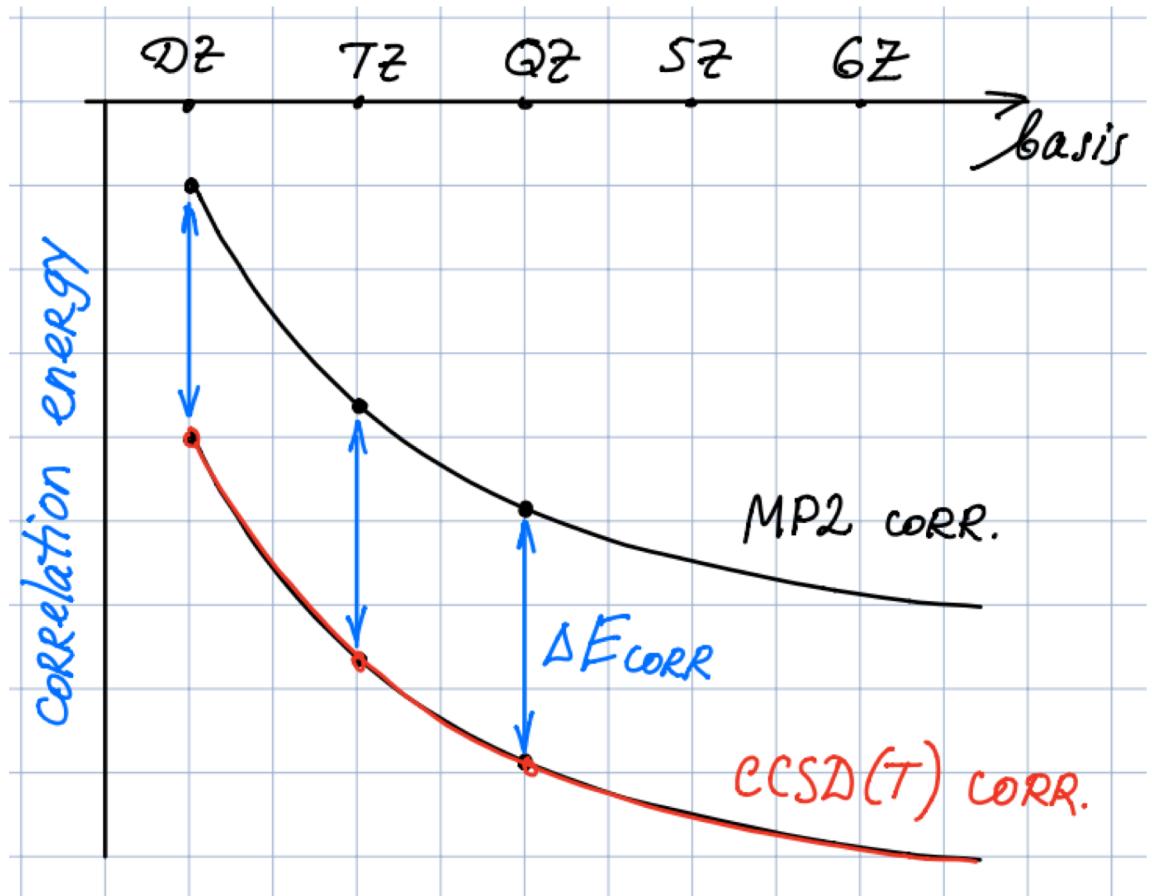
assume we can perform HF in DZ, TZ, QZ

however, only CCSD(T)/DZ is practical

Can we still estimate CCSD(T)/CBS energy?



Energy additivity scheme



Compare MP2 and CCSD(T) correlation energies in different basis sets.
Often, the correlation energy difference remains fairly constant in different bases.

Why?

MP2 captures the effect from pairs of electrons located on the same MO →
Large basis set is needed

CC captures the remaining correlation of disjoint electrons – most of it is already present in a small DZ or TZ basis

Then we can extrapolate the CCSD(T) correlation energy in e.g. QZ basis as:
 $E_{QZ}(\text{CCSD(T)}) = E_{DZ}(\text{CCSD(T)}) + [E_{QZ}(\text{MP2}) - E_{DZ}(\text{MP2})]$

Composite extrapolation schemes

Goal: to estimate “infinite correlation, infinite basis” limit efficiently. Generally, we need:

1. Select the geometry
2. Estimate the Hartree-Fock energy
3. Estimate the electron correlation energy
4. Estimate the energy from translation, rotation and vibration

Note: different properties converge differently with respect to basis set and correlation

Predefine the target accuracy of the extrapolation scheme (often, $\sim 1 \text{ kcal/mol} = \sim 4 \text{ kJ/mol}$)

Then, the errors in each step should be kept below the target accuracy

The error is defined as the change in energy between the present calculation and the “infinite correlation, infinite basis” limit

Composite extrapolation schemes

1. Select the geometry

geometries converge fast: HF/DZP often gives ~ 1 kcal/mol error; MP2/DZP or DFT/DZP are typically sufficient

2. Estimate the Hartree-Fock energy

error depends on the basis set; energy scales as $\sim \exp(-\sqrt{X})$; estimating HF limit is rarely the problem

3. Estimate the electron correlation energy

this is the main bottleneck! \rightarrow steep scaling of correlated methods and slow convergence of the correlation energy with respect to basis set

often, a combination of MP2 and CCSD(T) (via energy additivity scheme) is used

4. Estimate the energy from translation, rotation and vibration

rotation and translation: depend on geometry only; absolute energies are small; errors can be neglected

vibrations (ZPE energies): calculations are hard, but absolute energies are small, so large relative errors can be tolerated. Also, errors in frequencies are systematic and can be scaled

Gaussian-n models

Gn, n=1,2,3,4

Each model exists in several different versions, e.g. Gn(MP2), GnB3 etc

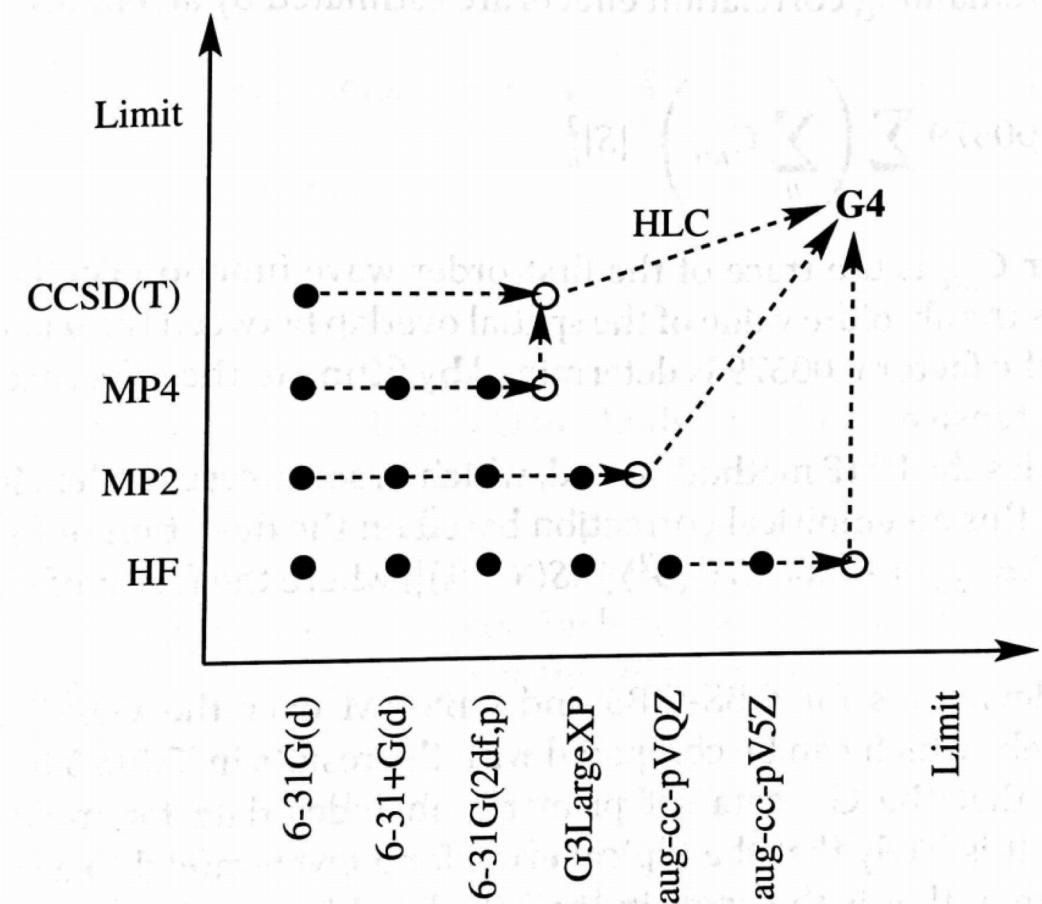
Performance and parameterization is based on experimental molecular data sets:
enthalpies of formation, ionization potentials and electron affinities

Performance: G2: ~6.2 kJ/mol;

G3: ~4.5 kJ/mol; G4: ~ 3.5 kJ/mol

G4:

- Geometry and frequencies at B3LYP/6-31G(2df,p)
- Spin-orbit corrections from experiments or other theoretical estimates
- Higher-level correction (HLC) includes empirical parameters fitted to the reference data ← HLC makes G4 non-size consistent



Basis set superposition error (BSSE)

Consider complexation of molecules A and B:

$$\Delta E_{\text{complex}} = E(\text{AB})^*_{ab} - E(\text{A})_a - E(\text{B})_b$$

* denotes optimized geometry of the complex

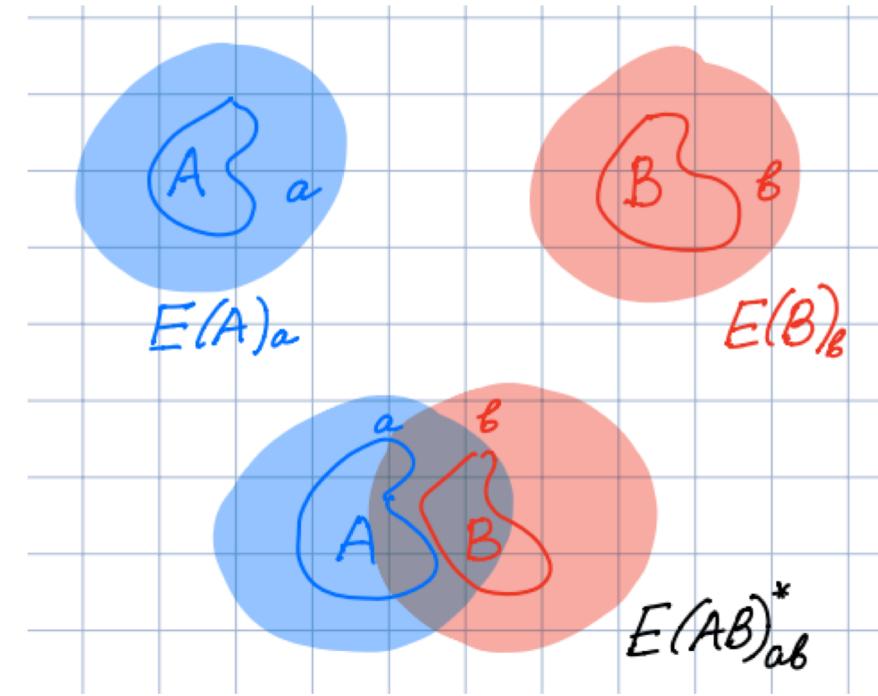
In complex AB, part A can borrow some basis functions from B, and vice versa →

Effectively, both A and B parts in AB are computed in a larger basis set (compared to individual molecules A and B) →

Computed $E(\text{AB})^*_{ab}$ is too low (too negative) →

$\Delta E_{\text{complex}}$ is too negative (binding is too strong)

This is basis set superposition error (BSSE). BSSE becomes smaller when basis set is increased (A and B do not need to borrow neighbor basis functions)

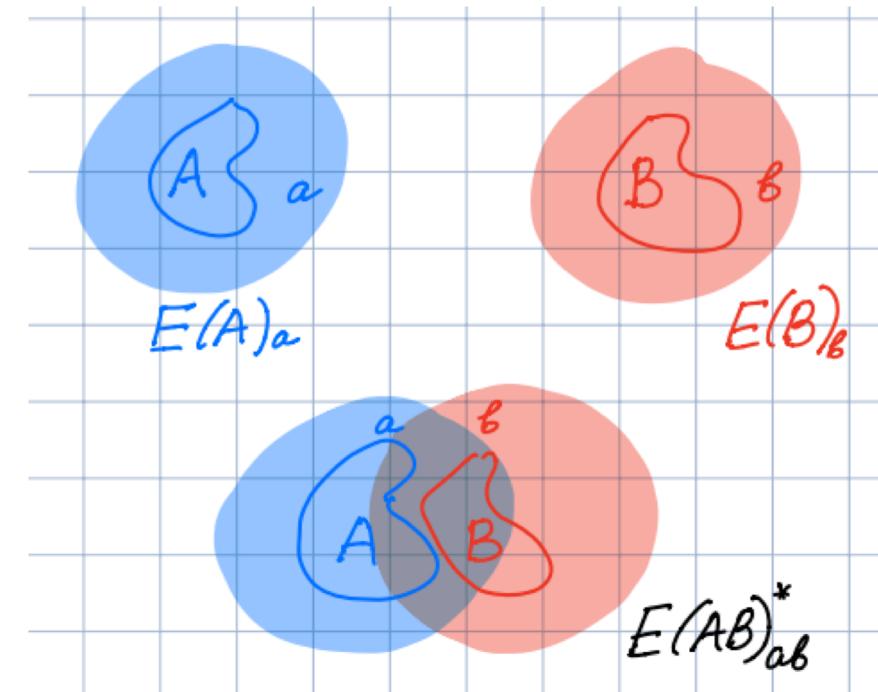
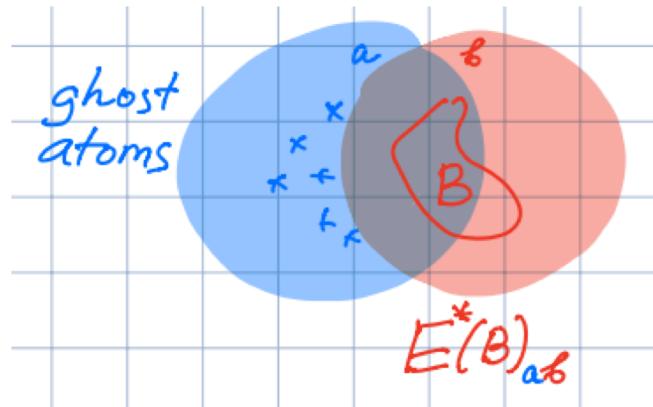
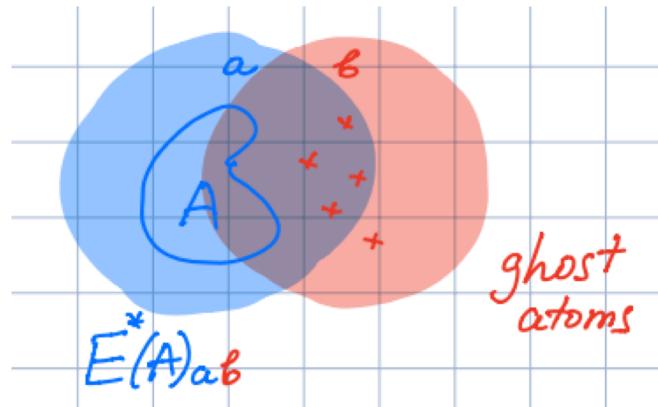


Basis set superposition error (BSSE)

$$\Delta E_{\text{complex}} = E(\text{AB})^*_{ab} - E(\text{A})_a - E(\text{B})_b$$

An approximate way to correct the basis set superposition error is the **Counterpoise (CP) correction**

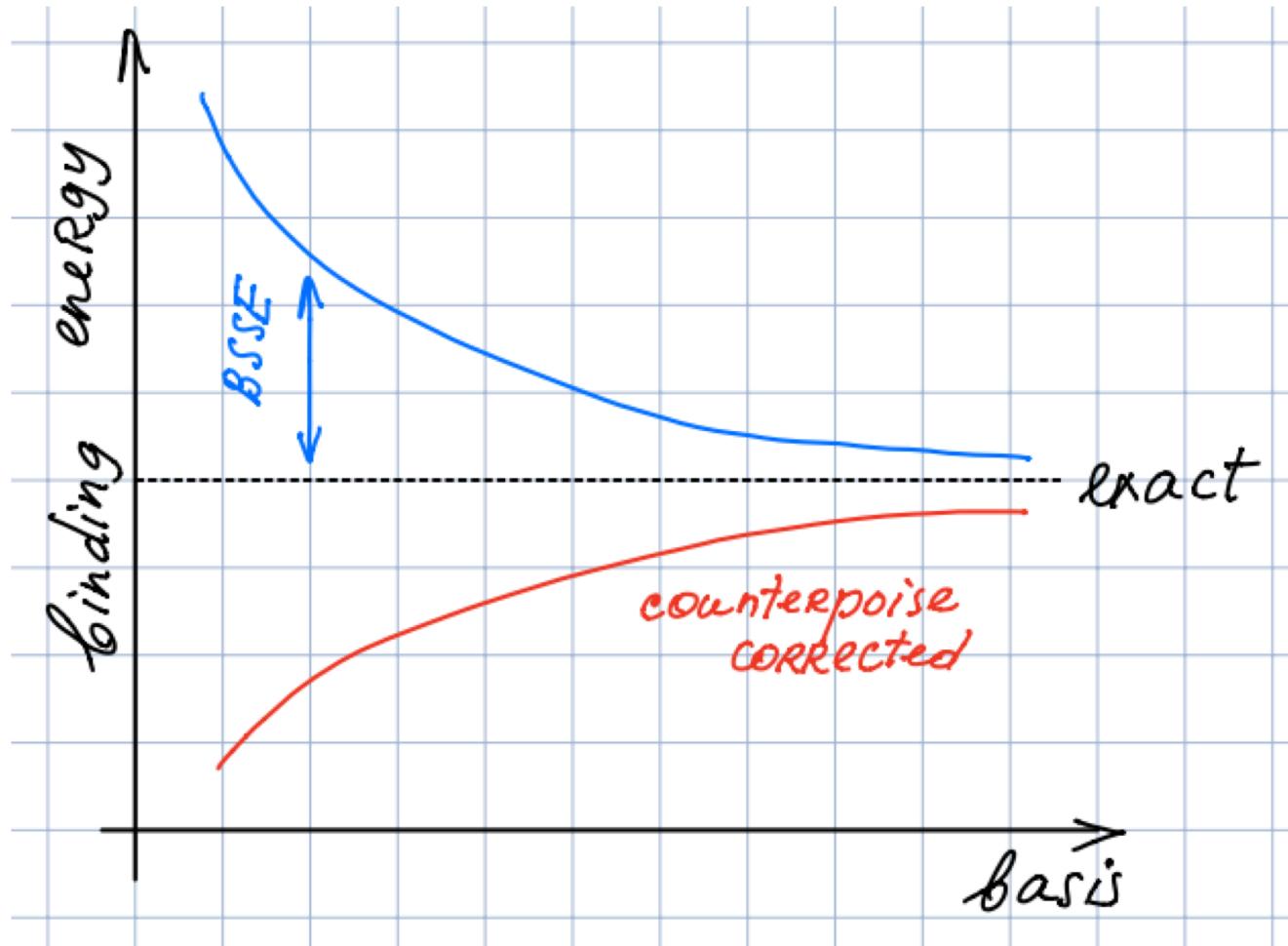
Let's correct monomer energies $E(\text{A})_a$ and $E(\text{B})_b$ by recomputing them in a joint ab basis



$$\Delta E_{\text{CP}} = E(\text{A})^*_{ab} - E(\text{A})^*_a + E(\text{B})^*_{ab} - E(\text{B})^*_b$$

CP corrected complexation energy is then
 $\Delta E_{\text{complex}} (\text{CP}) = \Delta E_{\text{complex}} + \Delta E_{\text{CP}}$

Basis set superposition error (BSSE)



CP correction tends to understabilize binding (now monomers A and B are too well described compared to the dimer)

Non CP-corrected and CP-corrected energies approach the exact binding energy from different sides.

Sometimes, reasonable estimate can be obtained by taking an average between the two schemes.