

Accelerating Quantum Computations of Chemistry Through Regularized Compressed Double Factorization

Oumarou Oumarou¹, Maximilian Scheurer¹, Robert M. Parrish², Edward G. Hohenstein², and Christian Gogolin¹

¹ Covestro Deutschland AG, Leverkusen 51373, Germany

² QC Ware Corporation, Palo Alto, CA 94301, USA

06/04/2024

We propose the regularized compressed double factorization (RC-DF) method to classically compute compressed representations of molecular Hamiltonians that enable efficient simulation with noisy intermediate scale (NISQ) and error corrected quantum algorithms. We find that already for small systems with 12 to 20 qubits, the resulting NISQ measurement scheme reduces the number of measurement bases by roughly a factor of three and the shot count to reach chemical accuracy by a factor of three to six compared to truncated double factorization (DF) and we see order of magnitude improvements over Pauli grouping schemes. We demonstrate the scalability of our approach by performing RC-DF on the Cpd I species of cytochrome P450 with 58 orbitals and find that using the resulting compressed Hamiltonian cuts the run time of qubitization and truncated DF based error corrected algorithms almost in half and even outperforms the lambda parameters achievable with tensor hypercontraction (THC) while at the same time reducing the CCSD(T) energy error heuristic by an order of magnitude.

The run time of algorithms on most of today’s noisy intermediate scale (NISQ) hardware platforms is largely independent of the anyway shallow circuit depth. Instead, it is mostly a function of the number of distinct circuits that need to be evaluated and the number of repeated circuit executions because re-programming the quantum device to perform a different circuit and the time for measurement and qubit reset dominate the

Robert M. Parrish: rob.parrish@qcware.com

Christian Gogolin: christian.gogolin@covestro.com

overall circuit execution time [1].

Many NISQ algorithms such as the variational quantum eigensolver (VQE) [2] are essentially methods to reduce circuit depth at the expense of requiring many repetitions, also called shots. In a similar fashion, error mitigation techniques [3, 4] create a tolerance for the noise of NISQ devices by further increasing the number of shots needed to obtain a final result. The total number of shots is thus often the limiting factor on the path towards quantum advantage.

This is a particularly pressing issue in quantum chemistry simulation. Here, molecular Hamiltonians, which in their second-quantized form have $O(n^4)$ terms, where n is the number of spatial orbitals, need to be measured with very high accuracy. Naive measurement schemes require an extremely fast growing number of distinct observables and total number of shots [5] for reaching the required accuracy.

Methods to cope with this problem fall in two broad classes. First, methods [6, 7, 8] which, starting from a decomposition of the observable into Pauli operators, group or otherwise combine these Pauli operators into sets that are jointly measurable with no or only minimal increase in circuit depth. Second, methods which yield a compressed and possibly approximate representation of the original Hamiltonian in the form of a tensor contraction. For fermionic second-quantized Hamiltonians, these are mainly density fitting [9], tensor hypercontraction (THC) [10, 11, 12], and double factorization (DF) [13, 14] (see [14] for a comparison). While the Pauli grouping methods are applicable to general qubit Hamiltonians, the second class of methods typically yields better performance when applicable [5].

arXiv:2212.07957v3 [quant-ph] 4 Jun 2024

These compressed representations also enable drastic resource reductions in leading fault tolerant algorithms for the simulation of chemistry based on linear combinations of unitaries (LCU) and qubitization [15, 12, 16, 17, 11]. Here run time is mainly a function of the so-called lambda parameter. Its precise definition depends on the algorithm and will be discussed later, but it can be thought of as a norm-like quantity that depends on the magnitude of the coefficients of the representation of the Hamiltonian. THC typically yields lower lambda parameters than existing DF schemes. The fact that some tensors in the THC decomposition are non-square and non-unitary causes other overheads and complications [17] which does not make THC a viable option for typical NISQ quantum algorithms.

In contrast, explicit double factorization (X-DF) and compressed double factorization (C-DF) [18, 19, 20, 21, 22] naturally yield a NISQ-friendly measurement scheme that only requires a linear depth orbital/Givens rotation circuit before the final measurements, is compatible with particle number post selection, and also an LCU representation of the Hamiltonian suitable for error corrected algorithms based on qubitization [23].

The X-DF measurement scheme reduces the number of distinct measurement bases to at most $n(n+1)/2$ and drastically decreases the number of shots to reach a target accuracy, when compared to Pauli-based schemes. The number of bases can be further reduced by truncating the X-DF representation of the Hamiltonian, thereby making the representation approximate. This can reduce the required number of shots, but the error resulting from the now approximate representation of the Hamiltonian quickly outweighs this. C-DF is designed to overcome this issue by performing a tighter least-squares numerical tensor fitting of the molecular Hamiltonian to truncated double-factorized form. By lifting a rank constraint in the equation defining the X-DF Hamiltonian and using the resulting additional freedom to improve the representation of the molecular Hamiltonian by means of parameter optimization starting from a truncated X-DF guess, it achieves lower approximation errors than truncated X-DF. However, when attempting practical deployment of C-DF in the context of quantum algorithms, one encounters an additional major barrier: the optimization of the C-

DF tensor fitting to minimize least squares error does not consider the variance properties of the resulting representation. In practice, this means that the variance of the resulting energy estimator can erratically fluctuate and can be orders of magnitude higher than the variance of the X-DF energy estimator and the approximation error of both X-DF and C-DF.

In this work we propose the regularized compressed double factorization method (RC-DF) to fix this. RC-DF uses the same functional form of the compressed Hamiltonian as C-DF but it adds a regularization term to the C-DF cost function that is used when optimizing the parameters of the compressed representation¹. The regularization term stabilizes the optimization and reduces the variance of the resulting NISQ energy estimator as well as the λ parameter determining the resources of fault tolerant quantum algorithms. We find that RC-DF consistently outperforms both previous double factorization schemes in terms of variance, approximation error, and lambda parameter and even yields lambda parameters lower than THC.

1 Comparison of factorization methods

We start from the well known form of the second-quantized electronic structure Hamiltonian

$$\hat{H} = E_c + \sum_{pq} (p|\hat{h}_c|q)\hat{E}_{pq} + \frac{1}{2} \sum_{pqrs} (pq|rs) (\hat{E}_{pq}\hat{E}_{rs} - \delta_{qr}\hat{E}_{ps}), \quad (1)$$

where

$$(p|\hat{h}_c|q) = \int \phi_p^*(r) \left(-\frac{1}{2} \nabla^2(r) - \sum_m \frac{Z_m}{r - r_m} \right) \phi_q(r) dr$$

$$(pq|rs) = \iint \phi_p^*(r_1) \phi_q(r_2) \frac{1}{r_{12}} \phi(r_1)_r^* \phi(r_2)_s dr_1 dr_2$$

are the symmetric one-electron integrals and the real and 8-fold symmetric two-electron integrals with Z_m and r_m the charges and positions of the

¹When working out the implications of RC-DF for fault tolerant quantum algorithms, we became aware that a similar erratic behavior of the λ parameter of THC had been observed in [11] and an L1 regularization has been proposed as a cure there.

nuclei and ϕ the spacial molecular orbitals, and $\hat{E}_{pq} := \hat{p}^\dagger \hat{q} + \hat{p} \hat{q}^\dagger$ is the singlet excitation operator. The exact X-DF representation of the Hamiltonian is determined by diagonalizing the modified one-electron integrals tensor \mathcal{F}_{pq} and doubly diagonalizing the two-electron integrals tensor to obtain

$$\mathcal{F}_{pq} := (p|\hat{h}_c|q) - \frac{1}{2} \sum_r (pr|qr) + \sum_r (pq|rr) \quad (2)$$

$$= \sum_k U_{pk}^\varnothing \mathcal{F}_k^\varnothing U_{qk}^\varnothing \quad (3)$$

and

$$(pq|rs) = \sum_{t=1}^{n_t} V_{pq}^t g_t V_{rs}^t \quad (4)$$

$$= \sum_{t=1}^{n_t} \sum_{kl=1}^n U_{pk}^t U_{qk}^t Z_{kl}^t U_{rl}^t U_{sl}^t, \quad (5)$$

where the U_{pk}^t result from diagonalizing the $V_{pq}^t = \sum_k U_{pk}^t \Lambda_k U_{qk}^t$ and consequently $Z_{kl}^t = \Lambda_k g_t \Lambda_l$ is, for every t , a symmetric outer product, hence of rank one, and the U_{pk}^t are unitary (in fact without loss of generality special orthogonal). The second factorization is possible whenever $(pq|rs)$ is real and 8-fold symmetric (as is always the case for non-relativistic Coulomb repulsion integrals), as this is enough to ensure that the V_{pq}^t are not only orthogonal but also real and symmetric for every t (see Lemma 1 in Appendix D). With n_t equal to the maximum number $n(n+1)/2$ of non-zero eigenvalues of $(pq|rs)$ the Hamiltonian can then be written exactly (see Appendix F for the full derivation) as

$$\begin{aligned} \hat{H} = \mathcal{E} - \frac{1}{2} \sum_k \mathcal{F}_k^\varnothing U^\varnothing (\hat{Z}_k + \hat{Z}_{\bar{k}}) (U^\varnothing)^\dagger \\ + \frac{1}{8} \sum_{t=1}^{n_t} \sum_{kl=1}^n Z_{kl}^t U^t (\hat{Z}_k \hat{Z}_l - \delta_{kl} + \hat{Z}_k \hat{Z}_{\bar{l}} + \hat{Z}_{\bar{k}} \hat{Z}_l + \hat{Z}_{\bar{k}} \hat{Z}_{\bar{l}} - \delta_{\bar{k}\bar{l}}) (U^t)^\dagger, \end{aligned} \quad (6)$$

where

$$\mathcal{E} = E_c + \sum_p (p|\hat{h}_c|p) + \frac{1}{2} \sum_{pq} (pp|qq) - \frac{1}{4} \sum_{pq} (pq|pq), \quad (7)$$

is independent of the state, and U^\varnothing and U^t rotate the orbitals for each t according to U_{pk}^t (See Fig. 1), $\hat{Z}_k, \hat{Z}_{\bar{k}}$ are respectively pauli operators on qubit $2k$ and $2k+1$, $k \in [0, n-1]$.

If the sum over t is ordered according to $|g_t|$, the Hamiltonian can be approximated with a truncated X-DF representation with fewer terms (also called leaves).

In (R)C-DF the rank-one constraint on the Z_{kl}^t is lifted and they are allowed to be arbitrary symmetric matrices. The orbital rotations U_{pk}^t and coefficients Z_{kl}^t are then obtained using a two-step gradient based optimization procedure by first exponentially parametrizing the orbital rotations $U_{pq}^t := \exp(X^t)_{pq}$ via anti-symmetric generators X_{pq}^t and then minimizing the squared Frobenius norm (in [16] this is called the incoher-

ent error)

$$\frac{1}{2} \left\| \underbrace{\sum_{t=1}^{n_t} \sum_{kl=1}^n U_{pk}^t U_{qk}^t Z_{kl}^t U_{rl}^t U_{sl}^t - (pq|rs)}_{\Delta_{pqrs}} \right\|_{\mathcal{F}}^2 \quad (8)$$

of the difference between the left and right hand side of (5) for some pre-set $n_t \leq n(n+1)/2$ starting from a truncated X-DF initial guess (for details see [18]).

Irrespective of whether a Hamiltonian representation of the form (6) was found via X-DF or (R)C-DF, the energy can be then be measured by means of quantum circuit with a linear gate depth overhead of the form shown in Fig. 1.

The most favorable resource estimates of quantum algorithms for error corrected simulation of chemistry with qubitization have been obtained with THC [17, 11]. In this context THC approximates the two-body part of the Hamiltonian ac-

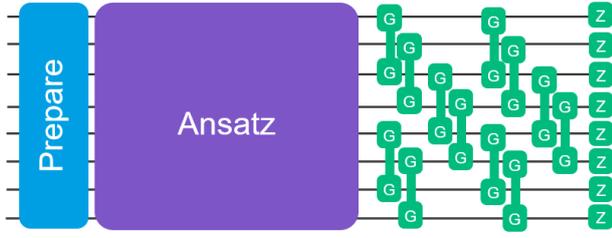


Figure 1: DF measurement scheme according to the LCU decomposition in (6). From each U^\emptyset and U^t the parameters of a square shaped fabric of givens gates G can be computed. The results of \hat{Z} and $\hat{Z} \otimes \hat{Z}$ measurements in these $n_t + 1$ distinct bases can then be contracted against the \mathcal{F}_k^\emptyset and Z_{kl}^t tensors to obtain an energy estimator.

cording to

$$(pq|rs) \approx \sum_{kl=1}^M \chi_p^k \chi_q^k \zeta_{kl} \chi_r^l \chi_s^l, \quad (9)$$

with $\sum_k^M \chi_p^k \chi_q^k \leq 1$ and $M \leq n^2$ the THC rank. Also here the symmetric ζ_{kl} and rectangular χ_p^k are found by means of minimizing the Frobenius norm error. When DF is used with qubitization [12, 16] it is usually presented and performed according to

$$\begin{aligned} (pq|rs) &= \sum_{t=1}^{n^2} L_{pq}^t L_{rs}^t \quad (10) \\ &= \frac{1}{2} \sum_{t=1}^{n^2} U^t \left(\sum_k \lambda_k^t (\hat{Z}_k + \hat{Z}_{\bar{k}}) \right)^2 (U^t)^\dagger, \quad (11) \end{aligned}$$

where the L_{pq}^t can be found with Cholesky decomposition or eigen decomposition as $L_{pq}^t = \sqrt{g_t} V_{pq}^t$ and the scalar $\lambda_k^t = \sqrt{g_t} \Lambda_k$ are the eigenvectors and the U^t the diagonalizing unitaries of the L_{pq}^t (here g_t , V_{pq}^t , and Λ_k refer to the quantities introduced in the context of X-DF above). For a NISQ measurement scheme such as that in Fig. 1 it makes no sense to partially discard a leaf (because the measurement data is available for all contributions from a leaf), but in qubitization truncation can be done on the level of setting individual λ_k^t equal to zero.

The (R)C-DF form of the Hamiltonian, in which Z_{kl}^t is no longer rank one, can be re-cast as a sum of operator squares similar to (10). By taking the matrix square root $W_{kl}^t := (\sqrt{Z^t})_{kl}$ so that $Z_{kl}^t = \sum_i^n W_{ki}^t W_{li}^t$ (we use the implementation in `scipy` of the algorithm from [24], which

works also for non-positive Z_{kl}^t matrices, in which case the W_{ki}^t come out complex, which is compatible with the scheme from [16]) one can write

$$\begin{aligned} (pq|rs) &= \sum_t^{n_t} \sum_i^n \left(\sum_k U_{pk}^t U_{qk}^t W_{ki}^t \right) \quad (12) \\ &\quad \times \left(\sum_l W_{il}^t U_{rl}^t U_{sl}^t \right). \end{aligned}$$

This allows one to run the algorithm of [16] with (R)C-DF Hamiltonians as input.

The block encoding of the qubitization method needs the Hamiltonian in the form of an LCU. The number of ancillary qubits and T gates is then determined by the number of terms of the LCU and a sort of normalization factor, called the lambda parameter [16, 17]. The result of (truncated) X-DF, C-DF, and RC-DF is itself an LCU with a lambda factor

$$\lambda_{\text{DF}}^{\text{LCU}} := \sum_k |\mathcal{F}_k^\emptyset| + \sum_{t=1}^{n_t} \left(\sum_{k<l} |Z_{kl}^t| + \frac{1}{4} \sum_k |Z_{kk}^t| \right). \quad (13)$$

Alternatively, because of (12), one can use the algorithm from [16], which achieves a contribution from the two-body part of the Hamiltonian to lambda of $1/4 \sum_t \| (L_{pq}^t)_{pq} \|_1^2$ for Hamiltonians of the form (10) (where $\| \cdot \|_1$ is the Schatten 1-norm) and using (12) one can thus obtain

$$\begin{aligned} \lambda_{\text{DF}}^{\text{Burg}} &:= \sum_k |\mathcal{F}_k^\emptyset| + \frac{1}{4} \sum_t^{n_t} \sum_i^n \left\| \left(\sum_k U_{pk}^t U_{qk}^t W_{ki}^t \right)_{pq} \right\|_1^2 \quad (14) \\ &= \sum_k |\mathcal{F}_k^\emptyset| + \frac{1}{4} \sum_t^{n_t} \sum_i^n \left(\sum_k |W_{ki}^t| \right)^2. \quad (15) \end{aligned}$$

The difference between $\lambda_{\text{DF}}^{\text{Burg}}$ and $\lambda_{\text{DF}}^{\text{LCU}}$ stems from the different LCU representation of the Hamiltonian. The latter uses the equation in (6) which is evidently an LCU and subsequently its $\| \cdot \|_1$ is (13). The former however uses the algorithm and LCU from (13) of [16]. Finally, for THC, Lee et al. [17] have obtained a lambda of

$$\lambda_{\text{THC}}^{\text{Lee}} := \sum_k |\mathcal{F}_k^\emptyset| + \frac{1}{2} \sum_{kl}^M |\zeta_{kl}|. \quad (16)$$

The precise run times of the algorithms corresponding to the different lambda values differ

and depend on factorization-specific quantities such as the THC rank M but they all scale like their respective lambda divided by the allowable phase estimation energy error times the sum of run times of certain circuit primitives plus logarithmic overheads. The differences between the lambda values have turned out to outweigh the influence of other factors when comparing algorithm run times for similar overall target accuracies [17].

2 Regularized compressed double factorization

While C-DF allows to reduce the number of leafs needed for good accuracy from close to n^2 for X-DF to roughly linear in n while maintaining an approximate but sufficiently accurate representation of the Hamiltonian, it turns out that the optimization of C-DF often converges to Z_{kl}^t tensors with very large entries. This is problematic since the variance of the NISQ energy estimator and both lambda parameters grow with the number and magnitude of $|Z_{kl}^t|$ values. To solve this issue, we propose to add to the C-DF cost function from (8) a regularization term penalizing large $|Z_{kl}^t|$ via a tensor of weights $\rho_{tkl} \geq 0$

$$\frac{1}{2} \|\Delta_{pqrs}\|_{\mathcal{F}}^2 + \sum_{tkl} \rho_{tkl} |Z_{kl}^t|^2, \quad (17)$$

We have tested both weighted L1 and L2 regularizations of the form $(\sum_{tkl} \rho_{tkl} |Z_{kl}^t|^\gamma)^{\frac{2}{\gamma}}$ with $\gamma \in 1, 2$ but concentrate on L2 regularization with uniform regularization strength $\rho_{tkl} = \rho$ in the rest of the main text.

As in C-DF, a joint optimization of the U_{pq}^t and Z_{pq}^t has unfavorable performance also with regularization, but the two-step optimization of C-DF proposed in [18] can be adopted to the regularized case. Further, for large n , a very expensive 6-index matrix inversion can be circumvented by carrying it out in a matrix-free manner with, e.g., a conjugate gradient algorithm (for details see Appendix A). We have found that this step benefits from the regularization (L2), as it improves the conditioning of the matrix. In RC-DF, initialization can be done either from X-DF truncated to the target number of C-DF leafs or one can start from the full X-DF factorization and put a high penalty on the leafs that are to

be truncated in the end. In practice, contrary to the difficulties reported in [11, 17] on converging THC, RC-DF seems to be rather well behaved. Convergence may take thousands of iterations, but we had no difficulty converging RC-DF in large active spaces to much tighter residual Frobenius norm errors (8) and coupled cluster with singles, doubles, and perturbative triples (CCSD(T)) energy errors than those reported for THC [11] (see Appendix A for further details of the optimization procedure).

3 Numerical results

Unless otherwise explicitly stated all results shown in the following were obtained with L2 regularization starting from a truncated X-DF guess and with a uniform regularization factor for all n_t leafs of $\rho_{tkl} =: \rho$.

We first investigate the advantages of RC-DF over other NISQ measurement schemes. The performance of any such scheme is determined by both the systematic error introduced in case the Hamiltonian is approximated and the variance, Var, of the estimator. We quantify the overall performance by the mean squared error and take the square root to obtain a quantity that has units of energy $\sqrt{\text{MSE}} := \sqrt{\langle \hat{H} - \hat{H}' \rangle^2 + \text{Var}}$, where \hat{H}' is the compressed Hamiltonian obtained with the respective flavor of double factorization and for the Pauli grouping based schemes $\hat{H}' = \hat{H}$.

The variance further depends on the state, the overall shot budget, and the shot distribution. In the main text we show data for the state being the complete active space configuration interaction (CASCI) ground state, but the plots look very similar for representative states along a VQE optimization trajectory. We consider two shot distribution schemes: A “uniform” distribution which divides the total number of shots uniformly among all bases in which measurements need to be performed, and an “according to weights” distribution which distributes the shots according to the L2 norm of the coefficients of each group of jointly measurable Pauli operators. We chose the overall shot budget so that the best method is able to achieve chemical accuracy of 10^{-3} Hartree.

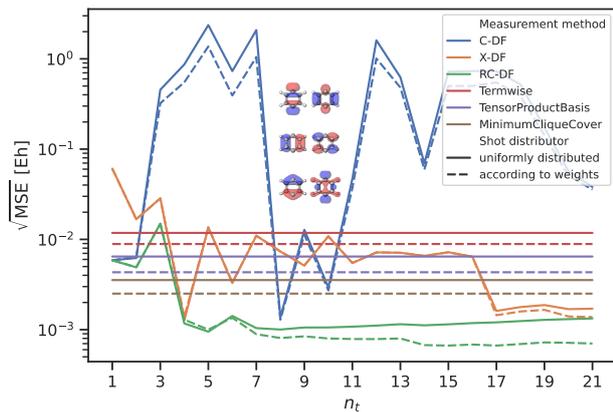


Figure 2: Performance of measurement schemes based on C-DF, X-DF, and RC-DF (with L2 regularization $\rho = 10^{-6}$) in comparison with the naive termwise Pauli scheme as well as the tensor product basis [28] and a minimum clique cover [6] based Pauli grouping methods as implemented in [29] for 3×10^5 shots each in computing a single point energy in the (6e, 6o) CASCI ground state of para-benzyne on 12 qubits. To reach a MSE on par with that of RC-DF at $n_t = 7$ would require over 1.1×10^6 shots with X-DF and $n_t = 17$ and about 3.8×10^6 shots with the best Pauli grouping scheme. We also compare shot distribution schemes (see main text) and find that “according to weights” improves performance.

3.1 NISQ measurement of the para-benzyne ground state

As a first test case we consider the CASCI ground state in a (6e, 6o) active space of FON-HF/cc-pVDZ (with fractional occupations determined with temperature $0.1[1/\text{Eh}]$ and Gaussian broadening within the active space) [25, 26, 27] orbitals of para-benzyne (see Fig. 2). The weighted shot distribution yields lower variance than uniform shot distribution in all cases by directing more shots to the more important first leafs. RC-DF beats the second best method (X-DF) by approximately a factor of five and while the maximum number $21 = 6(6+1)/2$ of leafs yields the lowest MSE, chemical accuracy is consistently achievable with RC-DF from $n_t = 7$ on. The MSE of C-DF fluctuates widely and while it by chance achieves a good variance and approximation error for 8 leafs, this is of limited use for practical applications. The results are virtually unchanged over a broad range of ρ values and good values for ρ can be found in a systematic way even when running on quantum hardware (see Appendix B).

3.2 NISQ measurement of the singlet-triplet gap of naphthalene

As a second test case, we investigate the singlet-triplet energy gap in a (10e, 10o) active space consisting of the π system of naphthalene constructed with AVAS [30] as implemented in PySCF [31, 32]. The reference state was computed at the HF/def2-SVP[33] level of theory. We only show data for double factorization based measurement schemes since the Pauli grouping based schemes become increasingly less competitive for larger systems. Since the “according to weights” method is significantly better than uniform shot distribution, we use it exclusively in this case. We compute the factorization once per n_t and then evaluate the energetically lowest singlet and triplet energies from the same decomposition. The MSE of the singlet-triplet energy gap is given by the square of the difference between the noiseless energy gaps Δ computed with the exact and the gap Δ' from the compressed Hamiltonian plus the sum of the variances of the singlet Var_S and triplet Var_T energies $\text{MSE} = (\Delta - \Delta')^2 + \text{Var}_S + \text{Var}_T$. We find the variances Var_S and Var_T of both states to be very similar, and for (R)C-DF, the MSE is dominated by the variance contributions for $n_t \geq 8$ whereas for X-DF, the systematic energy error remains high until $n_t \approx 25$. Surprisingly, RC-DF reaches chemical accuracy already at $n_t = 6$ with the assigned shot budget and is consistently more accurate than the other two factorization methods for all $n_t > 2$. By contrast, the variance of C-DF singlet-triplet gap estimations is quite erratic, as explained previously. While the accuracy in predicting the energy gap with X-DF is well controllable, an approximately 12 times larger shot budget and more leafs would be needed to reach chemical accuracy. This test case shows the reliable performance of RC-DF for medium-sized active spaces and indicates that the method could be employed for other chemical properties such as activation energies.

3.3 Combination with fluid fermionic fragments

We further explore how X-DF and RC-DF can be combined with the fluid fermionic fragments (FFF) [21] technique to further reduce shot budgets. The FFF technique exploits the fact

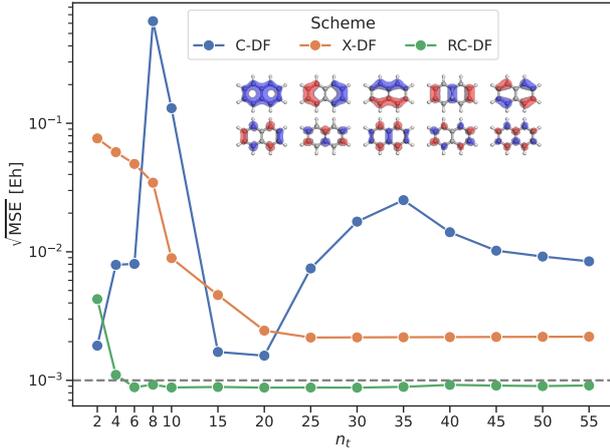


Figure 3: Square-root of the MSE for the singlet-triplet energy gap of naphthalene on 20 qubits (i.e., (10e, 10o) active space, shown in the inset). For each factorization method, 2×10^6 shots were used in total for the singlet and triplet state, distributed according to weights. A regularization factor of $\rho = 10^{-3}$ was used for RC-DF. RC-DF reaches chemical accuracy with only $n_t = 6$ leaves. X-DF would require at least $n_t = 25$ leaves and at least 12×10^6 shots to achieve the same accuracy as RC-DF with only $n_t = 6$ leaves.

that certain quadratic terms of the Hamiltonian, called fluid fermionic fragments, can be taken care of in different parts of the energy estimator. FFF minimizes the variance by optimizing how these terms are spread over the different possible locations (for more details see Appendix G). The variance can thereby either be approximated with that of a mock state whose variance can be classically efficiently computed or it can for example be estimated with part of the shot budget or from a classical shadow [34, 35], which can then also be used for the energy estimation. In any case, the final form of the FFF optimized energy estimator is then state dependent and optimized to have low variance for certain states. We find (see Figure 8 in Appendix G) that for the cases considered RC-DF and FFF nicely complement each other. Using RC-DF as initial point for FFF yields the lowest shot budgets and that the FFF optimization converges faster when started from RC-DF than from X-DF and that the state independent distribution of the fluid fermionic fragments corresponding to the Hamiltonian as written in (6) is a good initial guess for the FFF coefficients.

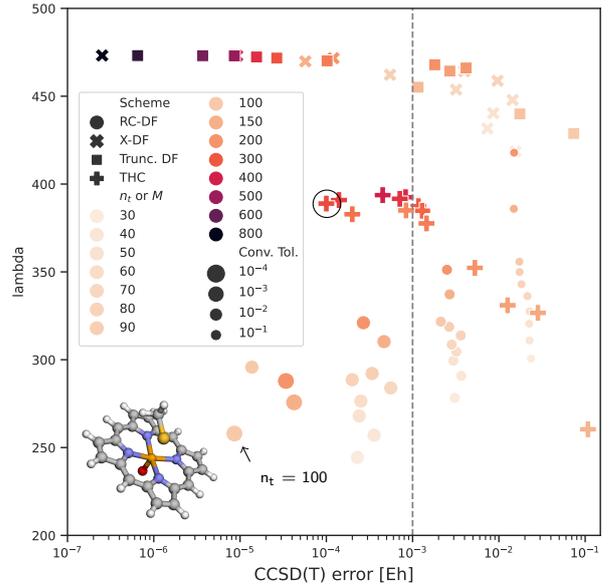


Figure 4: Comparison of the achievable CCSD(T) error heuristic and lambda values $\lambda_{\text{DF}}^{\text{Burg}}$ for the truncated DF method based on (10), X-DF and RC-DF, as well as $\lambda_{\text{THC}}^{\text{Lee}}$ for THC. The color scheme represents the number of leaves n_t for double factorization schemes, or the THC rank M . The active space Hamiltonian of the Cpd I model of cytochrome P450 and the data for THC and truncated DF were taken from [11]. The encircled THC data point was used for the resource estimates there. To compare different levels of convergence we vary the squared Frobenius norm error (8) at which we abort the RC-DF optimization (Conv. Tol.) and use $\rho = 10^{-3}$. The data is tabulated in Appendix E.

3.4 RC-DF for error corrected quantum computing

We now turn to exploring the usefulness of RC-DF for error corrected algorithms based on qubitization. As example we take the $(34\alpha + 29\beta e, 58o)$ active space of the Cpd I species of cytochrome P450 proposed in [11]. As this system is beyond the regime accessible with CASCI, we use, as in [17, 11], the CCSD(T) energy error as a heuristic to assess the quality of the compressed representation. Computational details can be found in Appendix E.

The aim is then to find compressed representation of the Hamiltonian that achieves both a low CCSD(T) error and a low lambda value. We find $\lambda_{\text{DF}}^{\text{Burg}} < \lambda_{\text{DF}}^{\text{LCU}}$ in all cases and thus only display and compare the former. As can be seen in Fig. 4, RC-DF outperforms both previous DF methods and THC by a substantial amount. Compared to

truncated DF or X-DF we can almost cut $\lambda_{\text{DF}}^{\text{Burg}}$ in half, and thereby also the run time of the quantum computer, while at the same time achieving a CCSD(T) error smaller than 5×10^{-5} , which was not reached with THC [11]. Further comparing with THC, we can reduce the CCSD(T) error by an order of magnitude and the lambda value by roughly one third or, alternatively, RC-DF can achieve a $\lambda_{\text{DF}}^{\text{Burg}}$ that is only 60% of λ_{THC} at comparable CCSD(T) error. This improvement becomes even more noteworthy when recognizing that when qubitizing the RC-DF Hamiltonian one need not worry about the complications caused by the non-orthogonal nature of THC that had to be worked around in [17] (see Section III of [17]). Converging the best RC-DF data point with $n_t = 100$ took 2944 L-BFGS iterations and approximately 11 hours on a single GPU and our not highly optimized JAX [36] based implementation.

To investigate the scaling of the achievable lambda values with system size we considered the hydrogen chain benchmark in the STO-6G basis previously proposed in [12, 17]. Also there we find that at $n = 100$ orbitals (corresponding to 100 hydrogen atoms) RC-DF achieves values of $\lambda_{\text{DF}}^{\text{Burg}}$ that are lower than previously reported values for λ_{THC} and we find, as for THC, an approximately linear scaling of $\lambda_{\text{DF}}^{\text{Burg}}$ with n if n_t is chosen such that the CCSD(T) error per particle is constant (see Appendix C for details).

It should be clear that the lambda values, while highly significant for determining the quantum resources of fault tolerant algorithms, are not the only relevant quantity and we do not claim that our analysis constitutes a comprehensive comparison of the quantum resources required for different methods. We also leave to future work the possibility of regularizing methods such as RC-DF or THC with quantities more directly related to the required quantum resources than the norm-like term in (17), which could more directly steer the optimization towards factorizations with low qubit or T gate count, as desired.

4 Conclusions

We proposed the regularized compressed double factorization (RC-DF) method which, from a unified framework yields both a NISQ compatible measurement scheme with only linear cir-

cuit overhead and can be used in conjunction with qubitization in error corrected quantum algorithms for the simulation of chemistry. We found that in both of these scenarios RC-DF leads to lower quantum run times when compared to previous double factorization (DF) and tensor hypercontraction (THC) schemes. Contrary to THC, the Hamiltonian in DF form can also be used to construct trotter schemes that need to alternate between a very small number of non-commuting operators. It will be interesting to compare the quantum resources (e.g. number of Toffoli gates) required for phase estimation based on THC and RC-DF via qubitization with RC-DF and trotterization.

In the NISQ setting, this advantage is a consequence of the fact that the regularization guides the optimization towards compressed representations of the Hamiltonian with smaller coefficients, which reduces the variance of the resulting energy estimator. In qubitization schemes, the smaller coefficients reduce the norm-like lambda parameter of the Hamiltonian on which the T gate count depends in a multiplicative fashion.

Avoiding a six-index intermediate quantity during the RC-DF optimization and adopting a two step gradient based scheme previously developed by some of us for non-regularized compressed DF, we were able to make RC-DF scale well into the regime where quantum computers may provide an advantage over classical methods. More work is needed to understand the precise scaling of RC-DF with active space and basis set size and to explore other options for regularization.

Data Availability

The data supporting the findings of this manuscript have been uploaded to Zenodo with the DOI 10.5281/zenodo.7866658 [37], including instructions on how to load the data using Python.

Acknowledgements

We thank Gian-Luca Anselmetti, Fotios Gkritsis, and Pauline Ollitrault, Artur Izmaylov and Seonghoon Choi for valuable discussions. RMP thanks Mario Motta and Jeff Cohen for many

discussions regarding C-DF. QC Ware Corp. acknowledges generous funding from Covestro for the undertaking of this project. Covestro acknowledges funding from the German Ministry for Education and Research (BMBF) under the funding program quantum technologies as part of project HFAK (13N15630).

Conflict of Interest: EGH and RMP own stock/options in QC Ware Corp.

References

- [1] T. E. O’Brien, G. Anselmetti, F. Gkritsis, V. E. Elfving, S. Polla, W. J. Huggins, O. Oumarou, K. Kechedzhi, D. Abanin, R. Acharya, I. Aleiner, R. Allen, T. I. Andersen, K. Anderson, M. Ansmann, F. Arute, K. Arya, A. Asfaw, J. Atalaya, D. Bacon, J. C. Bardin, A. Bengtsson, S. Boixo, G. Bortoli, A. Bourassa, J. Bovaird, L. Brill, M. Broughton, B. Buckley, D. A. Buell, T. Burger, B. Burkett, N. Bushnell, J. Campero, Y. Chen, Z. Chen, B. Chiaro, D. Chik, J. Cogan, R. Collins, P. Conner, W. Courtney, A. L. Crook, B. Curtin, D. M. Debroy, S. Demura, I. Drozdov, A. Dunsworth, C. Erickson, L. Faoro, E. Farhi, R. Fatemi, V. S. Ferreira, L. Flores Burgos, E. Forati, A. G. Fowler, B. Foxen, W. Giang, C. Gidney, D. Gilboa, M. Giustina, R. Gosula, A. Grajales Dau, J. A. Gross, S. Habegger, M. C. Hamilton, M. Hansen, M. P. Harrigan, S. D. Harrington, P. Heu, J. Hilton, M. R. Hoffmann, S. Hong, T. Huang, A. Huff, L. B. Ioffe, S. V. Isakov, J. Iveland, E. Jeffrey, Z. Jiang, C. Jones, P. Juhas, D. Kafri, J. Kelly, T. Khattar, M. Khezri, M. Kieferová, S. Kim, P. V. Klimov, A. R. Klots, R. Kothari, A. N. Korotkov, F. Kostritsa, J. M. Kreikebaum, D. Landhuis, P. Laptev, K. Lau, L. Laws, J. Lee, K. Lee, B. J. Lester, A. T. Lill, W. Liu, W. P. Livingston, A. Locharla, E. Lucero, F. D. Malone, S. Mandra, O. Martin, S. Martin, J. R. McClean, T. McCourt, M. McEwen, A. Megrant, X. Mi, A. Mieszala, K. C. Miao, M. Mohseni, S. Montazeri, A. Morvan, R. Movassagh, W. Mruczkiewicz, O. Naaman, M. Neeley, C. Neill, A. Nersisyan, H. Neven, M. Newman, J. H. Ng, A. Nguyen, M. Nguyen, M. Y. Niu, S. Omonije, A. Opremcak, A. Petukhov, R. Potter, L. P. Pryadko, C. Quintana, C. Rocque, P. Roushan, N. Saei, D. Sank, K. Sankaragomathi, K. J. Satzinger, H. F. Schurkus, C. Schuster, M. J. Shearn, A. Shorter, N. Shutty, V. Shvarts, J. Skrzuzny, V. Smelyanskiy, W. C. Smith, R. Somma, G. Sterling, D. Strain, M. Szalay, D. Thor, A. Torres, G. Vidal, B. Villalonga, C. Vollgraf Heidweiller, T. White, B. W. K. Woo, C. Xing, Z. J. Yao, P. Yeh, J. Yoo, G. Young, A. Zalcman, Y. Zhang, N. Zhu, N. Zobrist, C. Gogolin, R. Babbush, and N. C. Rubin. “Purification-based quantum error mitigation of pair-correlated electron simulations” (2022).
- [2] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J Love, Alán Aspuru-Guzik, and Jeremy L O’Brien. “A variational eigenvalue solver on a photonic quantum processor”. *Nature communications* **5**, 1–7 (2014).
- [3] Thomas E O’Brien, Stefano Polla, Nicholas C Rubin, William J Huggins, Sam McArdle, Sergio Boixo, Jarrod R McClean, and Ryan Babbush. “Error mitigation via verified phase estimation”. *PRX Quantum* **2**, 020317 (2021).
- [4] Zhenyu Cai, Ryan Babbush, Simon C. Benjamin, Suguru Endo, William J. Huggins, Ying Li, Jarrod R. McClean, and Thomas E. O’Brien. “Quantum error mitigation” (2022).
- [5] William J Huggins, Jarrod R McClean, Nicholas C Rubin, Zhang Jiang, Nathan Wiebe, K Birgitta Whaley, and Ryan Babbush. “Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers”. *npj Quantum Information* **7**, 1–9 (2021).
- [6] Vladyslav Verteletskyi, Tzu-Ching Yen, and Artur F Izmaylov. “Measurement optimization in the variational quantum eigensolver using a minimum clique cover”. *The Journal of chemical physics* **152**, 124114 (2020).
- [7] Laurin E Fischer, Daniel Miller, Francesco Tacchino, Panagiotis Kl Barkoutsos, Daniel J Egger, and Ivano Tavernelli. “Ancilla-free implementation of generalized

- measurements for qubits embedded in a qudit space”. *Physical Review Research* **4**, 033027 (2022).
- [8] Daniel Miller, Laurin E. Fischer, Igor O. Sokolov, Panagiotis Kl. Barkoutsos, and Ivano Tavernelli. “Hardware-tailored diagonalization circuits” (2022). [arXiv:2203.03646](https://arxiv.org/abs/2203.03646).
- [9] J. L. Whitten. “Coulombic potential energy integrals and approximations”. *The Journal of Chemical Physics* **58**, 4496–4501 (1973).
- [10] Edward G Hohenstein, Robert M Parrish, and Todd J Martínez. “Tensor hypercontraction density fitting. i. quartic scaling second-and third-order møller-pletset perturbation theory”. *The Journal of chemical physics* **137**, 044103 (2012).
- [11] Joshua J. Goings, Alec White, Joonho Lee, Christofer S. Tautermann, Matthias Degroote, Craig Gidney, Toru Shiozaki, Ryan Babbush, and Nicholas C. Rubin. “Reliably assessing the electronic structure of cytochrome p450 on today’s classical computers and tomorrow’s quantum computers”. *Proceedings of the National Academy of Sciences* **119**, e2203533119 (2022).
- [12] Dominic W Berry, Craig Gidney, Mario Motta, Jarrod R McClean, and Ryan Babbush. “Qubitization of arbitrary basis quantum chemistry leveraging sparsity and low rank factorization”. *Quantum* **3**, 208 (2019).
- [13] Edward G. Hohenstein, Oumarou Oumarou, Rachael Al-Saadon, Gian-Luca R. Anselmetti, Maximilian Scheurer, Christian Gogolin, and Robert M. Parrish. “Efficient quantum analytic nuclear gradients with double factorization”. *The Journal of Chemical Physics* **158**, 114119 (2023).
- [14] Jeffrey Cohn, Mario Motta, and Robert M. Parrish. “Quantum Filter Diagonalization with Compressed Double-Factorized Hamiltonians”. *PRX Quantum* **2**, 040352 (2021).
- [15] Ryan Babbush, Craig Gidney, Dominic W. Berry, Nathan Wiebe, Jarrod McClean, Alexandru Paler, Austin Fowler, and Hartmut Neven. “Encoding electronic spectra in quantum circuits with linear t complexity”. *Phys. Rev. X* **8**, 041015 (2018).
- [16] Vera von Burg, Guang Hao Low, Thomas Häner, Damian S. Steiger, Markus Reiher, Martin Roetteler, and Matthias Troyer. “Quantum computing enhanced computational catalysis”. *Phys. Rev. Res.* **3**, 033055 (2021).
- [17] Joonho Lee, Dominic W. Berry, Craig Gidney, William J. Huggins, Jarrod R. McClean, Nathan Wiebe, and Ryan Babbush. “Even more efficient quantum computations of chemistry through tensor hypercontraction”. *PRX Quantum* **2**, 030305 (2021).
- [18] Robert M. Parrish and Peter L. McMahon. “Quantum filter diagonalization: Quantum eigendecomposition without full quantum phase estimation” (2019). [arXiv:1909.08925](https://arxiv.org/abs/1909.08925).
- [19] Ignacio Loaiza, Alireza Marefat Khah, Nathan Wiebe, and Artur F Izmaylov. “Reducing molecular electronic hamiltonian simulation cost for linear combination of unitaries approaches”. *Quantum Science and Technology* (2022).
- [20] Tzu-Ching Yen and Artur F Izmaylov. “Cartan subalgebra approach to efficient measurements of quantum observables”. *PRX Quantum* **2**, 040320 (2021).
- [21] Seonghoon Choi, Ignacio Loaiza, and Artur F. Izmaylov. “Fluid fermionic fragments for optimizing quantum measurements of electronic Hamiltonians in the variational quantum eigensolver”. *Quantum* **7**, 889 (2023).
- [22] Artur F Izmaylov, Tzu-Ching Yen, Robert A Lang, and Vladyslav Verteletskyi. “Unitary partitioning approach to the measurement problem in the variational quantum eigensolver method”. *Journal of chemical theory and computation* **16**, 190–195 (2019).
- [23] Mario Motta, Erika Ye, Jarrod R McClean, Zhendong Li, Austin J Minnich, Ryan Babbush, and Garnet Kin Chan. “Low rank representations for quantum simulation of electronic structure”. *npj Quantum Information* **7**, 1–7 (2021).
- [24] Edvin Deadman, Nicholas J Higham, and Rui Ralha. “Blocked schur algorithms for computing the matrix square root”. In International Workshop on Ap-

- plied Parallel Computing. Pages 171–182. Springer (2012).
- [25] Mario Barbatti, Adélia JA Aquino, and Hans Lischka. “Ultrafast two-step process in the non-adiabatic relaxation of the ch₂ molecule”. *Molecular Physics* **104**, 1053–1060 (2006).
- [26] Bernhard Sellner, Mario Barbatti, Thomas Müller, Wolfgang Domcke, and Hans Lischka. “Ultrafast non-adiabatic dynamics of ethylene including rydberg states”. *Molecular physics* **111**, 2439–2450 (2013).
- [27] Thom H. Dunning. “Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen”. *J. Chem. Phys.* **90**, 1007–1023 (1989).
- [28] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M Chow, and Jay M Gambetta. “Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets”. *Nature* **549**, 242–246 (2017).
- [29] Ville Bergholm, Josh Izaac, Maria Schuld, Christian Gogolin, Shahnawaz Ahmed, Vishnu Ajith, M. Sohaib Alam, Guillermo Alonso-Linaje, B. AkashNarayanan, Ali Asadi, Juan Miguel Arrazola, Utkarsh Azad, Sam Banning, Carsten Blank, Thomas R Bromley, Benjamin A. Cordier, Jack Ceroni, Alain Delgado, Olivia Di Matteo, Amintor Dusko, Tanya Garg, Diego Guala, Anthony Hayes, Ryan Hill, Aroosa Ijaz, Theodor Isaacson, David Ittah, Soran Jahangiri, Praateek Jain, Edward Jiang, Ankit Khadelwal, Korbinian Kottmann, Robert A. Lang, Christina Lee, Thomas Loke, Angus Lowe, Keri McKiernan, Johannes Jakob Meyer, J. A. Montañez-Barrera, Romain Moyard, Zeyue Niu, Lee James O’Riordan, Steven Oud, Ashish Panigrahi, Chae-Yeun Park, Daniel Polatajko, Nicolás Quesada, Chase Roberts, Nahum Sá, Isidor Schoch, Borun Shi, Shuli Shu, Sukin Sim, Arshpreet Singh, Ingrid Strandberg, Jay Soni, Antal Száva, Slimane Thabet, Rodrigo A. Vargas-Hernández, Trevor Vincent, Nicola Vitucci, Maurice Weber, David Wierichs, Roeland Wiersema, Moritz Willmann, Vincent Wong, Shaoming Zhang, and Nathan Killoran. “PennyLane: Automatic differentiation of hybrid quantum-classical computations” (2018).
- [30] Elvira R. Sayfutyarova, Qiming Sun, Garnet Kin-Lic Chan, and Gerald Knizia. “Automated construction of molecular active spaces from atomic valence orbitals”. *Journal of Chemical Theory and Computation* **13**, 4063–4078 (2017).
- [31] Qiming Sun, Timothy C. Berkelbach, Nick S. Blunt, George H. Booth, Sheng Guo, Zhendong Li, Junzi Liu, James D. McClain, Elvira R. Sayfutyarova, Sandeep Sharma, Sebastian Wouters, and Garnet Kin-Lic Chan. “Pyscf: the python-based simulations of chemistry framework”. *WIREs Computational Molecular Science* **8**, e1340 (2018).
- [32] Qiming Sun, Xing Zhang, Samragni Banerjee, Peng Bao, Marc Barbry, Nick S. Blunt, Nikolay A. Bogdanov, George H. Booth, Jia Chen, Zhi-Hao Cui, Janus J. Eriksen, Yang Gao, Sheng Guo, Jan Hermann, Matthew R. Hermes, Kevin Koh, Peter Koval, Susi Lehtola, Zhendong Li, Junzi Liu, Narbe Mardirossian, James D. McClain, Mario Motta, Bastien Mussard, Hung Q. Pham, Artem Pulkin, Wirawan Purwanto, Paul J. Robinson, Enrico Ronca, Elvira R. Sayfutyarova, Maximilian Scheurer, Henry F. Schurkus, James E. T. Smith, Chong Sun, Shi-Ning Sun, Shiv Upadhyay, Lucas K. Wagner, Xiao Wang, Alec White, James Daniel Whitfield, Mark J. Williamson, Sebastian Wouters, Jun Yang, Jason M. Yu, Tianyu Zhu, Timothy C. Berkelbach, Sandeep Sharma, Alexander Yu. Sokolov, and Garnet Kin-Lic Chan. “Recent developments in the pyscf program package”. *The Journal of Chemical Physics* **153**, 024109 (2020).
- [33] Florian Weigend and Reinhart Ahlrichs. “Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for h to rn: Design and assessment of accuracy”. *Phys. Chem. Chem. Phys.* **7**, 3297 (2005).
- [34] Kianna Wan, William J. Huggins, Joonho Lee, and Ryan Babbush. “Matchgate shadows for fermionic quantum simulation”.

- [Communications in Mathematical Physics](#) **404**, 629–700 (2023).
- [35] Guang Hao Low. “Classical shadows of fermions with particle number symmetry” (2022). arXiv:2208.08964 [quant-ph].
- [36] Roy Frostig, Matthew Johnson, and Chris Leary. “Compiling machine learning programs via high-level tracing”. SysML (2018). url: <https://mlsys.org/Conferences/doc/2018/146.pdf>.
- [37] Oumarou Oumarou, Maximilian Scheurer, Robert M. Parrish, Edward G. Hohenstein, and Christian Gogolin. “Data for Accelerating Quantum Computations of Chemistry Through Regularized Compressed Double Factorization”. [Zenodo](#) (2023).
- [38] Ciyou Zhu, Richard H Byrd, Peihuang Lu, and Jorge Nocedal. “Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization”. [ACM Transactions on mathematical software \(TOMS\)](#) **23**, 550–560 (1997).
- [39] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. [Nature Methods](#) **17**, 261–272 (2020).
- [40] Joshua J. Goings, Alec White, Christofer S. Tautermann, Matthias Degroote, Craig Gidney, Toru Shiozaki, Ryan Babbush, and Nicholas C. Rubin. “Data for reliably assessing the electronic structure of cytochrome p450 on today’s classical computers and tomorrow’s quantum computers”. [Zenodo](#) (2022).
- [41] Dougal Maclaurin, David Duvenaud, and Ryan P Adams. “Autograd: Effortless gradients in numpy”. In ICML 2015 AutoML workshop. Volume 238. (2015). url: <https://indico.ijclab.in2p3.fr/event/2914/contributions/6483/subcontributions/180/attachments/6060/7185/automl-short.pdf>.

Appendix A: RC-DF optimization procedure

The RC-DF cost function is

$$C(X, Z) := \frac{1}{2} \|\Delta_{pqrs}\|_{\mathcal{F}}^2 + \sum_{tkl} \rho_{tkl} |Z_{kl}^t|^\gamma, \quad (18)$$

with regularization tensor ρ_{tkl} and $\gamma = 1$ in the L1 case and $\gamma = 2$ in the L2 case. Just like C-DF as outlined in [18], also for RC-DF it is advisable to use a nested two-step optimization process to minimize C with respect to X and Z following these steps:

1. Update the X_{pq}^t for fixed Z_{kl}^t . This can be done with a gradient based optimizer using the gradient

$$\frac{\partial C}{\partial X_{pq}^t} = \frac{\partial \|\Delta_{pqrs}\|_{\mathcal{F}}^2}{\partial X_{pq}^t} = \sum_{mn} \frac{\partial C}{\partial U_{mn}^t} \frac{\partial U_{mn}^t}{\partial X_{pq}^t}. \quad (19)$$

2. Determine the optimal Z_{kl}^t given the updated X_{pq}^t by solving

$$\frac{\partial C}{\partial Z_{kl}^t} = 0. \quad (20)$$

The gradient with respect to the orbital rotations' generators X_{pq}^t for a given Z_{pq}^t is independent of the regularization and stays unchanged compared to the original C-DF as the U_{pq}^t do not appear in the regularization contribution to the cost function. Therefore

$$\frac{\partial C}{\partial U_{mn}^t} = -4 \sum_{qrst} \Delta_{mqrst} U_{qn}^t Z_{nl}^t U_{rl}^t U_{sl}^t. \quad (21)$$

Let us now look at the second step. In the L2 case (20) yields

$$\frac{\partial C}{\partial Z_{kl}^t} = - \sum_{pqrs} \Delta_{pqrs} U_{pk}^t U_{qk}^t U_{rl}^t U_{sl}^t + \rho_{kl}^t Z_{kl}^t = 0, \quad (22)$$

by replacing Δ_{pqrs} by its expression in 8, we have

$$\sum_{pqrs} (pq|rs) U_{pk}^t U_{qk}^t U_{rl}^t U_{sl}^t = \rho_{kl}^t Z_{kl}^t + \sum_{omn} \left[\sum_p U_{pk}^t U_{pm}^o \right] \left[\sum_q U_{qk}^t U_{qm}^o \right] Z_{mn}^o \left[\sum_r U_{rl}^t U_{rn}^o \right] \left[\sum_s U_{sl}^t U_{sn}^o \right]. \quad (23)$$

Note that the left-hand side of the equation is independent of Z_{mn}^o i.e constant, and the right-hand side can be written as a linear combination of the unknowns Z_{mn}^o . Therefore, we have a system of linear equations of the form

$$b_{kl}^t((pq|rs), U_{pq}^t) = \sum_{omn} A_{omn}^{tkl}(U_{pq}^t, \rho_{kl}^t) Z_{mn}^o \quad (24)$$

with

$$b((pq|rs), U_{pq}^t)_{tkl} = \sum_{pqrs} (pq|rs) U_{pk}^t U_{qk}^t U_{rl}^t U_{sl}^t \quad (25)$$

and

$$A(U_{pq}^t, \rho_{kl}^t)_{tkl, omn} = \sum_{omn} \left(\left[\sum_p U_{pk}^t U_{pm}^o \right] \left[\sum_q U_{qk}^t U_{qm}^o \right] \left[\sum_r U_{rl}^t U_{rn}^o \right] \left[\sum_s U_{sl}^t U_{sn}^o \right] + \delta_{(omn, tkl)} \rho_{tkl} \right). \quad (26)$$

In the L1 case, (20) yields

$$\frac{\partial \mathcal{C}}{\partial Z_{kl}^t} = - \sum_{pqrs} \Delta_{pqrs} U_{pk}^t U_{qk}^t U_{rl}^t U_{sl}^t + \rho_{kl}^t \text{sign}(Z_{kl}^t) = 0, \quad (27)$$

which implies

$$\sum_{pqrs} (pq|rs) U_{pk}^t U_{qk}^t U_{rl}^t U_{sl}^t - \rho_{kl}^t \text{sign}(Z_{kl}^t) = \sum_{omn} \left[\sum_p U_{pk}^t U_{pm}^o \right] \left[\sum_q U_{qk}^t U_{qm}^o \right] Z_{mn}^o \left[\sum_r U_{rl}^t U_{rn}^o \right] \left[\sum_s U_{sl}^t U_{sn}^o \right] \quad (28)$$

which is again a system of equations of the form

$$b_{kl}^t \left((pq|rs), U_{pq}^t, \rho_{kl}^t, Z_{kl}^t \right) = \sum_{omn} A_{omn}^{tkl} (U_{pq}^t) Z_{mn}^o. \quad (29)$$

with

$$b_{kl}^t \left((pq|rs), U_{pq}^t, \rho_{kl}^t, Z_{kl}^t \right) = \sum_{pqrs} (pq|rs) U_{pk}^t U_{qk}^t U_{rl}^t U_{sl}^t - \rho_{kl}^t \text{sign}(Z_{kl}^t) \quad (30)$$

$$A_{omn}^{tkl} = \sum_{omn} \left[\sum_p U_{pk}^t U_{pm}^o \right] \left[\sum_q U_{qk}^t U_{qm}^o \right] \left[\sum_r U_{rl}^t U_{rn}^o \right] \left[\sum_s U_{sl}^t U_{sn}^o \right]. \quad (31)$$

So in both cases one can find the optimal Z_{kl}^t by pseudo-inverting $A_{tkl,omn}$. While pseudo-inverting the six-index tensor $A_{tkl,omn}$ to determine the Z_{kl}^t can be done for medium size systems, it is intractable for large systems. To circumvent this problem, the inversion can be carried out in a matrix-free manner with, e.g., a conjugate gradient algorithm. This procedure only requires the matrix-vector product and the matrix diagonal rather than the dense matrix. Lastly, any gradient-descent based optimization algorithm can be used. For the result presented in this work, we have used the L-BFGS [38] algorithm as implemented in SciPy [39]. All time critical routines for the steps above were just-in-time compiled and ran on an NVIDIA Tesla V100 with the help of JAX [36].

Appendix B: Tuning of the regularization

The regularization tensor ρ_{tkl} is a hyper-parameter that controls how much effort is put on achieving small $|Z_{pq}^t|$ versus reducing the Frobenius norm error. A good balance must be found to obtain both a small systematic energy error and a small variance and lambda value. We concentrate only the case that after truncating some number of n_t leafs the regularization is chosen uniformly $\rho_{tkl} =: \rho$.

In the main text we have quantified the overall performance of the measurement scheme by means of the MSE, which combines the systematic error in the energy because of the approximation of the Hamiltonian and the statistical error due to variance and shot noise. Here let us look at those two contributions separately. We consider the same data underlying Figure 2 from the main text.

In Figure 5 we show only the systematic error that results from RC-DF approximating the two body part of the Hamiltonian and represents what is achievable in the limit of infinitely many shots. The energy approximation error grows roughly linearly with $\rho/10$ and when the number of leafs is increased it seems to become easier for the optimizer to find Z_{kl}^t that are not too much distorted by the regularization and thus represent the $(pq|rs)$ tensor well leading to a smaller energy error.

In Figure 6 we show the standard deviation $\sqrt{\text{Var}}$ of the ground state energy estimator, quantifying how far from the infinite shot budget limit a measured energy value is likely to lie. Larger regularization factors ρ manifestly reduce the standard deviation. If the shot distribution takes into account the weight of the leafs this effect is greatly enhanced. Very strong regularization does not seem to help. Increasing or decreasing the total shot budget would simply move the data point up and down according

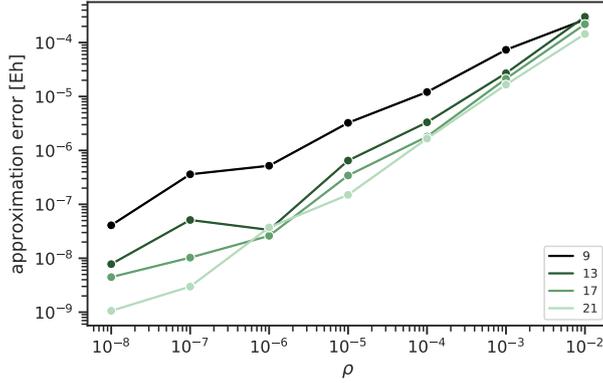


Figure 5: RC-DF systematic ground state energy approximation error of para-benzyne as a function of the regularization factor ρ with different numbers of leaves n_t .

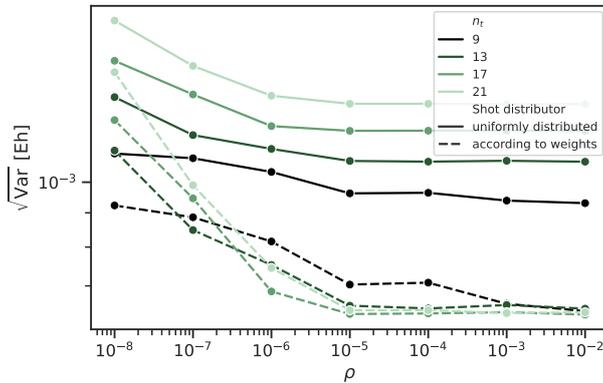


Figure 6: Standard deviation of the ground state energy estimator of para-benzyne for a budget of 300.000 shots as a function of ρ for different numbers of leaves n_t .

to the well known one over square root raw. As can be nicely seen there is a rather large window between $\rho \in [10^{-6}, 10^{-2}]$ where both errors (and thus their sum) is well below 10^{-3} . In practice one find a good ρ by estimating the standard deviation on a quantum device starting from a comparably large regularization and then reduce the regularization to reduce the systematically error until the variance becomes too large to be compensated for by the affordable shot budget.

Appendix C: Lambda scaling with system size

To investigate the scaling of the lambda parameters achievable with double factorization in the large system size limit we use the hydrogen chain benchmark previously employed in the context of qubitization combined with low rank factorization in [12] and with THC in [17]. In this benchmark system neutral hydrogen atoms are placed with 1.4 Bohr distance on a line and represented in the STO-6G basis, which has one spatial orbital per hydrogen so that the size of the complete active space n is equal to the number of hydrogen atoms.

The lambda factor of the qubitization scheme from [12] is the sum of a one-body contribution λ_T and a two-body contribution λ_W and is strongly dominated by the latter. Between $n = 10$ and $n = 100$ the authors find that λ_W grows roughly proportional to $n^{2.5}$ to values up to about 2×10^5 (see Fig 11(a) in [12]).

In [17] a THC based qubitization scheme is proposed that improves over what the authors call the “naïve” approach. In Fig 20 of that work the two-body contribution λ_2 of these two schemes is

compared, again for $10 \leq n \leq 100$. The “naïve” λ_2 is found to grow approximately proportional to $n^{3.16}$ to values well beyond 5×10^6 . The λ_2 achievable with the non-orthogonal basis THC based qubitization proposed in [17] (which corresponds to the two-body part of $\lambda_{\text{THC}}^{\text{Lee}}$) is found to only grow approximately proportionally to $n^{1.16}$ and reaches approximately 4×10^2 at $n = 100$. Fig 9 of the same work compares the full lambda values (including the one-body contribution) for different factorization methods including THC (the main focus of that work and the most competitive of the factorization methods in this plot) and the Cholesky based DF described around (10). The parameters of these methods are chosen to yield a CCSD(T) correlation energy error per atom of at most 50 micro Hartree, amounting to 0.5×10^{-3} Hartree at $n = 100$. The authors find a scaling of roughly $n^{1.88}$ for DF and of $n^{1.11}$ for THC and at $n = 100$ values of $\lambda > 2 \times 10^3$ for DF and $\lambda \approx 5 \times 10^2$ for THC.

With a regularization of $\rho = 5 \times 10^{-5}$ and even just a linear number of leafs $n_t = \lfloor (n+1)/2 \rfloor$ (red line in Fig. 7) we find that RC-DF (with two norm regularization $\gamma = 2$) is able to yield a constant Frobenius norm error and an absolute CCSD(T) error per atom $|\Delta_{\text{CCSD(T)}}|/n$ in line with the 50 micro Hartree per atom used in Fig 9 of [17] (see the green line in Fig. 7b). At these parameters $\lambda_{\text{RC-DF}}^{\text{Burg}}$ is found to scale approximately like $n^{1.08 \pm 0.10}$ (fit through the values for $70 \leq n \leq 100$) and reaches approximately 2.5×10^2 at $n = 100$, roughly a factor of two better than the THC results from [17]. Compared to X-DF (which requires n_t to grow faster than linearly to obtain acceptable accuracy (see the blue and yellow lines in Fig. 7b), RC-DF yields roughly one order of magnitude lower lambda values at $n = 100$, mostly owing to the fact that we find that the X-DF lambda values scale roughly quadratic with n . λ^{Burg} and λ^{LCU} seem to have a similar scaling for all double factorization schemes considered here.

Appendix D: Necessary and sufficient condition for the symmetry of the V_{pq}^t

In this section we present a proof that shows that 8-fold symmetry of the $(pq|rs)$ tensor is a sufficient condition for the V_{pq}^t to be symmetric for every t where the eigenvalue $g_t \neq 0$. Hence, the Z_{pq}^t are real, which is required for the X-DF procedure to work, and the U_{pq}^t are orthogonal (and thus can be chosen to be special orthogonal without loss of generality), which is essential for their implementation on a quantum computer by means of a fabric of givens rotations. Slightly abusing notation, in the following lemma, we use $(pq|rs)$ for a four index tensor that does not necessarily arise from electron overlap integrals, but has the stated properties.

Lemma 1. *Let $(pq|rs)$ be any real, symmetric, i.e., $(pq|rs) = (rs|pq)$ tensor of shape $n \times n \times n \times n$. By grouping the indices pq and rs , let $(g_t)_t$ be its n^2 eigenvalues, V_{pq}^t its diagonalizing unitary and let $T_+ = \{t : g_t > 0\}$, $T_- = \{t : g_t < 0\}$, and $T := T_+ \cup T_-$. Then $(pq|rs)$ can be written in the form*

$$(pq|rs) = \sum_{t \in T} V_{pq}^t g^t V_{rs}^t. \quad (32)$$

Further, if and only if $(pq|rs)$ is in addition 8-fold symmetric, i.e., $(qp|rs) = (qp|sr) = (pq|sr) = (pq|rs) = (rs|pq) = (sr|pq) = (sr|qp) = (rs|qp)$ the matrices $(V_{pq}^t)_{pq}$ are symmetric, i.e., $V_{pq}^t = V_{qp}^t \forall t \in T$, and $|T| \leq n(n+1)/2$.

Proof. That symmetry of the V_{pq}^t implies 8-fold symmetry of $(qp|rs)$ can be verified directly from (32) and whenever the V_{pq}^t are symmetric, since they are also orthogonal for different values of t , there can be at most $n(n+1)/2$ of them and thus the bond on the size $|T|$ of T holds.

To show that 8-fold symmetry implies symmetry of all V_{pq}^t with $t \in T$ let us first define

$$(pq|rs)_{\pm} := \sum_{t \in T_{\pm}} V_{pq}^t g^t V_{rs}^t \quad (33)$$

$$|(pq|rs)| := \sum_{t \in T} V_{pq}^t |g^t| V_{rs}^t. \quad (34)$$

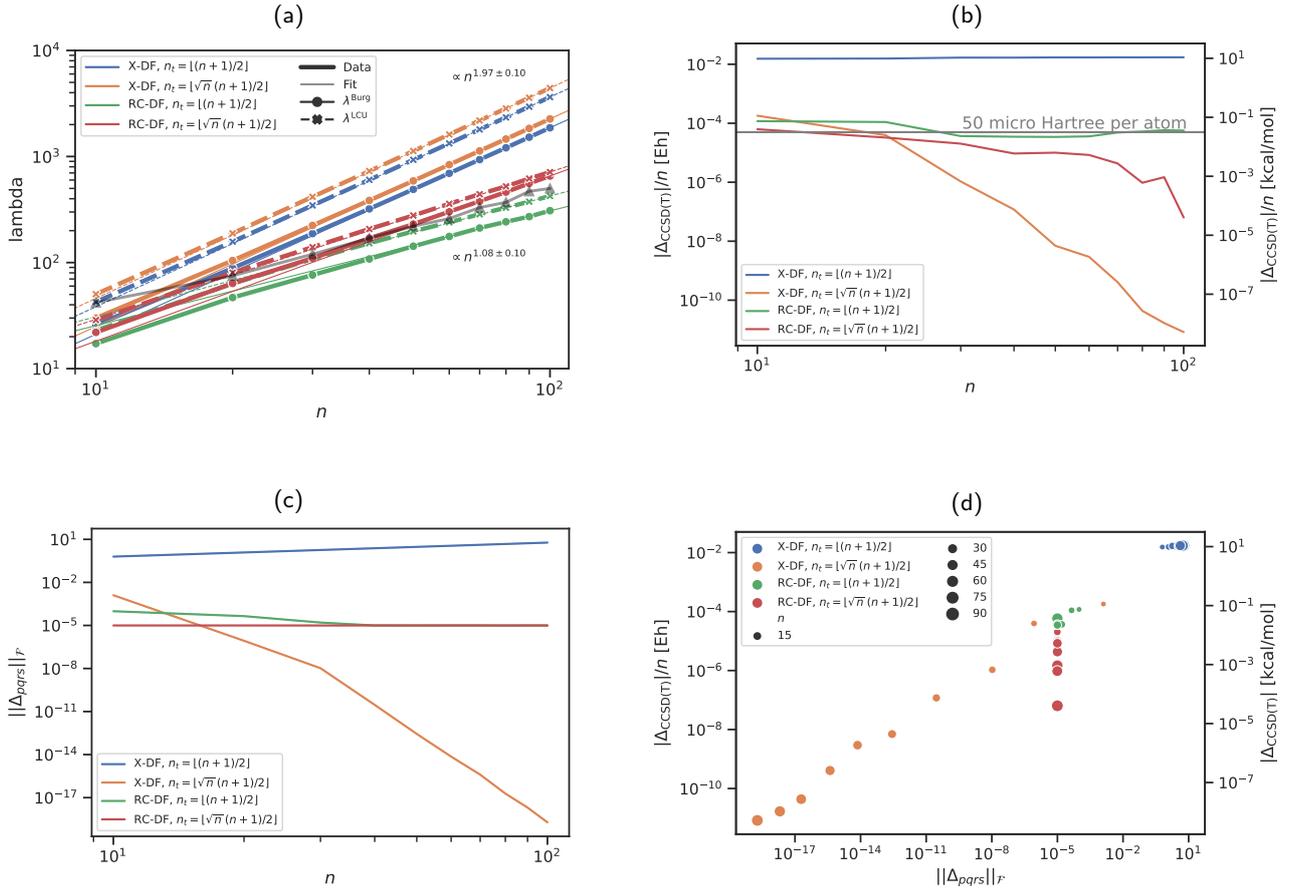


Figure 7: Comparison of lambda values achievable with X-DF and two-norm ($\gamma = 2$) RC-DF with regularization $\rho = 5 \times 10^{-5}$ and convergence tolerance $\|\Delta_{pqrs}\|_{\mathcal{F}} \leq 10^{-5}$ for the hydrogen chain benchmark in the STO-6G basis previously use in [12, 17] (panel a). The black triangles are the lambda values eyeballed from Fig. 9 of [17]. The thin lines are power law fits to the date for $70 \leq n \leq 100$ used to estimate the exponents stated in the text and displayed in the figure. Two different scalings of n_t with n were chosen such that the CCSD(T) error per particle $|\Delta_{\text{CCSD(T)}}|/n$ (panel b) is approximately below 50 micro Hartree per atom for all cases except X-DF with linear number of leafs. The Frobenius norm error of the two-body tensor $\|\Delta_{pqrs}\|_{\mathcal{F}}$ (panel c) seems to be well correlated with $|\Delta_{\text{CCSD(T)}}|$ for X-DF (panel d) but their magnitudes can differ by several orders for other methods, so care needs to be taken when drawing conclusions base on $\|\Delta_{pqrs}\|_{\mathcal{F}}$ (The RC-DF data points are on a vertical line because $\|\Delta_{pqrs}\|_{\mathcal{F}} \leq 10^{-5}$ was used as an abort criterion for the RC-DF optimization).

so that we have $(pq|rs) \pm |(pq|rs)| = 2(pq|rs)_\pm$. Since we also have $|(pq|rs)| = (pq|rs) R$ with

$$R := \sum_{t \in T} V_{pq}^t \text{sign}(g^t) V_{rs}^t. \quad (35)$$

symmetric and invertible, 8-fold symmetry of $(pq|rs)$ implies 8-fold symmetry of $|(pq|rs)|$ and thereby 8-fold symmetry of both $(pq|rs)_+$ and $(pq|rs)_-$ individually. The symmetry $(pq|rs)_\pm = (qp|rs)_\pm$ then implies that

$$\sum_{t \in T_\pm} (V_{pq}^t V_{rs}^t - V_{qp}^t V_{rs}^t) g^t = 0 \quad \forall p, q, r, s. \quad (36)$$

For $p = r$ and $q = s$ we can specialized this to

$$\sum_{t \in T_\pm} g^t V_{pq}^t V_{qp}^t = \sum_{t \in T_\pm} g^t (V_{pq}^t)^2 \quad \forall p, q. \quad (37)$$

The symmetry $(pq|pq)_\pm = (qp|qp)_\pm$ together with (32) implies

$$\sum_{t \in T_\pm} g_t (V_{pq}^t)^2 = \sum_{t \in T_\pm} g_t (V_{qp}^t)^2 \quad (38)$$

and thus the left and right hand side of (37) can be identified to be the left and right hand side of the Cauchy-Schwarz inequalities of the inner product of the vectors $(\sqrt{g_t^*} V_{pq}^t)_{t \in T_\pm}$ and $(\sqrt{g_t} V_{qp}^t)_{t \in T_\pm}$:

$$\left(\sum_{t \in T_\pm} (\sqrt{g_t^*} V_{pq}^t)^* \sqrt{g_t} V_{qp}^t \right)^2 = \quad (39)$$

$$\left(\sum_{t \in T_\pm} g_t V_{pq}^t V_{qp}^t \right)^2 \leq \sum_{t \in T_\pm} |g_t| (V_{pq}^t)^2 \sum_{t \in T_\pm} |g_t| (V_{qp}^t)^2 \quad (40)$$

$$= \sum_{t \in T_\pm} \pm |g_t| (V_{pq}^t)^2 \sum_{t \in T_\pm} \pm |g_t| (V_{qp}^t)^2 \quad (41)$$

$$= \sum_{t \in T_\pm} g_t (V_{pq}^t)^2 \sum_{t \in T_\pm} g_t (V_{qp}^t)^2 \quad (42)$$

$$= \left(\sum_{t \in T_\pm} g_t (V_{qp}^t)^2 \right)^2 \quad \forall p, q, \quad (43)$$

where in the last step we have used (38). Thus (37) implies that these Cauchy-Schwarz inequalities are indeed fulfilled with equality, which is the case if and only if the two vectors in the Cauchy-Schwarz inequality are identical up to a prefactor, i.e., for each p, q there must exists a scalar α such that

$$\sqrt{g_t^*} V_{pq}^t = \alpha \sqrt{g_t} V_{qp}^t \quad \forall t \in T_\pm \quad (44)$$

$$\Leftrightarrow V_{pq}^t = \pm \alpha V_{qp}^t \quad \forall t \in T_\pm. \quad (45)$$

Since $(pq|rs) = (qp|rs)$, we further have

$$(pq|rs) = \sum_{t \in T_\pm} V_{pq}^t g^t V_{rs}^t = \pm \alpha \sum_{t \in T_{pm}} V_{qp}^t g^t V_{rs}^t \quad (46)$$

$$= \sum_{t \in T_\pm} V_{qp}^t g^t V_{rs}^t = (qp|rs), \quad (47)$$

which is possible only if $\pm \alpha = 1$. Thus we have as claimed $V_{pq}^t = V_{qp}^t \quad \forall t \in T$.

□

Appendix E: Computational methodology for Cpd I test case

The active space integrals for Cpd I were obtained from the deposited data [40] of Ref. [11]. The system under consideration here is labelled “X” in Ref. [11], and consists of a $(34\alpha+29\beta, 58o)$ active space. We factorized the two-body integrals using RC-DF with $n_t \in \{30, 40, 50, 60, 70, 80, 90, 100, 150, 200\}$ and additional $n_t = 400, 800$ with X-DF. For RC-DF, we employed $\rho = 10^{-3}$ and varied the convergence tolerance from 10^{-1} to 10^{-4} . From the factorized two-body integrals, a CCSD(T) energy was computed as described in Ref. [11] using the `chemftr` Python library (<https://github.com/ncrubin/chemftr>) interfaced with PySCF [31, 32]. The CCSD(T) energy error is the energy difference of the CCSD(T) energies with exact two-body integrals and the two-body integrals reconstructed from the factorization. For both factorization schemes, we evaluated $\lambda_{\text{DF}}^{\text{Burg}}$ and $\lambda_{\text{DF}}^{\text{LCU}}$. Since the data for the truncated DF scheme in Ref. [11] are not shown in the paper, we recomputed the factorization using `chemftr` and checked that for exact factorizations, the lambda parameters agree exactly with our X-DF implementation. The RC-DF and X-DF results are shown in Tables 1 and 2, respectively. In addition, Table 3 contains the recomputed results for truncated DF from Ref. [11]. The factorized Hamiltonians, together with the resulting energy errors and lambda factors were deposited on Zenodo [37].

Appendix F: Derivation of the double factorized Hamiltonian in terms of Pauli operators

Inserting (5) into (1) we have

$$H = E_c + \sum_{pq} (p|\hat{\kappa}|q) E_{pq}^+ + \frac{1}{2} \sum_{tkl} Z_{kl}^t U^t E_{kk}^+ E_{ll}^+ U^{\dagger t} \quad (48)$$

with

$$(p|\hat{\kappa}|q) = (p|\hat{h}_c|q) - \frac{1}{2} \sum_r (pr|qr). \quad (49)$$

Using the Jordan Wigner mapping we can write

$$E_{kk}^+ = I - \frac{\hat{Z}_k + \hat{Z}_{\bar{k}}}{2} \quad (50)$$

and using the following identity

$$E_{kk}^+ E_{ll}^+ = -I + E_{kk}^+ + E_{ll}^+ + \frac{1}{4} (\hat{Z}_k + \hat{Z}_{\bar{k}}) (\hat{Z}_l + \hat{Z}_{\bar{l}}) \quad (51)$$

yields

$$H = E_c - \frac{1}{2} \sum_{tkl} Z_{kl}^t + \sum_{pq} (p|\hat{\kappa}|q) E_{pq}^+ + \sum_{tk} \sum_l Z_{kl}^t U^t E_{ll}^+ U^{\dagger t} + \frac{1}{8} \sum_{tkl} Z_{kl}^t U^t (\hat{Z}_k + \hat{Z}_{\bar{k}}) (\hat{Z}_l + \hat{Z}_{\bar{l}}) U^{\dagger t}. \quad (52)$$

Our aim is now to sort terms according to whether they contain an even or odd number of \hat{Z} operators to arrive at expression (6). Using again (5) we can rewrite $\sum_{tk} \sum_l Z_{kl}^t U^t E_{ll}^+ U^{\dagger t} = \sum_{pq} \sum_{tk} \sum_l U_{pl}^t U_{ql}^t Z_{kl}^t E_{pq}^+$ and hence we have

$$(pq|rr) = \sum_{tkl} U_{pk}^t U_{qk}^t Z_{kl}^t U_{rl}^t{}^2 \quad (53)$$

$$\implies \sum_r (pq|rr) = \sum_{tkl} U_{pk}^t U_{qk}^t Z_{kl}^t. \quad (54)$$

Using these expressions and the shorthand \mathcal{F}_{pq} defined in (3) we can rewrite (52) to read

$$H = E_c - \frac{1}{2} \sum_{tkl} Z_{kl}^t + \sum_{pq} ((p|\hat{\kappa}|q) + \sum_{tk} \sum_l U_{pl}^t U_{ql}^t Z_{kl}^t) E_{pq}^+ + \frac{1}{8} \sum_{tkl} Z_{kl}^t U^t (\hat{Z}_k + \hat{Z}_{\bar{k}}) (\hat{Z}_l + \hat{Z}_{\bar{l}}) U^{t\dagger} \quad (55)$$

$$= E_c - \frac{1}{2} \sum_{tkl} Z_{kl}^t + \sum_{pq} ((p|\hat{\kappa}|q) + \sum_r (pq|rr)) E_{pq}^+ + \frac{1}{8} \sum_{tkl} Z_{kl}^t U^t (\hat{Z}_k + \hat{Z}_{\bar{k}}) (\hat{Z}_l + \hat{Z}_{\bar{l}}) U^{t\dagger} \quad (56)$$

$$= E_c - \frac{1}{2} \sum_{tkl} Z_{kl}^t + \sum_{pq} \mathcal{F}_{pq} E_{pq}^+ + \frac{1}{8} \sum_{tkl} Z_{kl}^t U^t (\hat{Z}_k + \hat{Z}_{\bar{k}}) (\hat{Z}_l + \hat{Z}_{\bar{l}}) U^{t\dagger} \quad (57)$$

$$= E_c - \frac{1}{2} \sum_{tkl} Z_{kl}^t + \sum_k \mathcal{F}_k^\varnothing U^{\varnothing\dagger} E_k^+ U^\varnothing + \frac{1}{8} \sum_{tkl} Z_{kl}^t U^t (\hat{Z}_k + \hat{Z}_{\bar{k}}) (\hat{Z}_l + \hat{Z}_{\bar{l}}) U^{t\dagger} \quad (58)$$

Replacing E_k^+ according to (50) and pulling out the $k = l$ terms from the last sum we arrive at

$$H = E_c - \frac{1}{2} \sum_{tkl} Z_{kl}^t + \sum_p \mathcal{F}_p^\varnothing + \frac{1}{8} \sum_{tk} Z_{kk}^t + \sum_k \mathcal{F}_k^\varnothing U^{\varnothing\dagger} (Z_k + Z_{\bar{k}}) U^\varnothing + \quad (59)$$

$$\frac{1}{8} \sum_{tkl} Z_{kl}^t U^t (\hat{Z}_k \hat{Z}_l - \delta_{kl} + \hat{Z}_k \hat{Z}_{\bar{l}} + \hat{Z}_{\bar{k}} \hat{Z}_l + \hat{Z}_{\bar{k}} \hat{Z}_{\bar{l}} - \delta_{\bar{k}\bar{l}}) U^{t\dagger}. \quad (60)$$

Using 69 and 71, we write the total offset in the above equation as:

$$\mathcal{E} = E_c - \frac{1}{2} \sum_{tkl} Z_{kl}^t + \sum_p \mathcal{F}_p^\varnothing + \frac{1}{8} \sum_{tk} Z_{kk}^t \quad (61)$$

$$= E_c - \frac{1}{2} \sum_{pq} (pp|qq) + \sum_p \mathcal{F}_p^\varnothing + \frac{1}{8} \sum_{pq} (pq|pq) \quad (62)$$

Since $\text{trace}(\mathcal{F}_{pq}) = \sum_p \mathcal{F}_p^\varnothing$, we have:

$$\sum_p \mathcal{F}_p^\varnothing = \text{trace}(\mathcal{F}_{pq}) \quad (63)$$

$$= \sum_p \mathcal{F}_{pp} = \sum_p (p|\hat{\kappa}|p) + \sum_r (pp|rr) \quad (64)$$

$$= \sum_p ((p|\hat{h}_c|p) - \frac{1}{2} \sum_r (pr|pr) + \sum_r (pp|rr)) \quad (65)$$

As a result we have:

$$\mathcal{E} = E_c - \frac{1}{2} \sum_{pq} (pp|qq) + \sum_p ((p|\hat{h}_c|p) - \frac{1}{2} \sum_r (pr|pr) + \sum_r (pp|rr)) + \frac{1}{8} \sum_{pq} (pq|pq) \quad (66)$$

$$= E_c - \frac{1}{2} \sum_{pq} (pp|qq) + \sum_p (p|\hat{h}_c|p) - \frac{1}{2} \sum_{pr} (pr|pr) + \sum_{pr} (pp|rr) + \frac{1}{8} \sum_{pq} (pq|pq) \quad (67)$$

$$= E_c + \frac{1}{2} \sum_{pq} (pp|qq) + \sum_p (p|\hat{h}_c|p) - \frac{1}{4} \sum_{pr} (pr|pr) \quad (68)$$

Using (5) again we can identify the the second term in the last expression to be

$$(pp|qq) = \sum_{tkl} U_{pk}^t{}^2 Z_{kl}^t U_{ql}^t{}^2 \quad (69)$$

$$\implies \sum_{pq} (pp|qq) = \sum_{tkl} Z_{kl}^t. \quad (70)$$

Similarly

Moreover, since for $k = l \implies \hat{Z}_k \hat{Z}_l = I$, we have an extra offset of $\sum_{tk} Z_{kl}^t$:

$$(pq|pq) = \sum_{tkl} U_{pk}^t U_{qk}^t Z_{kl}^t U_{pl}^t U_{ql}^t \quad (71)$$

$$\implies \sum_{pq} (pq|pq) = \sum_{tkl} \sum_p U_{pk}^t U_{pl}^t Z_{kl}^t \sum_q U_{qk}^t U_{ql}^t \quad (72)$$

$$= \sum_{tkl} \delta_{kl} Z_{kl}^t \delta_{kl} \quad (73)$$

$$= \sum_{tk} Z_{kk}^t \quad (74)$$

Appendix G: Comparison of RC-DF and FFF

The Fluid Fermionic Fragments (FFF) method is based on the fact that some contributions to the Hamiltonian can be moved back and fourth freely between the second and third term of the electronic structure Hamiltonian as written in (1). In the fermionic picture, these "fluid" parts of the Hamiltonian correspond to terms that are quadratic in the creation and annihilation operators (see (9) and (10) in Ref [21]) and which, after diagonalization of the quadratic part contribute to the terms proportional to particle number operators, and which under Jordan Wigner yield Pauli \hat{Z} operators.

Here we present the FFF method in the qubit picture. To that end, starting from (1), we first factorize the $(pq|rs)$ part of the Hamiltonian only, which yields:

$$\hat{H} = E_0 + \sum_{pq} (p|\hat{\kappa}|q) E_{pq}^+ + \frac{1}{2} \sum_{pqrs} (pq|rs) E_{pq}^+ E_{rs}^+ \quad (75)$$

$$= E_0 + \sum_{pq} (p|\hat{\kappa}|q) E_{pq}^+ + \frac{1}{2} \sum_t U^t \left(\sum_{kl} Z_{kl}^t E_{kk}^+ E_{ll}^+ \right) U^{t\dagger} \quad (76)$$

$$= E_0 + \sum_{pq} (p|\hat{\kappa}|q) E_{pq}^+ + \frac{1}{2} \sum_t U^t \left(\sum_{kl} Z_{kl}^t (-I + E_{kk}^+ + E_{ll}^+ + \frac{1}{4} (\hat{Z}_k + \hat{Z}_{\bar{k}}) (\hat{Z}_l + \hat{Z}_{\bar{l}})) \right) U^{t\dagger} \quad (77)$$

$$= E_0 - \frac{1}{2} \sum_{tkl} Z_{kl}^t + \sum_{pq} (p|\hat{\kappa}|q) E_{pq}^+ + \sum_t U^t \left(\sum_k \left(\sum_l Z_{kl}^t \right) E_{kk}^+ \right) U^{t\dagger} \quad (78)$$

$$+ \frac{1}{8} \sum_t U^t \left(\sum_{kl} Z_{kl}^t ((\hat{Z}_k + \hat{Z}_{\bar{k}}) (\hat{Z}_l + \hat{Z}_{\bar{l}})) \right) U^{t\dagger} \quad (79)$$

Now we can add and subtract terms of the form $U^t (\sum_k c_k^t E_{kk}^+) U^{t\dagger} = \sum_k c_k^t \sum_{pq} u_{pk}^t u_{qk}^t E_{pq}^+$ with c_k^t the FFF coefficients to obtain

$$\hat{H} = E_0 - \frac{1}{2} \sum_{tkl} Z_{kl}^t + \sum_{pq} (p|\hat{\kappa}|q) E_{pq}^+ - \sum_t U^t \left(\sum_k c_k^t E_{kk}^+ \right) U^{t\dagger} + \sum_t U^t \left(\sum_k \left(\sum_l Z_{kl}^t + c_k^t \right) E_{kk}^+ \right) U^{t\dagger} \quad (80)$$

$$+ \frac{1}{8} \sum_t U^t \left(\sum_{kl} Z_{kl}^t ((\hat{Z}_k + \hat{Z}_{\bar{k}}) (\hat{Z}_l + \hat{Z}_{\bar{l}})) \right) U^{t\dagger} \quad (81)$$

$$= E_0 - \frac{1}{2} \sum_{tkl} Z_{kl}^t + \sum_{pq} ((p|\hat{\kappa}|q) - \sum_t \sum_k u_{pk}^t u_{qk}^t c_k^t) E_{pq}^+ + \sum_t U^t \left(\sum_k \left(\sum_l Z_{kl}^t + c_k^t \right) E_{kk}^+ \right) U^{t\dagger} \quad (82)$$

$$+ \frac{1}{8} \sum_t U^t \left(\sum_{kl} Z_{kl}^t ((\hat{Z}_k + \hat{Z}_{\bar{k}}) (\hat{Z}_l + \hat{Z}_{\bar{l}})) \right) U^{t\dagger} \quad (83)$$

$$= \mathcal{E}' + U^{\prime\varnothing\dagger} \sum_k c_k^{\varnothing} \hat{Z}_k U^{\prime\varnothing} - \frac{1}{2} \sum_t U^t \left(\sum_k \left(\sum_l Z_{kl}^t + c_k^t \right) \hat{Z}_k \right) U^{t\dagger} \quad (84)$$

$$+ \frac{1}{8} \sum_t U^t \left(\sum_{kl} Z_{kl}^t (\hat{Z}_k \hat{Z}_l - \delta_{kl} + \hat{Z}_k \hat{Z}_{\bar{l}} + \hat{Z}_{\bar{k}} \hat{Z}_l + \hat{Z}_{\bar{k}} \hat{Z}_{\bar{l}} - \delta_{\bar{k}\bar{l}}) \right) U^{t\dagger} \quad (85)$$

with

$$\mathcal{E}' = E_0 - \frac{1}{2} \sum_{tkl} Z_{kl}^t + \sum_k \mathcal{C}_{kk} + \sum_{tk} \left(\sum_l Z_{kl}^t + c_k^t \right) + \frac{1}{4} \sum_{tk} Z_{kk}^t \quad (86)$$

$$\mathcal{C}_{pq} = (p|\hat{\kappa}|q) - \sum_t \sum_k u_{pk}^t u_{qk}^t c_k^t \quad (87)$$

and U'^{\varnothing} and $\mathcal{C}_k^{\varnothing}$ are the diagonalizing unitaries and eigenvalues of \mathcal{C}_{pq} .

The case all $c_k^t = 0$ corresponds to how X-DF was introduced in [18] and this was taken as the prior art benchmark in [21]. The case

$$c_k^t = - \sum_l Z_{kl}^t \quad (88)$$

corresponds to the way we wrote the Hamiltonian in (6), with no single qubit \hat{Z} contributions in the two body leafs. It turns out that neither of these choices is optimal with respect to variance and thus shot count and optimizing the c_k^t can further yield improvements,

Optimization can be done with a gradient based optimizer and in [21] as well as here we used **LBFGSB** as implemented in `scipy`. The number of shots is then optimized, via a proxy state, using a nested loop where in each iteration the coefficients c_k^t are updated using the partial derivative at fixed shot distribution then the shots are optimally distributed according to the variances computed with the new c_k^t in the proxy state. For simplicity and better comparability with the results from [21] we use the exact ground state as the proxy state. This is not efficient but [21] found little difference between using the real ground state and an approximate proxy state for which variances can be computed efficiently. We compute the needed derivatives by means of a fully auto-differentiable code that computes the variances as a function of the c_k^t [29, 36, 41].

For some cases for which we have performed simulations (see Figure 8) we find that all $c_k^t = 0$ and/or X-DF with the c_k^t corresponding to (6) is a local minimum and hence gradient based optimization of the c_k^t does not work. We consistently found good final shot counts from random uniformly distributed within $[0, 1[$ initializations. Alternatively one can initialize from coefficients according to (88) which lead to faster convergence but seems to yield the same final or very similar shot budgets. Overall we find that combining RC-Df with FFF yields the lowest shot budgets. The term mapping used in (6) is significantly better than choosing all $c_k^t = 0$ and RC-DF (with and without FFF) consistently outperforms X-DF.

Table 1: RC-DF performance summary for Cpd I

Conv. Tol.	n_t	$\ \Delta_{pqrs}\ _{\mathcal{F}}$	CCSD(T) error [mEh]	$\lambda_{DF}^{\text{Burg}}$	$\lambda_{DF}^{\text{LCU}}$
10^{-1}	30	0.4472	-23.5907	300.6	456.5
	40	0.4412	-22.6134	311.0	468.6
	50	0.4382	-22.4729	320.6	478.0
	60	0.4415	-22.7342	327.5	478.1
	70	0.4452	-21.4039	336.3	483.7
	80	0.4352	-18.2827	343.0	490.5
	90	0.4310	-17.3526	349.9	496.1
	100	0.4278	-17.2627	355.8	498.1
	150	0.4232	-14.8979	385.9	521.0
	200	0.4258	-15.0266	417.9	545.2
10^{-2}	30	0.1411	-3.0769	278.2	430.4
	40	0.1411	-3.6537	290.8	448.5
	50	0.1408	-2.9621	299.5	462.7
	60	0.1414	-3.2143	304.6	470.0
	70	0.1406	-2.8366	308.6	475.4
	80	0.1407	-3.6169	313.8	482.2
	90	0.1407	-2.6580	318.7	489.2
	100	0.1409	-2.1238	321.7	492.1
	150	0.1410	-2.6875	337.2	508.8
	200	0.1413	-2.5034	351.2	524.2
10^{-3}	30	0.0447	-0.2292	244.4	386.6
	40	0.0447	-0.3581	257.1	394.1
	50	0.0447	-0.2405	268.0	409.6
	60	0.0447	-0.2515	276.5	422.8
	70	0.0447	-0.5585	284.0	434.3
	80	0.0447	-0.1995	288.6	442.6
	90	0.0447	-0.3395	292.2	450.4
	100	0.0447	-0.0137	295.8	456.6
	150	0.0446	-0.4667	310.3	481.2
	200	0.0446	-0.2690	321.1	499.3
10^{-4}	100	0.0141	0.0086	258.0	390.8
	150	0.0141	-0.0423	275.7	414.4
	200	0.0141	-0.0340	287.9	435.3

Table 2: X-DF performance summary for Cpd I

n_t	$\ \Delta_{pqrs}\ _{\mathcal{F}}$	CCSD(T) error [mEh]	$\lambda_{\text{DF}}^{\text{Burg}}$	$\lambda_{\text{DF}}^{\text{LCU}}$
30	1.01455	15.5688	418.5	759.2
40	0.72606	-7.3444	431.6	784.9
50	0.58478	-8.5971	440.3	802.0
60	0.47298	-14.4322	447.9	816.8
70	0.38068	-3.1890	453.7	828.2
80	0.29495	-9.6247	458.8	838.3
90	0.21062	0.5473	462.2	844.9
100	0.15487	-3.9773	464.3	848.9
150	0.04288	-0.0570	469.8	859.8
200	0.01609	0.1208	471.7	863.4
400	0.00126	0.0092	473.0	866.0
800	0.00001	0.0003	473.2	866.3

Table 3: Truncated DF performance summary for Cpd I^{a)}

n_t ^{b)}	threshold ^{c)}	$\ \Delta_{pqrs}\ _{\mathcal{F}}$	CCSD(T) error [mEh]	$\lambda_{\text{DF}}^{\text{Burg}}$
99	0.07500	0.4909	-73.8021	428.8
114	0.05000	0.3631	-17.4890	439.9
131	0.02500	0.1892	1.1578	455.1
178	0.01000	0.0836	2.6993	464.3
194	0.00750	0.0644	4.1638	466.1
207	0.00500	0.0457	1.8040	467.9
241	0.00250	0.0245	-0.1019	470.1
309	0.00100	0.0106	0.0266	471.7
364	0.00050	0.0054	0.0155	472.3
505	0.00010	0.0012	-0.0086	473.0
568	0.00005	0.0006	0.0037	473.1
702	0.00001	0.0001	0.0007	473.1

^{a)} Recomputed with `chemftr`.

^{b)} Referred to as L in Refs. [12, 17].

^{c)} Eigenvector screening threshold with which the accuracy of the factorization is tuned, see Ref. [12].

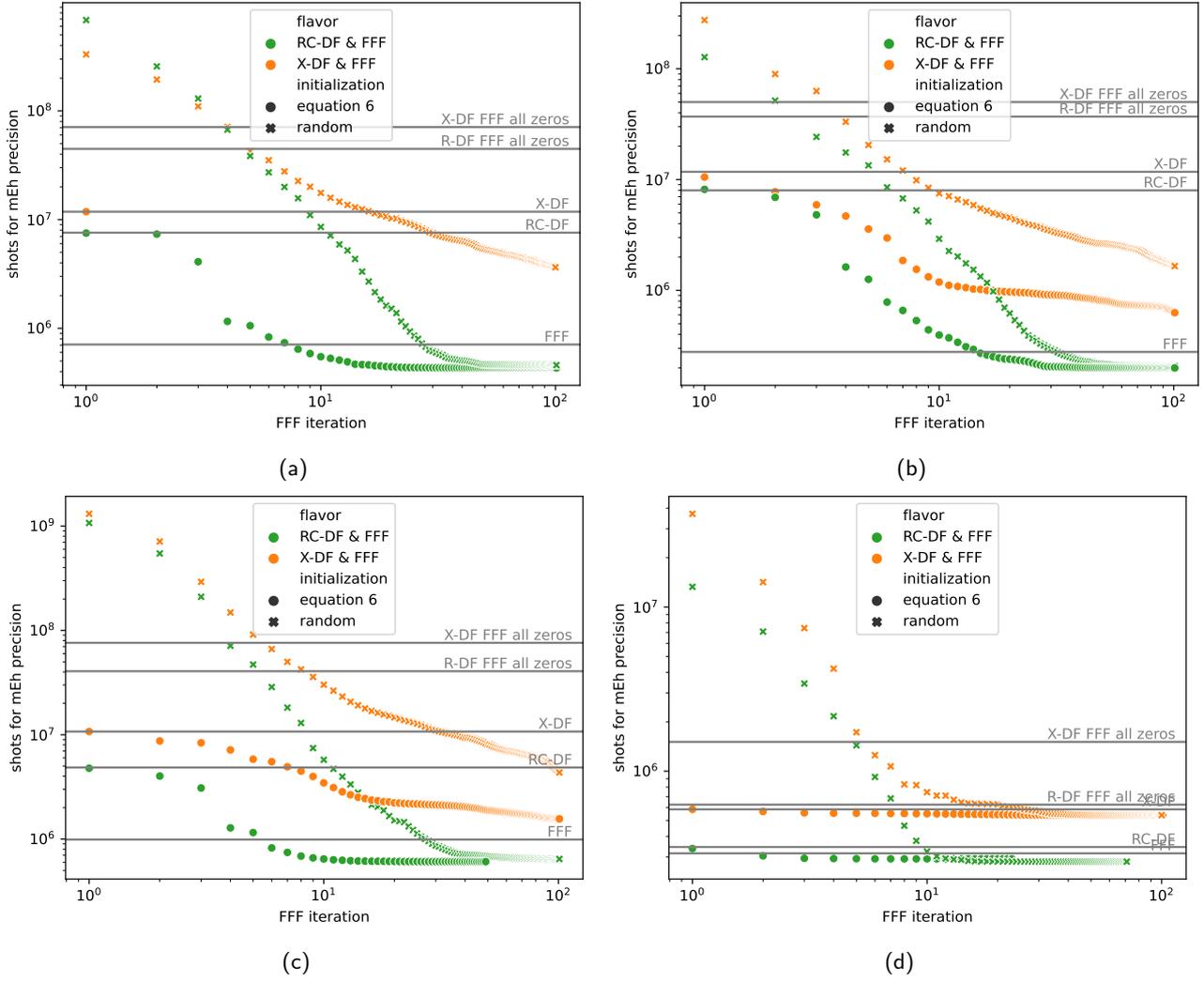


Figure 8: Number of shots to reach mili Hartree precision with optimal shot distribution for combinations of X-DF and RC-DF with the fluid fermionic fragments (FFF) method for (a) H_2O with $n_t = 24$ and $(10e, 7o)$ active space, (b) HF with $n_t = 16$ and $(10e, 6o)$ active space, (c) NH_3 with $n_t = 28$ and $(10e, 8o)$ active space and (d) H_4 with $n_t = 10$ and $(4e, 4o)$ active space (all with RHF orbitals in the STO-3G basis). The gray horizontal lines are the number of shots when the Hamiltonians are measured as written in (85) with all FFF coefficients c_k^t equal to zero (X-DF/RC-DF all zero) or as written in (6) (X-DF/RC-DF) and the best original results of FFF from [21]. The examples (a) to (c) were specifically picked because the gap between the plain RC-DF shot budget and the FFF shot budget from [21] were large. For other cases, plain RC-DF already yields similar results to FFF. The crosses/dots show how the number of shots decreases during optimization of the c_k^t FFF coefficients (with the true ground state taken as proxy state for simplicity) in (85) after initializing them randomly according to a uniform distribution within $[0,1[$ such that the initial Hamiltonian coincides with (6).