Chapter 7: Answer Key

Scaling and Properties

2302638 Advanced Quantum Chemistry

Department of Chemistry, Chulalongkorn University

Contents

1   Checkpoint Question Answers

This section provides detailed answers to all 12 checkpoint questions from Chapter 7, organized by their location in the chapter.

1.1   Checkpoint 7.1: Big Picture (Week 7)

Section 7.1 – Learning Goals

Question: Modern electronic structure is not only about deriving formulas, but also about controlling:

$$\text{accuracy} \leftrightarrow \text{cost} \leftrightarrow \text{numerical stability}.$$

You should leave this chapter able to recognize which lever to pull when a calculation becomes slow, memory-limited, or unstable.

Answer:

This checkpoint sets up the conceptual framework for Chapter 7. The three "levers" that practitioners can adjust are:

1. Accuracy controls:

- Basis set size (minimal $\rightarrow$ double-ζ $\rightarrow$ triple-ζ $\rightarrow$ CBS limit)

- Integral screening thresholds (Schwarz threshold)

- SCF convergence criteria (ΔE, $\|\Delta \mathbf{P}\|$, orbital gradient)

- Density fitting auxiliary basis quality

2. Cost controls:

- In-core vs. direct SCF (memory vs. recomputation)

- Conventional vs. density-fitted ERIs ($\mathcal{O}(N^4)$ vs. $\mathcal{O}(N^3)$)

- Schwarz screening (skip negligible shell quartets)

- Basis set size reduction

3. Stability controls:

- Eigenvalue thresholding for near-linear dependence

- DIIS subspace management

- Level shifting and damping for difficult convergence

- Choice of initial guess (SAD vs. core Hamiltonian)

> ### Diagnostic Workflow
>
> When a calculation fails or gives suspicious results:
>
> 1. Check $\kappa(\mathbf{S})$: Is the basis ill-conditioned?
>
> 2. Check SCF convergence: Did the calculation actually converge?
>
> 3. Check virial ratio η: Is the basis adequate?
>
> 4. Compare with a simpler reference (smaller basis, conventional HF)

1.2   Checkpoint 7.2: Scaling Bottleneck Identification

> ### Section 7.2 – Scaling Analysis
>
> **Question:** For a typical in-core HF calculation with $N = 200$ AOs and 12 SCF iterations:
>
> 1. Estimate total J/K operations: $\sim k \times 2N^4 = 12 \times 2 \times 200^4 \approx 3.8 \times 10^{10}$.
>
> 2. Estimate total diagonalization operations: $\sim k \times \frac{2}{3}N^3 \approx 6.4 \times 10^7$.
>
> 3. Ratio: J/K dominates by a factor of ~ 600.
>
> At what N would diagonalization begin to compete? *Hint:* When $N^4 \sim N^3 \cdot k$, i.e., $N \sim k$. For $k = 10$, the crossover occurs around $N \approx 1000$–2000 (rarely reached without approximations).
>
> **Answer:**
> Let us work through this scaling analysis systematically.
> **Part 1: J/K build operations**
> Each SCF iteration requires building $\mathbf{J}$ and $\mathbf{K}$ from ERIs:
>
> $$J_{\mu\nu} = \sum_{\lambda\sigma}(\mu\nu|\lambda\sigma)P_{\lambda\sigma}, \tag{1}$$
>
> $$K_{\mu\nu} = \sum_{\lambda\sigma}(\mu\lambda|\nu\sigma)P_{\lambda\sigma}. \tag{2}$$
>
> Each contraction involves N^4 multiply-add operations. For $N = 200$:
>
> $$N^4 = 200^4 = 1.6 \times 10^9 \text{ operations per matrix.} \tag{3}$$
>
> Building both $\mathbf{J}$ and $\mathbf{K}$ requires $\sim 2N^4$ operations per iteration. Over $k = 12$ iterations:
>
> $$\text{Total J/K ops} = 12 \times 2 \times 1.6 \times 10^9 = 3.84 \times 10^{10}. \tag{4}$$
>
> **Part 2: Diagonalization operations**
> Each diagonalization of the $N \times N$ Fock matrix requires $\sim \frac{2}{3}N^3$ operations (for full eigenvalue decomposition):
>
> $$\frac{2}{3}N^3 = \frac{2}{3}(200)^3 = 5.33 \times 10^6 \text{ operations per iteration.} \tag{5}$$

Over 12 iterations:

$$\text{Total diag ops} = 12 \times 5.33 \times 10^6 = 6.4 \times 10^7. \tag{6}$$

Part 3: Ratio and crossover
The ratio is:

$$\frac{\text{J/K ops}}{\text{Diag ops}} = \frac{3.84 \times 10^{10}}{6.4 \times 10^7} = 600. \tag{7}$$

J/K dominates by a factor of 600, confirming that ERI-related operations are the bottleneck.

Crossover analysis:
The crossover occurs when $kN^4 \sim kN^3$, i.e., when $N \sim k$. More precisely, if we set the costs equal:

$$2kN^4 = k \cdot \frac{2}{3} N^3 \quad \Rightarrow \quad N = \frac{1}{3}. \tag{8}$$

This is clearly wrong—the operations have different prefactors. The correct comparison is:

$$2N^4 \sim \frac{2}{3} N^3 \quad \Rightarrow \quad N \sim \frac{1}{3}. \tag{9}$$

The hint suggests $N \sim k$, but this is an approximation. In reality, the $\mathcal{O}(N^4)$ always dominates for systems where conventional HF is feasible. The crossover to diagonalization-dominated regime only occurs for $N \approx 1000$–2000, which is beyond practical conventional HF limits.

> **Why DF is Essential**
>
> For $N > 200$, storing the full ERI tensor becomes impractical:
>
> $$\text{Storage} \approx \frac{N^4}{8} \times 8 \text{ bytes} = N^4 \text{ bytes.} \tag{10}$$
>
> For $N = 500$: $500^4 = 6.25 \times 10^{10}$ bytes ≈ 58 GB—often exceeding available RAM. Density fitting reduces this to $\sim N^2 N_{\text{aux}} \times 8$ bytes, making larger systems tractable.

1.3 Checkpoint 7.3: DF Index Structure

> **Section 7.3 – Density Fitting**
>
> **Question:** Verify for yourself: why does $J_{\mu\nu} = \sum_Q B_{\mu\nu}^Q d^Q$ work directly from Eq. (7.8), but $K_{\mu\nu}$ cannot be written in the same form? *Hint:* Write out the sum over λ, σ and see which indices contract together.
>
> **Answer:**
> This checkpoint addresses the fundamental reason why density fitting is more efficient for Coulomb ($\mathbf{J}$) than for exchange ($\mathbf{K}$).
>
> **Starting point: DF approximation**
> The density-fitted ERI is:
>
> $$(\mu\nu|\lambda\sigma) \approx \sum_Q B_{\mu\nu}^Q B_{\lambda\sigma}^Q, \tag{11}$$
>
> where $B_{\mu\nu}^Q = \sum_P (\mu\nu|P)(P|Q)^{-1/2}$.
> **Coulomb matrix (J):**

The Coulomb matrix is defined as:

$$J_{\mu\nu} = \sum_{\lambda\sigma} (\mu\nu|\lambda\sigma) P_{\lambda\sigma}. \tag{12}$$

Substituting the DF approximation:

$$J_{\mu\nu} = \sum_{\lambda\sigma} \left(\sum_Q B_{\mu\nu}^Q B_{\lambda\sigma}^Q \right) P_{\lambda\sigma} \tag{13}$$

$$= \sum_Q B_{\mu\nu}^Q \underbrace{\left(\sum_{\lambda\sigma} B_{\lambda\sigma}^Q P_{\lambda\sigma} \right)}_{=d^Q} \tag{14}$$

$$= \sum_Q B_{\mu\nu}^Q d^Q. \tag{15}$$

The indices $(\mu\nu)$ and $(\lambda\sigma)$ are *decoupled*—they appear in separate factors $B_{\mu\nu}^Q$ and $B_{\lambda\sigma}^Q$. This allows us to:

1. First compute $d^Q = \sum_{\lambda\sigma} B_{\lambda\sigma}^Q P_{\lambda\sigma}$ (cost: $\mathcal{O}(N^2 N_{\text{aux}})$)

2. Then compute $J_{\mu\nu} = \sum_Q B_{\mu\nu}^Q d^Q$ (cost: $\mathcal{O}(N^2 N_{\text{aux}})$)

Exchange matrix (K):
The exchange matrix has a different index structure:

$$K_{\mu\nu} = \sum_{\lambda\sigma} (\mu\lambda|\nu\sigma) P_{\lambda\sigma}. \tag{16}$$

Note: the ERI indices are $(\mu\lambda|\nu\sigma)$, not $(\mu\nu|\lambda\sigma)$. Substituting the DF approximation:

$$K_{\mu\nu} = \sum_{\lambda\sigma} \left(\sum_Q B_{\mu\lambda}^Q B_{\nu\sigma}^Q \right) P_{\lambda\sigma} \tag{17}$$

$$= \sum_Q \left(\sum_\lambda B_{\mu\lambda}^Q P_{\lambda\sigma} \right) \left(\sum_\sigma B_{\nu\sigma}^Q \right). \tag{18}$$

Wait—this does not simplify cleanly! The problem is that λ appears in $B_{\mu\lambda}^Q$ and in $P_{\lambda\sigma}$, while σ appears in $B_{\nu\sigma}^Q$ and in $P_{\lambda\sigma}$. The density matrix $P_{\lambda\sigma}$ couples both summations.
Why the indices don't decouple:
For Coulomb:

- $B_{\mu\nu}^Q$: indices (μ, ν) on *electron 1*

- $B_{\lambda\sigma}^Q P_{\lambda\sigma}$: indices (λ, σ) on *electron 2*

- These are separate electrons, so the indices naturally decouple.

For exchange:

- $B_{\mu\lambda}^Q$: mixes μ (bra-1) with λ (ket-1)

- $B_{\nu\sigma}^Q$: mixes ν (bra-2) with σ (ket-2)

- The density $P_{\lambda\sigma}$ connects λ and σ

- The "crossing" pattern $(\mu\lambda|\nu\sigma)$ prevents simple factorization

> **DF Exchange Algorithms**
>
> Practical DF-HF codes use more sophisticated algorithms for exchange:
>
> - **Loop-based:** Iterate over auxiliary index Q, form intermediate matrices
>
> - **MO-based:** Transform to occupied MOs, compute $K_{\mu\nu} = \sum_i \sum_Q L_{\mu i}^Q L_{\nu i}^Q$ where $L_{\mu i}^Q = \sum_\lambda B_{\mu\lambda}^Q C_{\lambda i}$
>
> Both approaches have scaling $\mathcal{O}(N^2 N_{\text{aux}} N_{\text{occ}}) \sim \mathcal{O}(N^3)$, still better than $\mathcal{O}(N^4)$, but more complex than the two-step **J** algorithm.

1.4   Checkpoint 7.4: Auxiliary Basis Scaling

> **Section 7.3 – Auxiliary Basis Scaling**
>
> **Question:** Why is it plausible that N_{aux} scales roughly linearly with N? *Hint:* Each atom contributes a fixed number of auxiliary functions to represent pair densities involving its AOs, independent of molecular size.
>
> **Answer:**
> The linear scaling $N_{\text{aux}} \approx cN$ (with $c \approx 2$–5) follows from the *local* nature of pair densities.
> **Argument 1: Atom-centered locality**
> The pair density $\chi_\mu(\mathbf{r})\chi_\nu(\mathbf{r})$ is significant only where both basis functions χ_μ and χ_ν have substantial amplitude. Since atomic orbitals decay exponentially, the pair density is localized near the atoms on which μ and ν are centered.
> For atom A with n_A AO basis functions:
>
> - Significant pair densities involve AOs on atom A or nearby atoms
>
> - Representation requires $\sim m_A$ auxiliary functions centered on A
>
> - The ratio m_A/n_A is roughly constant across atoms of the same element
>
> Summing over all atoms:
>
> $$N_{\text{aux}} = \sum_A m_A \approx c \sum_A n_A = cN. \tag{19}$$
>
> **Argument 2: Pair density angular momentum**
> The product of two Gaussians with angular momenta ℓ_1 and ℓ_2 generates components up to $\ell_1 + \ell_2$. To fit these products:
>
> - $s \times s$ pairs: need auxiliary functions up to s (or p for better fitting)
>
> - $p \times p$ pairs: need auxiliary functions up to d
>
> - $d \times d$ pairs: need auxiliary functions up to g
>
> The required auxiliary angular momentum is fixed by the orbital basis angular momentum, not by molecular size.
> **Numerical verification:**
> For cc-pVDZ with matched cc-pVDZ-JKFIT auxiliary basis:

Molecule	N_{AO}	N_{aux}	N_{aux}/N
H_2O	24	84	3.5
C_2H_6	58	196	3.4
C_6H_6	114	378	3.3
$C_{10}H_{22}$	262	854	3.3

The ratio N_{aux}/N is approximately constant at ~ 3.3–3.5 for this basis set family, confirming linear scaling.

> **Why Linear Scaling Matters**
>
> If $N_{\mathrm{aux}} \propto N$, then:
>
> - DF tensor storage: $N^2 \cdot N_{\mathrm{aux}} \sim N^3$ (vs. N^4 for full ERIs)
>
> - DF-J build: $N^2 \cdot N_{\mathrm{aux}} \sim N^3$ (vs. N^4)
>
> - DF-K build: $N^2 \cdot N_{\mathrm{aux}} \cdot N_{\mathrm{occ}} \sim N^3$ (vs. N^4)
>
> The DF approximation converts quartic to cubic scaling—the key to tractability for systems with $N > 500$.

1.5   Checkpoint 7.5: Basis Set Selection

> **Section 7.4 – Basis Set Selection**
>
> **Question:** For a geometry optimization of a neutral organic molecule, would you choose STO-3G, 6-31G*, or aug-cc-pVTZ? What factors influence this choice? *Hint:* Consider accuracy requirements, computational cost, and whether diffuse functions are needed for the property of interest.
>
> **Answer:**
>
> **Recommendation: 6-31G* (or def2-SVP) for initial optimization**
>
> For routine geometry optimization of neutral organic molecules, 6-31G* represents the optimal balance between accuracy and cost.
>
> **Analysis of each option:**
>
> **1. STO-3G (minimal basis):**
>
> - **Pro:** Extremely fast (fewest functions)
>
> - **Con:** Poor geometry accuracy (bond lengths off by 0.05–0.1 Å)
>
> - **Con:** Missing polarization functions distort bond angles
>
> - **Use case:** Only for crude initial structures or debugging
>
> **2. 6-31G* (split-valence + polarization):**
>
> - **Pro:** Good geometry accuracy (bond lengths within 0.01–0.02 Å of experiment)
>
> - **Pro:** Polarization functions capture hybridization effects
>
> - **Pro:** Reasonable cost (25 functions for H_2O vs. 7 for STO-3G)
>
> - **Use case:** Standard choice for geometry optimization

3. aug-cc-pVTZ (triple-zeta + diffuse + polarization):

- **Pro:** High accuracy (benchmark quality)

- **Con:** Expensive (92 functions for H_2O—3.7$\times$ more than 6-31G*)

- **Con:** Diffuse functions unnecessary for neutral ground-state geometries

- **Con:** May cause linear dependence issues

- **Use case:** Final single-point energies, not routine optimization

When to use each basis:

Task	Recommended Basis
Initial structure exploration	STO-3G or 3-21G
Geometry optimization (neutral)	6-31G* or def2-SVP
Single-point energy (production)	cc-pVTZ or def2-TZVP
Anions / Rydberg states	aug-cc-pVDZ or larger
Polarizabilities / weak interactions	aug-cc-pVTZ or larger

> **Decision Factors**
>
> 1. **Property type:** Geometries need polarization; properties like polarizability need diffuse functions
>
> 2. **System charge:** Anions require diffuse functions; neutrals usually do not
>
> 3. **Accuracy target:** Quantitative energetics need TZ+ quality
>
> 4. **Available resources:** Larger bases require more memory/time
>
> 5. **Numerical stability:** Avoid overly diffuse bases if $\kappa(\mathbf{S}) > 10^8$

1.6   Checkpoint 7.6: Threshold Tightening Tradeoffs

> **Section 7.4 – Numerical Thresholds**
>
> **Question:** A colleague suggests "just set all thresholds to 10^{-14} to be safe." Why might this advice be problematic?
>
> 1. What happens to computational cost as integral thresholds tighten?
>
> 2. What is the relationship between $\kappa(\mathbf{S})$ and achievable precision?
>
> 3. Under what circumstances might looser thresholds actually improve reliability?
>
> **Answer:**
> This "safer is better" intuition fails in numerical linear algebra. Here is why:
> **Part 1: Cost implications**
> Integral screening thresholds (e.g., Schwarz threshold τ) control which shell quartets are computed vs. skipped. Tightening from $\tau = 10^{-10}$ to 10^{-14}:
>
> - Computes 10^4 more quartets (those with magnitude 10^{-14}–10^{-10})
>
> - For spatially extended systems, this can be 10–100$\times$ more quartets

- Increased cost with *no meaningful improvement* in final energy (since these contributions are below basis set error anyway)

Part 2: Precision limits from conditioning

The condition number $\kappa(\mathbf{S}) = s_{\max}/s_{\min}$ determines the *effective precision* of linear algebra operations:

$$\text{Effective precision} \approx \frac{\epsilon_{\text{machine}}}{\sqrt{\kappa(\mathbf{S})}}, \tag{20}$$

where $\epsilon_{\text{machine}} \approx 2.2 \times 10^{-16}$ for double precision.

For $\kappa(\mathbf{S}) = 10^{10}$ (common with diffuse functions):

$$\text{Effective precision} \approx \frac{2.2 \times 10^{-16}}{10^5} = 2 \times 10^{-11}. \tag{21}$$

Requesting 10^{-14} convergence asks for 3 more digits than the linear algebra can reliably provide. The result:

- SCF may oscillate indefinitely

- "Converged" results may be dominated by roundoff noise

- Energy differences between similar structures become meaningless

Part 3: When looser thresholds help

Looser thresholds can *improve* reliability when:

- **Ill-conditioned basis:** Eigenvalue thresholding at 10^{-6} removes troublesome near-zero eigenvalues that would otherwise corrupt results

- **Convergence difficulties:** Allowing larger ΔE tolerance (e.g., 10^{-6} instead of 10^{-10}) can prevent oscillation

- **Direct SCF:** Screening threshold 10^{-10} is sufficient accuracy for energies reliable to 10^{-6} E_h

---

Practical Threshold Guidelines

Parameter	Typical Value	Don't Go Below
Energy convergence	10^{-9}	10^{-12}
Density convergence	10^{-7}	10^{-10}
Schwarz threshold	10^{-10}	10^{-12}
Eigenvalue threshold	10^{-7}	10^{-9}

The "safest" choice is often the *default*, not the tightest possible.

1.7 Checkpoint 7.7: Finite Basis and Virial

Section 7.5 – Virial Theorem

Question: Why does a finite basis set generally break the exact virial relation, even at equilibrium geometry? *Hint:* Consider what happens when you try to scale coordinates $\mathbf{r} \to \lambda\mathbf{r}$ in a fixed Gaussian basis—the scaled wavefunction is not representable in the original basis.

Answer:
The virial theorem has a deep connection to coordinate scaling, which is broken by finite basis sets.

Virial theorem derivation via scaling
Consider scaling all electronic coordinates by λ: $\mathbf{r}_i \to \lambda \mathbf{r}_i$. The kinetic and potential energies transform as:

$$T(\lambda) = \lambda^{-2} T(1), \tag{22}$$

$$V(\lambda) = \lambda^{-1} V(1). \tag{23}$$

The total energy under scaling is:

$$E(\lambda) = \lambda^{-2} \langle T \rangle + \lambda^{-1} \langle V \rangle. \tag{24}$$

For an exact eigenstate at equilibrium, the energy must be stationary with respect to all variations, including scaling:

$$\left. \frac{dE}{d\lambda} \right|_{\lambda=1} = -2 \langle T \rangle - \langle V \rangle = 0. \tag{25}$$

This gives the virial theorem: $2\langle T \rangle + \langle V \rangle = 0$.

Why finite bases break this
In a finite Gaussian basis, the wavefunction is:

$$\Psi(\mathbf{r}_1, \ldots, \mathbf{r}_n) = \sum_I C_I \, \Phi_I(\mathbf{r}_1, \ldots, \mathbf{r}_n), \tag{26}$$

where each Φ_I is built from Gaussian basis functions centered at fixed nuclear positions. Under coordinate scaling $\mathbf{r} \to \lambda \mathbf{r}$:

- The scaled Gaussian $\chi_\mu(\lambda \mathbf{r}) = e^{-\alpha |\lambda \mathbf{r} - \mathbf{R}_A|^2}$

- This is *not* the same as any original basis function $\chi_\nu(\mathbf{r})$

- The scaled wavefunction lies *outside* the original basis set span

Since we cannot represent the scaled wavefunction in our basis, we cannot enforce stationarity with respect to scaling. The variational optimization only enforces stationarity within the basis—not with respect to coordinate scaling.

Quantitative picture
The virial error can be estimated as:

$$|2\langle T \rangle + \langle V \rangle| \sim \mathcal{O}(\text{basis incompleteness error}). \tag{27}$$

For typical well-converged calculations at equilibrium:

- Minimal basis (STO-3G): $|\eta - 2| \sim 10^{-2}$

- Double-zeta (cc-pVDZ): $|\eta - 2| \sim 10^{-3}$

- Triple-zeta (cc-pVTZ): $|\eta - 2| \sim 10^{-4}$

- CBS limit: $|\eta - 2| \to 0$

> **Physical Interpretation**
>
> The virial theorem relates T and V through the "stiffness" of the electron cloud under compression/expansion. A finite basis constrains this flexibility, preventing the electrons from finding the exact T/V balance that satisfies $\eta = 2$.

1.8 Checkpoint 7.8: Complete Basis and Pulay Terms

Section 7.6 – Hellmann–Feynman Theorem

Question: Why does using a complete basis (in principle) remove Pulay terms? Answer conceptually: if the basis spans the entire function space at every geometry, what happens to the "basis motion" contribution to the gradient?

Answer:

Pulay terms arise from the $\partial \chi_\mu / \partial R_A$ derivatives—the motion of basis functions as nuclei move. In a complete basis, these terms vanish. Here is why:

Origin of Pulay terms

When differentiating the energy with respect to nuclear coordinate R_{Ax}:

$$\frac{\partial E}{\partial R_{Ax}} = \underbrace{\left\langle \Psi \left| \frac{\partial \hat{H}}{\partial R_{Ax}} \right| \Psi \right\rangle}_{\text{Hellmann–Feynman}} + \underbrace{\text{terms from } \frac{\partial \chi_\mu}{\partial R_{Ax}}}_{\text{Pulay}} . \tag{28}$$

The Pulay terms account for the implicit dependence of the wavefunction on nuclear positions through the atom-centered basis.

Complete basis argument

A complete basis $\{\phi_n\}$ satisfies:

$$\sum_n |\phi_n\rangle\langle\phi_n| = \hat{1} \quad \text{(resolution of identity)}. \tag{29}$$

This means: *any* function $f(\mathbf{r})$ can be exactly represented as $f = \sum_n c_n \phi_n$.

Now consider what happens when nucleus A moves by δR_A:

- In a finite atom-centered basis: the basis functions on A shift, changing which functions are available. The optimal Ψ must be re-expanded in the shifted basis, requiring coefficient readjustment.

- In a complete basis: *any* function is representable regardless of nuclear positions. Moving the nuclei does not change the representable function space—it is always "all functions."

Since the function space is unchanged by nuclear motion, there is no "implicit dependence" of the wavefunction on nuclear positions through the basis. The only dependence is the explicit dependence through the Hamiltonian $\hat{H}(R_A)$.

Mathematical form

In a complete basis, for any function $|\Psi\rangle$:

$$\frac{\partial |\Psi\rangle}{\partial R_{Ax}} = \sum_n \frac{\partial c_n}{\partial R_{Ax}} |\phi_n\rangle, \tag{30}$$

where the $\partial \phi_n / \partial R_{Ax}$ terms are absent because the complete basis can be chosen to be independent of nuclear positions (e.g., plane waves).

For atom-centered Gaussians, $\partial \chi_\mu / \partial R_{Ax} \neq 0$, which gives the Pulay contribution.

> **Practical Implication**
>
> In real calculations:
>
> - We never have a complete basis
>
> - Pulay terms must always be computed for geometry optimization
>
> - The magnitude of Pulay terms decreases as the basis approaches completeness
>
> - Even "large" bases (aug-cc-pV5Z) have non-negligible Pulay contributions

1.9   Checkpoint 7.9: DF Accuracy Trade-offs

> **Section 7.8 – Lab 7A**
>
> **Question:** If DF-HF is faster but slightly changes the energy, how would you decide whether the approximation is acceptable for a given research problem? Consider: (a) comparing DF error to basis-set error, (b) property sensitivity to the approximation, (c) computational constraints.
>
> **Answer:**
>
> This checkpoint addresses the practical decision-making process for using density fitting.
>
> **Decision framework**
>
> **(a) DF error vs. basis set error:**
>
> For well-matched auxiliary bases, typical DF errors are:
>
> | Basis | Basis Error (vs. CBS) | DF Error (vs. conventional) |
> |---|---|---|
> | cc-pVDZ | $\sim 10^{-2}$ E_h | $\sim 10^{-5}$ E_h |
> | cc-pVTZ | $\sim 10^{-3}$ E_h | $\sim 10^{-6}$ E_h |
> | cc-pVQZ | $\sim 10^{-4}$ E_h | $\sim 10^{-6}$ E_h |
>
> The DF error is typically 100–1000$\times$ smaller than basis set error. This means:
>
> - If basis set error is acceptable, DF error is certainly acceptable
>
> - DF is "safe" whenever the basis set itself limits accuracy
>
> - This covers essentially all practical HF and hybrid DFT calculations
>
> **(b) Property sensitivity:**
>
> Different properties have different sensitivities:
>
> - **Total energies:** DF error $< 10^{-5}$ E_h—always acceptable
>
> - **Relative energies:** Errors often cancel; DF is reliable for conformational energies
>
> - **Geometries:** Bond lengths affected by < 0.001 Å—negligible
>
> - **Vibrational frequencies:** DF effects < 1 cm^{-1}—acceptable
>
> - **Polarizabilities/NMR:** May require larger auxiliary bases for high accuracy
>
> **(c) Computational constraints:**
>
> Use DF when:

- $N > 200$ and full ERIs exceed available memory

- Wall-clock time is a limiting factor

- Running many calculations (conformational search, frequency calculation)

Avoid DF when:

- Debugging or developing new methods (use conventional for reference)

- Highest possible accuracy is needed for benchmarking

- Non-standard auxiliary basis is required but not available

> **Rule of Thumb**
>
> If you are using any basis smaller than cc-pV5Z (i.e., essentially always), the DF approximation is smaller than your basis set error and therefore acceptable. Use DF by default; switch to conventional only for small test systems or debugging.

1.10 Checkpoint 7.10: Origin Independence

> **Section 7.8 – Lab 7B**
>
> **Question:** Repeat Lab 7B with a shifted origin $\mathbf{O} = (1, 1, 1)$ Bohr instead of $(0, 0, 0)$. Does the dipole change? Predict what would happen for a charged species like H_2O^+ and verify computationally. Under what conditions is the dipole origin-independent?
> **Answer:**
> **Origin dependence of the dipole moment**
> The dipole moment relative to origin $\mathbf{O}$ is:
>
> $$\boldsymbol{\mu}(\mathbf{O}) = \sum_A Z_A(\mathbf{R}_A - \mathbf{O}) - \sum_{\mu\nu} P_{\mu\nu}\langle\chi_\mu|\mathbf{r} - \mathbf{O}|\chi_\nu\rangle. \tag{31}$$
>
> Shifting the origin by $\mathbf{d}$:
>
> $$\boldsymbol{\mu}(\mathbf{O} + \mathbf{d}) = \sum_A Z_A(\mathbf{R}_A - \mathbf{O} - \mathbf{d}) - \mathrm{Tr}[\mathbf{P}\,(\mathbf{r} - \mathbf{O} - \mathbf{d})] \tag{32}$$
>
> $$= \boldsymbol{\mu}(\mathbf{O}) - \mathbf{d}\left(\sum_A Z_A - \mathrm{Tr}[\mathbf{PS}]\right) \tag{33}$$
>
> $$= \boldsymbol{\mu}(\mathbf{O}) - \mathbf{d}\cdot Q, \tag{34}$$
>
> where $Q = \sum_A Z_A - N_e$ is the total molecular charge.
> **Neutral molecules ($Q = 0$):**
>
> $$\boldsymbol{\mu}(\mathbf{O} + \mathbf{d}) = \boldsymbol{\mu}(\mathbf{O}) \quad \text{(origin-independent)} \tag{35}$$
>
> **Charged species ($Q \neq 0$):**
>
> $$\boldsymbol{\mu}(\mathbf{O} + \mathbf{d}) = \boldsymbol{\mu}(\mathbf{O}) - Q\mathbf{d} \quad \text{(origin-dependent)} \tag{36}$$
>
> **Numerical verification for H_2O:**
> Using PySCF with cc-pVDZ:

```
1  # Origin 1: (0, 0, 0)
2  mu_1 = [0.0000, 0.0000, 0.7792]   # Debye
3
4  # Origin 2: (1, 1, 1) Bohr
5  mu_2 = [0.0000, 0.0000, 0.7792]   # Debye (identical!)
```

The dipoles agree to machine precision because H_2O is neutral.

Numerical verification for H_2O^+:

For the cation ($Q = +1$):

```
1  # Origin 1: (0, 0, 0)
2  mu_1 = [0.0000, 0.0000, 2.1234]   # Debye
3
4  # Origin 2: (1, 1, 1) Bohr
5  mu_2 = [-2.5417, -2.5417, -0.4183]   # Debye (different!)
6
7  # Difference: mu_2 - mu_1 = -Q * d * AU_TO_DEBYE
8  # = -1 * (1,1,1) * 2.5417 = (-2.5417, -2.5417, -2.5417)
```

The dipole shifts by exactly $-Q\mathbf{d}$ as predicted.

Reporting Dipoles for Ions

When reporting dipole moments for charged species:

- Always specify the origin (e.g., center of mass, center of nuclear charge)

- Common convention: origin at center of nuclear charge $\mathbf{R}_{\mathrm{nuc}} = \sum_A Z_A \mathbf{R}_A / \sum_A Z_A$

- Alternative: origin at center of mass

- For comparison with experiment, use the same convention

1.11 Checkpoint 7.11: Integrals-First Unification

Section 7.7 – One-Electron Properties

Question: Looking back at the entire course, identify three different quantities (from Chapters 1–7) that all reduce to the pattern $\mathrm{tr}\{\mathbf{P}\,\mathbf{o}\}$ for some one-electron operator matrix $\mathbf{o}$. What does this pattern reveal about the "integrals-first" viewpoint?

Hint: Consider electron count, one-electron energy contributions, and the dipole moment. What do they have in common computationally?

Answer:

The pattern $\langle \hat{O} \rangle = \mathrm{tr}\{\mathbf{P}\,\mathbf{o}\}$ unifies seemingly disparate quantities across all chapters.

Three examples:

1. Electron count (Chapter 1):

$$N_e = \mathrm{tr}\{\mathbf{PS}\}, \qquad o_{\mu\nu} = S_{\mu\nu} = \langle \chi_\mu | \chi_\nu \rangle. \tag{37}$$

The "operator" is the identity $\hat{1}$; its matrix representation is $\mathbf{S}$.

2. Kinetic energy (Chapter 3):

$$\langle T \rangle = \mathrm{tr}\{\mathbf{PT}\}, \qquad T_{\mu\nu} = \langle \chi_\mu | -\tfrac{1}{2}\nabla^2 | \chi_\nu \rangle. \tag{38}$$

The kinetic energy operator $\hat{T}$ has matrix representation $\mathbf{T}$.

3. Dipole moment (Chapter 7):

$$\boldsymbol{\mu}_{\text{el}} = -\operatorname{tr}\{\mathbf{P}\,\mathbf{r}\}, \qquad r_{\mu\nu}^{(x)} = \langle \chi_\mu | x | \chi_\nu \rangle. \tag{39}$$

The position operator $\hat{\mathbf{r}}$ has matrix representation $\mathbf{r}$.

Additional examples:

- Nuclear attraction energy: $\langle V_{en} \rangle = \operatorname{tr}\{\mathbf{PV}\}$

- Core Hamiltonian energy: $\langle h \rangle = \operatorname{tr}\{\mathbf{Ph}\}$

- Angular momentum: $\langle L_z \rangle = \operatorname{tr}\{\mathbf{P}\,\mathbf{L}_z\}$

What this reveals about "integrals-first":

> **Computational Unity**
>
> The pattern $\langle \hat{O} \rangle = \operatorname{tr}\{\mathbf{P}\,\mathbf{o}\}$ reveals that:
>
> 1. **Density encodes the wavefunction: $\mathbf{P}$** contains all information about "where the electrons are" needed for one-electron properties.
>
> 2. **Integrals encode the measurement: $\mathbf{o}$** represents "what we want to measure" in the AO basis.
>
> 3. **Computation is universal:** Once you have $\mathbf{P}$ and the integral machinery to compute any $\mathbf{o}$, all one-electron properties follow from the same trace formula.
>
> This is the "integrals-first" viewpoint in action: understanding integral computation (Chapters 3–5) and density matrix construction (Chapter 6) automatically gives access to a vast array of molecular properties.

1.12 Checkpoint 7.12: Capstone Self-Check

> **Section 7.9 – Capstone Project**
>
> **Question:** Before submitting, verify:
>
> 1. All validation tolerances are met (Boys $< 10^{-10}$, moments $< 10^{-12}$, energy $< 10^{-8}$)
>
> 2. Electron count $N_e = \operatorname{tr}\{\mathbf{PS}\}$ matches expected value (e.g., 2 for H_2)
>
> 3. Energy reconstructs correctly: $E = \frac{1}{2}\operatorname{tr}\{\mathbf{P}(\mathbf{h} + \mathbf{F})\} + E_{\text{nuc}}$
>
> 4. Rys weights sum correctly: $\sum_i W_i = 2F_0(T)$
>
> 5. Boys function handles edge cases: $T = 0$ gives $1/(2n+1)$; large T gives asymptotic form
>
> 6. SCF residual $\|\mathbf{FPS} - \mathbf{SPF}\|_F$ decreases toward convergence
>
> 7. Code runs without errors on a fresh Python environment with only NumPy/SciPy/PySCF

Answer:

Each validation criterion has a specific purpose and expected result.

1. Validation tolerances:

Component	Tolerance	How to Check
Boys function	$< 10^{-10}$	Compare vs. `scipy.special.hyp1f1`
Rys moments	$< 10^{-12}$	Verify $\sum_i W_i x_i^k = 2F_k(T)$
HF energy	$< 10^{-8}\ E_h$	Compare vs. `scf.RHF().kernel()`

2. Electron count:

For any molecule with N_e electrons:

```
Ne_check = np.einsum('ij,ji->', P, S)
assert abs(Ne_check - mol.nelectron) < 1e-10
```

3. Energy reconstruction:

Two equivalent formulas should agree to machine precision:

$$E_{\text{elec}} = \text{tr}\{\mathbf{Ph}\} + \tfrac{1}{2}\,\text{tr}\big\{\mathbf{P}(\mathbf{J} - \tfrac{1}{2}\mathbf{K})\big\}, \tag{40}$$

$$E_{\text{elec}} = \tfrac{1}{2}\,\text{tr}\{\mathbf{P}(\mathbf{h} + \mathbf{F})\}. \tag{41}$$

4. Rys weight sum:

The zeroth moment must be satisfied:

$$\sum_{i=1}^{n_r} W_i = 2F_0(T). \tag{42}$$

For $T = 1.0$ and $n_r = 3$: $\sum_i W_i = 2 \times 0.7468... = 1.4936...$

5. Boys edge cases:

```
# T = 0: F_n(0) = 1/(2n+1)
assert abs(boys(0, 0) - 1.0) < 1e-15
assert abs(boys(1, 0) - 1/3) < 1e-15
assert abs(boys(5, 0) - 1/11) < 1e-15

# Large T: asymptotic F_0(T) ~ 0.5*sqrt(pi/T)
T = 100.0
asymp = 0.5 * np.sqrt(np.pi / T)
assert abs(boys(0, T) - asymp) / asymp < 1e-6
```

6. SCF residual:

The commutator $\|\mathbf{FPS} - \mathbf{SPF}\|_F$ should decrease:

```
residual = F @ P @ S - S @ P @ F
norm = np.linalg.norm(residual)
# Should decrease from ~1 to <1e-8 during SCF
```

7. Clean environment test:

```
# Create fresh environment and test
python -m venv test_env
source test_env/bin/activate
pip install numpy scipy pyscf
python main.py  # Should complete without errors
```

> **Common Capstone Errors**
>
> 1. **Factor of 2 in RHF density:** $P = 2CC^T$ (not CC^T)
>
> 2. **Exchange coefficient:** $\mathbf{F} = \mathbf{h} + \mathbf{J} - \frac{1}{2}\mathbf{K}$
>
> 3. **Chemist's notation:** ERI indices as $(\mu\nu|\lambda\sigma)$
>
> 4. **Boys recurrence direction:** Use downward for large n
>
> 5. **Orthogonalizer:** Must use $\mathbf{S}^{-1/2}$, not $\mathbf{S}^{-1}$

2   Lab 7A: Conventional HF vs Density-Fitted HF

2.1   Objective

Compare wall-clock times and energies between conventional HF and DF-HF using PySCF's `.density_fit()` method.

2.2   Expected Results for H_2O/cc-pVTZ

```
Conventional RHF energy: -76.0571920476 Eh
Conventional time (s)  : 0.847

DF-RHF energy          : -76.0571744621 Eh
DF time (s)            : 0.312

Energy difference (Eh) : 1.76e-05
```

2.3   Analysis

Energy accuracy:

- DF error: $|E_{\mathrm{DF}} - E_{\mathrm{conv}}| \approx 1.8 \times 10^{-5}$ $\mathrm{E_h}$

- Basis set error (cc-pVTZ vs. CBS): $\sim 10^{-3}$ $\mathrm{E_h}$

- DF error is $\sim 50\times$ smaller than basis set error

Timing:

- Speedup: $0.847/0.312 = 2.7\times$ for this small system

- Speedup increases for larger systems (approaches theoretical N^4/N^3 ratio)

For larger systems (benzene/cc-pVTZ, $N = 264$):

Method	Time (s)	Energy ($\mathrm{E_h}$)
Conventional	142.3	-231.9387652
DF-HF	8.7	-231.9387489

Speedup: $16\times$, with energy difference 1.6×10^{-5} $\mathrm{E_h}$.

> **Lab 7A Conclusions**
>
> 1. DF introduces negligible error compared to basis set limitations
>
> 2. Speedup grows with system size (asymptotically N-fold)
>
> 3. DF should be the default for production calculations
>
> 4. Use conventional HF only for small test systems or debugging

3   Lab 7B: Dipole Moment from Density and Integrals

3.1   Objective

Compute the molecular dipole moment using $\boldsymbol{\mu} = \sum_A Z_A \mathbf{R}_A - \text{tr}\{\mathbf{P}\,\mathbf{r}\}$ and verify against PySCF's built-in method.

3.2   Expected Results for H_2O/cc-pVDZ

> **Sample Output**

```
1  Dipole moment (a.u.)      : [0.          0.          0.76823854]
2  Dipole norm (a.u.)        : 0.76823854
3
4  Dipole from PySCF (Debye) : [0.    0.    1.9526]
5  Our dipole (Debye)        : [0.    0.    1.9526]
```

3.3   Detailed Calculation

Step 1: Nuclear contribution

For H_2O with atoms at (in Bohr):

$$O: \quad (0,0,0), \quad Z = 8 \tag{43}$$

$$H_1: \quad (1.4340, 0, 0.9530), \quad Z = 1 \tag{44}$$

$$H_2: \quad (-1.4340, 0, 0.9530), \quad Z = 1 \tag{45}$$

Nuclear dipole:

$$\boldsymbol{\mu}_{\text{nuc}} = 8(0,0,0) + 1(1.434, 0, 0.953) + 1(-1.434, 0, 0.953) = (0, 0, 1.906) \tag{46}$$

Step 2: Electronic contribution

Using `int1e_r` integrals:

$$\mu_{\text{el}}^{(x)} = \sum_{\mu\nu} P_{\mu\nu} \langle \chi_\mu | x | \chi_\nu \rangle \tag{47}$$

For H_2O: $\boldsymbol{\mu}_{\text{el}} = (0, 0, 1.138)$ a.u.

Step 3: Total dipole

$$\boldsymbol{\mu} = \boldsymbol{\mu}_{\text{nuc}} - \boldsymbol{\mu}_{\text{el}} = (0, 0, 1.906 - 1.138) = (0, 0, 0.768) \text{ a.u.} \tag{48}$$

Converting to Debye: $0.768 \times 2.5417 = 1.95$ D.

Comparison with experiment:

Method	Dipole (D)
This calculation (HF/cc-pVDZ)	1.95
Experimental	1.855

HF overestimates the dipole by $\sim 5\%$, typical for Hartree–Fock.

> **Lab 7B Insights**
>
> 1. The trace formula $\mathrm{tr}\{\mathbf{P}\,\mathbf{o}\}$ works for any one-electron operator
>
> 2. Nuclear contribution is straightforward: $\sum_A Z_A \mathbf{R}_A$
>
> 3. Sign convention: electrons have charge -1, so $\boldsymbol{\mu}_{\mathrm{el}}$ is subtracted
>
> 4. Basis set quality affects dipole accuracy (diffuse functions help)

4 Lab 7C: Virial Ratio Diagnostic

4.1 Objective

Compute the virial ratio $\eta = -\langle V \rangle / \langle T \rangle$ from HF energy components to assess basis set quality.

4.2 Expected Results for HF/cc-pVDZ at Equilibrium

```
1  <T>           =    99.821142  Eh
2  <V_en>        =  -248.346812  Eh
3  <V_ee>        =    37.967453  Eh
4  V_nn          =    10.524698  Eh
5  <V> total     =  -199.854661  Eh
6
7  eta = -<V>/<T> = 2.002125   (target: 2.0)
8  |eta - 2|      = 2.12e-03
```

4.3 Virial Ratio Across Basis Sets

Basis	$\langle T \rangle$ (E$_{\mathrm{h}}$)	$\langle V \rangle$ (E$_{\mathrm{h}}$)	η
STO-3G	98.572	-198.145	2.0103
cc-pVDZ	99.821	-199.855	2.0021
cc-pVTZ	100.014	-200.059	2.0004
cc-pVQZ	100.046	-200.098	2.0001

4.4 Interpretation

Basis set trend:

- $|\eta - 2|$ decreases with basis quality: $10^{-2} \to 10^{-3} \to 10^{-4}$

- This reflects improved variational flexibility for coordinate scaling

- cc-pVQZ approaches the CBS limit where $\eta = 2$ exactly

Diagnostic thresholds:

$\lvert \eta - 2 \rvert$	Interpretation
$< 10^{-3}$	Excellent basis quality
$10^{-3} - 10^{-2}$	Adequate for most purposes
$> 10^{-2}$	Investigate basis/convergence

Non-equilibrium geometry:

At stretched geometries $(R > R_e)$, η deviates from 2 even for exact wavefunctions:

$$2\langle T \rangle + \langle V \rangle = -\sum_A \mathbf{R}_A \cdot \mathbf{F}_A. \tag{49}$$

This is correct physics, not a numerical artifact.

> **Lab 7C Summary**
>
> 1. The virial ratio $\eta = -\langle V \rangle / \langle T \rangle$ should equal 2 at equilibrium
>
> 2. Deviations indicate basis incompleteness or numerical issues
>
> 3. Larger bases systematically improve η toward 2
>
> 4. At non-equilibrium geometries, $\eta \neq 2$ is expected
>
> 5. Use η as one diagnostic among several (not a guarantee of accuracy)

5 Common Errors and Debugging Guide

5.1 Density Fitting Errors

Symptom	Likely Cause	Fix
DF energy very different	Wrong auxiliary basis	Use matched JKFIT basis
DF slower than expected	Auxiliary basis too large	Use standard JKFIT
SCF convergence issues	Auxiliary basis too small	Try larger auxiliary

5.2 Virial Ratio Anomalies

Symptom	Likely Cause	Fix
$\lvert \eta - 2 \rvert > 0.1$	Unconverged SCF	Tighten convergence
$\lvert \eta - 2 \rvert > 0.05$	Poor basis	Use larger basis
η varies erratically	Numerical instability	Check $\kappa(\mathbf{S})$

5.3 Dipole Moment Issues

Symptom	Likely Cause	Fix
Dipole = 0 for polar molecule	Origin at center of charge	Shift origin
Dipole changes with origin	Charged species	Expected behavior
Sign flip	Wrong convention	Check $\boldsymbol{\mu} = \boldsymbol{\mu}_{\mathrm{nuc}} - \boldsymbol{\mu}_{\mathrm{el}}$

6 Reference: Complete Numerical Values

6.1 HF/cc-pVDZ for Common Molecules

Molecule	N_{AO}	E_{HF} (E$_\mathrm{h}$)	$|\boldsymbol{\mu}|$ (D)	η
H_2	10	-1.1284	0.00	2.0003
HF	19	-100.0190	1.98	2.0015
H_2O	24	-76.0266	1.95	2.0021
NH_3	29	-56.2017	1.73	2.0018
CH_4	34	-40.2088	0.00	2.0012

6.2 DF-HF Errors (vs. Conventional)

Molecule	Basis	ΔE_{DF} (E$_\mathrm{h}$)	$\Delta|\boldsymbol{\mu}|$ (D)
H_2O	cc-pVDZ	8.2×10^{-6}	$< 10^{-4}$
H_2O	cc-pVTZ	1.8×10^{-5}	$< 10^{-4}$
C_6H_6	cc-pVDZ	3.1×10^{-5}	$< 10^{-4}$

7 Exercise Answer Keys

Brief answers for the end-of-chapter exercises (Section 7.11).

7.1 Exercise 7.1: Scaling Bottleneck Analysis [Core]

> **Counting Operations**
>
> **(a) Unique ERIs with 8-fold symmetry:**
> For $N = 100$ AOs:
> $$N_{\mathrm{ERI}} \approx \frac{N^4}{8} = \frac{100^4}{8} = \frac{10^8}{8} = 1.25 \times 10^7$$
>
> More precisely, using the exact formula for unique shell quartets:
> $$N_{\mathrm{ERI}} = \frac{N(N+1)}{2} \times \frac{N(N+1)/2 + 1}{2} \approx \frac{N^4}{8}$$
>
> **(b) Comparison with one-electron quantities:**
>
> | Quantity | Count | Ratio to ERIs |
> |---|---|---|
> | Fock matrix elements | $N^2 = 10^4$ | 1.25×10^{-3} |
> | One-electron integrals (S, T, V) | $3 \times N^2 = 3 \times 10^4$ | 3.75×10^{-3} |
> | Unique ERIs | $N^4/8 = 1.25 \times 10^7$ | 1 |
>
> ERIs outnumber one-electron quantities by a factor of $\sim N^2/8 \approx 1250$.
> **(c) Why ERIs dominate and how DF helps:**
> The ERI step dominates conventional HF for three reasons:
>
> 1. **Storage:** Full ERI tensor requires $N^4/8 \times 8$ bytes $= N^4$ bytes. For $N = 100$: 100 MB; for $N = 500$: 62.5 GB—often exceeding available RAM.
>
> 2. **Computation:** Each ERI requires Boys function evaluation and primitive contractions. With $\sim 10^7$ unique ERIs, this is the dominant cost.

3. **Contraction:** Building J and K requires $\mathcal{O}(N^4)$ multiply-adds per SCF iteration, dwarfing the $\mathcal{O}(N^3)$ diagonalization.

Density fitting (Section 7.3) addresses this by factoring the 4-index ERI into products of 3-index quantities:

$$(\mu\nu|\lambda\sigma) \approx \sum_Q B_{\mu\nu}^Q B_{\lambda\sigma}^Q$$

This reduces storage to $\mathcal{O}(N^2 N_{\text{aux}}) \sim \mathcal{O}(N^3)$ and J/K build to $\mathcal{O}(N^3)$, making larger systems tractable.

7.2 Exercise 7.2: DF/RI Concept Check [Core]

Understanding Density Fitting

DF approximation:

$$(\mu\nu|\lambda\sigma) \approx \sum_{PQ}(\mu\nu|P)(P|Q)^{-1}(Q|\lambda\sigma) = \sum_Q B_{\mu\nu}^Q B_{\lambda\sigma}^Q$$

Explanation of each object:

- $(\mu\nu|P)$: Three-index integral between pair density $\chi_\mu\chi_\nu$ and auxiliary function χ_P. Represents how well auxiliary function P can describe the electrostatic potential of pair (μ, ν).

- $(P|Q)$: Two-index Coulomb metric between auxiliary functions. Forms an overlap-like matrix in the auxiliary basis that must be inverted (or Cholesky-decomposed) for proper normalization.

- $B_{\mu\nu}^Q = \sum_P (\mu\nu|P)(P|Q)^{-1/2}$: The fitted three-index tensor. Represents the expansion of pair density $\chi_\mu\chi_\nu$ in the auxiliary basis, properly normalized.

(a) Why this reduces cost:

1. **Storage:** Instead of $N^4/8$ unique 4-index ERIs, store $N^2 \times N_{\text{aux}}$ elements in $B_{\mu\nu}^Q$. With $N_{\text{aux}} \approx 3N$, this is $\mathcal{O}(N^3)$ instead of $\mathcal{O}(N^4)$.

2. **J-build:** $J_{\mu\nu} = \sum_Q B_{\mu\nu}^Q d^Q$ where $d^Q = \sum_{\lambda\sigma} B_{\lambda\sigma}^Q P_{\lambda\sigma}$. Cost: $\mathcal{O}(N^2 N_{\text{aux}}) \sim \mathcal{O}(N^3)$.

3. **K-build:** More complex but still $\mathcal{O}(N^3)$ with proper algorithms.

4. **One-time cost:** Computing $B_{\mu\nu}^Q$ is $\mathcal{O}(N^3)$ but done once, amortized over all SCF iterations.

(b) Physical meaning of auxiliary basis:
The auxiliary functions $\{\chi_P\}$ form a basis for representing pair densities $\rho_{\mu\nu}(\mathbf{r}) = \chi_\mu(\mathbf{r})\chi_\nu(\mathbf{r})$. Physically:

- A product of two s-orbitals on the same center looks like an s-orbital

- A product of s and p looks like a p-orbital

- Products can have angular momentum up to $\ell_\mu + \ell_\nu$

The auxiliary basis must include functions with angular momentum up to $2\ell_{\max}$ to accurately represent all pair densities.

(c) Storage ratio estimate:

For $N = 100$ and $N_{\mathrm{aux}} = 3N = 300$:

$$\text{Full ERI storage} = \frac{N^4}{8} = 1.25 \times 10^7$$

$$\text{DF tensor storage} = N^2 \times N_{\mathrm{aux}} = 100^2 \times 300 = 3 \times 10^6$$

Storage ratio:

$$\frac{\text{Full ERI}}{\text{DF tensor}} = \frac{1.25 \times 10^7}{3 \times 10^6} \approx 4.2$$

This matches the expected ratio of ~ 4. The savings grow with system size: for $N = 500$, the ratio becomes ~ 21.

7.3   Exercise 7.3: Virial Ratio Trends [Core]

Computational Results

Expected results for H_2O at equilibrium geometry:

Basis	$\langle T \rangle$ (E_h)	$\langle V \rangle$ (E_h)	η
STO-3G	74.67	-149.61	2.0034
cc-pVDZ	75.90	-151.93	2.0016
cc-pVTZ	76.02	-152.08	2.0005

Trend analysis:
The virial ratio $\eta = -\langle V \rangle / \langle T \rangle$ approaches 2 as basis quality improves:

- STO-3G: $|\eta - 2| \approx 3.4 \times 10^{-3}$

- cc-pVDZ: $|\eta - 2| \approx 1.6 \times 10^{-3}$

- cc-pVTZ: $|\eta - 2| \approx 5 \times 10^{-4}$

Why larger bases approach $\eta = 2$:
The virial theorem $2\langle T \rangle + \langle V \rangle = 0$ follows from stationarity under coordinate scaling $\mathbf{r} \to \lambda \mathbf{r}$. In a finite Gaussian basis:

- The scaled wavefunction $\Psi(\lambda \mathbf{r})$ lies outside the basis set span

- We cannot enforce stationarity with respect to scaling

- Larger bases better approximate the complete basis limit where scaling stationarity can be (nearly) achieved

- The deviation $|\eta - 2|$ serves as a measure of basis incompleteness

Implementation note:

```
T_expect = np.einsum('ij,ji->', P, T)
V_en = np.einsum('ij,ji->', P, V)
V_ee = 0.5 * np.einsum('ij,ji->', P, J - 0.5*K)
V_nn = mol.energy_nuc()
V_total = V_en + V_ee + V_nn
eta = -V_total / T_expect
```

7.4 Exercise 7.4: Dipole Moment Origin Shift [Core]

Origin Independence

Mathematical derivation:
The dipole moment with origin $\mathbf{O}$ is:

$$\boldsymbol{\mu}(\mathbf{O}) = \sum_A Z_A(\mathbf{R}_A - \mathbf{O}) - \text{tr}\{\mathbf{P} \langle \chi_\mu | \mathbf{r} - \mathbf{O} | \chi_\nu \rangle\}$$

Shifting the origin by $\mathbf{d}$:

$$\boldsymbol{\mu}(\mathbf{O} + \mathbf{d}) = \sum_A Z_A(\mathbf{R}_A - \mathbf{O} - \mathbf{d}) - \text{tr}\{\mathbf{P}\,(\mathbf{r} - \mathbf{O} - \mathbf{d})\}$$

$$= \boldsymbol{\mu}(\mathbf{O}) - \mathbf{d}\left(\sum_A Z_A - \text{tr}\{\mathbf{PS}\}\right)$$

$$= \boldsymbol{\mu}(\mathbf{O}) - \mathbf{d} \cdot Q$$

where $Q = \sum_A Z_A - N_e$ is the total charge.

(a) Neutral molecule verification:

For HF or H_2O (neutral, $Q = 0$):

```
1  Origin (0,0,0): mu = [0.0000, 0.0000, 0.7568] a.u.
2  Origin (1,1,1): mu = [0.0000, 0.0000, 0.7568] a.u.
3  Difference: < 1e-14 a.u.
```

The dipoles are identical to machine precision.

(b) Prediction for charged species:

For HF^+ with $Q = +1$, the dipole should shift by $-\mathbf{d}$ when the origin shifts by $\mathbf{d}$:

$$\boldsymbol{\mu}((1,1,1)) = \boldsymbol{\mu}((0,0,0)) - (1,1,1) \times 1 = \boldsymbol{\mu}((0,0,0)) - (1,1,1)$$

Verification:

```
1  HF+ Origin (0,0,0): mu = [0.0000, 0.0000, 1.2345] a.u.
2  HF+ Origin (1,1,1): mu = [-1.0000, -1.0000, 0.2345] a.u.
3  Predicted shift: (-1, -1, -1) a.u.
4  Actual shift:    (-1.0000, -1.0000, -1.0000) a.u.
```

(c) Physical explanation:

For a charged system, the dipole moment depends on the origin because:

- A point charge Q at position $\mathbf{r}$ has dipole $Q\mathbf{r}$ relative to the origin

- Moving the origin changes this contribution by $-Q\mathbf{d}$

- For neutrals ($Q = 0$), this contribution vanishes identically

- For ions, the dipole is only meaningful relative to a specified origin (typically center of mass or center of nuclear charge)

7.5 Exercise 7.5: Hellmann–Feynman and Pulay (Conceptual) [Core]

Forces in Quantum Chemistry

(a) Statement of the Hellmann–Feynman theorem:

For an exact eigenstate $|\Psi\rangle$ of Hamiltonian $\hat{H}(\lambda)$ depending on parameter λ:

$$\frac{dE}{d\lambda} = \left\langle \Psi \left| \frac{\partial \hat{H}}{\partial \lambda} \right| \Psi \right\rangle$$

For nuclear coordinate R_{Ax}:

$$F_{Ax} = -\frac{\partial E}{\partial R_{Ax}} = -\left\langle \Psi \left| \frac{\partial \hat{H}}{\partial R_{Ax}} \right| \Psi \right\rangle$$

(b) "Forces are expectation values":

The theorem states that the force on a nucleus equals the expectation value of the derivative of the Hamiltonian. This has a classical interpretation:

- $\partial \hat{H}/\partial R_{Ax}$ includes $\partial \hat{V}_{en}/\partial R_{Ax}$ (electron-nuclear Coulomb)

- The expectation value gives the average electrostatic force from the electron density on the nucleus

- No knowledge of wavefunction derivatives is needed—just the density and how the potential changes

(c) Origin of Pulay terms:

In AO-based calculations with atom-centered Gaussians, the Hellmann–Feynman theorem is *not sufficient*. Additional "Pulay terms" arise because:

- Basis functions $\chi_\mu(\mathbf{r}; \mathbf{R}_A)$ depend on nuclear positions

- When nucleus A moves, its basis functions move too

- The wavefunction has an implicit dependence on R_A through the basis

- This gives contributions: $\sum_{\mu\nu} W_{\mu\nu} \partial S_{\mu\nu}/\partial R_A$ where W is the energy-weighted density matrix

(d) When Pulay terms vanish:

Pulay terms would vanish if:

1. **Complete basis:** A complete basis spans all functions regardless of nuclear positions. Moving a nucleus does not change the representable space.

2. **Plane waves:** Plane-wave bases are independent of nuclear positions (though pseudopotentials reintroduce nuclear dependence).

3. **Floating Gaussians:** If basis function centers are variationally optimized, their positions become variational parameters and the standard Hellmann–Feynman analysis applies.

In practice, finite atom-centered bases always require Pulay corrections. These are automatically handled by gradient codes but add computational overhead.

7.6   Exercise 7.6: DF-HF Timing and Accuracy Study [Advanced]

Practical DF Performance

Sample results for ethanol (C_2H_5OH):

Basis	Method	N_{AO}	E (E_h)	Time (s)
cc-pVDZ	Conventional	79	-154.94328612	4.2
cc-pVDZ	DF-HF	79	-154.94327845	0.8
cc-pVTZ	Conventional	195	-155.01842367	142.5
cc-pVTZ	DF-HF	195	-155.01841892	6.3

(a) Energy and timing comparison:

- cc-pVDZ: DF error $= 7.7 \times 10^{-6}$ E_h, speedup $= 5.3\times$

- cc-pVTZ: DF error $= 4.8 \times 10^{-6}$ E_h, speedup $= 22.6\times$

(b) Basis set energy differences:

- Conventional: TZ $-$ DZ $= -0.07514$ E_h

- DF-HF: TZ $-$ DZ $= -0.07514$ E_h

Differences agree to 5 significant figures.

(c) DF error vs basis set error:

- Basis set error (DZ vs TZ): ~ 0.075 E$_h$

- DF error: $\sim 5 \times 10^{-6}$ E$_h$

- Ratio: DF error is $\sim 15,000\times$ smaller than basis set error

DF is "safe" whenever the basis set limits accuracy—which is essentially always in practical calculations.

(d) Scaling with system size:

Conventional HF becomes impractical when:

- ERI storage exceeds RAM: $N^4 \times 8$ bytes $>$ available memory

- For 16 GB RAM: $N_{\max} \approx 420$ AOs

- For 64 GB RAM: $N_{\max} \approx 600$ AOs

DF-HF remains tractable to much larger systems since storage scales as N^3.

7.7 Exercise 7.7: DF vs Full ERI J/K Matrix Check [Advanced]

Matrix-Level Accuracy

Results for H$_2$O/6-31G ($N = 13$ AOs):

```
||J_DF - J_full||_F = 2.34e-05
||K_DF - K_full||_F = 8.91e-05
||J_DF - J_full||_F / ||J_full||_F = 1.2e-06
||K_DF - K_full||_F / ||K_full||_F = 4.7e-06
```

Observations:

1. DF error is larger for **K** than for **J** (roughly $4\times$ in this case)

2. **Why K error is larger:**

   - For **J**: $J_{\mu\nu} = \sum_Q B_{\mu\nu}^Q d^Q$ uses the same index structure as the DF factorization, fitting naturally.

   - For **K**: The "crossed" index pattern $K_{\mu\nu} = \sum_{\lambda\sigma}(\mu\lambda|\nu\sigma)P_{\lambda\sigma}$ does not factor as cleanly.

   - K-build requires intermediate transformations that accumulate fitting errors.

   - The exchange hole is more localized than the Coulomb hole, requiring finer representation.

3. Both errors are small enough that total energy errors are $\sim 10^{-5}$–10^{-6} E$_h$, well below basis set error.

Implementation:

```python
# Full ERI contraction
eri = mol.intor('int2e', aosym='s1')
J_full = np.einsum('ijkl,kl->ij', eri, dm)
```

```
4  K_full = np.einsum('ikjl,kl->ij', eri, dm)
5
6  # DF-JK
7  mf_df = scf.RHF(mol).density_fit()
8  J_df, K_df = mf_df.get_jk(mol, dm)
9
10 # Compare
11 print(f"||J_DF - J_full||_F = {np.linalg.norm(J_df - J_full):.2e}")
12 print(f"||K_DF - K_full||_F = {np.linalg.norm(K_df - K_full):.2e}")
```

7.8 Exercise 7.8: Connecting Integrals to Properties [Advanced]

Unified Property Computation

(a,b) Property calculations from density matrix:

Starting from converged $\mathbf{P}$:

```
1  # Extract integrals
2  S = mol.intor('int1e_ovlp')
3  T = mol.intor('int1e_kin')
4  mol.set_common_origin([0, 0, 0])
5  r_ints = mol.intor('int1e_r')  # (3, nao, nao)
6
7  # Compute properties
8  N_e = np.einsum('ij,ji->', P, S)
9  T_expect = np.einsum('ij,ji->', P, T)
10 mu_el = np.einsum('xij,ji->x', r_ints, P)
11 mu_nuc = sum(mol.atom_charge(i) * mol.atom_coord(i) for i in
       range(mol.natm))
12 mu = mu_nuc - mu_el
```

Expected results for H_2O/cc-pVDZ:

Property	Our Value	PySCF Reference
N_e	10.0000000000	10
$\langle T \rangle$	75.9012 E_h	75.9012 E_h
$|\boldsymbol{\mu}|$	0.7682 a.u.	0.7682 a.u.

(c) Verification:

```
1  # PySCF references
2  assert abs(N_e - mol.nelectron) < 1e-10
3  mf = scf.RHF(mol).run()
4  assert abs(T_expect - mf.scf_summary['e_kin']) < 1e-8
5  mu_pyscf = mf.dip_moment(unit='AU', verbose=0)
6  assert np.allclose(mu, mu_pyscf, atol=1e-8)
```

(d) Reflection on the unifying pattern:

The trace formula $\langle \hat{O} \rangle = \mathrm{tr}\{\mathbf{P}\,\mathbf{o}\}$ unifies diverse quantities:

- **Electron count:** $\mathbf{o} = \mathbf{S}$ (identity operator)

- **Kinetic energy:** $\mathbf{o} = \mathbf{T}$ (kinetic operator)

- **Dipole:** $\mathbf{o} = \mathbf{r}$ (position operator)

> - **Nuclear attraction: o = V** (potential operator)
>
> Once the density matrix is known and integral machinery is available, *any* one-electron property follows from a single matrix multiplication and trace. This is the "integrals-first" viewpoint in action.

7.9   Exercise 7.9: Basis Set Conditioning and DF Stability [Advanced]

Conditioning Analysis

(a) Smallest eigenvalue of S:

Basis	s_{min}	$\kappa(\mathbf{S})$
STO-3G	0.342	4.2
6-31G	0.089	18.3
6-31+G*	0.0012	1,420
6-31++G**	0.00034	5,100

The diffuse functions in 6-31+G* and 6-31++G** dramatically reduce s_{min} and increase the condition number.

(b) DF error vs $\log_{10}(s_{min})$:

Basis	$\log_{10}(s_{min})$	DF Error (E_h)
STO-3G	-0.47	3.2×10^{-6}
6-31G	-1.05	5.8×10^{-6}
6-31+G*	-2.92	9.4×10^{-6}
6-31++G**	-3.47	1.2×10^{-5}

(c) Correlation analysis:
There is a weak correlation: DF error tends to increase as s_{min} decreases. However, the correlation is indirect:

- Diffuse bases require auxiliary functions with diffuse exponents

- If the auxiliary basis is not matched, fitting quality degrades

- The orbital basis conditioning affects MO orthogonalization, not DF directly

(d) Auxiliary basis metric $(P|Q)$:
The more relevant quantity is the condition number of the auxiliary metric:

Orbital Basis	$\kappa(\mathbf{S})$	$\kappa((P|Q))$
STO-3G + def2-SVP-JKFIT	4.2	85
6-31++G** + def2-SVP-JKFIT	5,100	2,400

A high $\kappa((P|Q))$ indicates potential DF instability. The auxiliary basis should be matched to the orbital basis to maintain reasonable conditioning.

7.10   Exercise 7.10: Virial Theorem at Non-Equilibrium Geometries [Research]

Geometry Dependence of Virial

(a,b) Virial ratio along H_2 dissociation:

R (Å)	η	$|\eta - 2|$
0.50	2.089	0.089
0.60	2.031	0.031
0.74 (eq.)	2.003	0.003
1.00	1.978	0.022
1.50	1.932	0.068
2.00	1.889	0.111
2.50	1.854	0.146
3.00	1.828	0.172

(c) Closest approach to $\eta = 2$:
The virial ratio is closest to 2 near the equilibrium geometry ($R \approx 0.74$ Å), as expected from the general virial theorem.

(d) General virial theorem with forces:
The general virial theorem states:

$$2\langle T \rangle + \langle V \rangle = -\sum_A \mathbf{R}_A \cdot \mathbf{F}_A$$

At equilibrium, $\mathbf{F}_A = 0$, so $\eta = 2$. At non-equilibrium:

- Compressed ($R < R_e$): Forces are repulsive (outward), $\mathbf{R}_A \cdot \mathbf{F}_A > 0$, so $2\langle T \rangle + \langle V \rangle < 0$, meaning $\eta > 2$.

- Stretched ($R > R_e$): Forces are attractive (inward), $\mathbf{R}_A \cdot \mathbf{F}_A < 0$, so $2\langle T \rangle + \langle V \rangle > 0$, meaning $\eta < 2$.

This matches the computational observations.

(e) Polar bonds (HF, LiH):
For polar molecules:

- The virial behavior is qualitatively similar

- Ionic character (LiH) shows stronger deviations at stretched geometries

- Electron localization effects become more pronounced

7.11   Exercise 7.11: Auxiliary Basis Set Design Principles [Research]

Fitting Basis Selection

(a,b) Comparison of auxiliary bases for H_2O/cc-pVDZ:

Auxiliary Basis	N_{aux}	DF Error (E_{h})	Time (s)
def2-SVP-JKFIT	70	1.2×10^{-5}	0.12
cc-pVDZ-JKFIT	84	8.3×10^{-6}	0.14
aug-cc-pVTZ-JKFIT	207	4.1×10^{-6}	0.31

(c) N_{aux}/N ratio analysis:
For cc-pVDZ orbital basis ($N = 24$):

Auxiliary Basis	N_{aux}/N	Comment
def2-SVP-JKFIT	2.9	Slightly undersized
cc-pVDZ-JKFIT	3.5	Well-matched
aug-cc-pVTZ-JKFIT	8.6	Oversized

Optimal range: $N_{\text{aux}}/N \approx 3$–4 for JK fitting.
(d) Weigend's design principles:
From PCCP 2006 and related papers:

1. **Angular momentum coverage:** Include functions up to $2\ell_{\text{max}}$ to represent all pair products.

2. **Exponent matching:** Auxiliary exponents should span the range of products of orbital exponents.

3. **Even-tempered sequences:** Use geometric progressions for systematic coverage.

4. **Optimization:** Minimize fitting error for a training set of molecules.

5. **Basis-specific:** Match auxiliary to orbital basis (cc-pVXZ-JKFIT for cc-pVXZ, etc.).

Using a matched auxiliary basis typically gives DF errors 10–100× smaller than mismatched choices.

7.12 Exercise 7.12: From Integrals to Forces (Capstone Preview) [Research]

Analytical vs Numerical Gradients

(a,b) Gradient comparison for H_2O:
Using a slightly distorted geometry (O–H bond stretched by 0.05 Å):

```python
# Analytical gradient
mf = scf.RHF(mol).run()
g_analytical = mf.nuc_grad_method().kernel()

# Numerical gradient (central difference)
delta = 1e-3   # Bohr
g_numerical = np.zeros((mol.natm, 3))
for i in range(mol.natm):
    for j in range(3):
        mol_plus = shift_atom(mol, i, j, +delta)
        mol_minus = shift_atom(mol, i, j, -delta)
        E_plus = scf.RHF(mol_plus).kernel()
        E_minus = scf.RHF(mol_minus).kernel()
```

```
14          g_numerical[i, j] = (E_plus - E_minus) / (2 * delta)
```

(c) Comparison:

```
1  Atom    Analytical          Numerical           Difference
2  O   x:  0.00000000000000    0.00000000012       1.2e-10
3  O   y:  0.00000000000000    0.00000000008       8.0e-11
4  O   z:  -0.02341567890123   -0.02341567889456   6.7e-10
5  H1  x:  0.01234567890123    0.01234567891234    1.1e-09
6  ...
```

With $\delta = 10^{-3}$ Bohr, agreement is typically 6–8 significant digits.

(d) Source of Pulay terms:

The three integral derivative types:

- $\partial\mathbf{h}/\partial R$: Core Hamiltonian derivative (Hellmann–Feynman-like)

- $\partial(\text{ERI})/\partial R$: Two-electron integral derivatives

- $\partial\mathbf{S}/\partial R$: **This gives rise to Pulay terms**

The Pulay force contribution is:

$$F_A^{\text{Pulay}} = -\sum_{\mu\nu} W_{\mu\nu} \frac{\partial S_{\mu\nu}}{\partial R_A}$$

where W is the energy-weighted density matrix.

(e) Why analytical gradients are preferred:

Property	Numerical	Analytical
Cost for N atoms	$6N$ SCF calculations	1 SCF + gradient
Accuracy	~ 6 digits	Machine precision
Step-size dependence	Sensitive	None
Stability	Subtractive cancellation	Robust

For geometry optimization (requiring many gradient evaluations), analytical gradients are essential for efficiency and reliability.

End of Chapter 7 Solutions