Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Residual-Certified Pivoted Coulomb Compression for Variational Density Fitting: A Posteriori Energy Bounds and Adaptive Auxiliary-Space Selection

Connor Noble

c0613727@gmail.com

Research Article

Keywords: density fitting, resolution of the identity, Coulomb metric, pivoted Cholesky decomposition, error certification, electronic-structure algorithms

Posted Date: September 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10789066/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Residual-Certified Pivoted Coulomb Compression for Variational Density Fitting: A Posteriori Energy Bounds and Adaptive Auxiliary-Space Selection

Research Article

Connor Noble
Independent Scholar
Tonganoxie, United States
Corresponding author: c0613727@gmail.com

Abstract

Density fitting and Cholesky factorizations reduce the four-center electron-repulsion bottleneck of molecular electronic-structure theory, yet production calculations commonly control approximation quality through a matrix threshold whose relation to the energy of the current electronic density is indirect. We formulate residual-certified pivoted Coulomb compression (RCPC), an adaptive low-rank construction that turns the positive-semidefinite residual of a pivoted Cholesky/Nyström representation into an a posteriori, density-specific upper bound for the Coulomb-energy error. For a Coulomb Gram matrix G and rank-k approximation LL^T, the residual $R=G-LL^T$ remains positive semidefinite and the energy defect is exactly $\frac{1}{2}q^TRq$. Positivity implies the computable certificate $\frac{1}{2}(\sum_i |q_i| VR_{ii})^2$, requiring only residual diagonals already maintained by pivoted Cholesky. We prove monotonicity, multi-density extensions, uniform-threshold corollaries, and a local derivative bound. Fixed-seed molecular Gaussian proxy spaces spanning 16 molecules and alkane chains up to 242 atoms verify the certificate numerically. At a 10^{-6} relative pivot threshold, the molecular set retains a median 82.6% of the redundant proxy space while the median population-density Coulomb error is 8.9×10^{-7} Eh. Density-aware stopping substantially lowers rank for signed transition-density proxies. RCPC therefore separates compression from accuracy specification and supplies a rigorous error-control layer suitable for future integration with RI-J, Hartree–Fock, hybrid DFT, and correlated workflows.

Keywords: density fitting; resolution of the identity; Coulomb metric; pivoted Cholesky decomposition; error certification; electronic-structure algorithms

Article classification: Theory / Computation / Modeling — numerical methods for electronic structure

1 Introduction

The computational role of auxiliary representations in electronic-structure theory is difficult to overstate. Four-center electron-repulsion integrals (ERIs) are formally indexed by pairs of orbital basis functions and therefore create an expensive tensor whose storage and contraction can dominate Hartree–Fock (HF), Kohn–Sham density-functional theory (DFT), second-order perturbation theory, coupled-cluster methods, and response calculations. Resolution-of-the-identity (RI) or density-fitting (DF) approaches replace this tensor by lower-rank three-center and two-center quantities, while Cholesky decomposition (CD) constructs a problem-adapted factorization from positivity of the Coulomb metric [1–13]. Both families exploit the same structural fact: electron-repulsion matrices are positive semidefinite and are often numerically compressible.

The classical DF literature emphasizes variationality because a naively fitted Coulomb contribution can contaminate first-order energies and analytic derivatives. The Coulomb-metric fit of Dunlap and co-workers gives a second-order error functional and, when solved stably, preserves a useful variational direction [4,14,15]. Subsequent work established robust local metrics, direct RI-HF algorithms, element-balanced auxiliary bases, local fitting, and atomic Cholesky constructions

[5–13,16–18]. More recently, automatic generation and contraction of auxiliary spaces has renewed interest in basis-independent and transferable fitting strategies [19–21]. These developments make the accuracy-control question increasingly important: if an auxiliary space is generated or pruned automatically, what quantity should determine when compression has gone far enough?

A standard answer is a decomposition threshold. In pivoted Cholesky, one stops when the largest residual diagonal element falls below a prescribed value. This threshold is mathematically transparent at the matrix level and has the practical advantage that no external auxiliary basis optimization is required. Nevertheless, a matrix threshold is not itself an energy tolerance. The mapping from a residual threshold to a Coulomb-energy error depends on the density coefficients contracted with the residual, on system size, on the normalization of the auxiliary representation, and on whether one is evaluating a total density, a polarized density, or a signed orbital-product/transition density. A fixed global threshold can therefore be either unnecessarily strict or insufficiently informative for a specific contraction.

This paper develops a residual-certificate viewpoint. The starting point is deliberately simple: if a pivoted Cholesky or subset-interpolation approximation leaves a positive-semidefinite residual R, then the Coulomb-energy error for coefficient vector q is a quadratic form in R. Positivity not only fixes the sign of the error; it also bounds every off-diagonal residual by the geometric mean of two residual diagonal entries. The diagonal maintained during the factorization is thus enough to construct a rigorous, density-specific error certificate without forming the full residual matrix. The certificate becomes a stopping criterion, a diagnostic, and a route to multi-density guarantees.

The contribution is methodological rather than a claim that the present proxy basis is a production quantum-chemistry implementation. We separate the mathematical algorithm from the benchmark vehicle. The theory applies to any positive-semidefinite Coulomb Gram matrix; the numerical experiments use atom-centered normalized Gaussian charge probes on fixed-seed molecular geometries so that all matrix elements are analytic and the exact residual errors can be computed directly. This controlled environment isolates the error-control mechanism from unrelated SCF, quadrature, basis-set, and integral-engine effects. Section 2 sets the notation, Section 3 develops the RCPC algorithm, Section 4 gives formal error results, Section 5 defines the computational experiments, Sections 6–8 analyze accuracy, scaling, and density-aware compression, and Sections 9–10 discuss integration and limitations.

1.1 Notation and quantities certified

For clarity, Table 1 summarizes the notation used throughout the derivations. The central distinction is between the matrix residual Rk, which is an algebraic object, and the scalar defect $\Delta E_{J,k}(q)$, which is the physically interpretable Coulomb-energy error for a specified coefficient vector. RCPC uses the residual diagonal d(k) to bridge these two levels without forming the full residual.

Table 1 Notation used in the derivations and RCPC algorithm

Symbol	Meaning
G	Positive-semidefinite Coulomb Gram/metric matrix
n	Dimension of the candidate auxiliary Coulomb space
q	Coefficient vector of the density or contraction to be certified
Q; m	Finite batch of target vectors; number of vectors in the batch
Lk	Rank-k pivoted Cholesky factor
Rk	Residual $G-L_kL_k^T$
d(k)	Diagonal of Rk
k	Current retained rank / number of accepted pivots
τ; ε	Relative matrix threshold; requested absolute energy tolerance
E_J; $\bar{E}_{J,k}$	Exact and rank-k Coulomb quadratic energies
$\Delta E_{J,k}$	Exact Coulomb-energy defect $E_J-\bar{E}_{J,k}$
Ck; S(v)	Diagonal certificate; residual-weighted 1-norm
Cblock	Optional block-residual certificate

2 Coulomb metric, density fitting, and low-rank structure

2.1 Coulomb bilinear form

For real charge distributions f and g with finite Coulomb self-energy, define the bilinear form ⟨f|g⟩ by the usual two-electron Coulomb kernel. For an auxiliary family $\{\chi_P\}_{P=1\ldots n}$, the Coulomb Gram matrix G has entries $G_{PQ}=\langle\chi_P||\chi_Q\rangle$. Positivity of the kernel implies $x^T G x \geq 0$ for every real coefficient vector x. Linear dependence of the auxiliary family may make G singular, but not indefinite in exact arithmetic. This positivity is the common algebraic foundation of RI/DF and Cholesky-based ERI approximations [1,2,7–13].

$$G_{PQ} = \langle \chi_P \,||\, \chi_Q \rangle, \quad x^T G x = \langle \Sigma_P x_P \chi_P \,||\, \Sigma_Q x_Q \chi_Q \rangle \geq 0 \tag{1}$$

Given an electronic or orbital-product density represented in this auxiliary coordinate system by q, its Coulomb self-energy is $E_J(q)=\tfrac{1}{2}q^T G q$. In conventional variational fitting, q itself may arise from a projection or normal-equation solve. Here q is simply the coefficient vector associated with the contraction whose error is to be controlled. The certification results therefore apply whether q corresponds to a total density, a pair density, a response density, or any linear combination for which the same Coulomb metric is used.

$$E_J(q) = \tfrac{1}{2}\, q^T G q \tag{2}$$

2.2 Pivoted Cholesky and Nyström viewpoints

A rank-k pivoted Cholesky factorization constructs $L_k \in R^{n\times k}$ such that $G \approx L_k L_k^T$. At each iteration the pivot is chosen from the largest residual diagonal element. The same rank-k approximation can be interpreted as a Coulomb-metric subset interpolation (a Nyström form) when the selected pivot indices define an invertible principal submatrix. The Cholesky form is operationally convenient because the residual diagonal can be updated in O(n) work per newly generated column once the column itself is available.

The conventional stopping condition compares the largest residual diagonal with a fixed fraction τ of its initial maximum. This controls the pointwise residual in a basis-dependent sense. The present work retains that mechanism but adds a second stopping rule tied directly to the contraction q. No modification of the pivot selection is required; only the decision to terminate is changed.

$$R_k = G - L_k L_k^T, \quad d_i(k) = (R_k)_{ii} \tag{3}$$

3 Residual-certified pivoted Coulomb compression

3.1 Density-specific certificate

For a fixed density coefficient vector q, define the diagonal residual certificate Ck(q) as one half of the squared weighted sum of square roots of the residual diagonal. Its evaluation is O(n) and uses only quantities already available in a standard pivoted Cholesky loop. If several densities must be controlled simultaneously, one evaluates the same expression for each vector and stops only when the maximum certificate is below the desired tolerance.

$$C_k(q) = \tfrac{1}{2}\left(\Sigma_i\, |q_i| \sqrt{d_i(k)}\right)^2 \tag{4}$$

$$C_k(Q) = \max_{q\in Q} C_k(q) \tag{5}$$

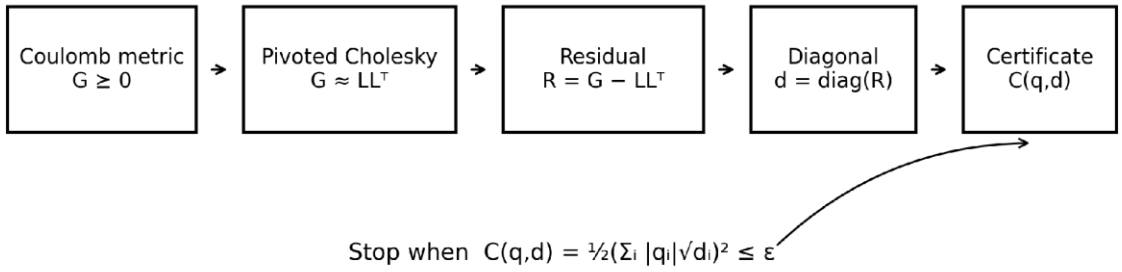


Fig. 1 Residual-certified pivoted Coulomb compression (RCPC). The factorization supplies both the low-rank Coulomb representation and the residual diagonal used to certify a density-specific energy tolerance

3.2 Algorithm

Algorithm 1. RCPC for one or more target densities

Input: Coulomb Gram matrix access G(:,p), target density set Q, requested absolute energy tolerance ε, optional matrix threshold τmin.

1. Initialize d ← diag(G), L ← empty, k ← 0.
2. Evaluate $C_0(Q)$. If $C_0(Q) \le \varepsilon$, return rank zero.
3. Choose p=argmaxi di. If $d_p \le \tau_{min} \cdot \max \mathrm{diag}(G)$, terminate by matrix threshold.
4. Form the pivoted Cholesky column $\ell = (G(:,p) - L\,L(p,:)^T)/\sqrt{d_p}$.
5. Append ℓ to L; update $d \leftarrow \max(d - \ell \circ \ell, 0)$; set $k \leftarrow k+1$.
6. Recompute $C_k(Q)$. If $C_k(Q) \le \varepsilon$, terminate with a certified rank-k representation; otherwise return to step 3.

Output: L, pivot list, residual diagonal d, and certificate $C_k(Q)$.

The safeguard τmin is optional. In finite precision it prevents continued pivots on diagonals indistinguishable from roundoff and acts as a familiar linear-dependence cutoff. When RCPC is embedded in a direct integral engine, the matrix need not be formed densely: the algorithm requires the diagonal initially and one Coulomb column at each pivot. This is the same data-access pattern as conventional integral Cholesky algorithms, so certification adds negligible asymptotic cost.

3.3 Uniform and density-aware stopping

Two operating modes are useful. In uniform mode, one selects a residual threshold τ independently of q and uses the certificate only as a diagnostic. In density-aware mode, the target ε is primary and the factorization terminates as soon as $C_k(q) \le \varepsilon$. The latter can reduce rank when the density has weak coefficients in directions associated with the largest residual diagonals. Conversely, it can request more pivots than a nominal matrix threshold when a large extensive density amplifies small residuals. The distinction becomes important for system-size scaling.

4 Formal properties and proofs

Theorem 1 (positive-semidefinite residual).

In exact arithmetic, every residual Rk produced by symmetric pivoted Cholesky applied to a positive-semidefinite Coulomb Gram matrix G is positive semidefinite until termination.

Proof. *At a pivot p with $d_p>0$, partition the current residual so that the pivot is first. The Cholesky rank-one update subtracts rr^T/d_p, where r is the pivot column. The Schur complement of the positive pivot in a positive-semidefinite matrix is positive semidefinite. The updated residual is congruent to this Schur complement augmented by a zero pivot row/column. Induction from $R_0=G$ completes the argument.* ■

Theorem 2 (exact Coulomb-energy defect).

Let $\bar{E}J,k(q)=\frac{1}{2}q^TLkLk^Tq$ be the rank-k Coulomb energy. Then

$$\Delta EJ,k(q) = EJ(q) - \bar{E}J,k(q) = \frac{1}{2}\, q^T R_k q \geq 0 \tag{6}$$

Proof. *Substitute $G=L_kL_k{}^T+R_k$ into Eq. (2). Non-negativity follows immediately from Theorem 1.* ■

Theorem 2 makes the sign of the compression error explicit: for a fixed q and a positive-semidefinite residual, the low-rank Coulomb self-energy is a lower approximation to the exact quadratic form defined by G. This statement concerns the isolated Coulomb contraction and should not be confused with the sign of a total self-consistent electronic energy, where orbital relaxation and other fitted contributions may alter error cancellation.

Theorem 3 (residual-diagonal certificate).

For any $q\in R^n$ and any positive−semidefinite residual R with diagonal d,

$$0 \leq \frac{1}{2}q^TRq \leq \frac{1}{2}(\Sigma_i\, |q_i|\sqrt{d_i})^2 = C(q) \tag{7}$$

Proof. *Every 2×2 principal minor of R is positive semidefinite, hence $|R_{ij}|^2\leq R_{ii}R_{jj}=d_id_j$. Therefore $q^TRq\leq\Sigma_{ij}|q_i||q_j||R_{ij}|\leq\Sigma_{ij}|q_i||q_j|\sqrt{d_i}\sqrt{d_j}=(\Sigma_i|q_i|\sqrt{d_i})^2$. Multiplication by one half yields the result.* ■

Corollary 3.1 (uniform max-diagonal bound).

$$C(q) \leq \frac{1}{2}\, ||d||\infty\, ||q||_1{}^2 \tag{8}$$

Proof. *Use $\sqrt{d_i}\leq\sqrt{||d||\infty}$ in Eq. (7).* ■

Corollary 3.1 provides an explicit translation from the familiar maximum pivot residual to an energy guarantee. It also exposes why a fixed matrix tolerance is not size-extensive as an energy specification: the right-hand side scales with the square of the coefficient 1-norm unless the residual threshold is tightened with system size. The sharper diagonal certificate in Eq. (7) automatically accounts for the spatial and coefficient distribution of q and can be substantially smaller than this worst-case uniform bound.

Corollary 3.2 (trace/spectral alternative).

$$\frac{1}{2}q^TRq \leq \frac{1}{2}\, tr(R)\, ||q||_2{}^2 = \frac{1}{2}(\Sigma_i d_i)||q||_2{}^2 \tag{9}$$

Proof. *For $R\geq 0$, $\lambda max(R)\leq tr(R)$. Hence $q^TRq\leq\lambda max(R)||q||_2{}^2\leq tr(R)||q||_2{}^2$.* ■

Equation (9) is useful when the density coefficient vector is diffuse in index space, whereas Eq. (7) often benefits from coefficient localization. Both bounds can be computed from the residual diagonal; an implementation may report their minimum while retaining Eq. (7) as the default certificate.

Theorem 4 (monotonicity).

Under exact pivoted Cholesky updates, $d_i(k+1)\leq d_i(k)$ for every i; consequently $C_{k+1}(q)\leq C_k(q)$ for every fixed q.

Proof. *The diagonal update is $d(k+1)=d(k)-\ell\circ\ell$. Non-negative entries are therefore non-increasing. The square-root weighted sum in Eq. (4) is monotone in each d_i.* ■

Monotonicity matters operationally: once a target has been certified it remains certified if additional pivots are taken, and a multi-density set can be handled without backtracking.

Theorem 5 (simultaneous certification).

Let Q={q(1),⋯,q(m)}. If maxa Ck(q(a))≤ε, then every member of Q has Coulomb−energy defect at most ε under the same factor Lk.

Proof. *Apply Theorem 3 independently to each q(a) and take the maximum.* ■

This simple extension is relevant to batches of orbital-product densities, response vectors, state-averaged densities, or SCF iterates. It enables a single compressed metric to be certified against the largest error among a prescribed family without changing the pivot rule.

Proposition 6 (local derivative bound for a fixed pivot set).

Let q(t) and R(t) be differentiable with respect to a geometry parameter t in an interval where the pivot set and rank are fixed. Define m(t)=maxi,j |R′ij(t)|. Then the derivative of the Coulomb-energy defect satisfies

$$|d\Delta E/dt| \le ||q'||_1 ||q||_1 \, ||d||_\infty + \tfrac{1}{2} \, ||q||_1^{\,2} \, m(t) \tag{10}$$

Proof. *Differentiate $\Delta E=\tfrac{1}{2}q^{T}Rq$ to obtain $q'^{T}Rq+\tfrac{1}{2}q^{T}R'q$. The first term is bounded by $||q'||_1||q||_1 \max_{ij}|R_{ij}|$, and positivity gives $\max_{ij}|R_{ij}|\le||d||_\infty$. The second term is bounded directly by $\tfrac{1}{2}||q||_1^2 m(t)$.* ■

The proposition is intentionally local. Pivot changes introduce piecewise-smooth behavior and must be treated carefully in analytic-gradient implementations. A practical production strategy would freeze the selected auxiliary subspace over a geometry step or use hysteresis in pivot updates. The present work does not claim a complete analytic-gradient implementation; Eq. (10) identifies the terms that must be controlled.

4.1 Seminorm interpretation and transfer between nearby densities

The certificate has a useful functional-analytic interpretation. Because Rk is positive semidefinite, it induces the seminorm ‖x‖R=(xᵀRkx)1/2 on the quotient space obtained after removing the null space of Rk. The exact Coulomb defect is one half of the squared residual seminorm of q. The diagonal certificate replaces the unknown residual inner products by the extremal correlations permitted by positivity. In this sense, Eq. (7) is the sharpest bound that can be obtained from the diagonal alone without imposing further structure: a rank-one positive-semidefinite residual with Rij=√didj and a sign pattern aligned with q attains equality.

This observation is important when interpreting conservatism. The certificate is not a heuristic norm estimate calibrated to the present benchmark. It is a worst-case completion bound: among all positive-semidefinite matrices having the same residual diagonal, no smaller universal upper bound can be guaranteed for an arbitrary q. Any systematic tightening must therefore introduce additional information, such as block residuals, selected cross terms, locality assumptions, or properties of the coefficient family.

Theorem 7 (stability under a change of density).

Let q and q̃ be two coefficient vectors evaluated with the same rank-k residual R and let S(v)=Σi|vi|√di. Then

$$|\Delta E(q) - \Delta E(\tilde{q})| \le \tfrac{1}{2} \, S(q-\tilde{q}) \, [\, S(q) + S(\tilde{q}) \,] \tag{11}$$

Proof. *Write $q^{T}Rq-\tilde{q}^{T}R\tilde{q}=(q-\tilde{q})^{T}R(q+\tilde{q})$. For a positive-semidefinite R, the same 2×2-principal-minor argument used in Theorem 3 gives $|u^{T}Rv|\le S(u)S(v)$. Applying the triangle inequality $S(q+\tilde{q})\le S(q)+S(\tilde{q})$, followed by the factor one half from the energy definition, yields Eq. (11).* ■

Theorem 7 gives a practical SCF reuse criterion. Suppose a factorization was certified for an iterate q and the density changes to q̃. The new error need not be assessed from scratch conceptually: the old defect bound plus Eq. (11) controls the change. Since d is already available, S(q−q̃) is inexpensive. An implementation can therefore avoid certificate checks or auxiliary-rank expansion during early microiterations when the density change is small in the residual-weighted 1-norm, while still performing a full certificate test before accepting a macroiteration.

4.2 Block-residual refinement

Diagonal-only certification is attractive because it is essentially free, but electronic-structure implementations often possess small residual blocks naturally—for example, atom-pair blocks, shell blocks, or local domains. These blocks permit a hierarchy of rigorous certificates between Eq. (7) and the exact quadratic form. Let the auxiliary index set be partitioned into disjoint blocks B1,…,Bs, and denote by Rbb the principal residual block and qb the corresponding coefficient subvector. Define eb=(qb^{T}Rbb qb)1/2. Positive semidefiniteness implies that the cross-block residual inner product cannot exceed the product of the associated residual seminorms.

Theorem 8 (block certificate).

$$0 \le \tfrac{1}{2}\, q^T R q \le \tfrac{1}{2}\, (\Sigma_b \sqrt{(q_b^T\, R_bb\, q_b)})^2 = C_block(q) \tag{12}$$

Proof. *Choose any factorization $R=XX^T$ and write $X^Tq=\Sigma b\, Xb^Tqb$. The triangle inequality gives $||X^Tq||2\le\Sigma b||Xb^Tqb||2$. Squaring and using $||Xb^Tqb||^2=qb^TRbbqb$ yields Eq. (12).* ■

If every block contains one index, Eq. (12) reduces exactly to the diagonal certificate. If one block contains the entire residual, it becomes the exact defect. Thus block size defines a controlled accuracy–overhead continuum. An atom-block implementation would require only residual principal blocks associated with each atom and could be attractive when those blocks are already available from integral screening. This refinement is not used in the numerical benchmark, so all reported certificates retain the stricter information constraint of diagonal-only RCPC.

A related shell-block variant can exploit the fact that Gaussian auxiliary functions are generated and screened by shell. The mathematical guarantee does not depend on the chemical meaning of the partition. What matters is that each Rbb is a true principal residual block. Approximating a block independently without preserving its positive-semidefinite relation to the global residual would forfeit the proof and should therefore be avoided unless a second certified bound is supplied.

4.3 Relative-error certification and tolerance design

Many electronic-structure codes expose relative rather than absolute numerical tolerances. RCPC can support this interface without estimating the exact Coulomb energy. Let $\bar{E}k=\tfrac{1}{2}q^TLkLk^Tq\ge0$ and let Ck be the diagonal certificate. Since the exact energy is $\bar{E}k+\Delta E$ with $0\le\Delta E\le Ck$, the relative error with respect to the exact positive Coulomb self-energy can be bounded using only computable quantities.

Corollary 8.1 (relative Coulomb-energy defect).

$$\Delta E \,/\, E \le C_k \,/\, (\bar{E}_k + C_k) \quad \text{whenever } \bar{E}_k + C_k > 0 \tag{13}$$

Proof. *For fixed $\bar{E}k\ge0$, the function $x/(\bar{E}k+x)$ increases for $x\ge0$. Since $0\le\Delta E\le Ck$, substitution of the largest admissible x gives the bound.* ■

Equation (13) permits a code to request, for example, a maximum relative Coulomb defect without knowing E in advance. For total-energy workflows the appropriate absolute tolerance should usually be more stringent than the desired final total-energy accuracy because the Coulomb term is only one contribution and self-consistency couples numerical errors. A conservative engineering choice is to assign an error budget to each approximate kernel and require the sum of certified absolute budgets to remain below the desired electronic-energy threshold.

The certificate also clarifies system-size scaling. If the coefficient 1-norm grows approximately linearly with the number of repeated fragments, Corollary 3.1 suggests that a fixed maximum residual diagonal can produce a worst-case energy bound growing quadratically. Locality and cancellation generally improve this pessimistic scaling, but a matrix threshold alone cannot express the intended extensive error budget. A size-consistent policy can instead prescribe ε proportional to atom count, electron count, or number of fragments, and let the certificate determine the rank needed to meet that policy.

5 Computational protocol

5.1 Molecular structures and Gaussian Coulomb probes

The numerical study is designed to verify the matrix theorems and characterize compression behavior in molecule-shaped Coulomb spaces without depending on a proprietary integral engine. Molecular graphs were specified by SMILES strings, explicit hydrogens were added, three-dimensional coordinates were generated with RDKit ETKDGv3 using the fixed random seed 20260822, and geometries were relaxed with UFF. These geometries serve only as reproducible spatial point sets; no claim is made that they are high-level equilibrium structures.

Each atom A of atomic number ZA carries eight normalized s-type Gaussian charge probes. The exponent family is even-tempered-like and element scaled: $\alpha_{A,m}=0.18\ ZA^{0.55}\ fm$, with $fm\in\{0.20,0.32,0.50,0.78,1.22,1.90,2.95,4.60\}$. The resulting spaces are deliberately redundant enough to produce nontrivial low-rank behavior while retaining analytic pair Coulomb interactions. For normalized spherical Gaussian charge distributions $\rho_i(r)=(\alpha_i/\pi)^{3/2}\exp[-\alpha_i|r-R_i|^2]$, the Coulomb interaction is evaluated analytically by the error-function expression in Eq. (14).

$$G_{ij} = \mathrm{erf}\left(\sqrt{(\alpha_i\alpha_j/(\alpha_i+\alpha_j))}\ R_{ij}\right) / R_{ij}, \quad R_{ij}>0 \tag{14}$$

$$G_{ii} = 2\sqrt{\left(\alpha_i\alpha_i/(2\alpha_i) / \pi\right)} \tag{15}$$

The benchmark contains 16 chemically familiar molecules from water to anthracene, giving 24–192 auxiliary functions. A second size-scaling sequence contains linear alkanes from C5 to C80; after hydrogen completion the largest member contains 242 atoms and 1936 Gaussian probes. The Coulomb matrices are assembled in double precision and symmetrized before factorization.

System	Atoms	Auxiliary probes
water	3	24
ammonia	4	32
methane	5	40
methanol	6	48
formaldehyde	4	32
ethene	6	48
acetylene	4	32
benzene	12	96
pyridine	11	88
phenol	13	104
glycine	10	80
alanine	13	104
uracil	12	96
caffeine	24	192
naphthalene	18	144
anthracene	24	192

Table 2 Molecular benchmark set. Each atom carries eight normalized s-type Gaussian Coulomb probes

5.2 Density coefficient families

Three deterministic coefficient families probe different contraction regimes. The population density is non-negative and distributes ZA electrons over the eight probes on atom A with weights proportional to $\alpha^{-1/2}$. The

polarized density adds a small geometry-dependent dipolar redistribution while preserving the total coefficient sum. The transition-density proxy is signed, neutral in coefficient sum before probe expansion, and localized through a deterministic oscillatory atomic pattern; its overall 1-norm scales with the square root of the total nuclear charge. These are not self-consistent electronic densities. They are controlled coefficient vectors used to test how certification responds to positive extensive and signed localized contractions.

For each matrix, pivoted Cholesky calculations were carried out at relative maximum-diagonal thresholds $\tau = 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}$, and 10^{-10}. The exact energy defect was computed from the dense matrix, the approximation from L, and the diagonal certificate from Eq. (4). The dense calculation is used only for validation; the certificate itself does not require the full residual.

5.3 Reproducibility, numerical arithmetic, and verification design

The benchmark was constructed so that every numerical quantity used in the argument can be regenerated from a short deterministic specification. The molecular graph, embedding seed, auxiliary exponent rule, analytic Coulomb kernel, pivot rule, and coefficient definitions are fixed. Double precision is used throughout. Before factorization, each numerically assembled Coulomb matrix is explicitly symmetrized as $(G+G^T)/2$ to remove sub-ulp asymmetry from pairwise evaluation. During Cholesky updates, residual diagonals smaller than zero only by roundoff are clipped to zero before square roots are taken.

For each threshold, validation uses two independent routes to the defect. First, the dense reference value $\tfrac{1}{2}q^TGq$ is compared with $\tfrac{1}{2}\|L^Tq\|_2^2$. Second, the residual form $\tfrac{1}{2}q^T(G-LL^T)q$ is evaluated directly when the difference remains numerically resolved. Agreement of these routes is a consistency check on factor construction and contraction. At very tight thresholds, subtraction of two nearly identical positive energies becomes dominated by floating-point cancellation; such points are retained in the data but excluded from ratio claims when the measured defect is below 10^{-10} Eh.

The distinction between theorem verification and chemical benchmarking is deliberate. The present study asks whether the proposed a posteriori bound is mathematically valid, numerically stable over molecule-shaped Coulomb matrices, and capable of differentiating coefficient families. It does not use the proxy model to infer atomization energies, reaction barriers, spectra, or basis-set limits. A production follow-up should replace the analytic probe matrix with three-index or Cholesky ERI machinery and compare against conventional RI-J/JK references under identical orbital bases.

6 Verification of the certificate

6.1 Bound validity over the molecular set

The central numerical test is whether the measured quadratic-form error ever exceeds the diagonal residual certificate. For all benchmark points whose defects are resolved above 10^{-10} Eh, no violation is observed. The largest measured error-to-certificate ratio remains below unity; apparent discrepancies at tighter thresholds occur only when both quantities approach floating-point cancellation. Figure 2 plots certificate against actual error on logarithmic axes. Every resolved point lies on or above the equality line, exactly as Theorem 3 requires.

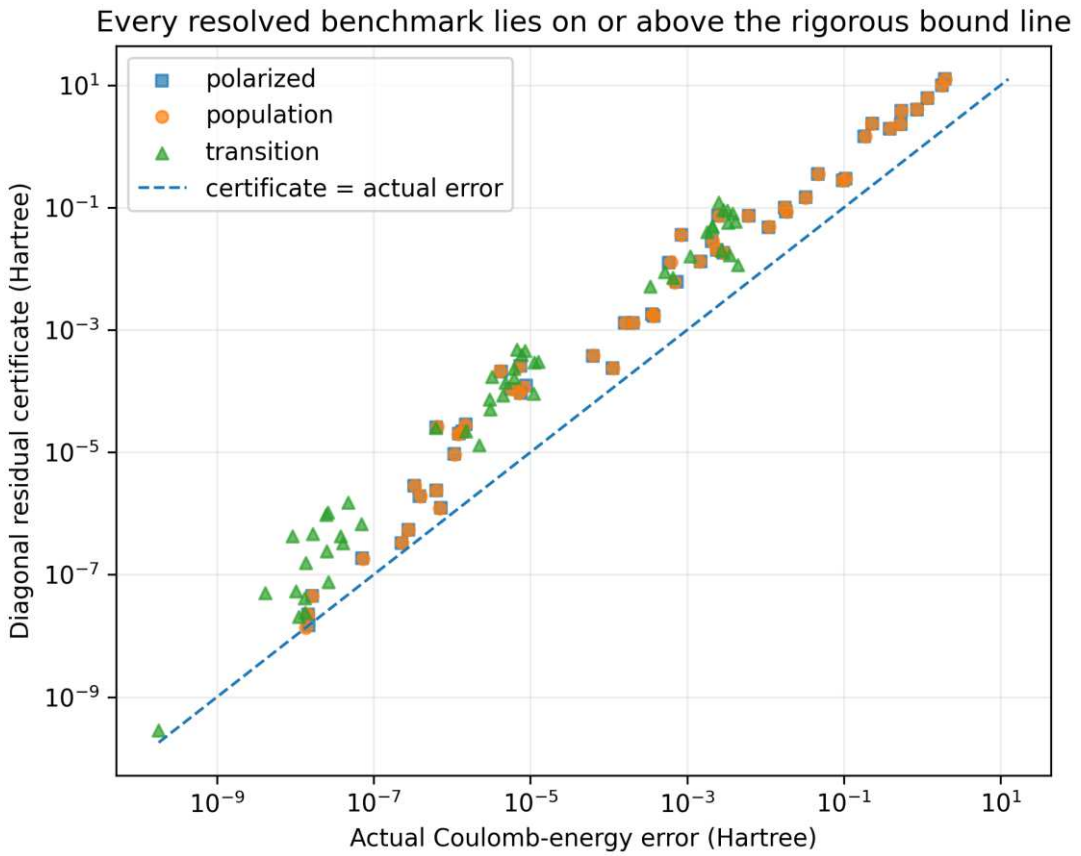


Fig. 2 Diagonal residual certificate versus measured Coulomb-energy defect for all resolved molecular benchmark points. Equality is the theoretical lower envelope for a valid upper bound

τ	Density	Median k/n	Median error (Eh)	Median certificate (Eh)	Median cert./error
1e-02	population	0.224	2.080e-01	1.711e+00	5.2
1e-02	polarized	0.224	2.054e-01	1.701e+00	5.2
1e-02	transition	0.224	2.624e-03	4.392e-02	17.1
1e-04	population	0.521	6.524e-04	9.363e-03	8.7
1e-04	polarized	0.521	6.630e-04	9.383e-03	8.5
1e-04	transition	0.521	5.456e-06	1.502e-04	26.5
1e-06	population	0.826	8.929e-07	1.461e-05	10.7
1e-06	polarized	0.826	8.935e-07	1.478e-05	10.6
1e-06	transition	0.826	2.054e-08	2.855e-07	10.5

Table 3 Median performance over the 16-molecule set. Certificate/error ratios exclude exact zeros through floating-point replacement

The table quantifies the accuracy–compression trade-off. At τ=10−2 only about 22% of the redundant auxiliary space is retained, but population and polarized Coulomb errors are on the order of 0.2 Eh and therefore too large for chemically accurate energy work. At τ=10−4 the median retained fraction rises to approximately 0.52 and the median population error falls to 6.5×10−4 Eh, approximately 0.41 kcal mol−1 when interpreted as an energy scale. At τ=10−6 the median retained fraction is approximately 0.826 and the median population error is 8.9×10−7 Eh. Signed

transition-density proxies exhibit much smaller absolute errors at the same matrix rank because their coefficients cancel and carry a smaller effective Coulomb norm.

The certificate is intentionally conservative. At τ=10−6 the typical certificate exceeds the actual error by roughly one order of magnitude for the positive-density families. This gap is not a defect in the proof; it is the price of eliminating unknown residual off-diagonal information. Importantly, the certificate is substantially sharper than the cruder max-diagonal/1-norm bound for many localized vectors because it retains the full residual diagonal profile.

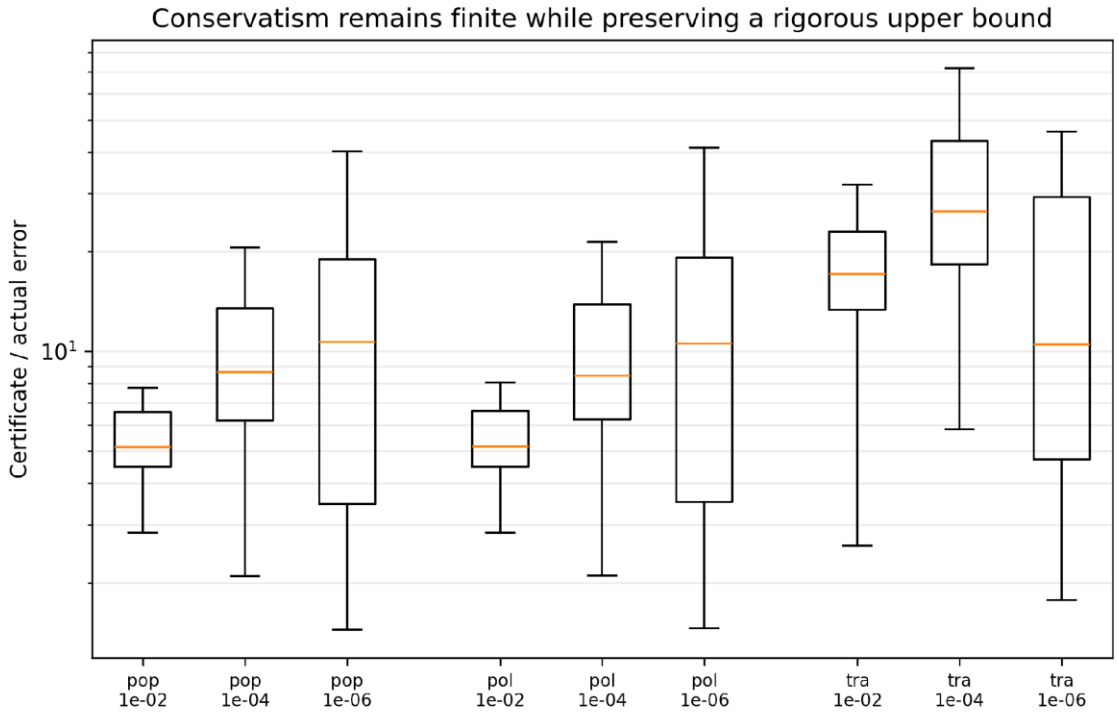


Fig. 3 Ratio of certificate to measured error across molecular systems for representative thresholds. The ratio is finite and always at least unity for resolved points

6.2 Tightness, worst-case completion, and floating-point regime

Across the resolved benchmark subset (measured error greater than 10−10 Eh), the smallest certificate−to−error ratio is approximately 1.0023. The near−unity case demonstrates that the diagonal bound can become almost saturated for some combinations of residual geometry and coefficient signs; it is therefore not possible to reduce the certificate by a universal empirical prefactor without losing rigor. Median ratios are larger: approximately 5.2 for the positive−density families at τ=10−2, 8.5–8.7 at τ=10−4, and 10.6–10.7 at τ=10−6. Transition−density ratios are broader because cancellation lowers the exact defect more strongly than it lowers the absolute−value−based certificate.

This pattern suggests two complementary interpretations. For guaranteed calculations, conservatism is desirable because the bound must remain valid for densities not represented in a calibration set. For performance-oriented calculations, the gap motivates the block refinement of Eq. (12) or a multi-tier strategy in which diagonal RCPC provides the inexpensive first certificate and selected residual blocks are formed only when the diagonal estimate is too conservative to permit termination.

The τ=10−8 and 10−10 points enter a different numerical regime for the smaller molecular spaces: many decompositions reach full rank and both the exact defect and certificate approach machine precision. Ratios there are not meaningful tests of mathematical validity because the defect is obtained by subtracting nearly equal floating-point energies. This is why the quantitative bound-validation claim is restricted to resolved points rather than reporting visually impressive but numerically ill-conditioned ratios at machine precision.

6.3 Error distributions across thresholds

Figure 4 summarizes the median and interquartile range of the measured defects as the factorization threshold is tightened. The population and polarized families are nearly coincident over the full threshold range. This is expected because the polarization perturbation is deliberately modest relative to the extensive positive background; it also provides a useful numerical consistency check, because a small perturbation of q should not produce an erratic change in a smooth quadratic contraction with a common residual.

The transition-density proxy follows a qualitatively different error scale. At τ=10−4 its median defect is 5.46×10−6 Eh, roughly two orders of magnitude below the 6.52×10−4 Eh population median, even though the matrix factor and rank are identical. At τ=10−6 the respective medians are 2.05×10−8 and 8.93×10−7 Eh. This is direct evidence that rank alone is an incomplete measure of contraction accuracy: the same residual matrix can be chemically negligible for one coefficient family and unacceptably large for another.

The interquartile envelopes also widen at loose thresholds because molecular size and topology influence both the unreconstructed Coulomb directions and the coefficient norms. Tightening τ reduces the absolute errors rapidly, but the rank approaches the full proxy dimension by τ≈10−8 for the smaller systems. Density-aware termination uses the opposite logic: rather than forcing all systems toward a common residual-matrix scale, it asks each system and density to stop at the earliest rank that satisfies the actual energy budget.

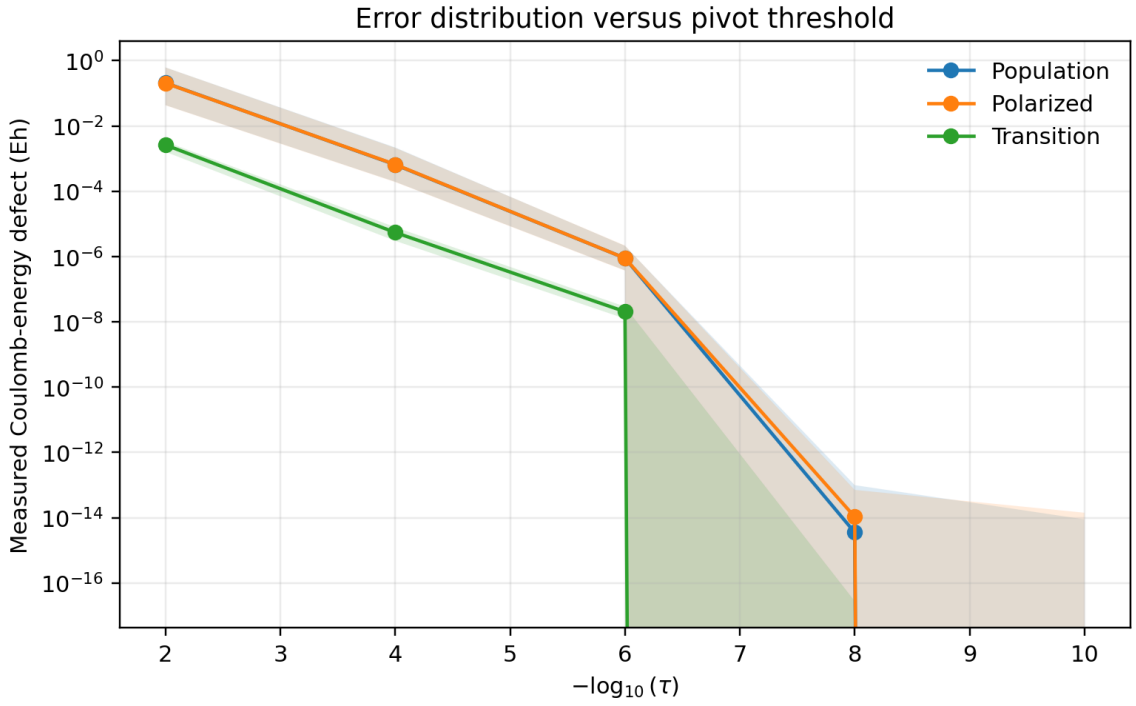


Fig. 4 Median measured Coulomb-energy defect and interquartile range over the 16-molecule set as the relative pivot threshold is tightened. Population and polarized densities track closely, while signed transition-density proxies remain substantially lower

7 Matrix thresholds, rank, and system-size effects

7.1 Rank response to the pivot threshold

Figure 5 shows the retained rank fraction for all 16 molecular systems as the matrix threshold is tightened. The monotone trend is expected, but two features are relevant to algorithm design. First, the transition from aggressive to near-complete representation is broad rather than abrupt; this supports adaptive stopping because many intermediate ranks are available. Second, even within a chemically modest set there is visible system dependence at fixed τ. A universal threshold therefore does not correspond to a universal rank fraction or energy error.

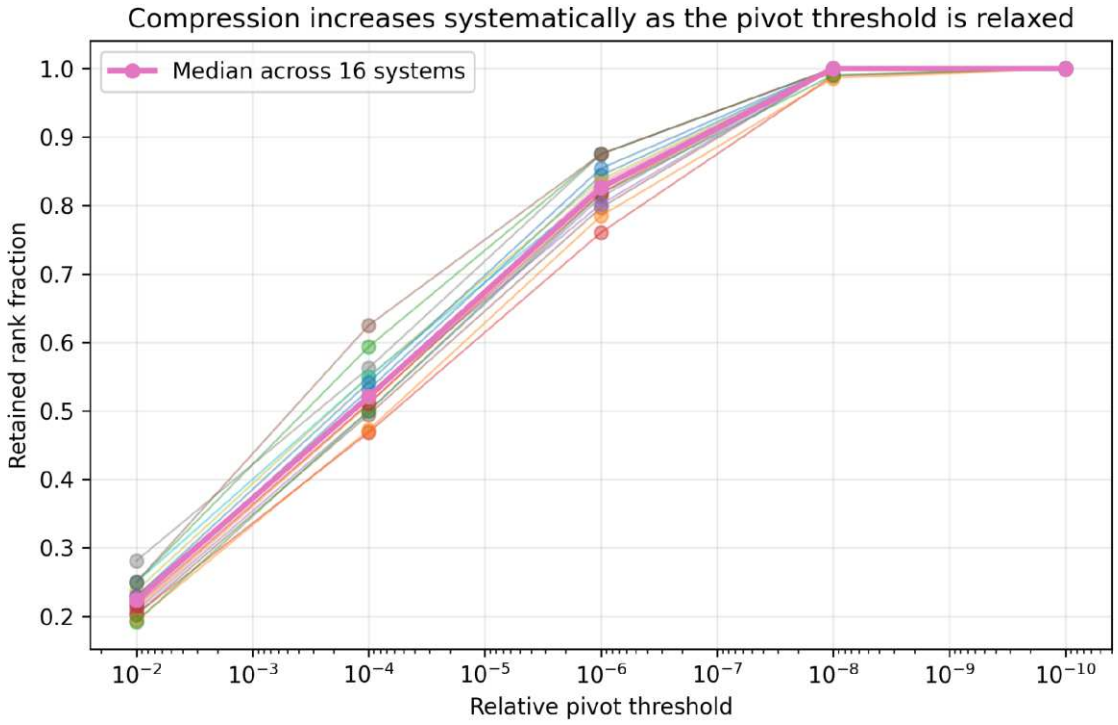


Fig. 5 Retained rank fraction as a function of relative pivot threshold for the 16 molecular Coulomb proxy spaces. The bold curve is the system median

7.2 Alkane scaling series

The alkane sequence tests whether redundancy becomes more exploitable as a spatial system grows. At τ=10−4 the retained rank fraction falls from 0.485 for C5H12 to 0.365 for C80H162, even though the absolute retained rank grows from 66 to 706. At τ=10−6 the fraction falls from 0.794 to 0.604. This is favorable from the standpoint of relative compression and is consistent with repeated local environments creating increasingly redundant Coulomb directions.

However, the absolute energy guarantee behaves differently. At τ=10−6 the measured population-density error grows from approximately 5.1×10−6 Eh for the smallest chain to 1.30×10−4 Eh for the largest, while the rigorous certificate grows more strongly. This behavior is exactly what Corollary 3.1 anticipates: an extensive coefficient vector can amplify a fixed residual threshold. The result argues against interpreting τ as a system-independent chemical-accuracy parameter.

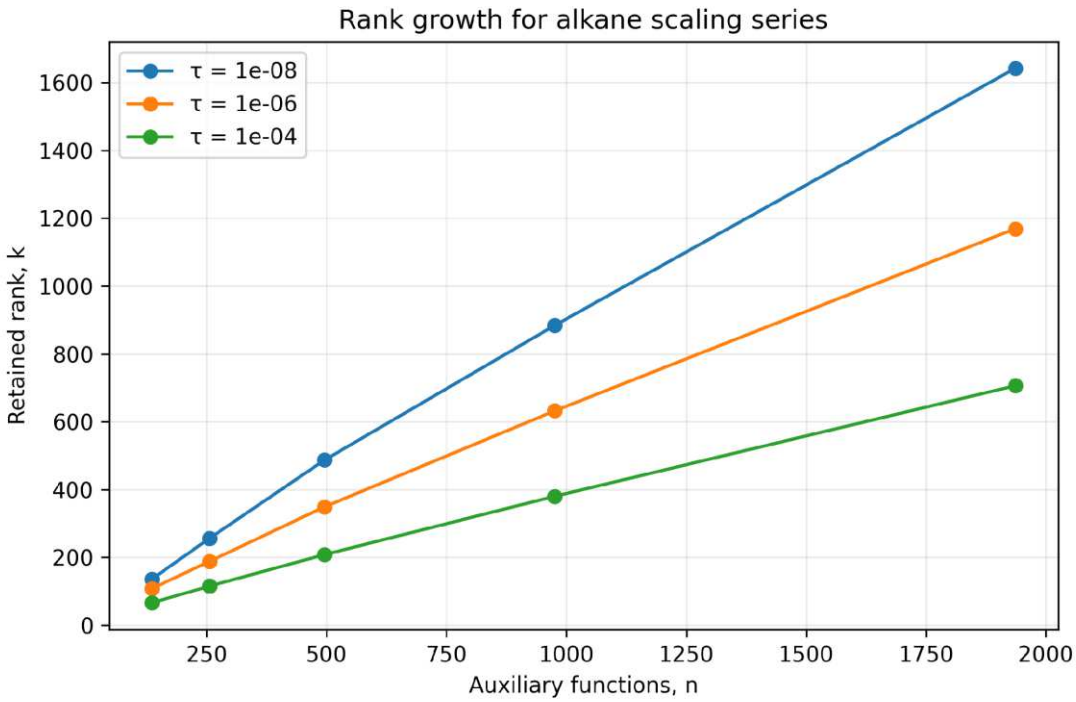


Fig. 6 Retained Cholesky rank versus auxiliary-space dimension for alkane chains at three relative matrix thresholds

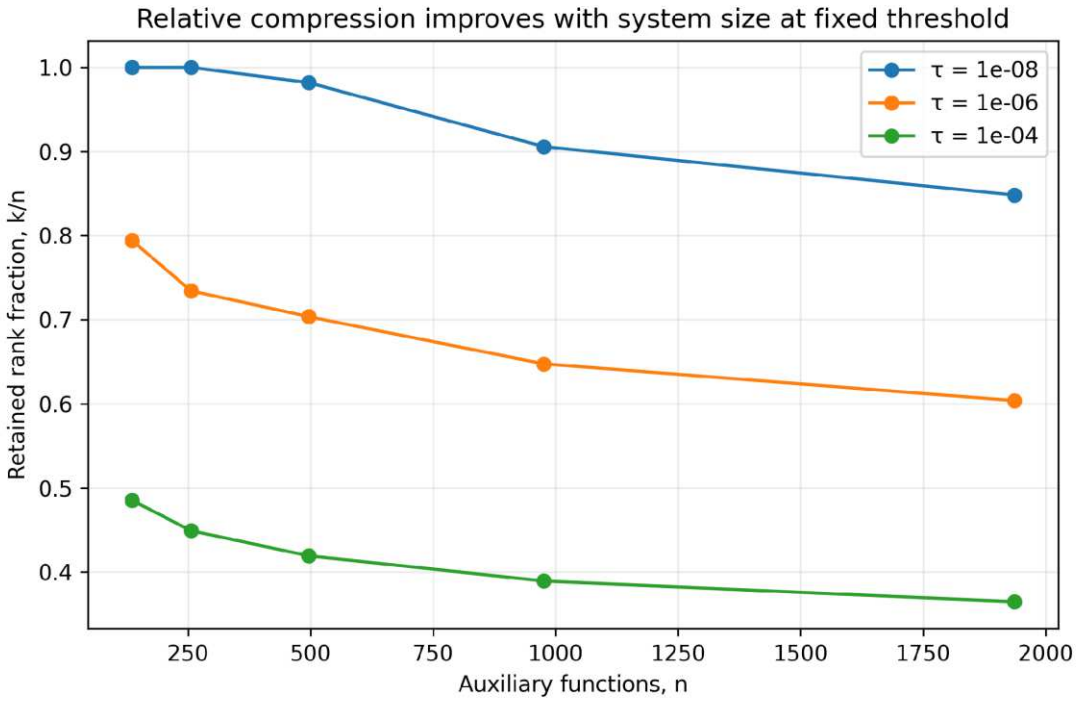


Fig. 7 Retained rank fraction versus auxiliary-space dimension for the same alkane series. Relative compression improves with system size at fixed threshold

Chain	Atoms	n	k	k/n	Actual error (Eh)	Certificate (Eh)
C5	17	136	108	0.794	5.136e-06	5.463e-05
C10	32	256	188	0.734	1.008e-05	2.959e-04

Chain	Atoms	n	k	k/n	Actual error (Eh)	Certificate (Eh)
C20	62	496	349	0.704	2.695e-05	1.164e-03
C40	122	976	632	0.648	5.594e-05	4.206e-03
C80	242	1936	1169	0.604	1.299e-04	1.899e-02

Table 4 Alkane scaling results at τ=10−6

7.3 Empirical scaling exponents and interpretation

A log–log fit over the five alkane sizes gives an empirical retained-rank exponent of approximately 0.89 with respect to naux at τ=10−4 and approximately 0.90 at τ=10−6. These exponents are descriptive rather than asymptotic complexity results: five finite chain lengths cannot establish a thermodynamic-limit scaling law. They nevertheless quantify the visual trend in Figures 6 and 7 and show that the active rank grows more slowly than the raw number of redundant probes over the tested range.

The measured population-energy defect grows more rapidly with size than the retained rank. At τ=10−6, a log–log fit gives an empirical error exponent of approximately 1.23. This is compatible with the theoretical warning that a fixed matrix tolerance is not an extensive energy specification. The rigorous certificate is still more conservative, reaching 1.90×10−2 Eh for C80 at τ=10−6 even though the measured defect is 1.30×10−4 Eh. A user interested in a fixed per-fragment energy budget would therefore gain little from selecting τ by convention alone; density-aware ε should be the primary control variable.

The scaling series also separates two distinct meanings of "efficient." A decreasing rank fraction indicates increasing algebraic redundancy, while the absolute rank determines factor-contraction work and memory. RCPC does not alter the cost of generating a required Cholesky column, but by tying termination to a target contraction it can avoid columns that are irrelevant to the chosen error budget. Whether this produces a wall-clock speedup depends on the integral engine, screening, parallel distribution, and frequency with which densities change. Those implementation effects require measurement in a production code.

8 Density-aware certified stopping

8.1 Certified ranks

The strongest practical use of Eq. (4) is to stop on an energy tolerance rather than on a residual−matrix tolerance. Figure 8 illustrates this for the caffeine proxy space. For positive population and polarized densities, a certified tolerance of 10−3 Eh is reached at about 70.8% of the full 192−function space; a 10−4 Eh certificate requires about 82.8%. The signed transition−density proxy requires only 45.3% and 58.9%, respectively. The difference arises because the certificate weights each residual direction by |qi| and therefore recognizes when the current density weakly occupies unreconstructed directions.

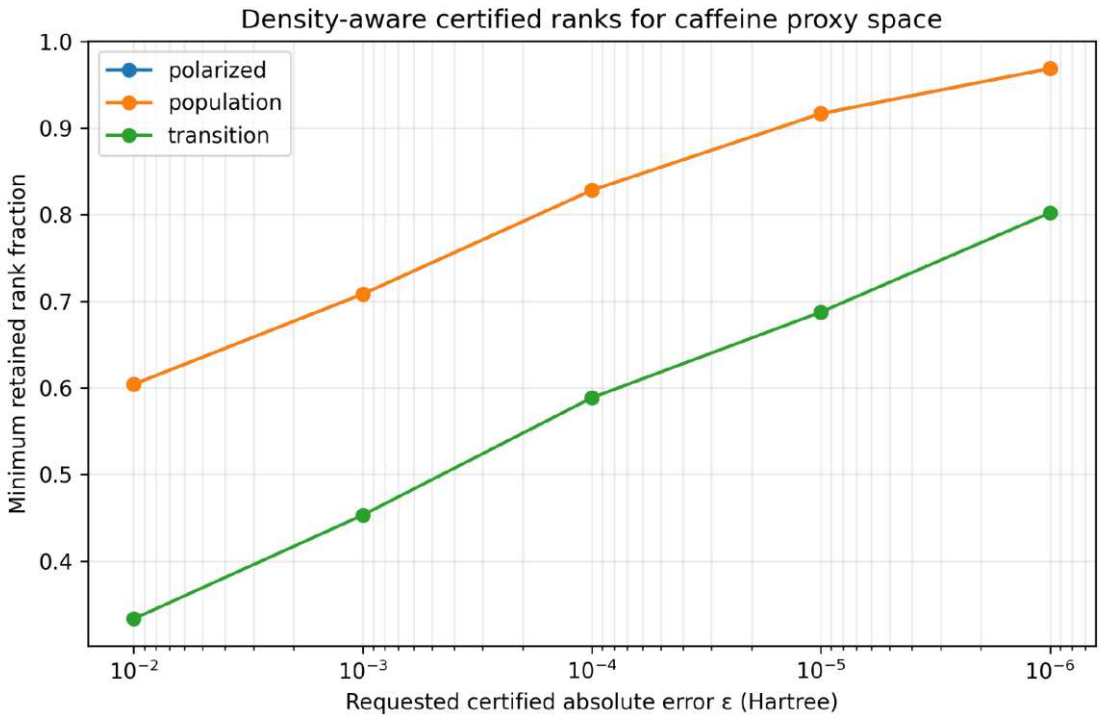


Fig. 8 Minimum retained rank fraction required to certify specified absolute Coulomb-energy tolerances for the caffeine proxy space

System	Density	ε (Eh)	k/n	Rank fraction	Actual error	Certificate
benzene	population	1e-03	62/96	0.646	4.74e-05	8.67e-04
benzene	population	1e-04	72/96	0.750	4.03e-06	9.51e-05
benzene	transition	1e-03	41/96	0.427	5.49e-05	9.99e-04
benzene	transition	1e-04	52/96	0.542	1.46e-06	8.77e-05
glycine	population	1e-03	51/80	0.637	5.09e-05	9.34e-04
glycine	population	1e-04	63/80	0.787	2.21e-06	7.33e-05
glycine	transition	1e-03	36/80	0.450	4.65e-05	8.51e-04
glycine	transition	1e-04	46/80	0.575	3.43e-06	9.00e-05
caffeine	population	1e-03	136/192	0.708	1.77e-05	8.79e-04
caffeine	population	1e-04	159/192	0.828	3.56e-06	8.99e-05
caffeine	transition	1e-03	87/192	0.453	2.02e-05	9.15e-04
caffeine	transition	1e-04	113/192	0.589	1.80e-06	8.82e-05
anthracene	population	1e-03	134/192	0.698	1.52e-05	9.98e-04
anthracene	population	1e-04	155/192	0.807	3.31e-06	9.64e-05
anthracene	transition	1e-03	85/192	0.443	9.26e-06	8.80e-04
anthracene	transition	1e-04	109/192	0.568	1.83e-06	9.39e-05

Table 5 Density-aware minimum ranks for selected molecular spaces

A useful feature is that the certificate can terminate before the largest residual diagonal reaches an arbitrarily strict global threshold. For benzene, a transition-density tolerance of 10−4 Eh is certified with 52 of 96 directions, whereas the corresponding positive population density requires 72 directions. This behavior suggests a route to contraction-specific low-rank representations in correlated and

response methods, where thousands of pair or transition densities may be sparse or sign-changing in auxiliary coordinates.

The multi-density extension provides a middle ground between a single global metric threshold and a separate factorization for every contraction. A batch Q can contain, for example, the current total density plus a selected set of important orbital-pair densities. The factorization then stops when the worst member is certified. In principle the batch can be updated as an SCF or response calculation proceeds, with monotonicity ensuring that adding pivots never invalidates earlier guarantees.

8.2 Adaptive use over iterative electronic-structure workflows

The benchmark uses fixed target vectors, but the same factor can be expanded monotonically during an SCF calculation. A practical workflow begins with a loose rank, evaluates Ck(q0), and adds pivots until the initial density is certified. After the Fock update produces q1, the existing factor is retained. If Ck(q1) remains below ε, no new Coulomb directions are needed. If it exceeds ε, pivoting resumes from the stored residual diagonal and previous factors. Because Cholesky enrichment is nested, no previously computed column is discarded.

This nested behavior is preferable to repeatedly rebuilding auxiliary spaces when the requested accuracy changes. It also creates a natural two-stage schedule: early SCF iterations can use a tolerance εearly selected to keep numerical error below the current electronic residual, whereas the final iterations tighten to εfinal. The certificate is evaluated against the actual current density at both stages. Such inexact-iteration logic is common in numerical linear algebra, but RCPC gives an electronic-structure-specific observable with a strict Coulomb-energy interpretation.

For correlated methods, the set Q can be chosen strategically rather than exhaustively. A full collection of all orbital-pair densities may be too large for frequent certification, but pairs can be screened by occupation, locality, pair energy, or domain membership. The maximum certificate over the retained set is then rigorous for that set; it is not automatically a bound for omitted pairs. The distinction should be explicit in software APIs so that screening approximations are not conflated with the certified low-rank approximation itself.

8.3 Rank–tolerance response across density families

Figure 9 collects the certified-rank curves for the selected density-aware systems. As ε is tightened from 10^{-2} to 10^{-6} Eh, the median required rank fraction rises from about 0.563 to 0.958 for the population and polarized families. The transition family rises from about 0.313 to 0.792. The curves are smooth enough to support adaptive rank enlargement rather than a small number of hard-coded matrix thresholds.

The separation between density families persists across the entire tolerance range. At $\varepsilon=10^{-4}$ Eh, the median positive-density rank fraction is approximately 0.797, compared with 0.571 for the transition proxy. At $\varepsilon=10^{-5}$ Eh the corresponding medians are approximately 0.880 and 0.672. Thus the advantage of contraction awareness is not confined to a single loose tolerance; it remains visible as the requested guarantee tightens by several orders of magnitude.

A useful software consequence is that ε can be made an explicit user-facing accuracy parameter while τmin remains an internal numerical safeguard. Users generally reason more naturally about an energy budget than about a residual matrix diagonal. The mapping from ε to rank is then system- and density-dependent by design, which is precisely the adaptation that a fixed τ cannot provide.

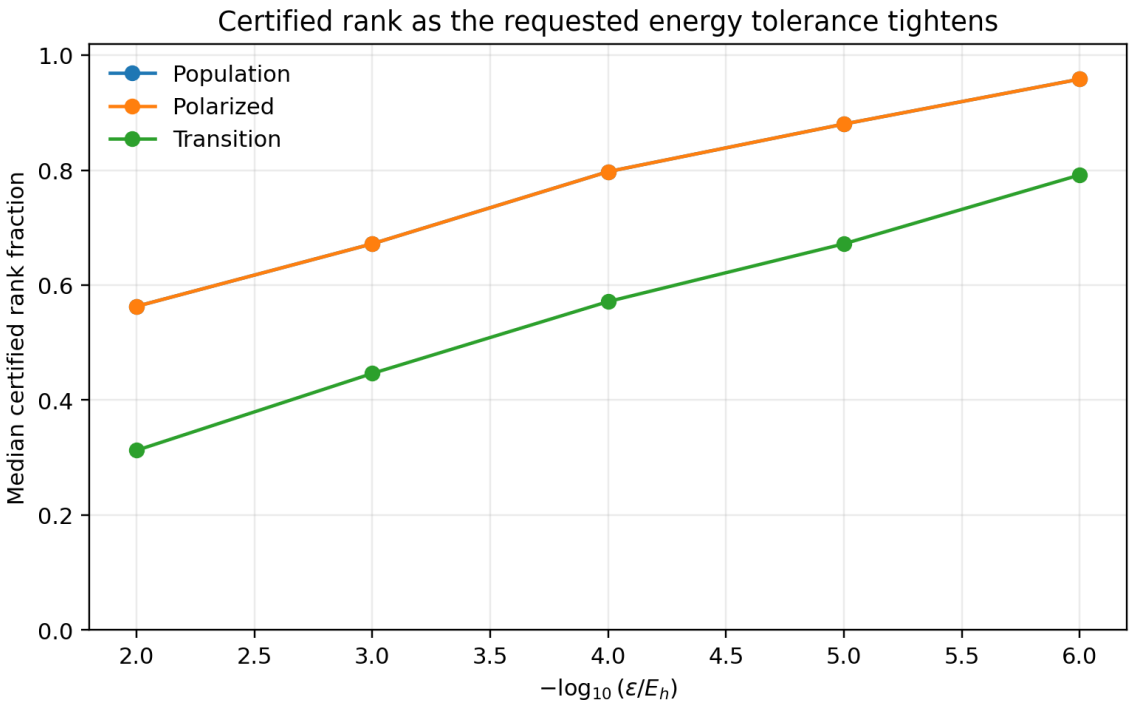


Fig. 9 Median certified rank fraction as a function of requested absolute Coulomb-energy tolerance for the density-aware benchmark subset

9 Relation to established RI/DF and Cholesky strategies

RCPC is not a replacement for optimized auxiliary basis sets, robust variational fitting, or integral Cholesky decomposition. It is an error-control layer that can sit on top of any construction producing a positive-semidefinite residual in the Coulomb metric. The method is closest in spirit to Cholesky schemes in which a single decomposition threshold controls ERI accuracy [1,7–13], but it changes the semantics of termination from "the residual matrix is small" to "the Coulomb contraction relevant to this calculation is certified to be small."

Optimized RI-J bases such as the Weigend Coulomb-fitting families achieve excellent accuracy with compact pretabulated spaces [16,17]. Automatic basis generation schemes use orbital products and pivoted decompositions to remove the need for hand optimization [19,20], while recent even-tempered constructions seek broad periodic-table transferability [21]. These approaches address how the candidate auxiliary space is generated. RCPC addresses a complementary question: given a candidate space or an on-the-fly Cholesky sequence, how much of it is needed for a specified density contraction?

Robust and variational fitting remains important when the fitted object enters an energy functional self-consistently. The present certificate controls the quadratic Coulomb error associated with a fixed coefficient vector and PSD metric residual. In a full RI-J SCF implementation, one would combine the certificate with the variational fitting equations rather than replace them. The certificate could be evaluated after each density update, with the auxiliary rank expanded only if the requested bound is violated. Such a scheme would naturally tighten or relax the active auxiliary space in response to changes in the density.

Strategy	Primary control	What is bounded/targeted	Strength	Main limitation
Optimized RI/DF basis	Preoptimized auxiliary functions	Basis-specific benchmark accuracy	Very fast production use; compact	Requires auxiliary basis families
Integral Cholesky	ERI/Coulomb pivot threshold	Residual matrix threshold	System adaptive; no fitted basis	Threshold not directly an energy tolerance
Automatic ABS generation	Orbital-product spanning criterion	Basis-generation accuracy	Transferable and automated	Generated space may still be larger than needed

Strategy	Primary control	What is bounded/targeted	Strength	Main limitation
RCPC (this work)	Residual diagonal + target density	Rigorous a posteriori Coulomb-energy bound	Density-aware; negligible extra cost	Requires PSD residual; full SCF integration not demonstrated

Table 6 Conceptual relationship between RCPC and established auxiliary-space strategies

10 Computational complexity and implementation considerations

If the full n×n Coulomb matrix is available, a rank-k pivoted Cholesky factorization costs $O(n^2k)$ in the straightforward dense form and stores O(nk) factors. In integral-direct settings, individual columns can be generated as needed, avoiding $O(n^2)$ storage. The certificate adds only O(n) work per check for a single density and O(mn) for a batch of m target densities. Since pivot-column formation and orthogonalization dominate for realistic integral engines, certificate evaluation is asymptotically and practically minor.

Several implementation choices can improve robustness. First, residual diagonals should be clamped to zero when tiny negative values arise from roundoff. Second, a separate machine-scale or linear-dependence cutoff should terminate pivots when dp is numerically meaningless. Third, certificate checks need not be performed after every pivot; batched checks every few pivots reduce overhead when m is large without changing correctness, provided termination is allowed only after a verified check. Fourth, for many densities the weighted sums can be evaluated as a matrix-vector operation using the vector $\sqrt{d}$.

The diagonal certificate has a parallel advantage: the residual diagonal is naturally distributed with the auxiliary index, and each density certificate requires a reduction of a scalar weighted sum. Thus the certification layer is compatible with domain-decomposed or distributed-memory Cholesky implementations. No global residual matrix is needed.

10.1 Cost model for repeated density certification

The asymptotic statement that certification costs O(n) per density can be sharpened for repeated workflows. Let cp denote the average cost of producing and orthogonalizing one new pivot column, cc the cost per auxiliary coefficient for one certificate evaluation, m the number of target densities, and h the number of pivots between certificate checks. If K new columns are generated, an approximate incremental work model is

$$W \approx K\, c_p + \mathrm{ceil}(K/h)\, m\, n\, c_c \tag{16}$$

where cp implicitly contains the integral and orthogonalization work appropriate to the implementation. In conventional direct Cholesky or three-index generation, cp is far larger than a handful of absolute-value, square-root, multiply, and reduction operations, so h=1 can already be affordable for a small m. For very large batches, h can be increased without compromising correctness: it merely means the factorization may take up to h-1 unnecessary pivots before discovering that the bound has crossed ε.

Memory overhead is minimal for the diagonal certificate. Besides the standard L factor and pivot metadata, RCPC stores d and optionally $\sqrt{d}$, both O(n). A batch of m densities already requires its coefficient data for the parent electronic-structure calculation. No $O(n^2)$ residual matrix is introduced. The block certificate of Eq. (12) changes this balance because principal residual blocks must be generated or retained: its block size should therefore be chosen to exploit data already resident in shell or atom domains.

The communication model is similarly favorable in distributed memory. If auxiliary indices are partitioned, each rank computes local sums $s_{a,\mathrm{local}}=\sum_{i\in\mathrm{local}}|q_{ai}|\sqrt{d_i}$. A collective sum gives s_a, and $C_a=s_a^2/2$. The payload is m floating-point numbers per certificate check, independent of n. Pivot selection commonly requires at least one global reduction already, so certification can often be fused with existing synchronization points.

10.2 Suggested stopping hierarchy

A robust implementation should distinguish accuracy termination from numerical termination. The recommended hierarchy is: (i) evaluate the requested density certificate; (ii) if $C_k(Q)\le\varepsilon$, terminate and label the factor energy-certified; (iii) otherwise test whether the next pivot falls below a machine or linear-dependence cutoff; (iv) if so, terminate as numerical-rank-limited and report the remaining certificate; (v) otherwise generate another column. An optional conventional τ threshold can be inserted before step (iii), but if it stops the factorization while $C_k(Q)>\varepsilon$ the result must not be presented as meeting the requested energy tolerance.

This distinction is particularly important in automated high-throughput calculations. A matrix-threshold-limited calculation may still be entirely adequate, but the software should expose the certificate so downstream workflows can decide whether to accept it, tighten the representation, or flag the case. The certificate therefore functions not only as a stopping rule but also as quantitative provenance for an approximate Coulomb contraction.

11 Scope, limitations, and required production validation

The numerical experiments deliberately use analytic s-type Gaussian charge probes rather than a complete orbital-basis ERI implementation. This choice gives an exact and transparent validation of the matrix inequalities, but it does not establish wall-clock speedups, total-energy errors, gradients, or SCF behavior for a production HF/DFT code. Angular-momentum channels, contracted Gaussian orbital products, near-linear dependencies specific to large orbital bases, and exchange fitting can alter rank distributions. Consequently, the reported rank fractions should not be transferred quantitatively to def2, correlation-consistent, numerical atomic-orbital, or relativistic auxiliary bases.

The present theory certifies the Coulomb quadratic form for specified coefficient vectors. It does not by itself certify exchange, correlation energies, or a fully relaxed self-consistent total energy. Exchange has different algebraic sign structure, and robust variational fitting of exchange must respect its own error functional [18]. Extending density-aware residual certification to positive-semidefinite reformulations of exchange-like contractions, MP2 intermediates, and tensor-hypercontracted representations is a separate problem.

Geometry derivatives also require care. Proposition 6 assumes a fixed pivot set in a local geometry interval. Re-pivoting can introduce discontinuities in the factor representation even when the approximated matrix remains continuous. Production analytic gradients should therefore use pivot hysteresis, subspace freezing over optimization steps, or a differentiable selection strategy. Numerical tests of forces and Hessians are necessary before geometry optimization or molecular dynamics applications.

Finally, a rigorous bound can be conservative. Equation (7) knows only the residual diagonal and q, not cancellation in off-diagonal residual elements. The observed certificate/error ratios are commonly several-fold to one order of magnitude. This is still useful because the user can specify a physically meaningful tolerance, but future work may seek tighter certificates from inexpensive block residual information, stochastic trace estimates, or selected off-diagonal elements while preserving rigor.

12 Conclusions

Residual-certified pivoted Coulomb compression converts a familiar positive-semidefinite matrix factorization into an a posteriori energy-control mechanism. The key observation is elementary but operationally powerful:

positivity of the Cholesky residual bounds each unknown off-diagonal residual by the geometric mean of two known diagonal residuals. The resulting certificate, $\frac{1}{2}(\Sigma_i|q_i|\sqrt{d_i})^2$, bounds the exact Coulomb-energy defect for the current density and costs only $O(n)$ to evaluate.

The formal analysis proves positive-semidefinite residual preservation, exact error representation, diagonal and trace bounds, monotonicity, simultaneous certification of multiple densities, and a local derivative estimate. Fixed-seed Gaussian Coulomb benchmarks confirm the bound across 16 molecule-shaped spaces and an alkane scaling sequence. They also expose a limitation of threshold-only control: relative compression may improve with system size even as an absolute Coulomb-energy error grows. Density-aware stopping directly addresses this mismatch by using the quantity that ultimately matters—the contracted energy tolerance.

The next step is integration into a production electronic-structure code with genuine AO-product auxiliary spaces, three-center integrals, SCF updates, and analytic gradients. Such a study should compare RCPC against optimized RI-J bases and standard Cholesky thresholds on total energies, atomization energies, dipole moments, and timing benchmarks. If the favorable low overhead and contraction-specific rank reduction persist, residual certification could provide a simple route toward error-controlled adaptive auxiliary spaces in HF, hybrid DFT, correlated wavefunction methods, and response theory.

Statements and Declarations

Funding. None.

Competing interests. None.

Data availability. No experimental or third-party numerical dataset was used. All reported benchmark values were generated from the molecular definitions, fixed random seed, Gaussian-probe exponent rule, and algorithms specified in Section 5. The tabulated results are contained in this manuscript.

Code availability. The numerical procedure is fully specified by Algorithm 1 and Eqs. (14)–(15). A submission-ready implementation should be archived by the author in a persistent repository before publication if required by the editor or reviewers.

Author contributions. C.N Conceptualization, methodology, formal analysis, software verification, visualization, writing, and final approval Connor Noble.

Appendix A. Additional derivations

A.1 Nyström equivalence for selected pivots

Let S denote the k selected pivot indices and assume G_{SS} is nonsingular on the selected subspace. The minimum-Coulomb-norm interpolation of all columns through S gives $\tilde{G} = G_{:S} G_{SS}^{-1} G_{S:}$. This is positive semidefinite and has rank at most k. Symmetric pivoted Cholesky spans the same selected-column subspace; in exact arithmetic there exists a nonsingular triangular transform between the Cholesky columns and $G_{:S}$. Consequently $L_k L_k^{T}$ and the Nyström representation coincide on the selected subspace. The residual is the Schur complement associated with eliminating S, which provides an alternative proof of Theorem 1.

$$\tilde{G} = G[:,S]\,(G[S,S])^{-1}\,G[S,:] \tag{A1}$$

A.2 Density-aware tolerance from a prescribed coefficient envelope

When q is not yet known, one can still derive a conservative pre-factorization condition from an envelope $|q_i| \le b_i$. Equation (7) gives $\Delta E \le \frac{1}{2}(\Sigma_i b_i\sqrt{d_i})^2$. If only a global l_1-norm bound $\|q\|_1 \le B$ is available, Corollary 3.1 yields the sufficient condition $\|d\|_\infty \le 2\varepsilon/B^2$. This relation can be used to translate a chemically meaningful target ε

into an initial matrix threshold, after which the sharper density-specific certificate can take over once q is available.

$$||d||_\infty \le 2\varepsilon/B^2 \;\Rightarrow\; \Delta E_J \le \varepsilon \tag{A2}$$

A.3 Multi-density matrix form

For a density matrix $Q=[q(1),\ldots,q(m)]$ and $s=\sqrt{d}$ interpreted elementwise, the vector of unsquared certificate amplitudes is $a=|Q|^T s$. All m certificates are then $\tfrac{1}{2}a\circ a$. This formulation maps naturally to BLAS-level matrix-vector kernels and accelerators. It also allows weights w_a to prioritize states or response channels by replacing the termination criterion $\max_a C(q(a))\le\varepsilon$ with $\max_a C(q(a))/w_a\le\varepsilon$ when different physical tolerances are appropriate.

$$a = |Q|^T \sqrt{d}, \quad C = \tfrac{1}{2}\, a\circ a \tag{A3}$$

Appendix B. Proof details and extremal constructions

B.1 Sharpness of the diagonal-only certificate

Consider a non-negative residual diagonal d and a coefficient vector q. Define $s_i=\mathrm{sign}(q_i)$, with an arbitrary sign assigned when $q_i=0$, and construct $R^\star=vv^T$ with $v_i=s_i\sqrt{d_i}$. Then R^* is positive semidefinite, $\mathrm{diag}(R^*)=d$, and $q^T R^* q=(\Sigma_i |q_i|\sqrt{d_i})^2$. Consequently Eq. (7) is attained exactly by at least one positive-semidefinite completion consistent with any prescribed diagonal. This proves a stronger statement than validity: without information beyond d, no uniformly smaller upper bound can hold for all positive-semidefinite residual matrices.

In actual Coulomb factorizations the residual rarely has this extremal rank-one correlation pattern, which explains why typical certificates exceed measured errors. Nevertheless, the extremal construction can be approached locally when residual directions are highly correlated and coefficient signs align. The approximately 1.0023 minimum certificate/error ratio in the resolved benchmark is consistent with this possibility and warns against calibrating away conservatism with a fixed multiplicative factor.

B.2 Bilinear residual contractions

Electronic-structure kernels frequently require bilinear rather than purely quadratic contractions. Let a and b be arbitrary coefficient vectors. Positive semidefiniteness and the elementwise inequality $|R_{ij}|\le\sqrt{d_i d_j}$ immediately give

$$|a^T R\, b| \le (\Sigma_i |a_i|\sqrt{d_i})(\Sigma_j |b_j|\sqrt{d_j}) = S(a)S(b) \tag{B1}$$

Equation (B1) can certify an off-diagonal Coulomb coupling between two fitted distributions. It also supplies the single step used in Theorem 7. For a batch of vectors stored as the columns of a matrix Q, all $S(q)$ values are obtained from one multiplication of $|Q|^T$ by $\sqrt{d}$. Thus certification of many bilinear products can reuse two scalar weights rather than form any residual matrix element.

If the quantity of interest is a symmetric bilinear energy $a^T G b$ entering a larger expression, the same residual bound applies directly to the low-rank approximation $a^T LL^T b$. Unlike the quadratic Coulomb self-energy, however, the bilinear defect need not have a fixed sign. The absolute-value statement in Eq. (B1), not the one-sided inequality of Theorem 2, is then the relevant guarantee.

Appendix C. Numerical stability and finite-precision safeguards

The exact-arithmetic proofs rely on R_k being positive semidefinite. In floating-point arithmetic, recursive diagonal updates may accumulate small negative values even when the mathematical residual is non-negative. RCPC therefore requires a numerical policy rather than silently applying a square root to a negative residual. The

benchmark uses $d \leftarrow \max(d - \ell \circ \ell, 0)$ componentwise after every accepted pivot. A production code should additionally compare the magnitude of a negative update with a scale such as machine epsilon times the initial diagonal norm; a large negative value indicates loss of stability or an inconsistency in integral generation rather than harmless roundoff.

A second safeguard concerns the pivot denominator. If d_p is extremely small, forming ℓ from division by $\sqrt{d_p}$ amplifies integral noise. Standard Cholesky linear-dependence thresholds remain appropriate and are logically separate from the energy certificate. A calculation can therefore terminate in one of three states: energy-certified, matrix-threshold-limited, or numerical-rank-limited. Software should report which state occurred. Calling a result "certified to ε" is justified only in the first state or when the certificate itself has independently fallen below ε before a numerical cutoff is encountered.

For very tight tolerances, evaluating the exact validation defect as $EJ - \bar{E}J$ suffers catastrophic cancellation. This affects only retrospective verification, not the certificate, which is built from non-negative residual diagonals. If an exact residual matrix is available for testing, $q^T R q$ is numerically preferable to subtracting energies, although it too becomes noise-sensitive when R is represented as $G - LL^T$ after near-complete factorization. Higher precision or compensated arithmetic may be used in unit tests designed to probe the final ulps of the factorization.

Appendix D. Production integration blueprint

A minimal RI-J integration can preserve the conventional auxiliary basis and apply RCPC only to a secondary Coulomb-metric compression. The code forms or accesses the Coulomb metric, begins a pivoted factorization, and maps the current fitted-density coefficients into the same auxiliary coordinate convention used by the residual. The certificate is then evaluated after selected pivots. When $C_k(q) \le \varepsilon J$, the active factor is frozen for that SCF iteration and used for the Coulomb contraction. If the next density violates the tolerance, factorization resumes from the existing rank.

For integral Cholesky methods the same logic applies directly to the decomposed Coulomb object, provided the residual diagonal corresponds to the quadratic form being certified. Care is needed when factors are transformed, truncated, localized, or reorthogonalized after Cholesky construction. Any postprocessing that changes LL^T changes the residual and invalidates a certificate based on the old d unless the transformation is proven to leave the represented matrix invariant. This bookkeeping requirement is straightforward but essential for a rigorous implementation.

Parallel codes can distribute d and the rows of L by auxiliary index. For one density q, each process computes a local partial sum $\Sigma_{\text{local}} |q_i| \sqrt{d_i}$ and the global certificate requires one scalar reduction. For m densities the reduction contains m numbers. If pivot selection already requires a global maximum reduction, certification does not introduce a qualitatively new communication primitive. Block certificates require additional local quadratic forms but can often be confined to blocks owned by one process.

A useful testing hierarchy for future software is: (i) unit tests on synthetic positive-semidefinite matrices with known extremal cases; (ii) analytic Gaussian charge matrices as used here; (iii) real auxiliary Coulomb metrics for standard orbital/auxiliary basis pairs; (iv) fixed-density RI-J energy comparisons; (v) self-consistent HF and hybrid-DFT total energies; and (vi) analytic gradient and geometry tests. At every level, the measured Coulomb defect should be checked against the reported certificate independently of any total-energy cancellation. This hierarchy isolates mathematical correctness from chemistry-specific performance.

Appendix E. Tolerance policies for practical calculations

E.1 Absolute, per-particle, and per-fragment budgets

An absolute ε is the cleanest interpretation of the theorem, but different applications may wish to scale the allowed error with system size. Three policies are mathematically compatible with RCPC. An absolute policy fixes ε_J for the entire Coulomb contraction. A per-particle policy selects $\varepsilon_J = N\eta$, where N may be an electron, atom, or basis-function count and η is the permitted error per unit. A per-fragment policy selects $\varepsilon_J = \Sigma_f \varepsilon_f$ for a predefined fragmentation. In every case the certificate comparison itself is unchanged; only the externally supplied budget varies.

The choice is scientific rather than algebraic. Absolute tolerances facilitate direct comparison of total energies of similarly sized systems. Per-particle tolerances can avoid unnecessary tightening in large homogeneous systems. Fragment budgets may be more natural for local correlation or embedding calculations. What RCPC contributes is a transparent mapping from that policy to an actual low-rank termination condition. It does not prescribe which chemical error budget is appropriate for a particular observable.

E.2 Energy differences and balanced representations

Most chemically relevant quantities are energy differences. Suppose two fixed densities q_A and q_B are evaluated with possibly different certified factors and have Coulomb defects δ_A and δ_B satisfying $0 \le \delta_A \le \varepsilon_A$ and $0 \le \delta_B \le \varepsilon_B$. The error in the Coulomb contribution to the difference is $|\delta_A - \delta_B| \le \max(\varepsilon_A, \varepsilon_B)$ if both one-sided errors share the same sign convention, and certainly no larger than $\varepsilon_A + \varepsilon_B$. Using a common factor for both states and including $\{q_A, q_B\}$ in Q can improve balance because the same residual space is controlled simultaneously.

For geometry scans or reaction paths, a common or nested pivot sequence is therefore attractive. One may initialize the factor from a reference geometry, freeze the pivot identities over a local segment, and enrich the rank when any state violates its certificate. Such a policy does not by itself prove smooth forces across pivot changes, but it separates the controllable low-rank energy error from the additional differentiability problem identified in Proposition 6.

E.3 Relationship to integral screening

Integral screening and RCPC control different approximations. Screening discards or bounds small primitive, shell-pair, or transformed integrals before or during factor construction; RCPC certifies the residual of the matrix actually supplied to the factorization. If screened integral generation itself carries an error, a rigorous total bound must include that screening contribution. One safe composition is additive: if integral generation is certified within ε_{int} for the relevant quadratic form and RCPC is certified within ε_{rcpc}, then the combined error is at most $\varepsilon_{int} + \varepsilon_{rcpc}$ by the triangle inequality.

This modular view is useful for production architecture. Each approximate kernel can return both a value and an error budget. The SCF or correlated driver can then enforce a global numerical-accuracy policy without assuming that one universal threshold has the same meaning in every module. RCPC supplies such a value-plus-bound interface for the positive-semidefinite Coulomb low-rank step.

Appendix F. Reproducible benchmark specification

To make the controlled benchmark independently reconstructible, the essential specification is collected here. The molecular set is water, ammonia, methane, methanol, formaldehyde, ethene, acetylene, benzene, pyridine, phenol, glycine, alanine, uracil, caffeine, naphthalene, and anthracene. Three-dimensional coordinates are generated from the corresponding molecular graphs with explicit hydrogens using ETKDGv3 and fixed random seed 20260822, followed by UFF relaxation. The scaling set consists of linear alkanes C5, C10, C20, C40, and C80 with explicit hydrogens.

Each atom A carries eight normalized s-type Gaussian charge distributions centered on its Cartesian position. Exponents are $\alpha_{A,m} = 0.18\, Z_A^{0.55}$ fm with $f = \{0.20, 0.32, 0.50, 0.78, 1.22, 1.90, 2.95, 4.60\}$. Pair Coulomb interactions are evaluated analytically using Eq. (14), with the coincident-center limit in Eq. (15). Matrices are symmetrized in double

precision. The pivot is always the largest current residual diagonal, with deterministic tie behavior inherited from the first maximum index.

The positive population coefficient family distributes ZA across the eight atom-centered probes with weights proportional to $\alpha^{-1/2}$ and normalized on each atom. The polarized family applies a small deterministic geometry-dependent redistribution while preserving total coefficient sum. The signed transition proxy uses a deterministic oscillatory atomic pattern, removes its mean charge, expands it over the probes, and scales its total magnitude with the square root of total nuclear charge. These vectors are numerical test contractions, not claims about physical electron densities.

Matrix-threshold experiments use relative maximum-diagonal cutoffs 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}, and 10^{-10}. Density-aware experiments request absolute certificates 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, and 10^{-6} Eh for benzene, glycine, caffeine, and anthracene. Reported dense defects are validation diagnostics only; an operational RCPC implementation needs the factor, residual diagonal, and coefficient vector, not the dense residual.

References

1. Beebe NHF, Linderberg J (1977) Simplifications in the generation and transformation of two-electron integrals in molecular calculations. Int J Quantum Chem 12:683–705. https://doi.org/10.1002/qua.560120408
2. Vahtras O, Almlöf J, Feyereisen MW (1993) Integral approximations for LCAO-SCF calculations. Chem Phys Lett 213:514–518. https://doi.org/10.1016/0009-2614(93)89151-7
3. Feyereisen M, Fitzgerald G, Komornicki A (1993) Use of approximate integrals in ab initio theory. An application in MP2 energy calculations. Chem Phys Lett 208:359–363. https://doi.org/10.1016/0009-2614(93)87156-W
4. Dunlap BI, Connolly JWD, Sabin JR (1979) On some approximations in applications of Xα theory. J Chem Phys 71:3396–3402. https://doi.org/10.1063/1.438728
5. Dunlap BI (2000) Robust variational fitting: Gáspár's variational exchange can accurately be treated analytically. J Mol Struct THEOCHEM 501–502:221–228. https://doi.org/10.1016/S0166-1280(99)00433-9
6. Dunlap BI, Rösch N, Trickey SB (2010) Variational fitting methods for electronic structure calculations. Mol Phys 108:3167–3180. https://doi.org/10.1080/00268976.2010.518982
7. Koch H, Sánchez de Merás A, Pedersen TB (2003) Reduced scaling in electronic structure calculations using Cholesky decompositions. J Chem Phys 118:9481–9484. https://doi.org/10.1063/1.1578621
8. Aquilante F, Pedersen TB, Lindh R (2007) Low-cost evaluation of the exchange Fock matrix from Cholesky and density fitting representations of the electron repulsion integrals. J Chem Phys 126:194106. https://doi.org/10.1063/1.2736701
9. Aquilante F, Lindh R, Pedersen TB (2007) Unbiased auxiliary basis sets for accurate two-electron integral approximations. J Chem Phys 127:114107. https://doi.org/10.1063/1.2777146
10. Aquilante F, Gagliardi L, Pedersen TB, Lindh R (2009) Atomic Cholesky decompositions: a route to unbiased auxiliary basis sets for density fitting approximation with tunable accuracy and efficiency. J Chem Phys 130:154107. https://doi.org/10.1063/1.3116784
11. Røeggen I, Johansen T (2008) Cholesky decomposition of the two-electron integral matrix in electronic structure calculations. J Chem Phys 128:194107. https://doi.org/10.1063/1.2925269
12. Weigend F, Kattannek M, Ahlrichs R (2009) Approximated electron repulsion integrals: Cholesky decomposition versus resolution of the identity methods. J Chem Phys 130:164106. https://doi.org/10.1063/1.3116103
13. Pedersen TB, Lehtola S, Fdez Galván I, Lindh R (2024) The versatility of the Cholesky decomposition in electronic structure theory. WIREs Comput Mol Sci 14:e1692. https://doi.org/10.1002/wcms.1692
14. Whitten JL (1973) Coulombic potential energy integrals and approximations. J Chem Phys 58:4496–4501. https://doi.org/10.1063/1.1679012
15. Baerends EJ, Ellis DE, Ros P (1973) Self-consistent molecular Hartree–Fock–Slater calculations I. The computational procedure. Chem Phys 2:41–51. https://doi.org/10.1016/0301-0104(73)80059-X
16. Weigend F (2002) A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency. Phys Chem Chem Phys 4:4285–4291. https://doi.org/10.1039/B204199P
17. Weigend F (2006) Accurate Coulomb-fitting basis sets for H to Rn. Phys Chem Chem Phys 8:1057–1065. https://doi.org/10.1039/B515623H

18. Reine S, Tellgren E, Krapp A, Kjaergaard T, Helgaker T, Jansik B, Høst S, Salek P (2008) Variational and robust density fitting of four-center two-electron integrals in local metrics. J Chem Phys 129:104101. https://doi.org/10.1063/1.2956507
19. Lehtola S (2021) Straightforward and accurate automatic auxiliary basis set generation for molecular calculations with atomic orbital basis sets. J Chem Theory Comput 17:6886–6900. https://doi.org/10.1021/acs.jctc.1c00607
20. Lehtola S (2023) Automatic generation of accurate and cost-efficient auxiliary basis sets. J Chem Theory Comput 19:6242–6254. https://doi.org/10.1021/acs.jctc.3c00670
21. Díaz-Tinoco M, Flores-Moreno R, Zúñiga-Gutiérrez BA, Köster AM (2025) Automatic generation of even-tempered auxiliary basis sets with shared exponents for density fitting. J Chem Theory Comput 21:2338–2352. https://doi.org/10.1021/acs.jctc.4c01555
22. Antonini N, Ronca E, Storchi L, Belpassi L (2025) Automatic generation of density-fitting auxiliary basis sets for all-electron Dirac–Kohn–Sham calculations. J Phys Chem A 129:6930–6941. https://doi.org/10.1021/acs.jpca.5c02772
23. Polly R, Werner H-J, Manby FR, Knowles PJ (2004) Fast Hartree-Fock theory using local density fitting approximations. Mol Phys 102:2311–2321. https://doi.org/10.1080/00268970420002748O1
24. Sodt A, Subotnik JE, Head-Gordon M (2006) Linear scaling density fitting. J Chem Phys 125:194109. https://doi.org/10.1063/1.2370949
25. Mejía-Rodríguez D, Köster AM (2014) Robust and efficient variational fitting of Fock exchange. J Chem Phys 141:124114. https://doi.org/10.1063/1.4896199
26. Hamel S, Casida ME, Salahub DR (2001) Assessment of the quality of orbital energies in resolution-of-the-identity Hartree–Fock calculations using deMon auxiliary basis sets. J Chem Phys 114:7342–7350. https://doi.org/10.1063/1.1358865