Lecture 15 · What we learned

Debrief of Experiment 10 and a synthesis of the course

Table of contents

1 Part 1 · Looking back at Experiment 10 . . . 1
1.1 The spread of choices . . . 1
1.2 What tended to work . . . 2
1.3 What made a good report . . . 2
2 Part 2 · What we learned . . . 2
2.1 Two approximations, revisited . . . 2
2.2 A hierarchy of errors . . . 3
2.3 The black box warning . . . 3
2.4 Where the field is going . . . 4
2.5 Course feedback . . . 4
2.6 Key points . . . 5
2.7 Study questions . . . 5
2.8 Further reading . . . 5
Bibliography . . . 5

i Learning goals

After this lecture you should be able to:

- look at the spread of methods and results in Experiment 10 and explain what made the good choices good;
- place every source of error you met in this course in a hierarchy, with typical sizes;
- explain why a quantum chemistry program can't be used as a black box;
- name three directions in which electronic-structure methods are developing, and what problem each addresses.

1 Part 1 · Looking back at Experiment 10

! Read this after you submit Experiment 10

This part discusses what the class typically chose and found in Experiment 10. If your instructor has released the class results, open the class results page alongside it.

1.1 The spread of choices

For the first time in the course, everyone ran something different. The class results usually show a wide spread: several functionals from different rungs, with and without dispersion, two or three basis sets, different grids. The computed bond energies typically spread over more than 10 kcal/mol around the reference of 16.2 kcal/mol. That spread is the most important result of the experiment. With the same three geometries and the same program, the choices alone moved the answer by more than any chemist would accept.

1.2 What tended to work

A few patterns are typical.

Dispersion made the biggest difference. The hexatriene ligand is large and polarizable, and it sits close to the two phosphines. Part of the binding is dispersion between the ligand and the rest of the complex, the same physics as the alkane and dimer reactions of Experiment 4. Functionals with no dispersion at all, B3LYP being the usual example, typically *underbind* the complex by many kcal/mol, far outside chemical accuracy. Adding -D4, or choosing a functional with VV10 or with dispersion built into its parametrization, usually closed most of that gap.

Modern functionals with good transition-metal benchmarks did well. Students who looked up metal–ligand bond energies in the benchmark literature (Lecture 14) and chose a dispersion-inclusive hybrid or meta-GGA that performs well there usually landed within a few kcal/mol of the reference, often closer.

Triple-ζ basis sets with the def2 ECP were the sensible choice. def2-TZVP fits the 10-minute limit for this complex with most functionals. def2-SVP is faster, but its basis-set superposition error makes the complex look more stable than it is. Combined with -D4, it can overbind: the double counting of Experiment 6 again. A def2-SVP result that lands close to 16.2 should make you suspicious, not happy.

The details mattered less here, but they still mattered. A palladium(0) d^{10} complex is closed-shell with a reasonable HOMO–LUMO gap, so the stability analysis usually found nothing to fix. That was not a wasted check: you didn't know it in advance. Grid effects on a 16 kcal/mol reaction energy are usually small, but meta-GGAs on SG-1 can shift the result by a few tenths of a kcal/mol.

1.3 What made a good report

The best reports had three things in common. They chose the method *before* computing, for stated reasons tied to benchmarks and to the course's evidence. Their methods paragraph was complete enough to rebuild every input. And they discussed the result honestly: the signed error, the likely sources of the remaining error, and what they would change with more computer time. A report that lands 2 kcal/mol away with a clear justification is better work than one that lands on 16.2 and can't explain why.

> ⚠ Common mistake: tuning to the reference
>
> It is tempting to try several functionals and keep the one closest to 16.2. In real consulting there is no reference value to tune to; that's why you were asked. A method chosen by matching the answer has no predictive power for the next molecule, the one without a reference.

2 Part 2 · What we learned

2.1 Two approximations, revisited

In Lecture 1 you met the idea that every calculation makes two approximations (see Lecture 1). Ten experiments later, you have seen each of them from several angles.

Approximating the wave function: the basis set. A finite basis can't represent the exact orbitals. The variational principle guarantees that a larger basis lowers the energy, and Experiment 1 showed reaction energies converging toward the complete-basis-set limit as ζ increases, at a cost that grows steeply. Experiment 6 showed the dark side: when the basis is small, its error can cancel the functional's error by luck, for one reaction and not for the next. For heavy elements, the basis comes with an effective core potential, which also brings in scalar relativistic effects (Experiments 8 and 10).

Approximating the Hamiltonian: the method. Hartree–Fock replaces electron–electron repulsion with a mean field and has no correlation. Kohn–Sham DFT is exact in principle, but every practical calculation uses an approximate E_{xc}. Jacob's ladder organizes the approximations by their ingredients (Experiment 2). Experiments 3–5 showed that the same functional can be excellent for one kind of reaction and poor for another, depending on error cancellation, dispersion, self-interaction error and static correlation. Experiments 8–10 added open shells and transition metals, where even the best functionals struggle.

Beyond those two, you met approximations that hide inside the program: the **integration grid** (Experiment 7), the **SCF solution** that may not be the lowest one (Experiment 8), the **convergence thresholds**, the **geometry** you didn't optimize, and the gap between an **electronic energy** and anything an experiment measures (Experiment 9).

2.2 A hierarchy of errors

The sources of error have very different typical sizes. Here they are in the order you are most likely to meet them, with the sizes you saw in this course.

Table 1: Sources of error in a DFT calculation, roughly in decreasing order of typical size.

Source of error	Typical size in this course	Where you saw it
Method (functional, or HF)	0.1–10 kcal/mol for good functionals; tens for hard cases; hundreds for HF atomizations	Experiments 3–5, 8–10
Basis set	a few kcal/mol from double- to triple-ζ; tens when a small basis is badly unbalanced	Experiments 1, 6
Missing physics (dispersion, spin–orbit, ZPE, thermal)	0.2–25 kcal/mol, depending on the reaction	Experiments 4, 9, 10
Wrong SCF solution	zero, or tens of kcal/mol: it is all or nothing	Experiment 8
Integration grid	0.01–0.5 kcal/mol in energies; more for weak interactions and frequencies	Experiment 7
SCF convergence threshold	negligible at 10^{-8}; noticeable only with loose thresholds	throughout

The general order is method > basis set > grid > SCF convergence. But the table has two warnings built in. First, the ranges overlap: a bad basis can be worse than a mediocre functional, and an unstable SCF solution can be the largest error of all. Second, several errors are **all or nothing**. A wrong spin state, a missing ECP, a forgotten dispersion correction or a wrong stoichiometric coefficient doesn't make the answer a bit worse; it makes it meaningless. Every one of them can bite, and the small ones bite exactly when the quantity you want is small, as the 0.28 kcal/mol well of Ar_2 showed.

2.3 The black box warning

Modern programs make a DFT calculation look easy. You draw a molecule, choose a functional from a menu, press Submit, and a number with eight decimals comes back. Nothing in that number tells you whether

- the functional is appropriate for the chemistry;
- the basis set is converged, or lucky;

- the grid is fine enough for the functional;
- the SCF found the lowest solution in the right spin state;
- the number is the quantity you meant to compare with.

Every experiment in this course was designed to expose one of these hidden choices. The program will not warn you about any of them. A quantum chemistry program is not a black box; it is a set of approximations you choose, whether you know it or not. The defaults are someone else's choices, made for someone else's problem, in some earlier version of the program. The practical rules that follow from this course are short:

1. Know which approximations you are making, and why.
2. Converge what can be converged (basis, grid, SCF), or at least check how far you are from convergence.
3. Choose the method with benchmark evidence for your kind of chemistry, not by habit or popularity.
4. Check the SCF solution and the spin state for anything open-shell or containing a metal.
5. Compare like with like: electronic energies with electronic references, enthalpies with enthalpies.
6. Report everything that changes the number.

When a problem is outside what you can check, for example strong multireference character or an unfamiliar class of transition-metal compounds, the right move is the one the experiment pages kept recommending: ask a theoretician.

2.4 Where the field is going

DFT has been the workhorse of computational chemistry for thirty years, and the approximations you used will remain in use for many more. Three directions are changing what is possible.

Machine-learned functionals. Instead of fitting a dozen parameters to a few hundred data points, as the Minnesota functionals did, one can train a neural network to represent the exchange–correlation energy density from the local ingredients of the density. The DM21 functional, for example, was trained on molecular data together with exact conditions for systems with fractional charge and spin, which target the delocalization and static-correlation errors of Lecture 9 [1]. Such functionals can be very accurate on their training domain. They also bring back, in a stronger form, the questions this course raised: how well they transfer to chemistry outside the training data, and how smooth, and therefore how grid-friendly, they are (Lecture 11).

Double hybrids. The fifth rung of Jacob's ladder adds a second-order perturbation correlation term computed from the virtual orbitals, as in B2PLYP [2]. In large benchmarks such as GMTKN55, the best double hybrids are the most accurate functionals overall [3], [4]. They cost about as much as MP2, more than a hybrid but far less than coupled cluster, and they need larger basis sets because the correlation part converges slowly with ζ, like the cc-pVnZ behavior you saw in Experiment 1.

Local correlation methods. The reference values in this course came from CCSD(T) at the complete-basis-set limit or from experiment. Canonical CCSD(T) scales as N^7, which limited it to small molecules. Local correlation methods exploit the fact that electron correlation is short-ranged. DLPNO-CCSD(T), for example, compresses the correlation space of each electron pair, and its cost grows nearly linearly with molecular size [5]. That makes coupled-cluster accuracy, or close to it, available for molecules with a hundred atoms or more. In practice it means two things for a DFT user: better reference values for benchmarking, and the option of checking a DFT result for your own system directly.

None of these removes the need for judgment. A machine-learned functional still needs a grid and a basis set, a double hybrid still needs a converged basis and a stable reference determinant, and a local correlation calculation has its own thresholds to report. The habits you built in this course transfer to all of them.

2.5 Course feedback

The last part of this lecture is for you. Before you leave, write down:

- **The muddiest point:** the idea from this course that is still the least clear to you.
- **The most useful experiment,** and why.
- **One thing to change** for next year's class.

Your instructor will collect these anonymously. They shape the next version of the course.

2.6 Key points

- In Experiment 10, the choices alone typically moved the answer by more than 10 kcal/mol. Dispersion was the biggest single factor; a triple-ζ basis with the def2 ECP was the sensible compromise; a small basis with -D4 can overbind.
- A good consulting report justifies the choice before computing, describes a reproducible setup, and discusses the result honestly.
- Every calculation approximates both the wave function (basis set) and the Hamiltonian (method), and also hides choices of grid, SCF solution, thresholds, geometry and the quantity compared.
- Typical error sizes: method > basis > grid > SCF convergence, but the ranges overlap, and several errors are all or nothing.
- A program is not a black box: defaults are someone else's choices.
- Machine-learned functionals, double hybrids and local correlation methods are extending what is possible, and each still needs the same care.

2.7 Study questions

1. Why does a dispersion correction matter so much for the palladium–hexatriene bond? Which experiment showed the same effect in a simpler system?
2. A classmate's def2-SVP result is closer to 16.2 than your def2-TZVP result with the same functional and -D4. Explain how this can happen and which result you would trust.
3. For each row of Table 1, name one experiment where that error was the largest one in the calculation.
4. Give two examples of "all or nothing" errors, and say how you would detect each.
5. Explain in your own words why defaults are a choice, using the grid as the example.
6. What problem do fractional-charge and fractional-spin conditions address in a machine-learned functional? Which class discussed that problem?
7. Why do double hybrids need larger basis sets than hybrids?
8. How could a local correlation method help you decide which functional to use for a new class of molecules?

2.8 Further reading

- P. Morgante and R. Peverati, "The devil in the details: a tutorial review on some undervalued aspects of density functional theory calculations" [6]: the paper this lab is built on.
- N. Mardirossian and M. Head-Gordon, "Thirty years of density functional theory in computational chemistry" [7].
- L. Goerigk et al., GMTKN55 [3], and N. Mehta, M. Casanova-Páez and L. Goerigk on double hybrids [4].
- J. Kirkpatrick et al., the DM21 machine-learned functional [1].
- C. Riplinger et al., DLPNO-CCSD(T) [5].
- A. J. Cohen, P. Mori-Sánchez and W. Yang, "Challenges for density functional theory" [8].

Bibliography

[1] J. Kirkpatrick, B. McMorrow, D. H. P. Turban, A. L. Gaunt, and others, "Pushing the frontiers of density functionals by solving the fractional electron problem," *Science*, vol. 374, pp. 1385–1389, 2021, doi: 10.1126/science.abj6511.

[2] S. Grimme, *J. Chem. Phys.*, vol. 124, p. 34108, 2006.

[3] L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, and S. Grimme, *Phys. Chem. Chem. Phys.*, vol. 19, pp. 32184–32215, 2017.

[4] N. Mehta, M. Casanova-Páez, and L. Goerigk, *Phys. Chem. Chem. Phys.*, vol. 20, pp. 23175–23194, 2018.

[5] C. Riplinger, B. Sandhoefer, A. Hansen, and F. Neese, "Natural triple excitations in local coupled cluster calculations with pair natural orbitals," *J. Chem. Phys.*, vol. 139, p. 134101, 2013, doi: 10.1063/1.4821834.

[6] P. Morgante and R. Peverati, "The devil in the details: A tutorial review on some undervalued aspects of density functional theory calculations," *Int. J. Quantum Chem.*, vol. 120, p. e26332, 2020.

[7] N. Mardirossian and M. Head-Gordon, *Mol. Phys.*, vol. 115, pp. 2315–2372, 2017.

[8] A. J. Cohen, P. Mori-Sánchez, and W. Yang, *Chem. Rev.*, vol. 112, pp. 289–320, 2012.