Assessing MP2 frozen natural orbitals in relativistic correlated electronic structure calculations

Xiang Yuan, Lucas Visscher, André Severo Pereira Gomes

Assessing MP2 frozen natural orbitals in relativistic correlated electronic structure calculations

Xiang Yuan,[1, 2, a)] Lucas Visscher,[2, b)] and André Severo Pereira Gomes[1, c)]
[1)]*Université de Lille, CNRS, UMR 8523 - PhLAM - Physique des Lasers, Atomes et Molécules, F-59000 Lille, France.*
[2)]*Department of Chemistry and Pharmaceutical Sciences, Faculty of Science, Vrije Universiteit Amsterdam, de Boelelaan 1083, 1081 HV Amsterdam, The Netherlands.*

The high computational scaling with basis set size and the number of correlated electrons is a bottleneck limiting applications of coupled cluster (CC) algorithms, in particular for calculations based on 2- or 4-component relativistic Hamiltonians, which often employ uncontracted basis sets. This problem may be alleviated by replacing canonical Hartree-Fock (HF) virtual orbitals by natural orbitals (NOs). In this paper we describe the implementation of a module for generating NOs for correlated wavefunctions, and in particular MP2 frozen natural orbitals (MP2FNOs), as a component of our novel implementation of relativistic coupled cluster theory for massively parallel architectures [J. Pototschnig et. al. *J. Chem. Theory Comput.* **17**, 5509, 2021]. Our implementation can manipulate complex or quaternion density matrices, thus allowing for the generation of both Kramers-restricted and Kramers-unrestricted MP2FNOs. Furthermore, NOs are re-expressed in the parent atomic orbital (AO) basis, allowing for generating CCSD NOs in AO basis for further analysis. By investigating the truncation errors of MP2FNOs for both the correlation energy and molecular properties-electric field gradients at the nuclei, electric dipole and quadrupole moments for hydrogen halides HX (X=F-Ts), and parity-violating energy differences for H_2Z_2 (Z=O-Se)-we find MP2FNOs accelerate the convergence of the correlation energy in a roughly uniform manner across the periodic table. It is possible to obtain reliable estimates for both energies and the molecular properties considered with VMO spaces truncated to about half the size of the full spaces.

I. INTRODUCTION

Understanding the electronic structure of large molecules or complexes containing heavy elements such as lanthanides or actinides is a problem of relevance for many technological applications. Examples are the nuclear fuel cycle, such as in the development of new extractants for separation processes[1–6], and the use of lanthanides as exceptionally strong single molecule magnets[7]. To model such materials, Density Functional Theory (DFT) has become the most widely used correlated electronic structure theory approach[8], even though it is difficult to systematically approach exact results with the currently available density functional approximations[8,9]. In the particular case of relativistic electronic structure calculations, DFT energies may even for closed-shell species strongly deviate from experimental or accurate theoretical results[10–12].

This also holds for molecular properties, recently Sunaga and Saue[13] reported that the performance of DFT for parity violation energy shift (PV) calculations – a property requiring a very accurate description of the electronic wave function near the nuclei – is somewhat disappointing, with deviations to CCSD being as large as 10%. These uncertainties in the performance of of DFT for heavy elements, especially for cases in which experimental values are absent or difficult to generate, calls for the use of state-of-the-art wavefunction methods to either provide accurate reference data, or be applied directly if the model sizes are small enough.

Among currently available approaches, coupled cluster (CC) theory serves as a "gold standard"[14,15] for its ability to give results approaching chemical accuracy for both correlation energies and properties. However, the main difficulty in employing CC theory in large-scale applications (such as molecules containing several hundred electrons) is its high computational scaling with respect to the size of the system (N). For example, without approximations CCSD and CCSD(T) approaches scale with $O(N^6)$ and $O(N^7)$, respectively, and also with approximations such as Laplace transforms[16], distance screening[17] and density fitting[18,19] the scaling and prefactors are still significantly higher than in a mean-field approach like DFT.

Another key ingredient in modeling heavy element species is the treatment of relativistic effects[20–22]. For cases in which a given molecular property is not particularly sensitive to effects such as spin-orbit coupling (SOC) (which is often so for molecular structural parameters such bond lengths and angles), or to contributions from electrons other than those in the valence (e.g. dipole moments), approximate treatments of relativity can be employed, through pseudopotential approaches[23,24], or by including only scalar relativistic effects. For more challenging applications, or higher precision, a more general framework can be based on the solution of the 4-component (4C) molecular Dirac equation[25–28]. Nowadays this approach is often made more tractable by solving the exact 2-component (X2C) equation that can be derived from the Dirac equation after a basis set discretization[29–39]. Both the original 4C approach as well as its X2C approximation can deliver accurate molecular properties across the periodic table, also for properties involving core electrons[10,13,40–45].

The one-electron functions (molecular spinors, for simplicity also referred to in the following as molecular orbitals) obtained from solving the 4C or X2C matrix equations, serve

[a)]Electronic mail: xiang.yuan@univ-lille.fr
[b)]Electronic mail: l.visscher@vu.nl
[c)]Electronic mail: andre.gomes@univ-lille.fr

then as a basis for a correlation treatment in the so-called no-pair approximation in which contributions to the correlation energy due to admixture of states with explicit electron-positron pairs are neglected[46]. As the computational cost of 2C approaches is lower than 4C ones, the choice between which treatment to use prior to the correlated treatment will depend on a case-by-case analysis of whether the additional cost of the latter will be offset by improvements in accuracy over the former.

Recently, Pototschnig *et al.*[47] described a new, efficient relativistic coupled cluster implementation based on ExaTENSOR[48], a distributed numerical tensor algebra library for GPU-accelerated HPC platforms. This code enabled the calculation of molecular properties with CCSD wave functions for systems such as $[(UO_2)(NO_3)_3]^-$, for which 200 electrons and around 1000 virtual molecular orbitals were included in the correlated treatment. Compared to nonrelativistic implementations, it is in particular the introduction of SOC which increases the computational cost (though not altering the overall scaling)[49]. Not only does this necessitate the use of complex algebra, it also makes use of contracted basis sets more difficult as one needs to be able to describe the differences in radial extent of spin-orbit split orbitals. This effectively doubles the number of functions that is needed to describe the *p*-, *d*- and *f*-type core orbitals[50] in the case of implementations based on expressing spinors in scalar basis sets[26,28,50]. In this framework another complication arises in how to define contractions for the small component part of the spinors that would respect the kinetic balance condition. We note that these problems do not arise in implementations capable of directly evaluating 2-electron integrals over 2C spinor basis functions[27,51].

The lack of efficient contraction schemes for scalar basis-derived implementations, combined with the fact that in many potential applications it may be required that all, or a significant part of core electrons (e.g. in core spectroscopies[45]) are treated explicitly, including the deep core orbitals that require many Gaussian type functions to be described correctly, leads to calculations having large virtual spaces. This directly affects the performance of CC-based approaches as these scale with the fourth power of the number of virtuals ($O(V^4)$). In practice one therefore often alleviates the computational effort to some extent by leaving out high energy virtuals that are mostly localized in the core, but convergence with respect of such an energy cut-off threshold is slow, especially when semi-core or core correlation need to be described as well.

In non-relativistic electronic structure theory it has long been recognized that, in contrast to canonical Hartree-Fock molecular orbitals (CMO), natural orbitals (NOs)–the eigenvectors of the one-body reduced density matrix (1RDM)[52,53]–provide a more compact and quickly converging orbital representation for describing post Hartree-Fock wavefunctions. Based on this observation, the idea of replacing the CMOs by NOs to reduce the size of the virtual space and thereby computational cost was introduced[54].

Rather than using an energy threshold as is done by CMOs, one may instead omit NOs which are likely to not strongly contribute to the total correlation energy from the virtual space. This is done by considering the magnitude of natural occupation numbers of approximate NOs, obtained from a method that quickly gives access to a reasonable approximation of the 1RDM of the correlated wave function. For this purpose, second order Møller–Plesset perturbation theory (MP2) is a particularly appealing approach to obtain the 1RDM and the approximate NOs because of its low scaling, non-iterative $O(N^5)$, and the ability to recover most of correlation effects. Within the natural orbital family of methods, the virtual frozen natural orbitals (FNOs) approach[54–56] has therefore become popular because of its clear concept and simple implementation. FNO theory has been thoroughly discussed for non-relativistic models such as configuration interaction (CI)[54], multi-configuration self-consistent field (MCSCF)[57,58] and coupled cluster[56,59]. Recently, Verma *et al.*[60] furthermore extended the FNOs algorithm to quantum computers.

The ability of virtual FNOs to reduce computational cost compared to CMOs is even more appealing in relativistic electronic structure calculations, in particular in the case of contracted basis sets or when correlating sub-valence electrons. The main goal of this paper is therefore to describe and showcase the implementation of MP2-based FNOs (MP2FNOs) within the framework of the new relativistic coupled cluster implementation for massively parallel, GPU-accelerated platforms[47]. While our code primarily aims at describing cases in which spin-orbit coupling is taken into account, we also demonstrate its use with a non-relativistic Hamiltonian.

Our second aim is to discuss the performance of MP2FNOs across the periodic table, by treating model systems containing elements ranging from first-row such as fluorine, to superheavy elements such as tennessine. Here, we shall focus in how well truncated FNO spaces describe both correlation energies as well as first-order properties such as the electric field gradient at the nuclei (EFGs), parity-violation energy shifts, and electric multipoles (dipole and quadrupole). The first two properties are chosen as representatives of properties for which a good description of wavefunctions in the region close the nuclei is important (even if only valence electrons are correlated), while the electric dipole and quadrupole moments are an example of properties for which the major contributions arise from valence electrons.

Besides being able to improve the efficiency of a calculation, natural orbitals are also interesting as tools for analysis, such as in estimating the multi-reference character of a system[61]. Their visualization in real space can furthermore provide insight into correlation effects, since by their one-particle nature they are easier to interpret than the wavefunction itself. Thus, another aim of our implementation was to provide a tool for obtaining natural orbitals for any correlated wavefunction, from a 1RDM. In this paper we shall make use of the analysis of the CCSD natural orbitals, and in subsequent work we plan to further explore the use of natural orbitals in visualization.

This article is organized as follows. In Sec. II the background of MP2FNOs theory is summarized. In Sec. III we described the details of implementation. All the sample calculations are presented and discussed in Sec.IV-V. And finally a brief summary is given in Sec. VI.

We note that upon completing this manuscript, we have

become aware of another implementation of the MP2FNOs approach for relativistic correlated methods, in the BAGH code[62]. While the main features of the MP2FNOs method are the same in both implementations we first note that our implementation fully exploits ExaTENSOR's single-node or distributed memory (multi-node) and GPU acceleration capabilities, and as such can be efficiently employed in systems ranging from local clusters to latest-generation supercomputers. Second, as it will be outlined below our implementation is capable of manipulating both complex and quaternion density matrices, thus allowing for the generation of both Kramers-restricted and Kramers-unrestricted MP2FNOs. Finally, it allows for re-expressing NOs in atomic orbital (AO) basis for further analysis.

II. THEORY

As the MP2FNO approach is well-known in a non-relatistic context and requires essentially no modification for application in a no-pair relativistic context, we will only provide a brief description. We apply the orbital-unrelaxed MP2 approximation (for working equations for the orbital-relaxed formalism in a relativistic context see reference 63). The second-order contribution to the occupied-virtual block of the density matrix from single excitations is zero for canonical orbitals[53] while the relaxation contributions to this block are ignored in orbital-unrelaxed MP2. This approximation thus decouples the occupied and virtual spaces and allows us to focus only on obtaining the virtual-virtual block of the density matrix. We want to keep occupied orbitals in their canonical form and therefore will not determine the occupied-occupied part of the density matrix, for which we simply retain the diagonal Hartree-Fock form $\boldsymbol{D}^{(0)}_{oo} = \mathbf{1}_{oo}$. With these approximations the second-order FNOs density matrix is given as

$$\boldsymbol{D}^{FNO} = \begin{bmatrix} \mathbf{1}_{oo} & \mathbf{0}_{ov} \\ \mathbf{0}_{ov} & \boldsymbol{D}^{(2)}_{vv} \end{bmatrix} \tag{1}$$

with the relevant matrix elements given as

$$D^{(2)}_{AB} = \frac{1}{2} \sum_{C,I,J} \frac{\langle AC||IJ\rangle \langle IJ||BC\rangle}{\varepsilon^{AC}_{IJ}\varepsilon^{BC}_{IJ}} \tag{2}$$

Here, following usual conventions *I, J* denote occupied spinors and *A, B* virtual spinors, and $\varepsilon^{AB}_{IJ} = \varepsilon_I + \varepsilon_J - \varepsilon_A - \varepsilon_B$, is the energy difference between canonical Hartree-Fock spinors. In our implementation, these matrices are obtained without imposing any time-reversal symmetry on the spinor set and are thus valid for both Kramers-restricted and Kramers-unrestricted Hartree-Fock procedures. If time-reversal[64] and spatial symmetry[65] are not present, these matrices are in general to be represented in complex algebra.

The simplest procedure to obtain natural spinors is to apply a diagonalization in complex algebra after $\boldsymbol{D}^{(2)}_{vv}$ is formed. However, for closed shell systems in which time-reversal symmetry is enforced for the orbitals at the mean-field level, as is the case in the DIRAC code, one would like to retain such Kramers pairing also for the natural spinors. This requires careful attention when degeneracies beyond the two-fold Kramers degeneracy are present, as complex diagonalization will arbitrarily mix the degenerate solutions such that Kramers pairing is not guaranteed.

In such cases, Kramers pairing of the natural spinors can be enforced by first transforming $\boldsymbol{D}^{(2)}_{vv}$ from complex to quaternion algebra $\boldsymbol{D}^{(Q)}_{vv}$,

$$\boldsymbol{D}^{(Q)}_{vv} = \boldsymbol{U}^{\dagger}\boldsymbol{D}^{(2)}_{vv}\boldsymbol{U} \tag{3}$$

where the tranformation matrix $\boldsymbol{U}$ is given by[66]

$$\boldsymbol{U} = \begin{pmatrix} \mathbb{I} & \check{j}\mathbb{I} \\ \check{j}\mathbb{I} & \mathbb{I} \end{pmatrix} \tag{4}$$

and $\mathbb{I}$ is a unit matrix of dimension $n \times n, n$ being the number of Kramers pairs. This transformation will block diagonalize $\boldsymbol{D}^{(2)}_{vv}$, leading to two decoupled matrix problems of half the original dimension in quaternion algebra.

$$\boldsymbol{D}^{(Q)}_{vv}\boldsymbol{V} = \boldsymbol{O}\boldsymbol{V} \tag{5}$$

Diagonalization of one of the two blocks provides a unique set (V) of quaternion FNOs in MO basis and their respective occupation numbers (O) from which the solutions for the other block can be generated via Kramers symmetry.

At this stage, we can employ the information from the occupation numbers to reduce the original virtual space $\boldsymbol{V}$ to $\bar{\boldsymbol{V}}$ by removing from $\boldsymbol{V}$ orbitals with occupation numbers lower than a user-defined threshold. The larger such threshold is, the smaller the remaining set will be. In the numerical examples presented in the following we will try to find the optimal balance between efficiency (smaller sets) and accuracy (recovering more of the correlation energy and better reproducing first order properties).

After the basis is truncated, it is convenient to recanonize the remaining orbitals. To this end, we perform two operations: first we transform the virtual-virtual block of the Fock matrix ($\boldsymbol{F}_{vv}$) into the truncated FNOs basis ($\bar{\boldsymbol{F}}_{vv}$) and then do a (quaternion re)diagonalisation to obtain a new set of canonical orbitals $\bar{\boldsymbol{W}}$ and orbital energies $\bar{\boldsymbol{\epsilon}}$.

$$\bar{\boldsymbol{F}} = \bar{\boldsymbol{V}}^{\dagger}\boldsymbol{F}\bar{\boldsymbol{V}} \tag{6}$$

$$\bar{\boldsymbol{F}}\bar{\boldsymbol{W}} = \bar{\boldsymbol{W}}\bar{\boldsymbol{\epsilon}} \tag{7}$$

The product $\bar{\boldsymbol{V}}\bar{\boldsymbol{W}}$ of these two transformation matrices gives the transformation that expresses the recanonized truncated natural orbital set in the atomic orbital(AO) basis. These orbitals can then be used in any subsequent correlated calculation as replacement to the original Hartree-Fock orbitals. In summary the transformation of the original Hartree-Fock orbitals $\boldsymbol{U}$ to the new set $\bar{\boldsymbol{U}}$ is thus given by

$$\boldsymbol{U}_{new} = [\boldsymbol{U}_{occ}, \bar{\boldsymbol{U}}_{vir}] \tag{8}$$

where

$$\bar{\boldsymbol{U}}_{vir} = \boldsymbol{U}_{vir}(\bar{\boldsymbol{V}}\bar{\boldsymbol{W}}) \tag{9}$$

$$\bar{\boldsymbol{U}}_{occ} = \boldsymbol{U}_{occ} \tag{10}$$

and the dimension of the rectangular matrix $\bar{\boldsymbol{V}}$ and that of the square matrix $\bar{\boldsymbol{U}}$ depends on the truncation threshold.

III. IMPLEMENTATION

The aforementioned algorithm has been implemented in the relativistic quantum chemistry package DIRAC[26], as part of the ExaCorr code[47]. Our implementation allows for calculations to be carried out either using a single-node configuration (for which ExaTENSOR provides OpenMP parallelization, on top of GPU offloading) or employing distributed memory in the case of multi-node runs.

Our implementation is structured following its three main tasks: one module deals with the construction of density matrices in the MO basis, another carries out the complex or quaternion diagonalization of density matrices, and the final module takes care of the construction of the recanonized MP2FNOs. The final step should be repeated if the truncation treshold is changed, the first two are independent of this threshold.

A. Density matrix construction

In Exacorr all computationally expensive operations such as the tensor contractions used in the determination of amplitudes or the construction of density matrices are offloaded to ExaTENSOR library.

In the case of MP2FNOs, we have created a dedicated driver for MP2 calculations, in which we (i) calculate and store in memory two electron repulsion integrals (ERIs) in AO basis with the aid of the InteRest library[67]; (ii) employ the standard Yoshimine scheme[68] (which scales as $(O(N^5))$ to transform the AO integrals to the MO basis. Only the direct and exchange contributions to the $\langle ij||ab\rangle$ (OOVV-type) integrals are calculated, which makes the integral transformation step much faster than the generation of a complete set of MO integrals as is normally done in ExaCorr. After antisymmetrization the $\langle ij||ab\rangle$ integrals are stored (in memory) as ExaTENSOR tensors, which may reside on a single node or be distributed over several nodes. Finally, (iii) the MP2 amplitudes tensor and the MP2 energy are determined and the density matrix $\boldsymbol{D}^{(2)}_{vv}$ is constructed according to equation 2.

B. Complex-quaternion transformation and diagonalization

The density matrix computed in Exacorr is generated in complex algebra. The complex-quaternion transformation is carried out following Eqn. 27 in ref. 69

$$^{Q}\gamma_{pq} = \mathrm{Re}(\gamma_{pq}) + \check{i}\mathrm{Im}(\gamma_{pq}) + \check{j}\mathrm{Re}(\gamma_{p\bar{q}}) + \check{k}\mathrm{Im}(\gamma_{p\bar{q}}), \quad (11)$$

in which lowercase symbols with (without) bars indicate the Kamers pairing of the original MO basis and the one-particle reduced density matrix (1RDM) is now indicated by γ for consistency of notation with that of ref. 69. This quaternion form can be diagonalized using the quaternion diagonalization routine provided by DIRAC and be back-transformed to complex representation by the routines provided in this module. If the original basis did not possess Kramers symmetry, the diagonalization is directly carried out in complex algebra. The resulting full set of FNOs is then stored on file for analysis and processing by module (C) described below. The module responsible for the quaternion transformation and diagonalization is constructed in such a way that it can also be used with wave function models other than MP2.

C. Selection and recanonization

The third module retrieves FNOs that have occupation numbers above treshold and transforms the original Fock matrix to the truncated virtual space. After recanonization (Eqn. 7) and transformation to AO-basis (Eqn. 10) the final reduced set of FNOs is stored on file. As this step takes virtually no time compared to the other steps in the procedure, this may be easily repeated to test out the effect of varying the treshold value on the generated recanonized orbitals and their energies.

D. Summary of the MP2FNO Algorithm

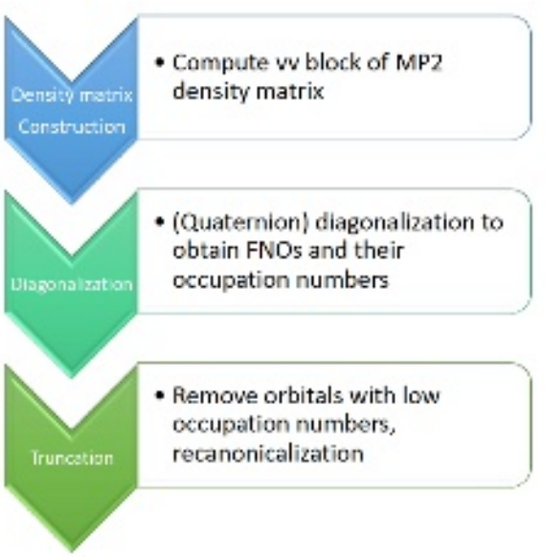


Figure 1. Work flow of a MP2 Frozen natural orbitals calculation.

The MP2FNO workflow is schematically represented in Figure 1, and consists of the following steps:

1. A calculation in the full basis set is performed to get a set of occupied and virtual molecular orbitals.

2. The virtual-virtual block of MP2 density matrix is generated. We note that in this step is possible to either consider the full virtual set, or already employ a preliminary screening in which for instance very high-energy virtuals are neglected.

3. Optional, only for restricted starting orbitals: Transform the complex $\boldsymbol{D}_{vv}^{(2)}$ into the quaternion representation, $\boldsymbol{D}_{vv}^{(Q)}$.

4. Diagonalize $\boldsymbol{D}_{vv}$ or $\boldsymbol{D}_{vv}^{(Q)}$ to determine FNOs and occupation numbers.

5. For a given threshold: truncate the FNO space and transform the Fock matrix to recanonize and write the orbitals out in the AO basis.

6. Use the recanonized FNOs in higher level wavefunction models such as coupled cluster.

IV. COMPUTATIONAL DETAILS

All MP2 and CC calculations are carried out with development versions (revisions `2e659d7, 0f8e9f2, 9e10bc667, 8b81f8a, e7d2d4d, ec415a5, 0ff0d6f, d70bbe2`) of the DIRAC code[26,70,71], and of the ExaTENSOR[48,72] library (revision `d304c03b7`), employing Dyall basis sets of triple-zeta quality (dyall.av3z)[73–75] for the heavy elements (Br, I, At, Ts, Se), and Dunning aug-cc-pVTZ sets[76–78] for (H, F, Cl, O, S), all of which are left uncontracted unless otherwise noted. The aug-cc-pCVTZ basis set was also employed in the HCl case[79,80], for investigating how the additional core correlating functions in aug-cc-pCVTZ affect the results. Finally, for investigating the convergence of expectation values for HI, we employed double-zeta quality Dyall[73,74] (dyall.av2z) and Dunning[76–78] (aug-cc-pVDZ) basis sets for I and H respectively.

Here we make use in all calculations of the exact two component Hamiltonian (X2C)[32], in which we include two-electron spin-orbit contributions via to the untransformed two-electron potential via atomic mean-field contributions calculated with the AMFI code[81–83]. In our calculations we have made use of Kramers symmetry in the generation of the molecular spinors[26]. As part of the supplementary information, we provide in table S2 results with different Hamiltonians (Dirac-Coulomb, X2C, spinfree X2C and Non-relativistic) for expectation values for the HTs molecule. We note the code is equally capable of handling orbitals obtained with the so-called molecular-mean-field approach, in which the transformation to 2C is carried out after a 4C mean-field calculation[39]. In the case of the HCl molecule, we have also employed the non-relativistic Hamiltonian (as specified by the `.NONREL` keyword), both with contracted and uncontracted bases sets.

The molecular structures employed in all calculations have been taken from the literature: In the case of hydrogen halides, from Huber[84] for HF to HI, from Gomes and Visscher[41] for HAt and from Thierfelder *et al.*[42] for HTs. The internuclear distances employed are thus H-F(0.9168 Å); H-Cl(1.27455 Å); H-Br(1.41443 Å); H-I(1.609 Å); H-At(1.722 Å); H-Ts(1.941 Å). For the chiral molecules H_2Z_2, the Z-Z bond length, H-Z bond length and H-Z-Z bond angle are taken from Table I of Laerdahl and Schwerdtfeger[85] and the dihedral angle is fixed at 45 degrees throughout the computation. All calculations have been carried out with a point charge nucleus model.

Besides calculating MP2, CCSD and CCSD(T) energies, we have obtained electric dipole moment (EDM), electric quadrupole moment (EQM) and electric field gradient (EFG) for the HX systems, and parity violation energy differences (PV) for H_2Z_2 systems. These properties have been obtained analytically for CCSD wavefunctions using the implementation described in ref. 69.

In the calculations, the size of the complete virtual spinor spaces is 156 (HF), 182 (HCl), 310 (HBr), 382 (HI), 544 (HAt) and 588 (HTs), 314 (H_2O_2), 366 (H_2S_2) and 622 (H_2Se_2). Unless otherwise noted, all electrons were correlated in the calculations. For the smaller systems HF, HCl, HBr, H_2O_2 and H_2S_2 we were able to perform calculations on a single node of the laboratory compute cluster in Lille. For HI, HAt, H_2Se_2 and HTs we employed, respectively, 32, 32, 49 and 64 nodes of the Summit supercomputer. For HI calculations with double-zeta quality basis sets, we were able to perform single-node calculations on the Jean Zay supercomputer.

The data presented in this manuscript is available at the Zenodo repository of reference 86.

V. RESULTS AND DISCUSSION

In this section we discuss the performance of the MP2FNOs approach for the CCSD correlation energy and molecular properties, and the impact of employing different occupation number thresholds on the results. In order to minimize computational cost on our systematic studies, we carried out CCSD(T) calculations only when discussing the effect of FNO and CMO truncation on HCl bond lengths.

For minimizing the bias in the CMO and FNO comparison, we report results in which we employ the closest possible number of CMO to that of FNO, such that for CMO we avoid truncations that would remove close-lying orbitals such as those belonging to a same atomic shell. In practice, this means CMO calculations generally contain a few more orbitals than FNO ones.

In figure S1 in the supplemental information, we present for the EFG of HI the difference between the approach taken here for CMO, and that of considering striclty the same number of CMO and FNO. We can see that arbitrarily truncating the virtual space for CMO can lead to large oscillations in the expectation values, a point that will be addressed in detail in the following. Additionally, from the comparison of Hamiltonians for HTs in table S2 in the supplemental information, we see that our choice of the X2C Hamiltonian is a suitable one for our purposes; while total energies obviously are very different between Hamiltonians, we see that for those taking into account SOC, relative energies (HOMO-LUMO gap) and expectation values are very close to each other. This is in stark contrast with scalar relativistic and non-relativistic calculations, which still show fairly good agreement with 2C/4C approaches for the quadrupole moment and even the HOMO-LUMO gap, but are completely off the mark for the EFG and

dipole moment. This is due to the very strong spin-orbit splitting in the valence p-shell of the superheavy tenessine.

A. Correlation Energy

CCSD correlation energies obtained with CMO and FNOs are displayed in Figure 2 for the hydrogen halides. In these, we show values computed for four values of natural occupation number threshold ($1.0\times10^{-3}, 1.0\times10^{-4}, 1.0\times10^{-5}, 1.0\times10^{-6}$), as well as the two extrema: a point without any virtual orbitals (corresponding to the Hartree-Fock solution), and another without any truncation of the virtual space. Since the number of virtuals varies for each system, we provide relative measures in terms of the percentage of the virtual space included in the calculations for each point, and likewise for the amount of correlation energy recovered at each point.

From the plots, it is clearly the case that FNOs show a more rapid convergence than CMO across the periodic table, as could be expected from the non-relativistic literature[56,59,87]. For example, to recover only 50% correlation energy already 40% to 50% of the virtual CMO space is required, while only 20% of the FNO space suffices. Also for the more realistic goal of attaining at least 90% of the correlation energy use of FNOs can reduce the required space by 10s of % thereby introducing significant savings in (memory) storage of tensors with virtual indexes, as well as in operation counts for contractions involving virtual indexes in the CCSD equations.

An interesting finding is that the area enclosed by two lines slightly decreases as we go down the periodic table from F to Ts. This means that the FNOs approach recovers less correlation, for a given truncation, as the number of electrons increases. For example, a threshold of 1.0×10^{-5} (the 4th point in Figure 2) means, for HF and HCl, that more than 60% of the virtual space is included in the calculation, whereas for HTs the same threshold only includes 43%; this means that for HTs virtual orbitals with low occupation numbers are still important for the correlation treatment.

Apart from testing the performance for single point at the potential energy surface, it is also important to verify the performance of FNOs at different geometric structures. To this end, in Figure 3 we present the potential energy curves around equilibrium for HCl at CCSD(T) level, comparing sets of FNOs and CMO virtuals that are truncated to the same number of virtuals (corresponding to about 50% of the complete virtual space). We see that, except for a global energy shift, the FNOs follow the curvature of full virtual space potential energy curve more closely than the CMO curve does.

The better agreement of FNOs relative to CMOs can be quantified by a comparison of spectroscopic constants for the three curves, shown in Table I. There, we report the equilibrium distance (R_e) and vibrational constant (ω_e), calculated with the LEVEL program[88]. Taking the full valence space result as a reference, the truncated orbitals overestimate the R_e but underestimate the ω_e. However, the error of truncated FNOs is 0.0037Å(R_e) and 17 cm^{-1} (ω_e), which is only half that of truncated CMOs.

As we correlated all electrons, but employed basis sets without specific core correlating functions, one may ask what would be the effect of including such functions. In table S1 of the supplemental information, we provide a comparison of correlation energies and expectation vales between the uncontracted aug-pVTZ and aug-pCVTZ basis sets for the HCl molecule without using virtual space truncation. These results clearly indicate that, although the additional high angular momentum core correlating functions are quite important for increasing the amount of correlation energy for the core and core-valence interactions, the two types of uncontracted basis sets agree well for the computed property values. Another interesting point to note is that, as can be seen from table S3 of the supplemental information, the percentage of correlation energy recovered at each virtual space truncation point is nearly the same for MP2 and CCSD. This indicates that the information obtained from the full virtual space MP2 calculation preceding the truncation can potentially be used to correct the CCSD correlation energy for the effect of truncation. While going beyond the scope of this paper, this point merits further investigation, in which also correction for truncation errors in higher order methods such as CCSD(T) could be investigated.

Table I. Spectroscopic constants of ground state of HCl

	R_e(Å)	ω_e (cm^{-1})
Exp	1.2746	2991
Untruncated Orbitals	1.2756	2986
Truncated CMO	1.2859	2942
Truncated FNO	1.2793	2969

B. Molecular properties

1. *Electric Dipole and Quadrupole Moments*

To asses performance of FNOs for molecular properties we first considered the molecular EDM and EQM because these do well characterize the overall electronic charge distribution within molecules. These properties sample the regions away from the nuclei, as is clear from their operator forms in Eqs. 12 and 13, respectively:

$$D_\mu = e\sum_i (\vec{r}_i)_\mu \tag{12}$$

$$Q_{\mu\nu} = e\sum_i ((\vec{r}_i)_\mu (\vec{r}_i)_\nu - r^2\delta_{\mu\nu}) \tag{13}$$

Figure 4 shows how the EDM and EQM correlation correction varies with the truncation of the virtual orbital spaces, with dotted and solid lines representing EDM and EQM, respectively. As before FNOs results are plotted with square markers. For these properties, the convergence is non-monotonic, unlike energy that we considered before. For both CMO and FNO truncation, the use of a high truncation value and consequently small virtual space (less than 30% virtual orbitals) may lead to a strong overestimation or even a wrong sign of the correlation correction to the molecular property.

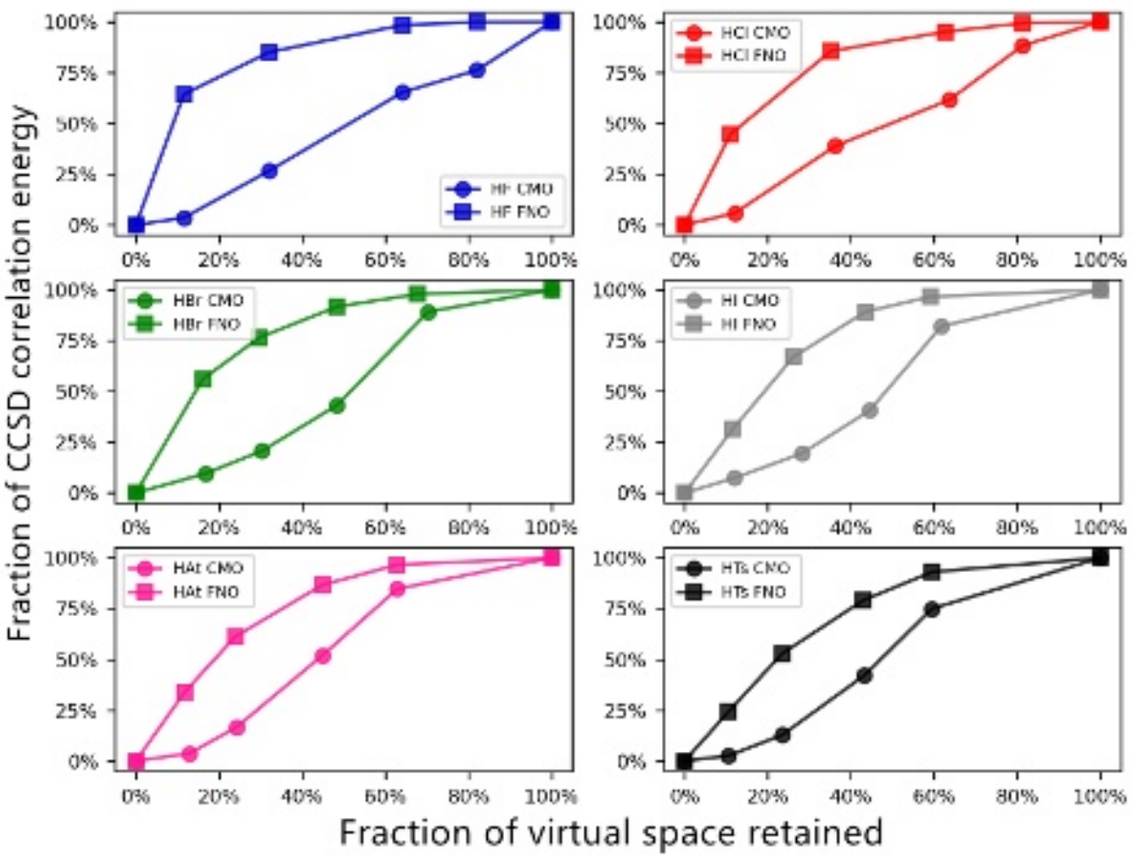


Figure 2. Convergence of the CCSD correlation energy with respect to the size of the virtual orbital space, for the X2C Hamiltonian. The X axis indicates the fraction of the virtual space retained, while the Y axis gives the fraction of the correlation energy recovered with respect to the value obtained with the untruncated virtual space.

Especially for CMO truncation, it is almost impossible to estimate the truncation error from a sequence of results for small virtual spaces. For FNO truncation oscillations are much less pronounced than for CMO, which indicates that also for these properties it is advantageous to use FNO truncation rather than CMO truncation. The similar performance for the correlation energy could thereby be used as a guideline. Taking the HF molecule as example, we note that with a threshold of 1.0×10^{-4}, the FNOs recover 72% of the correlation contribution to EDM while recovering 85% of the correlation energy. This indicates that the simplification of using a common cut-off for both properties could be a suitable strategy. In addition, we note that the performance of FNOs in EDM also applies to the heavier elements in the studied series. Even for HTs, setting the truncation threshold to 1.0×10^{-4} (using only 23% of the orbital space) already recovers 91% of the correlation contribution to the EDM.

The EQM shows a similar behaviour as the EDM, albeit with stronger oscillations. The largest oscillations for the EQM occur at the same position as for the EDM, but the amplitude thereof is much larger. This is again most pronounced in the CMO case. Taking HI as an example, retaining 12% of the CMO virtual space overestimates the correlation contribution to EDM by 150%, while the EQM contribution is even seven times too large. As for the EDMs, the oscillations resulting from FNO truncation are smaller and convergence is more smooth, which should make it possible to use a more aggressive cut-off strategy than is possible with CMOs.

2. *Electric field gradient*

To also consider the effect of truncation on properties that probe the regions close to atomic nuclei we now turn to the electric field gradients (EFGs) at the halogen nuclei. EFGs couple with nuclear quadrupole moments[89] for nuclei with spin greater or equal to one and are important in the analysis of nuclear magnetic resonance (NMR) experiments[90]. For

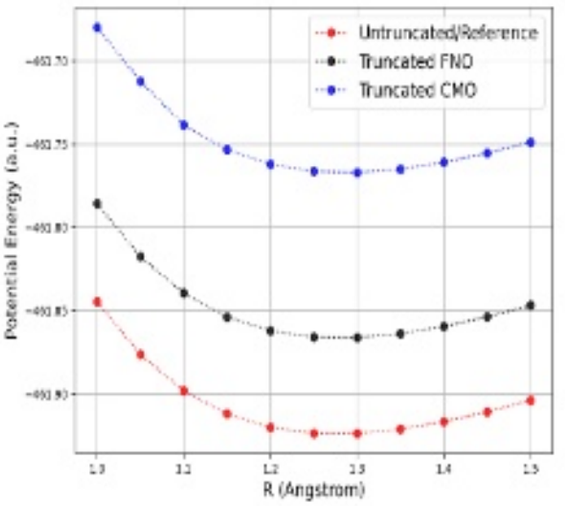


Figure 3. Potential energy curves of HCl at CCSD(T) level using untruncated orbitals (red) as reference, truncated FNO (black) and truncated CMO (blue), respectively. The X axis is the internuclear distance and the Y axis is the total energy in Hartrees.

EFGs both (semicore) correlation and SOC effects are known to be of importance[91] and they form therefore a good second test for the applicability of FNO truncation in relativistic calculations. The EFG is defined as the second derivative of the electric potential V(**R**) with respect to the nuclear position vector, taken at the nuclear position R_N

$$q_{\mu\nu}(\mathbf{R}_N) = -\frac{\partial V(\mathbf{R})}{\partial R_\mu \partial R_\nu}|_{\mathbf{R}=\mathbf{R}_N} \tag{14}$$

By introducing the EFG tensor operator[92]

$$\hat{q}^e_{\mu\nu} = \frac{3(\vec{r}-\vec{R}_N)_\mu(\vec{r}-\vec{R}_N)_\nu - |\vec{r}-\vec{R}_N|^2\delta_{\mu\nu}}{|\vec{r}-\vec{R}_N|^5} \tag{15}$$

the EFG can be expressed as the expectation value of a one-body operator which makes evaluating its value similar to computing the EDM and EQM.

For linear molecules it suffices to compute only the zz-component (with the z-axis chosen along the molecular bond) as the other non-zero parts of the tensor can then be determined by symmetry. Figure 5 plots this zz component of the EFG and shows for both FNO and CMO oscillations upon truncating the orbital spaces. In the case of correlating all electrons, we see that FNO truncation leads to worse performance than CMO truncation, due to even more rapid oscillations. Recalling that the only difference between computing the EFG and EDM properties is in the operators used, it is of interest to consider their differences. The EFG tensor operator scales as r^{-3} in contrast to the r^1 and r^2 scaling of the EDM and EQM operators. Density changes in the core region due to correlation are thus magnified by the EFG operator while they are hardly of influence for the EDM and EQM. Such changes may both come from core correlation as well as from correlating the valence electrons (through the tails of the valence orbitals in the core region). As core correlation may be harder to converge than valence correlation, we will separate the two effects by comparing all-electron and valence-only electrons calculations in the next subsection.

3. Convergence analysis for HCl and HI

The issue of oscillating convergence can be conveniently analyzed for the HCl molecule as chlorine is large enough to investigate the effect of core correlation, yet small enough to allow for quick calculations. Rather than taking only a few truncation values, we now systematically extend the size of the virtual space by adding individual FNOs and show the effect thereof on the EDM, EQM and EFG of HCl in Figure 6 for both all-electron and valence-only correlation calculations.

For the all-electron case, we see in the EDM and EQM plots for CMO truncation peaks when 13 and 23 orbitals are used. This is probably due to quasi-degeneracies, as the orbital energies of the 12th, 13th, 14th orbital are 0.42418, 0.43657 and 0.43673 Hartree, respectively. This set of virtual orbitals is primarily a diffuse chlorine p-type orbital shell that is split due to the formation of the hydrogen-chlorine bond as well as due to spin-orbit coupling. Taking only one of the two almost degenerate π-type orbitals in the correlation space, gives an unbalanced description and impacts these valence properties significantly. For the EFG this particular set of three CMOs is less important because diffuse orbitals only contribute indirectly to this property. For the EFG one may notice (Figure 6, lower left panel, blue line) a large oscillation around virtual orbital number 70 in the all-electron calculation. This is due to the more core-like p-orbitals 70, 71, and 72 with energies of respectively 35.63, 35.73 and 36.10 Hartree, that are impacting the correlation of core electrons. Restricting the correlation treatment to valence electrons only strongly reduces the effect of these virtuals (Figure 6, lower right panel).

The FNO curves for the EDM and EQM show only an initial oscillation that is followed by rather smooth convergence. This initial oscillation can also be traced to quasi-degenerate orbitals (but now in terms of occupation numbers). For the FNOs, the occupation number of the 2nd, 3rd, 4th, 5th orbital are 0.008692, 0.008676, 0.008508, 0.008491, respectively, and like with near-degenerate CMOs it appears recommendable to either include all or none of a degenerate set. For the EFG, the situtation is unfortunately more complicated, also with FNOs.

In the all-electron calculation values only stabilize after inclusion of about 50 orbitals, while in the valence-only calculation stable convergence is reached after addition of about 20 FNOs. Compared to the CMO truncation scheme, the advantage of FNO truncation appears to be absence of "late" oscillations (at low threshold values) that could cause artefacts in the CMO truncation schemes. Such oscillations are typical for EFGs, for which an indivdual orbital may provide a significant contribution but where the contribution of full,

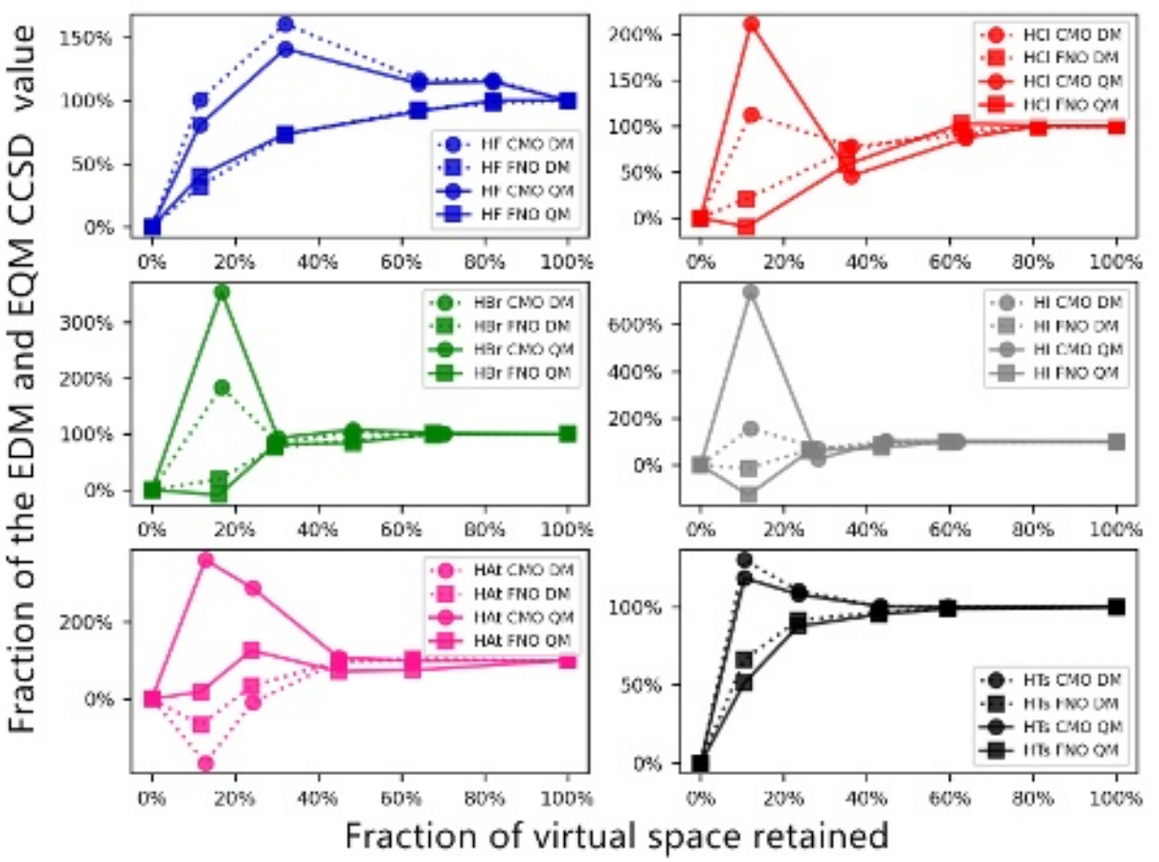


Figure 4. Convergence of the CCSD electric dipole and quadrupole moments with respect to the size of the virtual orbital space, for the X2C Hamiltonian. The X axis indicates the fraction of the virtual space retained, while the Y axis gives the fraction of the expectation values recovered with respect to the value obtained with the untruncated virtual space.

spherically symmetric shells of orbitals add up to zero due to symmetry. This is well-known for CMOs, but also holds for FNOs. An example are the EFG integrals of the 13th, 14th, 15th FNOs which evaluate to -78, 44 and 26 au. These large values get multiplied by very similar occupation numbers (resp. 7.9 10^{-4}, 7.0 10^{-4}, 7.0 10^{-4}) so that their contributions to the total EFG nearly cancel.

A similar argument can be used to rationalize the difference between the valence and all-electron calculations. In the correlation of core orbitals one effectively changes the relative occupation of the three $2p$ orbitals in the chlorine core by making their occupations slightly smaller than 2. Imbalances in the correlation get thereby for the EFG multiplied by large integral values (about 80 au for the $2p_{3/2}$ orbitals) which creates the large oscillation seen in these calculations. For EDMs and EQMs the integral values are much more alike and smaller in magnitude, leading to the observed smoother convergence.

To check if such oscillation problems are particular for the relativistic domain, or can also be found in the non-relativistic case, we have carried out the same analysis as above by employing a non-relativistic Hamiltonian, employing both contracted and uncontracted (valence) basis sets and correlating both valence and all electrons. The results are shown in Fig. 7.

From these, it can be seen that for the electric dipole and quadrupole moments, the non-relativistic calculations basically show the same oscillations as for X2C, and results for contracted and uncontracted basis sets are nearly indistinguishable from each other. For the EFG there is also no noticeable difference for the convergence patterns between X2C and non-relativistic results for uncontracted basis sets.

On the other hand, we observe significant differences between EFG calculations employing contracted and uncontracted basis sets. The first one is that, already at the Hartree-Fock level, there is a significant difference (nearly 10%) on the absolute value of the EFG, with contracted values underestimating uncontracted ones. Second, we note that with contracted basis sets, there are very few oscillations in FNO EFG values (even when correlating all electrons), and FNO results

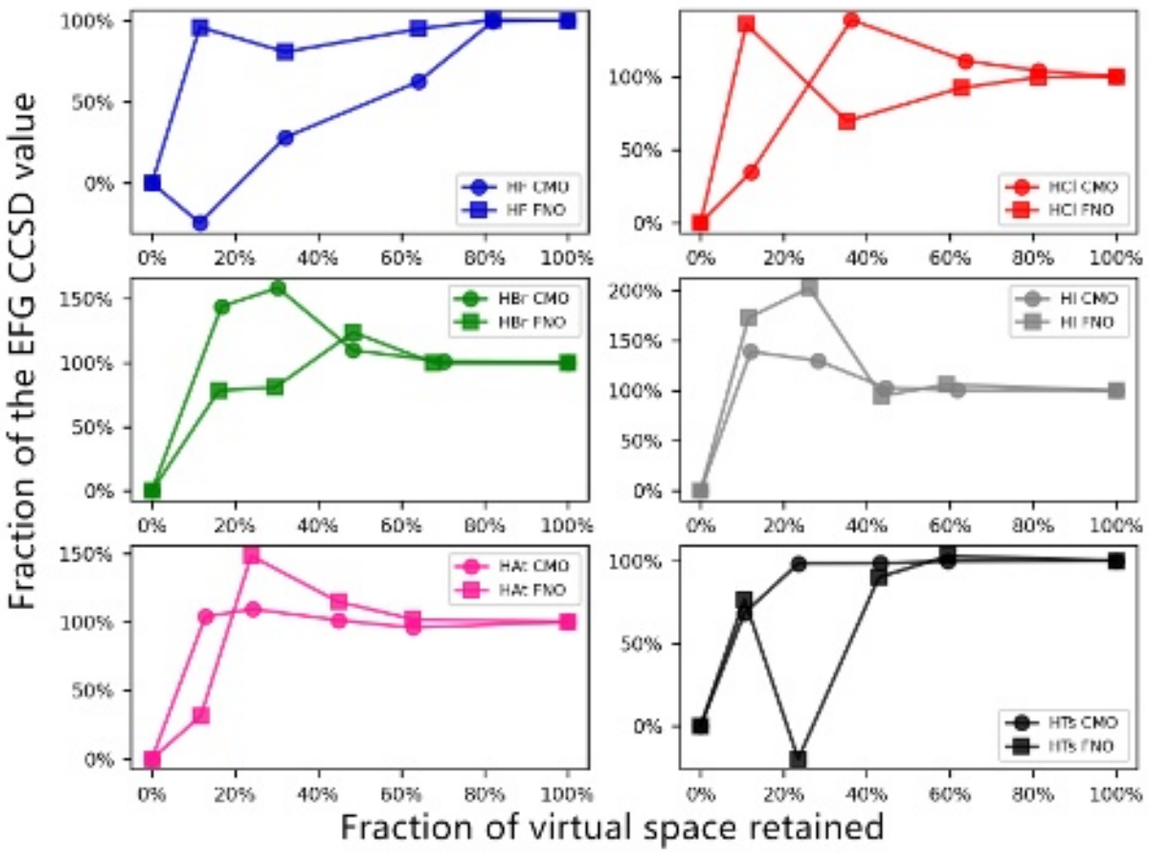


Figure 5. Convergence of the CCSD electric field gradient at the halogen nucleus with respect to the size of the virtual orbital space, for the X2C Hamiltonian. The X axis indicates the fraction of the virtual space retained, while the Y axis gives the fraction of the expectation value recovered with respect to the value obtained with the untruncated virtual space.

are already quite stable for a much lower number of virtuals than CMO results (CMO EFG values are still not completely converged at nearly full virtual spaces for the contracted sets).

This can be understood as a manifestation of the degree of atomic symmetry that is imposed by the contraction for the different orbital shells, particularly for the occupied core orbitals. In the contracted sets these are forced to maintain their atomic-like nature during the molecular calculation because electron correlation cannot as easily deform the orbitals as is possible when the basis set is uncontracted. This is underscored by the nearly identical behavior of the contracted all electron and valence calculations, as in the latter core orbitals are kept frozen.

To check if the convergence behavior mentioned above is also found for a heavier system, we carried out similar calculations for HI, but using an uncontracted double zeta basis set. It can be seen from Fig 8 that although HI has many more electrons, it still shows a very similar convergence as observed for HCl. For instance, to get a converged EFG value of HI, in the valence only computation, one just needs to correlate 20 FNOs.

Looking at these tests, we may conclude that for properties like EDM and EQM, which are dominated by valence electron contributions, one can safely use FNOs and their occupation numbers to cut off the orbital space, both in valence-only and all-electron correlation calculations. For sensitive properties for which core electrons may give large and nearly cancelling contributions, it is also with FNOs numerically more stable to correlate only the valence electrons.

C. Parity violation

Detection of parity violation (PV) effects, associated with weak force in atoms and molecules is an active field of research[13,93,94]. While this property can also be computed with perturbation theory starting from nonrelativistic theory[95], it is advantageous to use a relativistic quantum chemistry

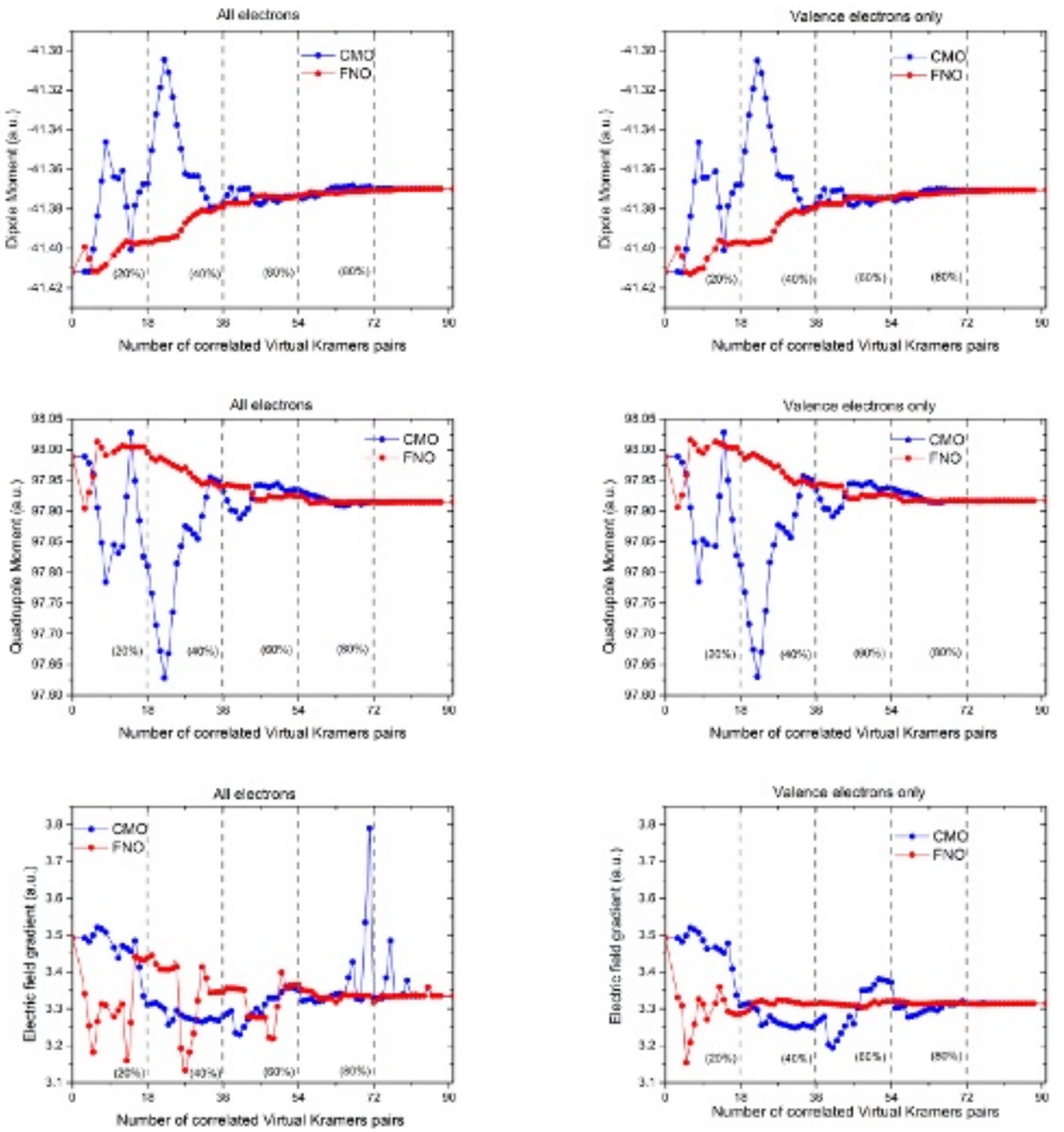


Figure 6. Effect of virtual space truncation on the expectation values (top: EDM; middle: EQM; bottom: EFG) of HCl, for the X2C Hamiltonian. The X axis gives the number of correlated virtual Kramers pairs, up to and including the full virtual space (the rightmost points in the graphs correspond to calculations with the full virtual space). Figures to the left correspond to calculations correlating all electrons, and to the right correlating valence electrons only. T.

framework because the PV energy can then be formulated as an expectation value of an effective one-body operator[85]:

$$E_{PV} = \sum_A \langle \Psi | \hat{H}^A_{PV} | \Psi \rangle \tag{16}$$

with

$$\hat{H}^A_{PV} = \frac{G_F}{2\sqrt{2}} Q^A_W \sum_i \gamma^5_i \rho^A(r_i) \tag{17}$$

and G_F=1.16637$\times 10^{-11}$MeV^{-2} being the Fermi coupling constant. The A and i label nuclei and electrons, respectively. The weak charge $Q^A_W = -N_A + Z_A(1 - 4sin^2\theta_\omega)$, where N_A

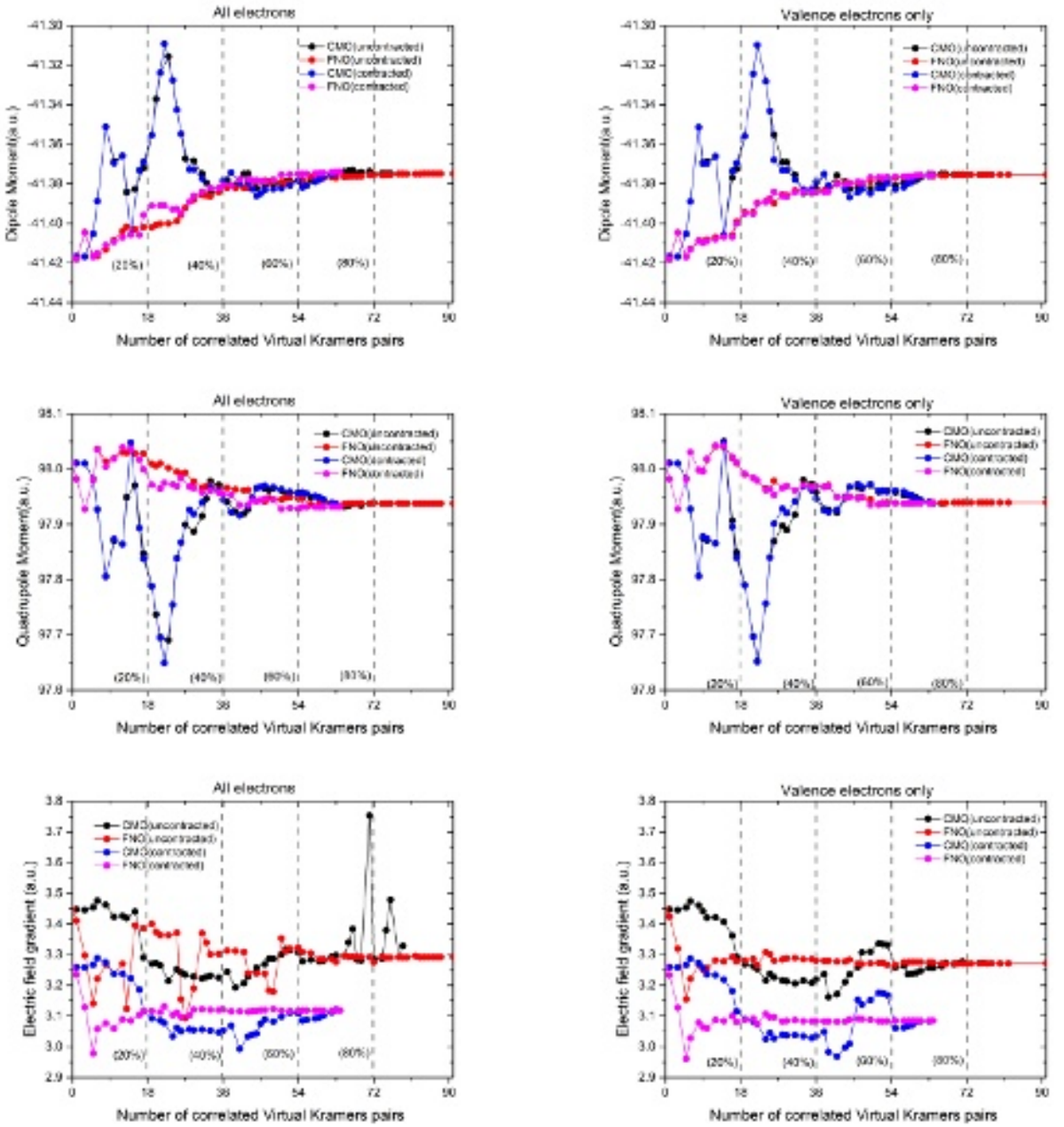


Figure 7. Effect of virtual space truncation on the expectation values (top: EDM; middle: EQM; bottom: EFG) of HCl, for the non-relativistic Hamiltonian and employing contracted and uncontracted Dunning basis sets. The X axis gives the number of correlated virtual Kramers pairs, up to and including the full virtual space (the rightmost points in the graphs correspond to calculations with the full virtual space). Figures to the left correspond to calculations correlating all electrons, and to the right correlating valence electrons only.

and Z_A is the number of neutrons and protons in each nucleus. θ_ω is Weinberg mixing angle, that is set to 0.2319 for $sin^2\theta_\omega$. ρ^A and γ_i^5 are normalized nucleon density and 4-dimensional chirality operator, respectively.

$$\gamma^5 = \begin{pmatrix} \mathbb{O} & \mathbb{I} \\ \mathbb{I} & \mathbb{O} \end{pmatrix} \tag{18}$$

Sunaga et al.[13] already investigated use of CMO truncation at CCSD level with two different threshold values and showed that truncation is well possible for this property. In current work, we also employ FNO truncation and test more thresholds for CMO truncation and its effect on the PV value.

The results of the two truncation schemes for the H_2Z_2 (Z=O, S, Se) molecules are displayed in Figure 9. We find

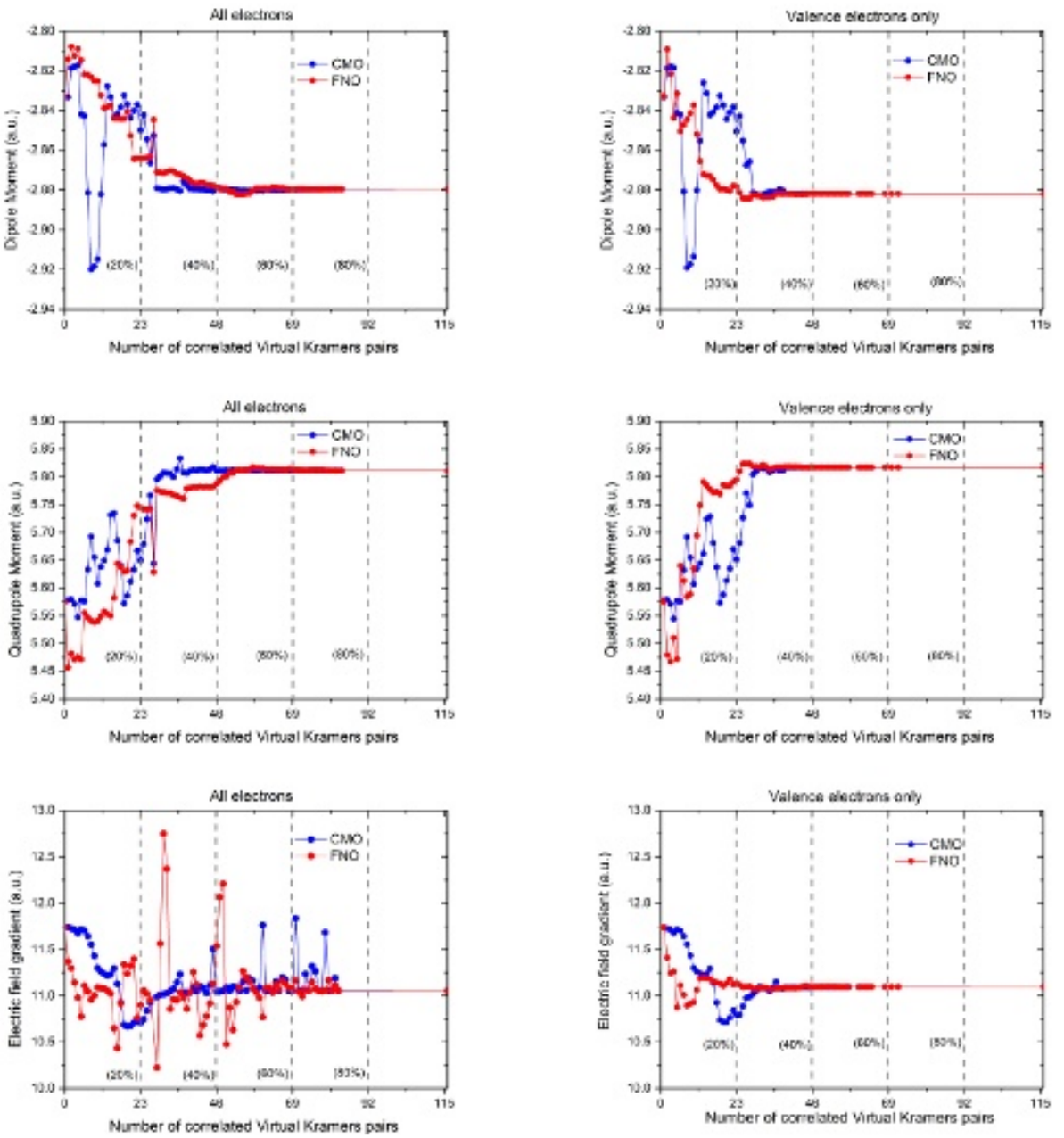


Figure 8. Effect of virtual space truncation on the expectation values (top: EDM; middle: EQM; bottom: EFG) of HI, for the X2C Hamiltonian. The X axis gives the number of correlated virtual Kramers pairs, up to and including the full virtual space (the rightmost points in the graphs correspond to calculations with the full virtual space). Figures to the left correspond to calculations correlating all electrons, and to the right correlating valence electrons only.

that both FNO and CMO truncation leads to strong oscillations like observed in the EFG case. Again, these are most pronounced for aggressive truncation in which more than 50% of the virtual orbital space is removed. Looking more closely at these systems we find that orbital energies (for CMO) or occupation numbers (for FNO) show a number of near-degeneracies. These orbitals give individually large but partially cancelling contributions. We have not investigated valence-only correlation explicitly for this property as core contributions are probably more important than for the EFG and it is more difficult to define a representative small test case. For our current tests we note that the convergence behaviour is consistent with results of reference 13, in which rather conservative CMO truncation tresholds of 100 and 500

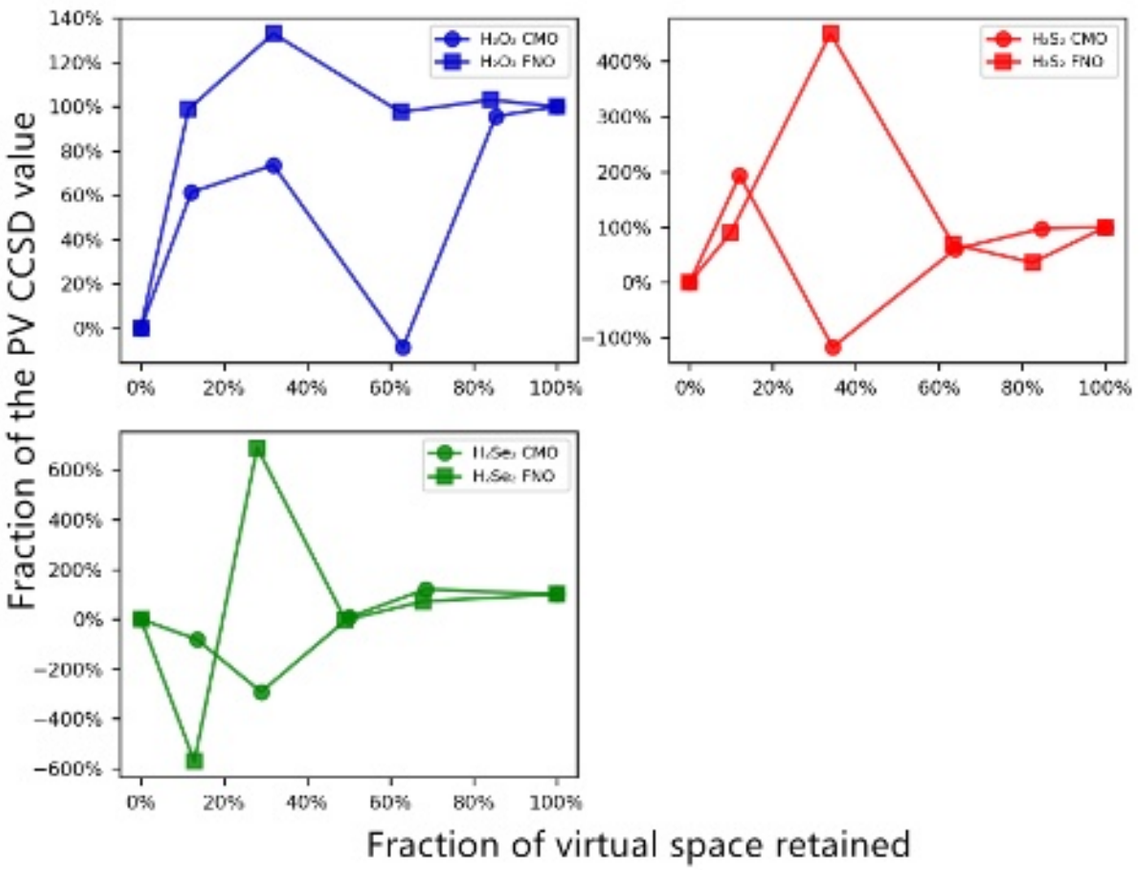


Figure 9. Convergence of the CCSD PV energy with respect to the size of the virtual orbital space, for the X2C Hamiltonian. The X axis indicates the fraction of the virtual space retained, while the Y axis gives the fraction of the expectation value recovered with respect to the value obtained with the untruncated virtual space.

Hartree were used for H_2Se_2. With both CMO and FNO truncation a 50 % reduction appears realistic, leading to a significant reduction of the computational effort. For instance, in the current test the H_2Se_2 calculation on 49 nodes of Summit with the full orbital space took 100 minutes while the 50 % reduced calculation too only 20 minutes.

Nevertheless, for the particular case of PV energy evaluation, FNO truncation does not appear to give a significant advantage over the simpler CMO truncation scheme. Apart from the sensitivity of PV operator on core correlation already observed in the EFG evaluation, another possible reason is our use of the unrelaxed MP2 density matrix. Shee, Visscher, and Saue[69] found for H_2S_2 that the contribution of orbital relaxation to the density matrix is significant when used to evaluate this property. It would be interesting to investigate how these contributions impact the generated FNOs are thereby the convergence of the FNO truncation scheme.

VI. CONCLUSIONS

In this work we describe the formulation and implementation of the MP2 frozen natural orbitals (MP2FNOs) method for relativistic electronic structure calculations, which is an appealing approach for truncating the large virtual orbital spaces typically associated with relativistic calculations, while retaining the high accuracy expected of wavefunction-based approaches such as coupled cluster.

This implementation was carried out in the massively parallel coupled cluster module of the DIRAC program, with the help of a framework to manipulate 1RDMs obtained from correlated wavefunction calculations. A particularity of our code is its ability to generate a set of canonical occupied and truncated virtual orbitals as well as natural orbitals in the atomic basis, so that these can be conveniently employed in post-Hartree-Fock calculations, and also used for analysis.

We employed our code to investigate the performance

of MP2FNOs for the calculation of correlation energies and ground-state first-order properties such as dipole and quadrupole moments, electric field gradient at the nuclei, and parity violation energy shifts. As model systems, we considered species containing elements from the first row up to and including the superheavy element tennessine.

We have found that though MP2FNOs are always capable of recovering more correlation energy than their canonical counterparts for a given truncation, this advantage is slightly diminished for the heaviest systems considered, containing Astatine and Tennessine. In spite of that, it is generally found that MP2FNOs can reduce virtual spaces to about 50% of their original size without significant errors in the energy.

Among the properties considered, a truncation of about 50% of the original virtual space has also been shown to provide values that are nearly converged to the results without truncation. That said, different properties exhibit very different convergence behavior; for valence properties such as the electric dipole moment MP2FNOs show a fairly smooth convergence to the reference values, whereas for the electric field gradient and in particular the parity violation energy shifts, for which regions close to the nuclei are important (and thus higher-lying virtuals are more important in the correlation treatment), significant variations on the calculated values are found for virtual spaces smaller than 50% of the original virtual space.

We have found that for EFGs, correlating only valence electrons provides a pragmatic solution that recovers the smooth convergence seen for valence properties by removing difficult to converge, individually large but in total nearly cancelling, contributions from the core electrons. Whether such a strategy can also work for the PV energy shifts remains to be investigated.

As a final point and perspective we note that improvement of the efficiency of the current scheme is well possible by implementing efficient approximate schemes like Cholesky decomposition and Laplace transforms to generate the MP2 1RDM. Work along these lines is in progress and should enable treatment of larger systems in the future.

SUPPLEMENTARY MATERIAL

See supplementary material for further details on the comparison of Hamiltonians for the HTs molecule, the influence of core correlation functions on the correlation energies and expectation values for the HCl molecule, the convergence of EFG for iodine in HI, and the fraction of MP2 and CCSD correlation energies recovered with different FNO truncations for the hydrogen halide series.

ACKNOWLEDGMENTS

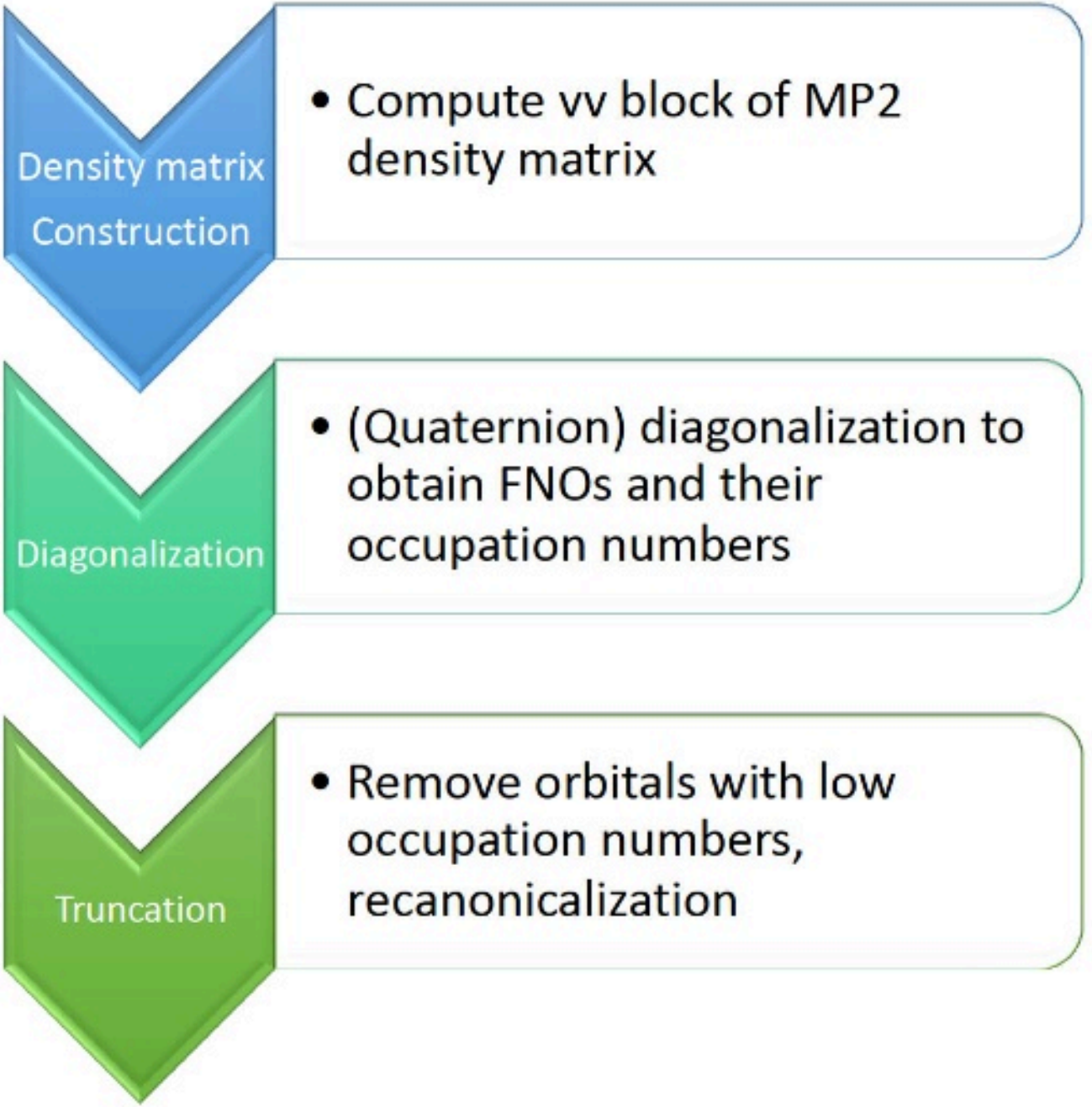
We acknowledge funding from projects Labex CaPPA (ANR-11-LABX-0005-01) and CompRIXS (ANR-19-CE29-0019, DFG JA 2329/6-1), the I-SITE ULNE project OVERSEE and MESONM International Associated Laboratory (LAI) (ANR-16-IDEX-0004), and support from the French national supercomputing facilities (grant DARI A0090801859, A0110801859).

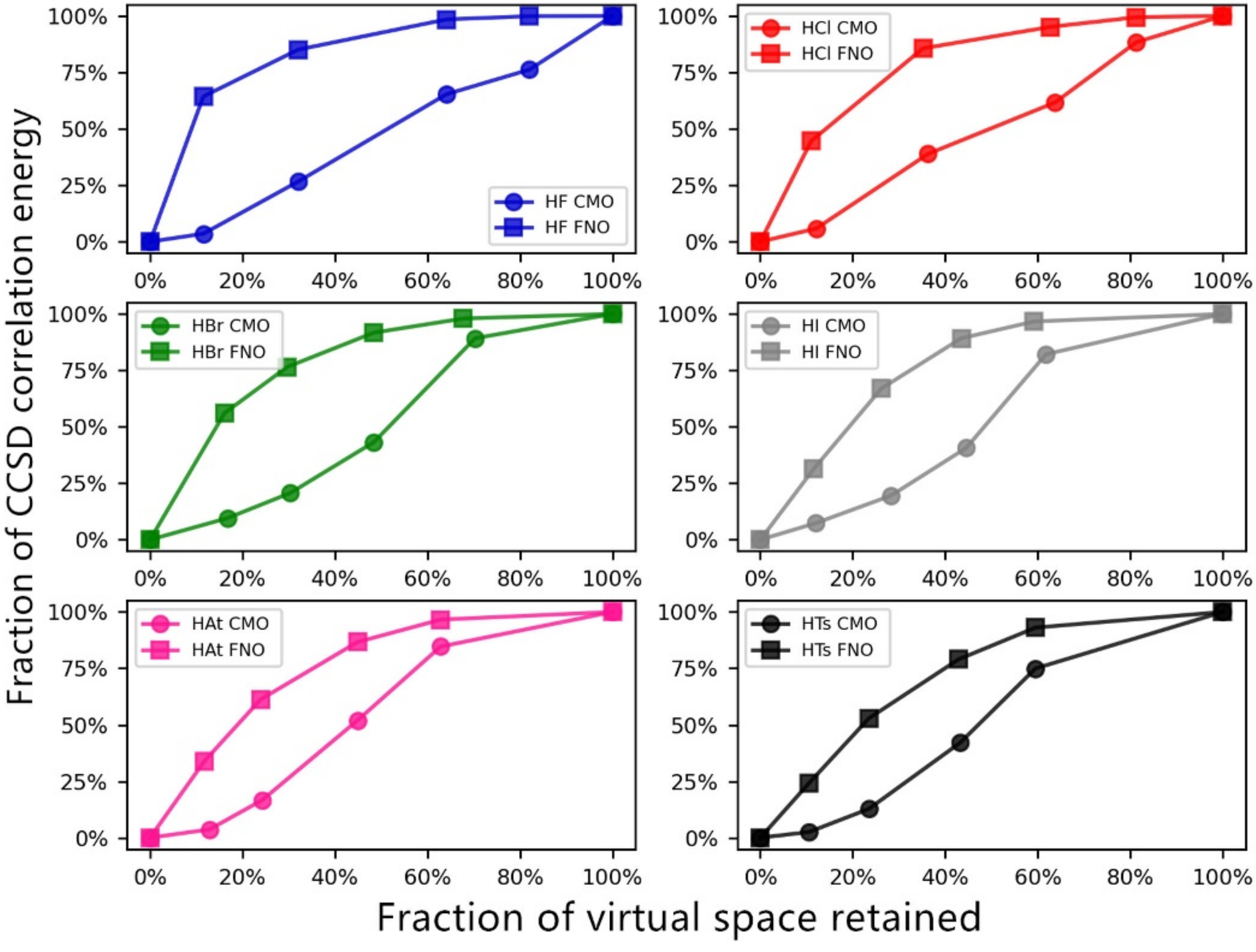
This research used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725.

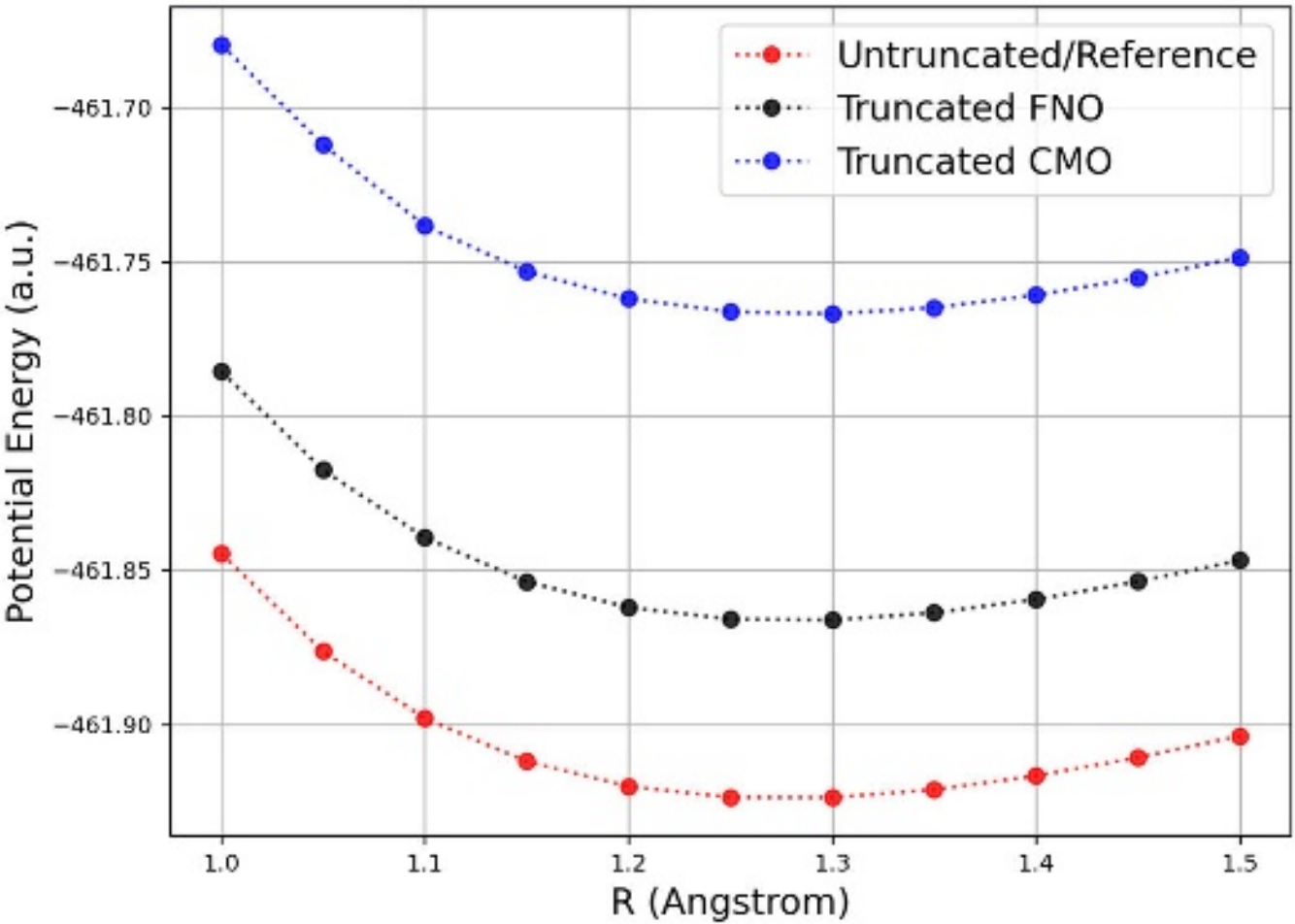
REFERENCES

[1]E. Acher, Y. H. Cherkaski, T. Dumas, C. Tamain, D. Guillaumont, N. Boubals, G. Javierre, C. Hennig, P. L. Solari, and M.-C. Charbonnel, Inorg. Chem. **55**, 5558 (2016).
[2]A. Leoncini, J. Huskens, and W. Verboom, Chem. Soc. Rev. **46**, 7229 (2017).
[3]X. Sun, H. Luo, and S. Dai, Chem. Rev. **112**, 2100 (2012).
[4]C. Berger, C. Marie, D. Guillaumont, C. Tamain, T. Dumas, T. Dirks, N. Boubals, E. Acher, M. Laszczyk, and L. Berthon, Inorg. Chem. **59**, 1823 (2020).
[5]J. M. Gogolski and M. P. Jensen, Sep. Sci. Technol **56**, 2775 (2020).
[6]C. Berger, E. Moreau, C. Marie, D. Guillaumont, A. Beillard, and L. Berthon, Solvent. Extr. Ion. Exc , 1 (2021).
[7]C. A. Gould, K. R. McClain, D. Reta, J. G. C. Kragskow, D. A. Marchiori, E. Lachman, E.-S. Choi, J. G. Analytis, R. D. Britt, N. F. Chilton, B. G. Harvey, and J. R. Long, Science **375**, 198 (2022).
[8]K. Burke, J. Chem. Phys. **136**, 150901 (2012).
[9]R. J. Bartlett, V. F. Lotrich, and I. V. Schweigert, J. Chem. Phys. **123**, 062205 (2005).
[10]S. Kervazo, F. Réal, F. Virot, A. S. P. Gomes, and V. Vallet, Inorg. Chem. **58**, 14507 (2019).
[11]L. E. Aebersold and A. K. Wilson, J. Phys. Chem. A **125**, 7029 (2021).
[12]F. Aquino, N. Govind, and J. Autschbach, J. Chem. Theory. Comput. **6**, 2669 (2010).
[13]A. Sunaga and T. Saue, Mol. Phys **119**, e1974592 (2021).
[14]T. Helgaker, J. Gauss, P. Jørgensen, and J. Olsen, J. Chem. Phys. **106**, 6430 (1997).
[15]K. L. Bak, J. Gauss, P. Jørgensen, J. Olsen, T. Helgaker, and J. F. Stanton, J. Chem. Phys. **114**, 6548 (2001).
[16]J. Almlöf, Chem. Phys. letters **181**, 319 (1991).
[17]M. Häser, Theor Chem Acc **87**, 147 (1993).
[18]E. J. Baerends, D. E. Ellis, and P. Ros, Chem. Phys. **2**, 41 (1973).
[19]M. Feyereisen, G. Fitzgerald, and A. Komornicki, Chem. Phys. Lett. **208**, 359 (1993).
[20]K. G. Dyall and K. Faegri, Jr., *Introduction to Relativistic Quantum Chemistry* (Oxford University Press, 2007).
[21]M. Reiher and A. Wolf, *Relativistic Quantum Chemistry* (Wiley, 2014).
[22]T. Saue and L. Visscher, in *Computational Methods in Lanthanide and Actinide Chemistry* (John Wiley & Sons Ltd, 2015) pp. 55–87.
[23]E. R. Batista, R. L. Martin, and P. Yang, in *Computational Methods in Lanthanide and Actinide Chemistry* (John Wiley & Sons Ltd, 2015) pp. 375–400.
[24]X. Cao and A. Weigand, in *Computational Methods in Lanthanide and Actinide Chemistry* (John Wiley & Sons Ltd, 2015) pp. 147–179.
[25]L. Visscher, O. Visser, P. J. Aerts, H. Merenga, and W. Nieuwpoort, Comput. Phys. Commun **81**, 120 (1994).
[26]T. Saue, R. Bast, A. S. P. Gomes, H. J. A. Jensen, I. A. Aucar, R. D. Remigio, K. G. Dyall, E. Eliav, E. Fasshauer, T. Fleig, L. Halbert, E. D. Hedegård, B. Helmich, C. R. Jacob, S. Knecht, J. K. Laerdahl, M. K. Nayak, J. M. H. Olsen, B. Senjean, A. Shee, and A. Sunaga, J. Chem. Phys. **152**, 204104 (2020).
[27]L. Belpassi, M. De Santis, H. M. Quiney, F. Tarantelli, and L. Storchi, J. Chem. Phys. **152**, 164118 (2020).
[28]M. Repisky, S. Komorovsky, M. Kadek, L. Konecny, U. Ekström, E. Malkin, M. Kaupp, K. Ruud, O. L. Malkina, and V. G. Malkin, J. Chem. Phys. **152**, 184101 (2020).
[29]W. Kutzelnigg and W. Liu, J. Chem. Phys. **123**, 241102 (2005).

[30]H. J. Aa. Jensen, (2005), *Douglas–Kroll the Easy Way*, Talk at Conference on Relativistic Effects in Heavy Elements - REHE, Mülheim, Germany, April, 2005. Available at `https://doi.org/10.6084/m9.figshare.12046158`.
[31]W. Liu and D. Peng, J. Chem. Phys. **125**, 044102 (2006).
[32]M. Iliaš and T. Saue, J. Chem. Phys. **126**, 064102 (2007).
[33]D. Peng, W. Liu, Y. Xiao, and L. Cheng, J. Chem. Phys. **127**, 104106 (2007).
[34]W. Liu and D. Peng, J. Chem. Phys. **131**, 031104 (2009).
[35]J. Liu, Y. Shen, A. Asthana, and L. Cheng, J. Chem. Phys. **148**, 034106 (2018).
[36]J. Liu and L. Cheng, J. Chem. Phys. **148**, 144108 (2018).
[37]L. Cheng, J. Chem. Phys. **151**, 104103 (2019).
[38]A. Asthana, J. Liu, and L. Cheng, J. Chem. Phys. **150**, 074102 (2019).
[39]J. Sikkema, L. Visscher, T. Saue, and M. Iliaš, J. Chem. Phys. **131**, 124116 (2009).
[40]K. Fægri and T. Saue, J. Chem. Phys. **115**, 2456 (2001).
[41]A. S. P. Gomes and L. Visscher, Chem. Phys. Lett. **399**, 1 (2004).
[42]C. Thierfelder, P. Schwerdtfeger, A. Koers, A. Borschevsky, and B. Fricke, Phys. Rev. A **80**, 022501 (2009).
[43]V. Pershina, in *Challenges and Advances in Computational Chemistry and Physics* (Springer Netherlands, 2010) pp. 451–520.
[44]J. Autschbach, N. Govind, R. Atta-Fynn, E. J. Bylaska, J. W. Weare, and W. A. de Jong, in *Computational Methods in Lanthanide and Actinide Chemistry* (John Wiley & Sons Ltd, 2015) pp. 299–342.
[45]L. Halbert, M. L. Vidal, A. Shee, S. Coriani, and A. S. P. Gomes, J. Chem. Theory Comput. **17**, 3583 (2021).
[46]J. Sucher, Phys. Rev. A **22**, 348 (1980).
[47]J. V. Pototschnig, A. Papadopoulos, D. I. Lyakh, M. Repisky, L. Halbert, A. S. P. Gomes, H. J. A. Jensen, and L. Visscher, J. Chem. Theory. Comput. **17**, 5509 (2021).
[48]D. I. Lyakh, Int. J. Quantum. Chem **119**, e25926 (2019).
[49]B. Helmich-Paris, M. Repisky, and L. Visscher, J. Chem. Phys. **145**, 014107 (2016), 1606.06498.
[50]L. Visscher, P. J. C. Aerts, and O. Visser, in *The Effects of Relativity in Atoms, Molecules, and the Solid State* (Springer US, 1991) pp. 197–205.
[51]L. Belpassi, F. Tarantelli, A. Sgamellotti, and H. M. Quiney, J. Chem. Phys. **122**, 184109 (2005).
[52]P.-O. Löwdin, Phys. Rev. **97**, 1474 (1955).
[53]E. R. Davidson, Rev. Mod. Phys. **44**, 451 (1972).
[54]T. L. Barr and E. R. Davidson, Phys. Rev. A **1**, 644 (1970).
[55]C. Sosa, J. Geertsen, G. W. Trucks, R. J. Bartlett, and J. A. Franz, Chem. Phys. Lett. **159**, 148 (1989).
[56]A. G. Taube and R. J. Bartlett, Collect. Czech. Chem. Commun. **70**, 837 (2005).
[57]H. J. A. Jensen, P. Jørgensen, H. Ågren, and J. Olsen, J. Chem. Phys. **88**, 3834 (1988).
[58]H. J. A. Jensen, P. Jørgensen, H. Ågren, and J. Olsen, J. Chem. Phys. **89**, 5354 (1988).
[59]A. G. Taube and R. J. Bartlett, J. Chem. Phys. **128**, 164101 (2008).
[60]P. Verma, L. Huntington, M. P. Coons, Y. Kawashima, T. Yamazaki, and A. Zaribafiyan, J. Chem. Phys. **155**, 034110 (2021).
[61]M. S. Gordon, M. W. Schmidt, G. M. Chaban, K. R. Glaesemann, W. J. Stevens, and C. Gonzalez, J. Chem. Phys. **110**, 4199 (1999).
[62]S. Chamoli, K. Surjuse, M. K. Nayak, and A. K. Dutta, arXiv preprint arXiv:2201.07752 (2022).
[63]J. N. van Stralen, L. Visscher, C. V. Larsen, and H. J. A. Jensen, Chem. Phys **311**, 81 (2005).
[64]H. A. Kramers, Proc. Acad. Amsterdam **33** (1930).
[65]L. Visscher, Chem. Phys. Lett. **253**, 20 (1996).
[66]T. Saue and H. J. A. Jensen, J. Chem. Phys. **111**, 6211 (1999).
[67]M. Repisky, S. Komorovsky, M. Kadek, L. Konecny, U. Ekström, E. Malkin, M. Kaupp, K. Ruud, O. L. Malkina, and V. G. Malkin, J. Chem. Phys. **152**, 184101 (2020).
[68]M. Yoshimine, J. Comput. Phys **11**, 449 (1973).
[69]A. Shee, L. Visscher, and T. Saue, J. Chem. Phys. **145**, 184107 (2016).
[70]H. J. A. Jensen, R. Bast, A. S. P. Gomes, T. Saue, L. Visscher, I. A. Aucar, V. Bakken, C. Chibueze, J. Creutzberg, K. G. Dyall, S. Dubillard, U. Ekström, E. Eliav, T. Enevoldsen, E. Faßhauer, T. Fleig, O. Fossgaard, L. Halbert, E. D. Hedegård, T. Helgaker, B. Helmich-Paris, J. Henriksson, M. van Horn, M. Iliaš, C. R. Jacob, S. Knecht, S. Komorovský, O. Kullie, J. K. Lærdahl, C. V. Larsen, Y. S. Lee, N. H. List, H. S. Nataraj, M. K. Nayak, P. Norman, G. Olejniczak, J. Olsen, J. M. H. Olsen, A. Papadopoulos, Y. C. Park, J. K. Pedersen, M. Pernpointner, J. V. Pototschnig, R. Di Remigio, M. Repiský, K. Ruud, P. Sałek, B. Schimmelpfennig, B. Senjean, A. Shee, J. Sikkema, A. Sunaga, A. J. Thorvaldsen, J. Thyssen, J. van Stralen, M. L. Vidal, S. Villaume, O. Visser, T. Winther, S. Yamamoto, and X. Yuan, "Dirac22," (2022).
[71]*DIRAC, a relativistic ab initio electronic structure program, Release DIRAC21, Written by Bast, R.; Gomes, A. S. P.; Saue, T.; Visscher, L.; Jensen, H. J. Aa.; Aucar, I. A.; Bakken, V.; Dyall, K. G.; Dubillard, S.; Ekström, U.; Eliav, E.; Enevoldsen, T.; Faßhauer, E.; Fleig, T.; Fossgaard, O.; Halbert, L.; Hedegård, E. D.; Helgaker, T.; Helmich-Paris, B.; Henriksson, J.; Iliaš, M.; Jacob, Ch. R.; Knecht, S.; Komorovský, S.; Kullie, O.; Lærdahl, J. K.; Larsen, C. V.; Lee, Y. S.; List, N. H.; Nataraj, H. S.; Nayak, M. K.; Norman, P.; Olejniczak, G.; Olsen, J.; Olsen, J. M. H.; Papadopoulos, A.; Park, Y. C.; Pedersen, J. K.; Pernpointner, M.; Pototschnig, J. V.; Di Remigio, R.; Repiský, M.; Ruud, K.; Sałek, P.; Schimmelpfennig, B.; Senjean, B.; Shee, A.; Sikkema, J.; Sunaga, A.; Thorvaldsen, A. J.; Thyssen, J.; van Stralen, J.; Vidal, M. L.; Villaume, S.; Visser, O.; Winther, T.; Yamamoto, S. see http://www.diracprogram.org* (2021).
[72]D. I. Lyakh, "Exatensor, a basic numerical tensor algebra library for distributed heterogeneous hpc platforms," `https://github.com/ORNL-QCI/ExaTENSOR`.
[73]K. G. Dyall, Theor Chem Acc **108**, 335 (2002).
[74]K. G. Dyall, Theor Chem Acc **115**, 441 (2006).
[75]K. G. Dyall, Theor Chem Acc **131**, 1172 (2012).
[76]T. H. Dunning, J. Chem. Phys. **90**, 1007 (1989).
[77]R. A. Kendall, T. H. Dunning, and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).
[78]D. E. Woon and T. H. Dunning, J. Chem. Phys. **98**, 1358 (1993).
[79]D. E. Woon and T. H. Dunning, J. Chem. Phys. **103**, 4572 (1995).
[80]K. A. Peterson and T. H. Dunning, J. Chem. Phys. **117**, 10548 (2002).
[81]B. A. Hess, C. M. Marian, U. Wahlgren, and O. Gropen, Chem. Phys. Lett. **251**, 365 (1996).
[82]C. M. Marian, in *Reviews in Computational Chemistry* (John Wiley & Sons, Inc., 2001) pp. 99–204.
[83]B. Schimmelpfennig, "*AMFI, an atomic mean-field spin-orbit integral program*," University of Stockholm, Stockholm, Sweden; 1999.
[84]K.-P. Huber, Molecular spectra and molecular structure (1979).
[85]J. K. Laerdahl and P. Schwerdtfeger, Phys. Rev. A **60**, 4439 (1999).
[86]X. Yuan, L. Visscher, and A. S. P. Gomes, "Dataset: Assessing mp2 frozen natural orbitals in relativistic correlated electronic structure calculations," `https://doi.org/10.5281/zenodo.5939520`.
[87]A. E. DePrince and C. D. Sherrill, J. Chem. Theory Comput. **9**, 2687 (2013).
[88]R. J. Le Roy, J. Quant. Spectrosc. Ra **186**, 167 (2017).
[89]V. Kello, Mol. Phys **89**, 127 (1996).
[90]J. Autschbach, S. Zheng, and R. W. Schurko, Concepts Magn. Reson. **36A**, 84 (2010).
[91]S. Larsson and P. Pyykkö, Chem. Phys. **101**, 355 (1986).
[92]L. Visscher, T. Enevoldsen, T. Saue, and J. Oddershede, J. Chem. Phys. **109**, 9677 (1998).
[93]R. Bast, A. Koers, A. S. P. Gomes, M. Iliaš, L. Visscher, P. Schwerdtfeger, and T. Saue, Phys. Chem. Chem. Phys. **13**, 864 (2011).
[94]G. Rauhut and P. Schwerdtfeger, Phys. Rev. A **103**, 042819 (2021).
[95]R. Berger and M. Quack, J. Chem. Phys. **112**, 3148 (2000).


Density matrix Construction
• Compute vv block of MP2 density matrix
Diagonalization
• (Quaternion) diagonalization to obtain FNOs and their occupation numbers
Truncation
• Remove orbitals with low occupation numbers, recanonicalization


Fraction of CCSD correlation energy
Fraction of virtual space retained
HF CMO
HF FNO
HCl CMO
HCl FNO
HBr CMO
HBr FNO
HI CMO
HI FNO
HAt CMO
HAt FNO
HTs CMO
HTs FNO
0%
20%
40%
60%
80%
100%
25%
50%
75%


Untruncated/Reference
Truncated FNO
Truncated CMO
−461.70
−461.75
−461.80
−461.85
−461.90
Potential Energy (a.u.)
1.0
1.1
1.2
1.3
1.4
1.5
R (Angstrom)

Fraction of the EDM and EQM CCSD value

Fraction of virtual space retained

HF CMO DM
HF FNO DM
HF CMO QM
HF FNO QM

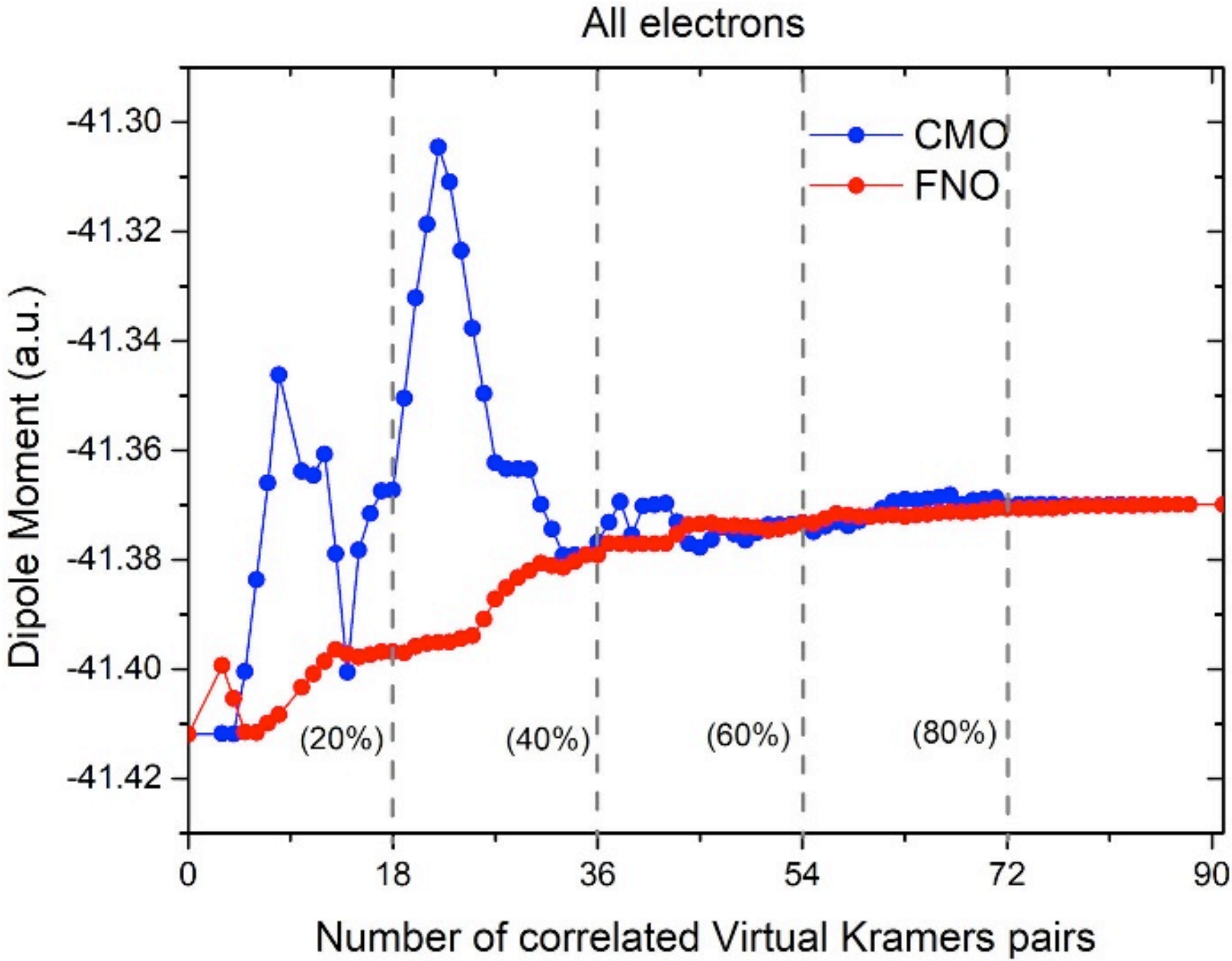
HCl CMO DM
HCl FNO DM
HCl CMO QM
HCl FNO QM

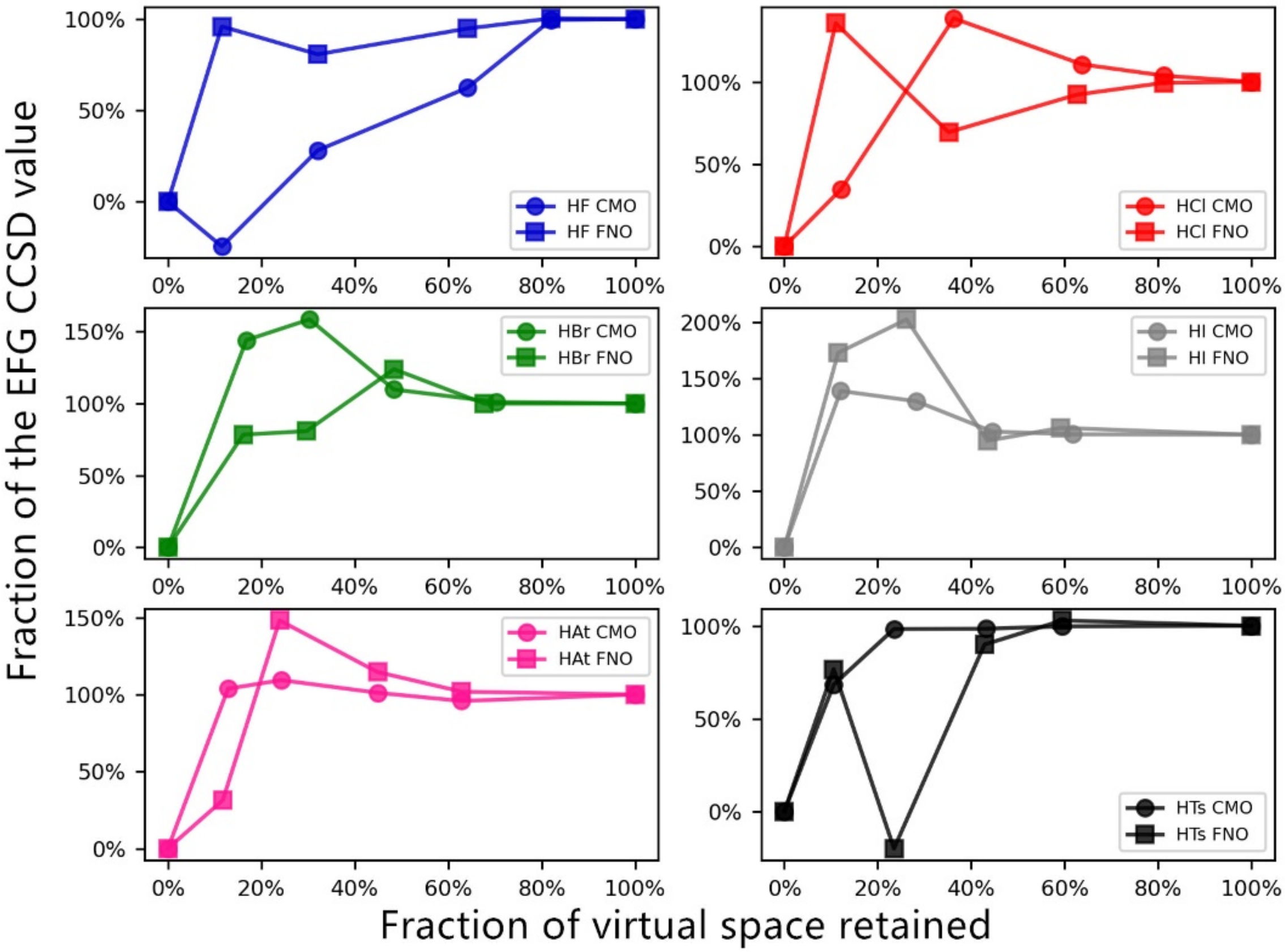
HBr CMO DM
HBr FNO DM
HBr CMO QM
HBr FNO QM

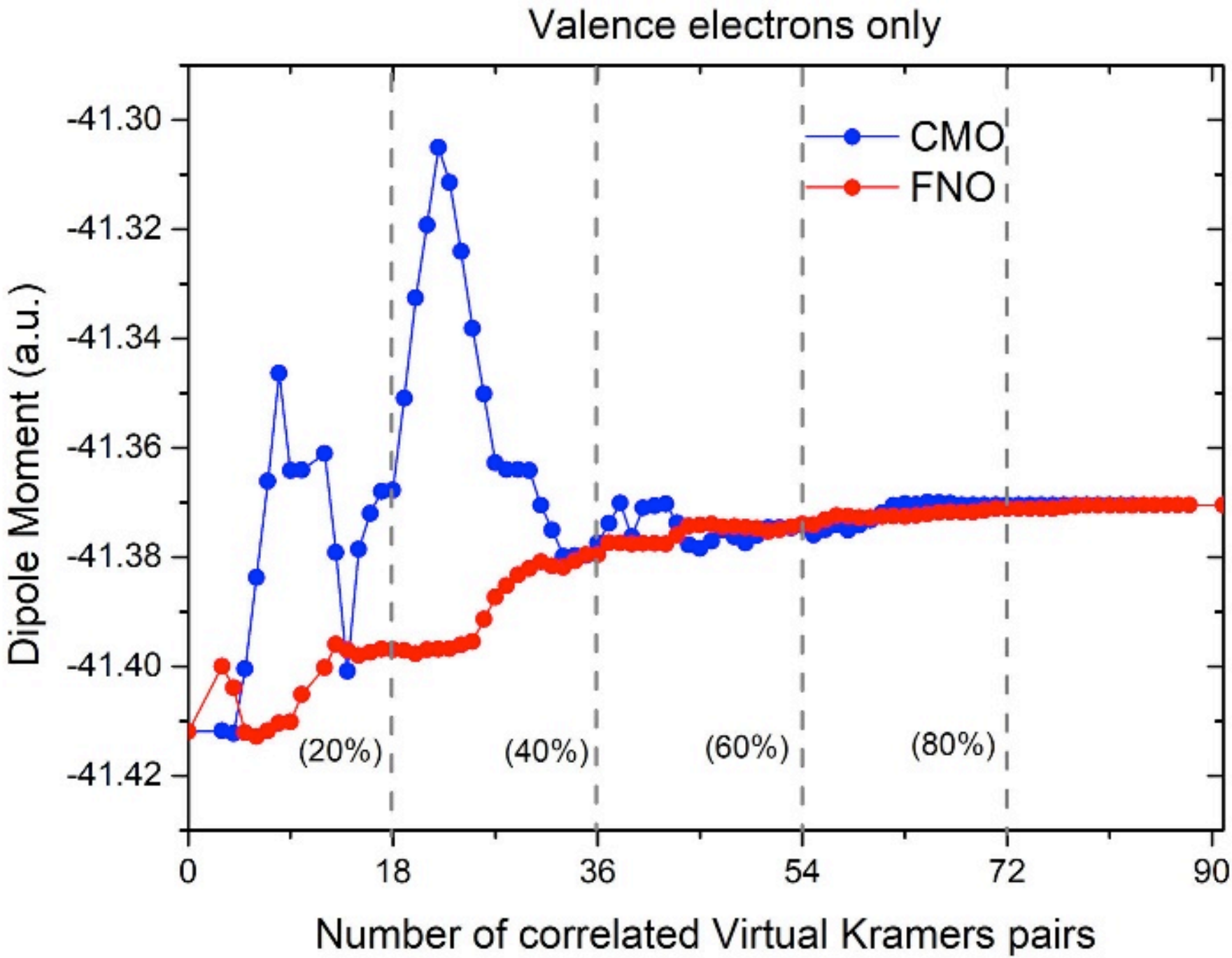
HI CMO DM
HI FNO DM
HI CMO QM
HI FNO QM

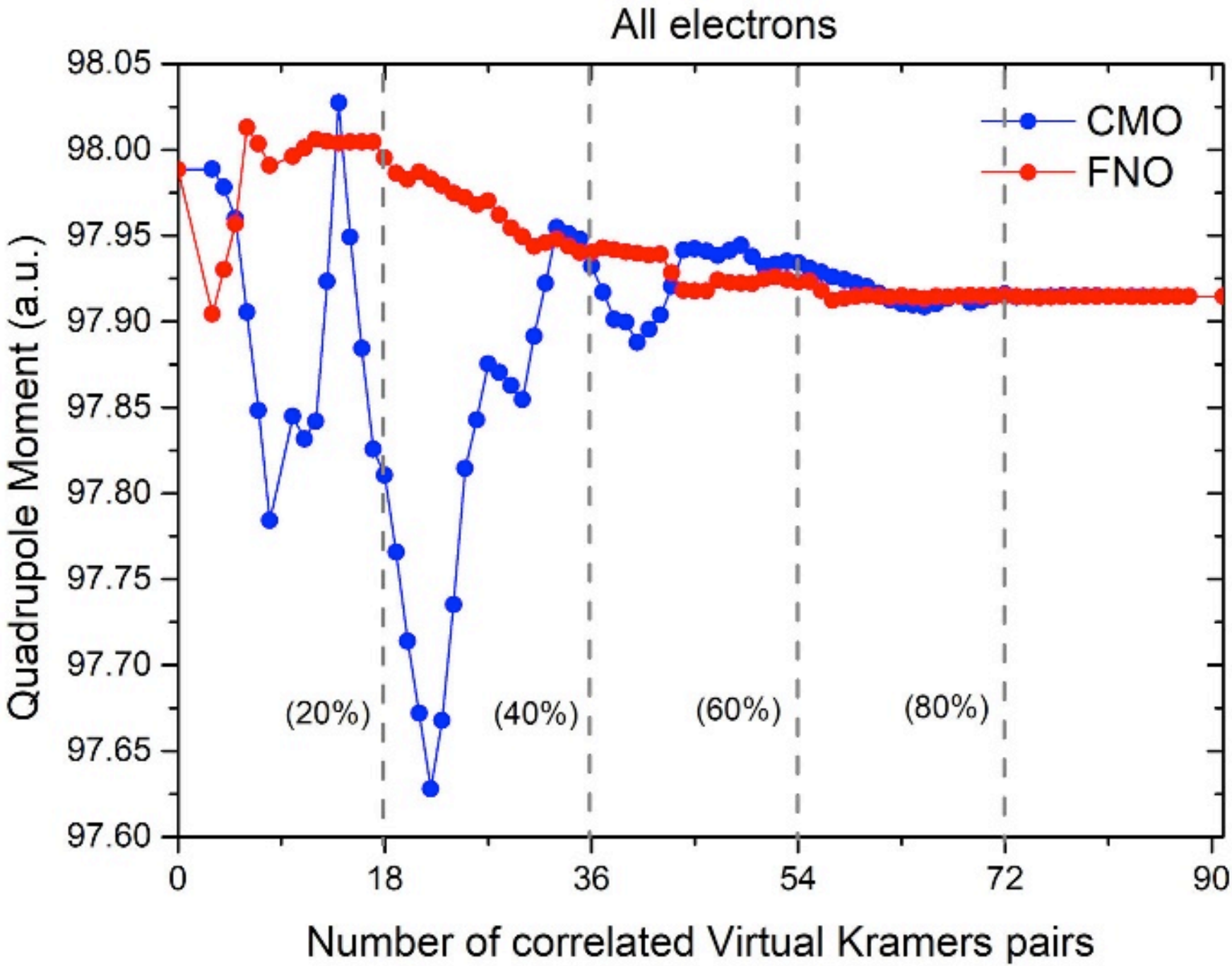
HAt CMO DM
HAt FNO DM
HAt CMO QM
HAt FNO QM

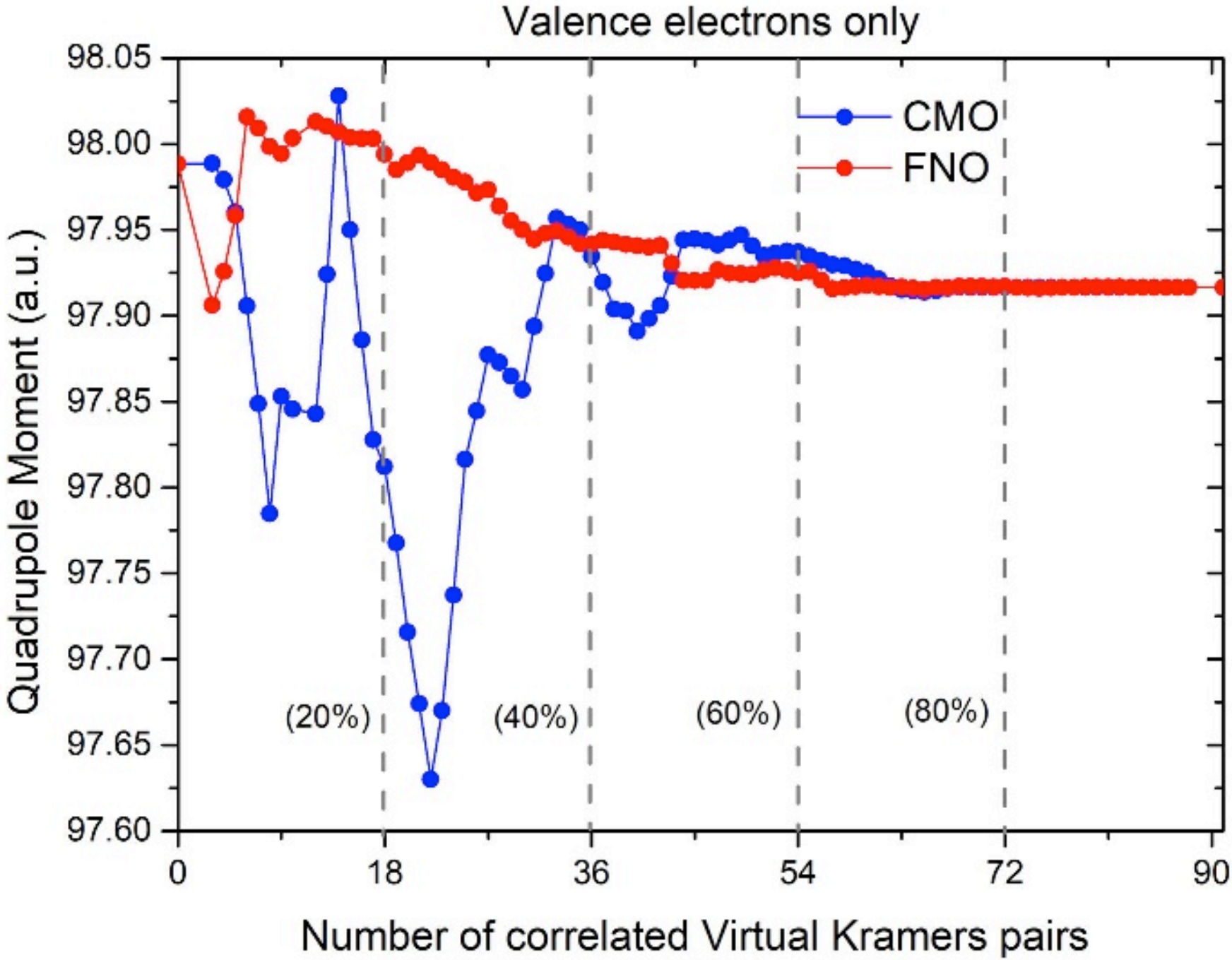
HTs CMO DM
HTs FNO DM
HTs CMO QM
HTs FNO QM

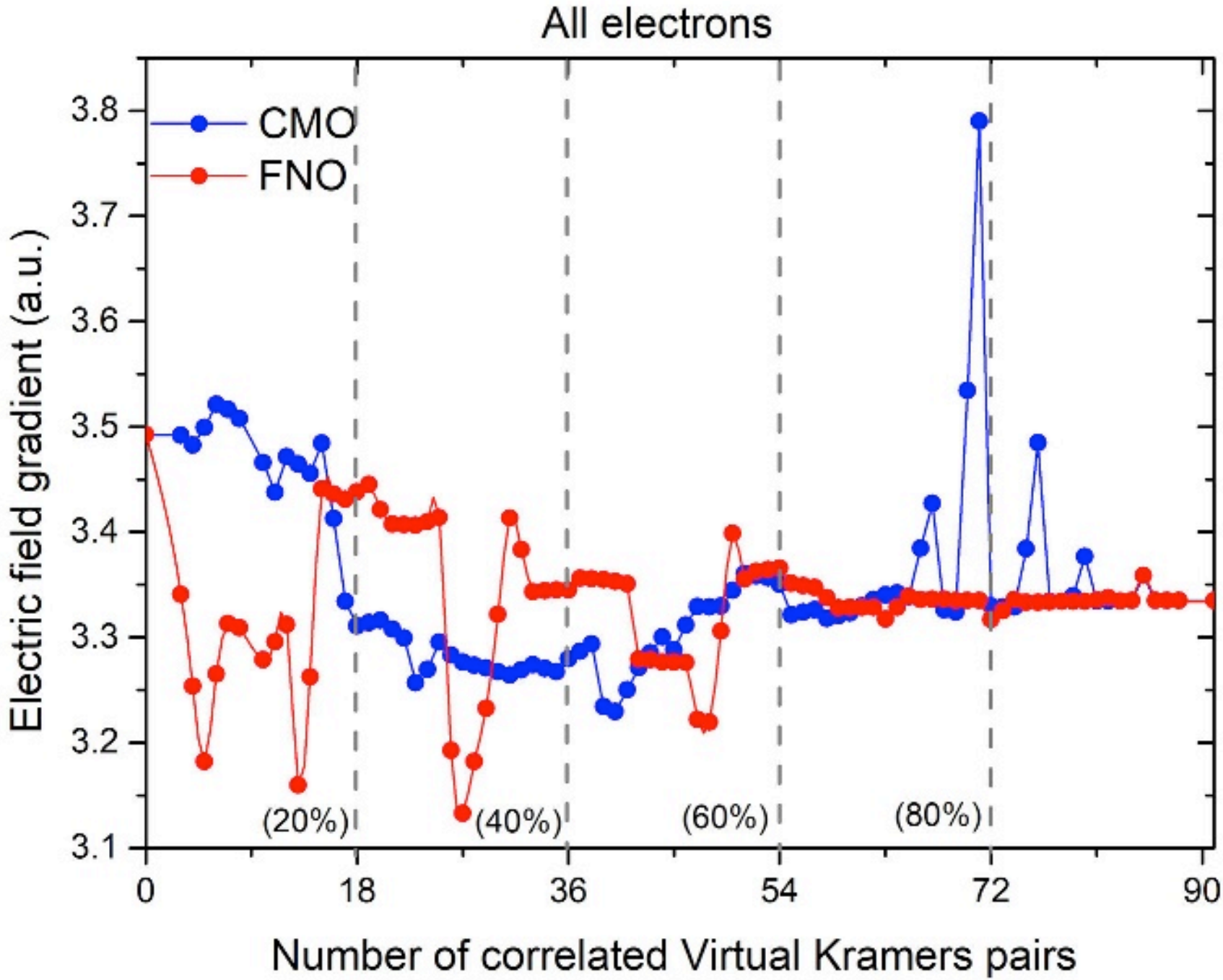

All electrons
CMO
FNO
-41.30
-41.32
-41.34
-41.36
-41.38
-41.40
-41.42
Dipole Moment (a.u.)
(20%)
(40%)
(60%)
(80%)
0
18
36
54
72
90
Number of correlated Virtual Kramers pairs

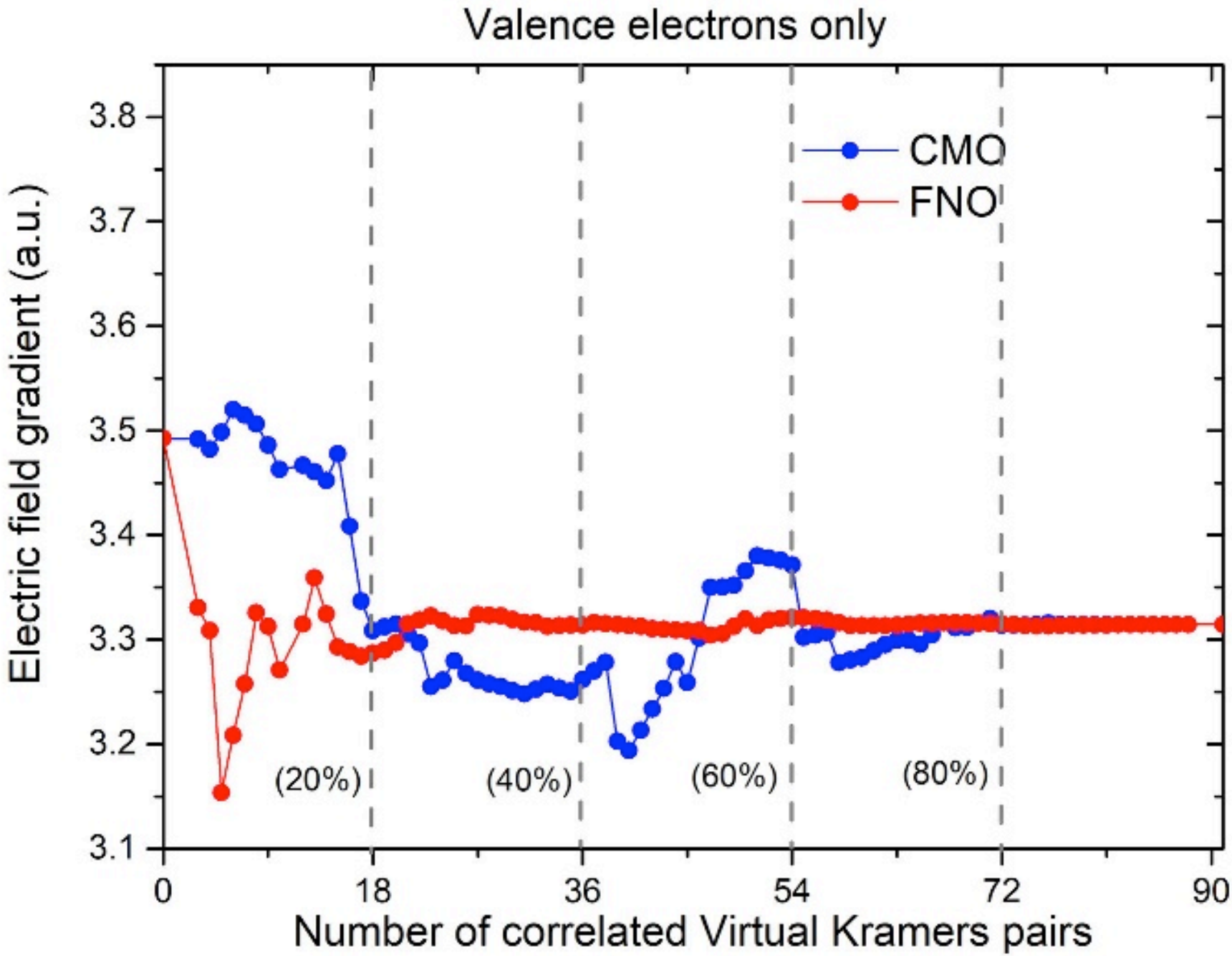

Fraction of the EFG CCSD value
Fraction of virtual space retained
HF CMO
HF FNO
HCl CMO
HCl FNO
HBr CMO
HBr FNO
HI CMO
HI FNO
HAt CMO
HAt FNO
HTs CMO
HTs FNO
0%
20%
40%
50%
60%
80%
100%
150%
200%

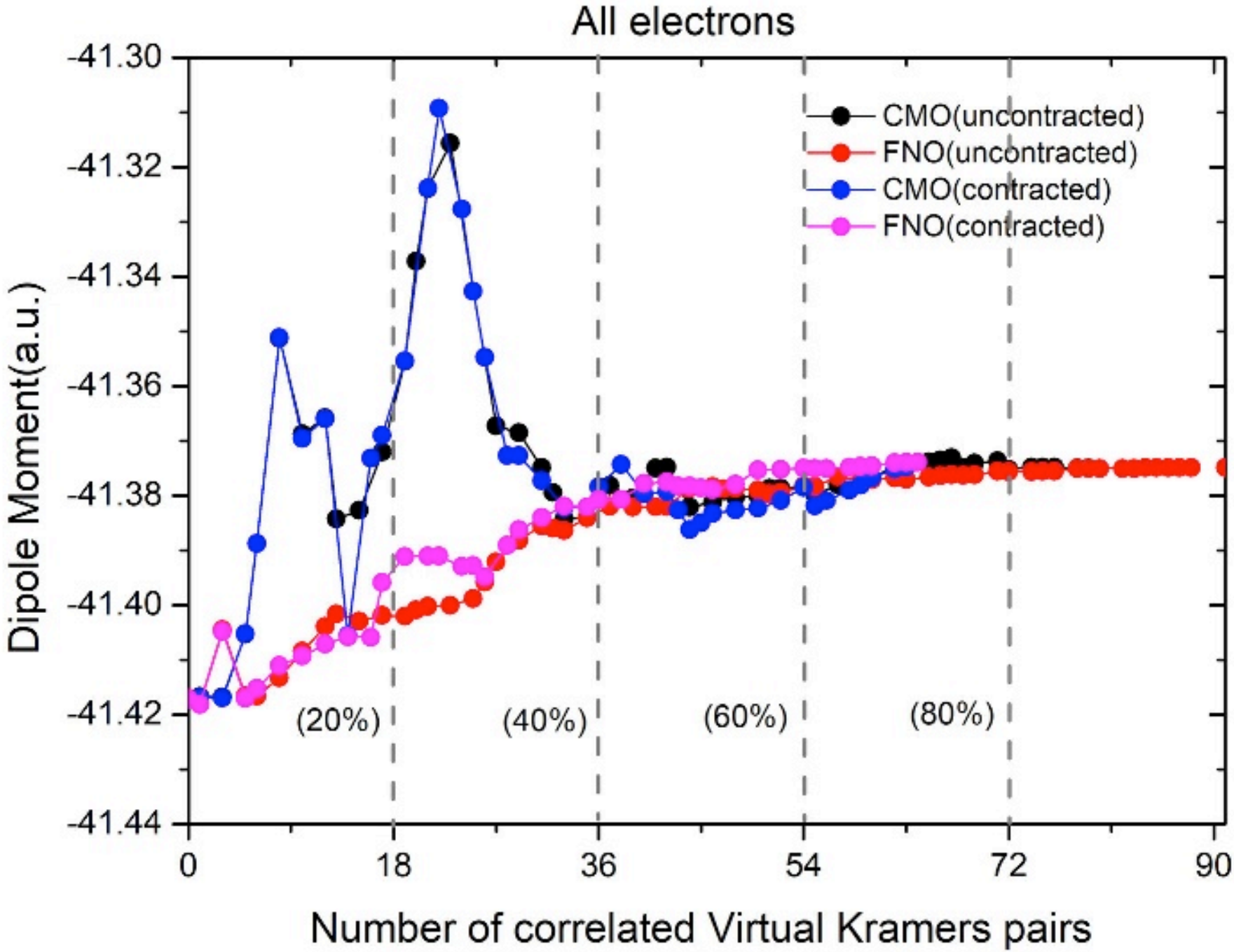

Valence electrons only
CMO
FNO
-41.30
-41.32
-41.34
-41.36
-41.38
-41.40
-41.42
Dipole Moment (a.u.)
(20%)
(40%)
(60%)
(80%)
0
18
36
54
72
90
Number of correlated Virtual Kramers pairs

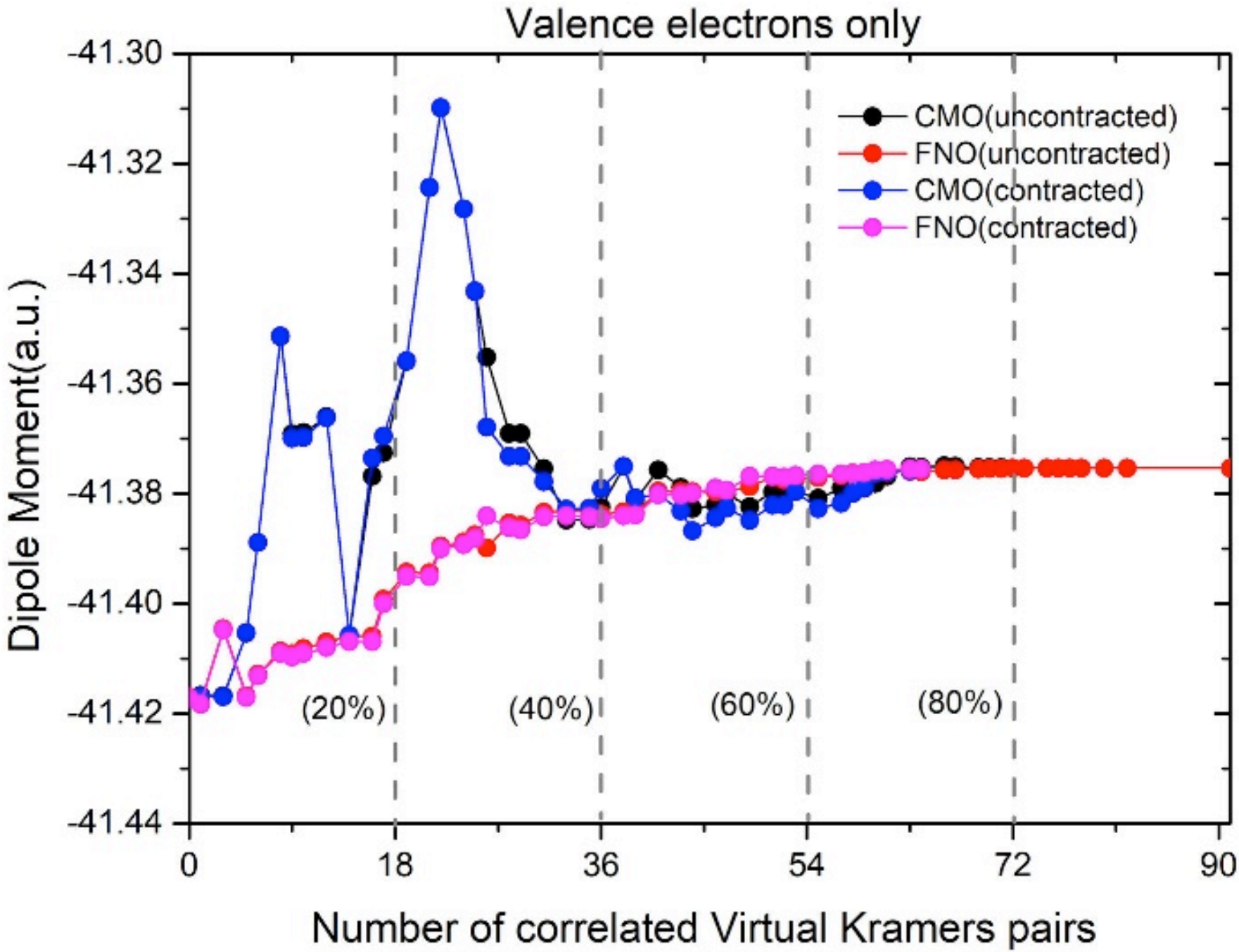

All electrons
98.05
98.00
97.95
97.90
97.85
97.80
97.75
97.70
97.65
97.60
Quadrupole Moment (a.u.)
CMO
FNO
(20%)
(40%)
(60%)
(80%)
0
18
36
54
72
90
Number of correlated Virtual Kramers pairs

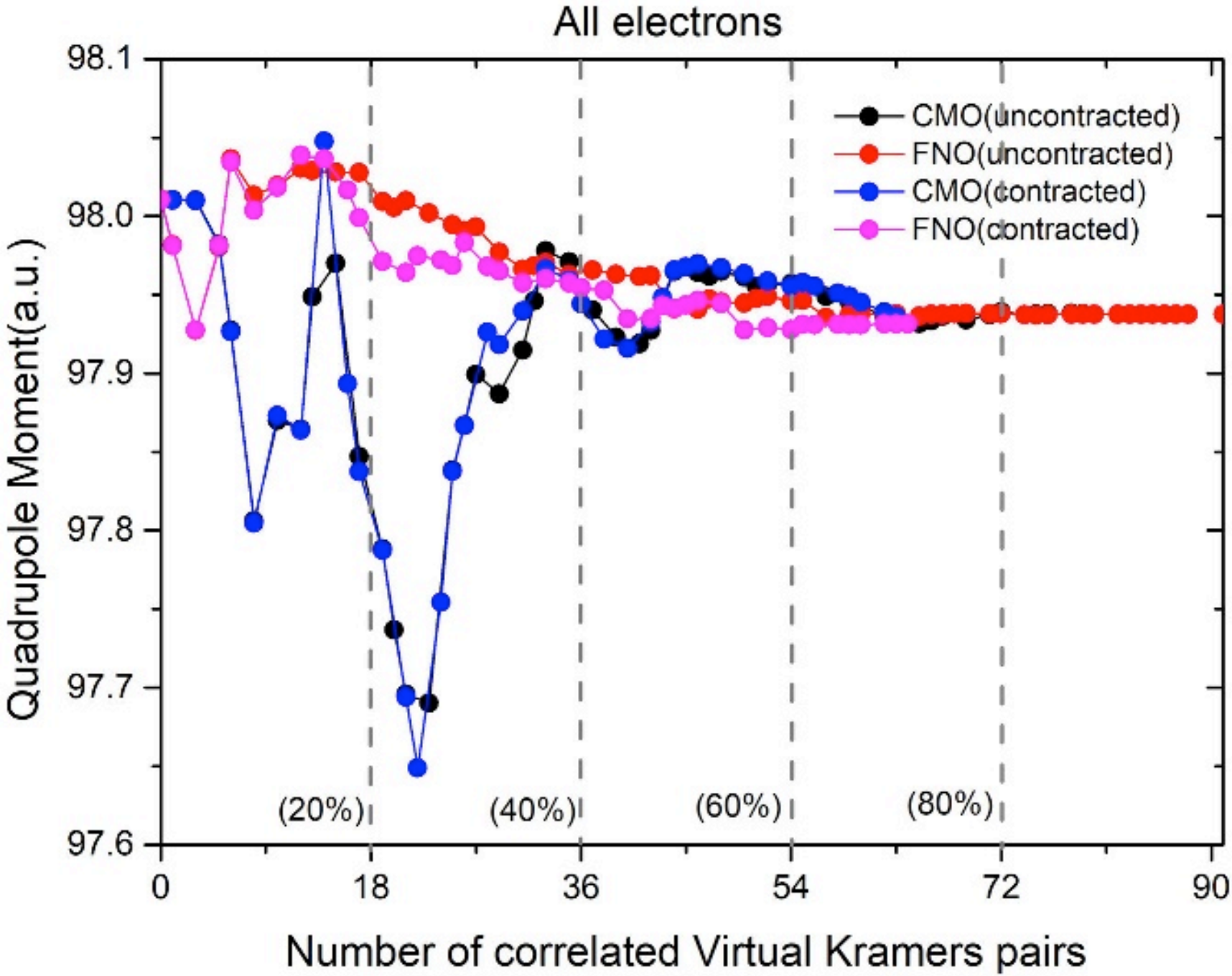

Valence electrons only
CMO
FNO
Quadrupole Moment (a.u.)
98.05
98.00
97.95
97.90
97.85
97.80
97.75
97.70
97.65
97.60
(20%)
(40%)
(60%)
(80%)
0
18
36
54
72
90
Number of correlated Virtual Kramers pairs

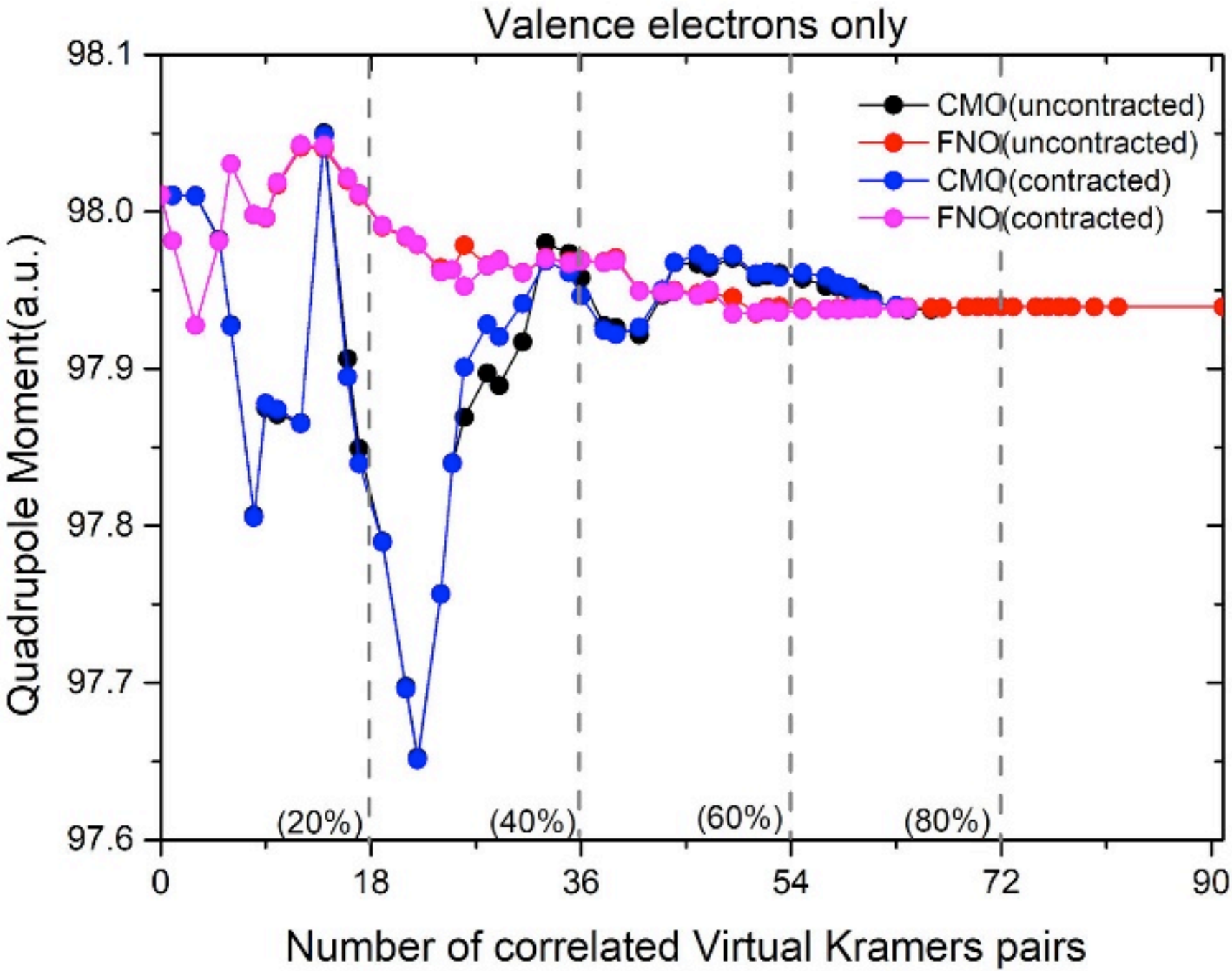

All electrons
CMO
FNO
Electric field gradient (a.u.)
3.8
3.7
3.6
3.5
3.4
3.3
3.2
3.1
0
18
36
54
72
90
(20%)
(40%)
(60%)
(80%)
Number of correlated Virtual Kramers pairs

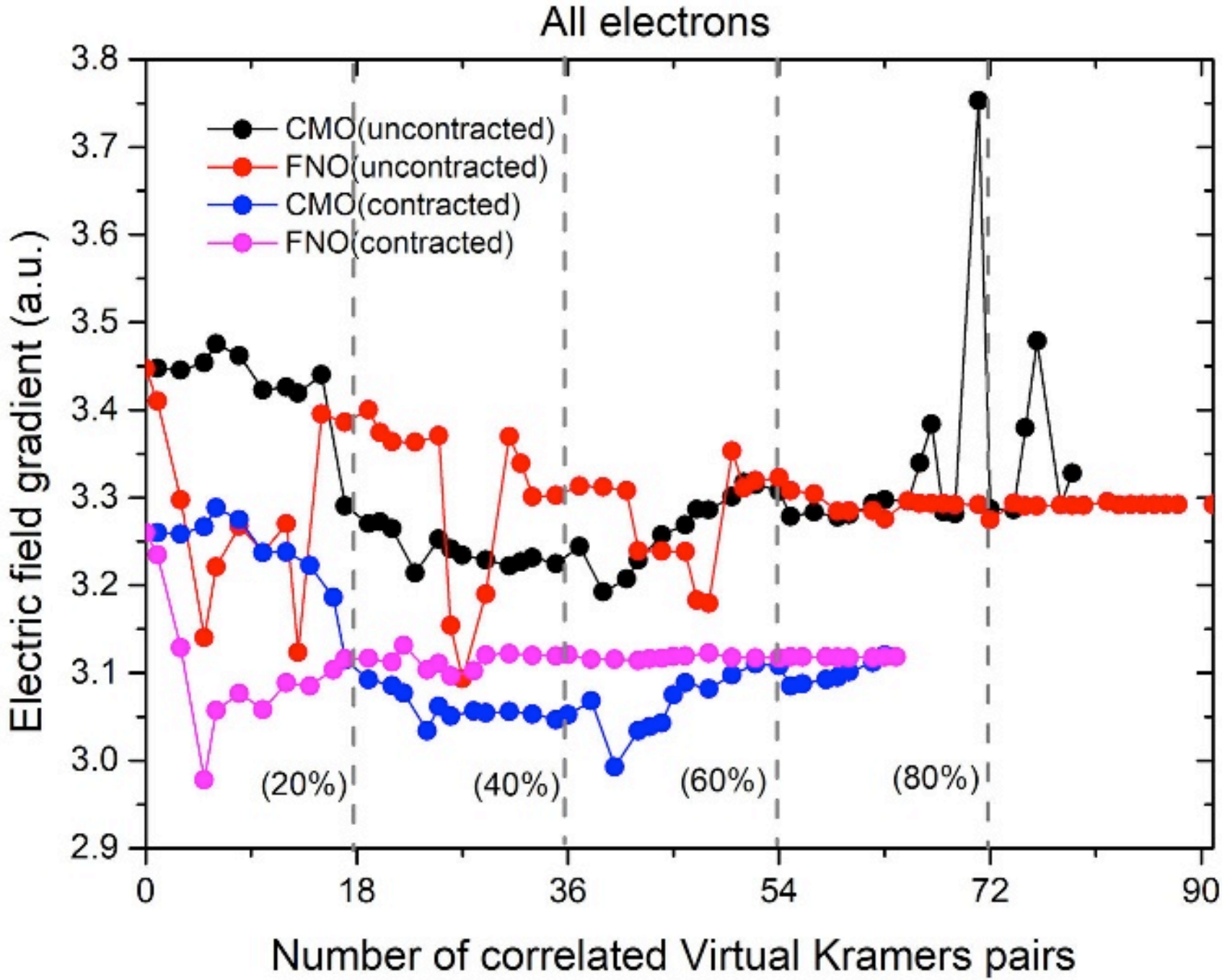

Valence electrons only
CMO
FNO
Electric field gradient (a.u.)
3.8
3.7
3.6
3.5
3.4
3.3
3.2
3.1
(20%)
(40%)
(60%)
(80%)
0
18
36
54
72
90
Number of correlated Virtual Kramers pairs

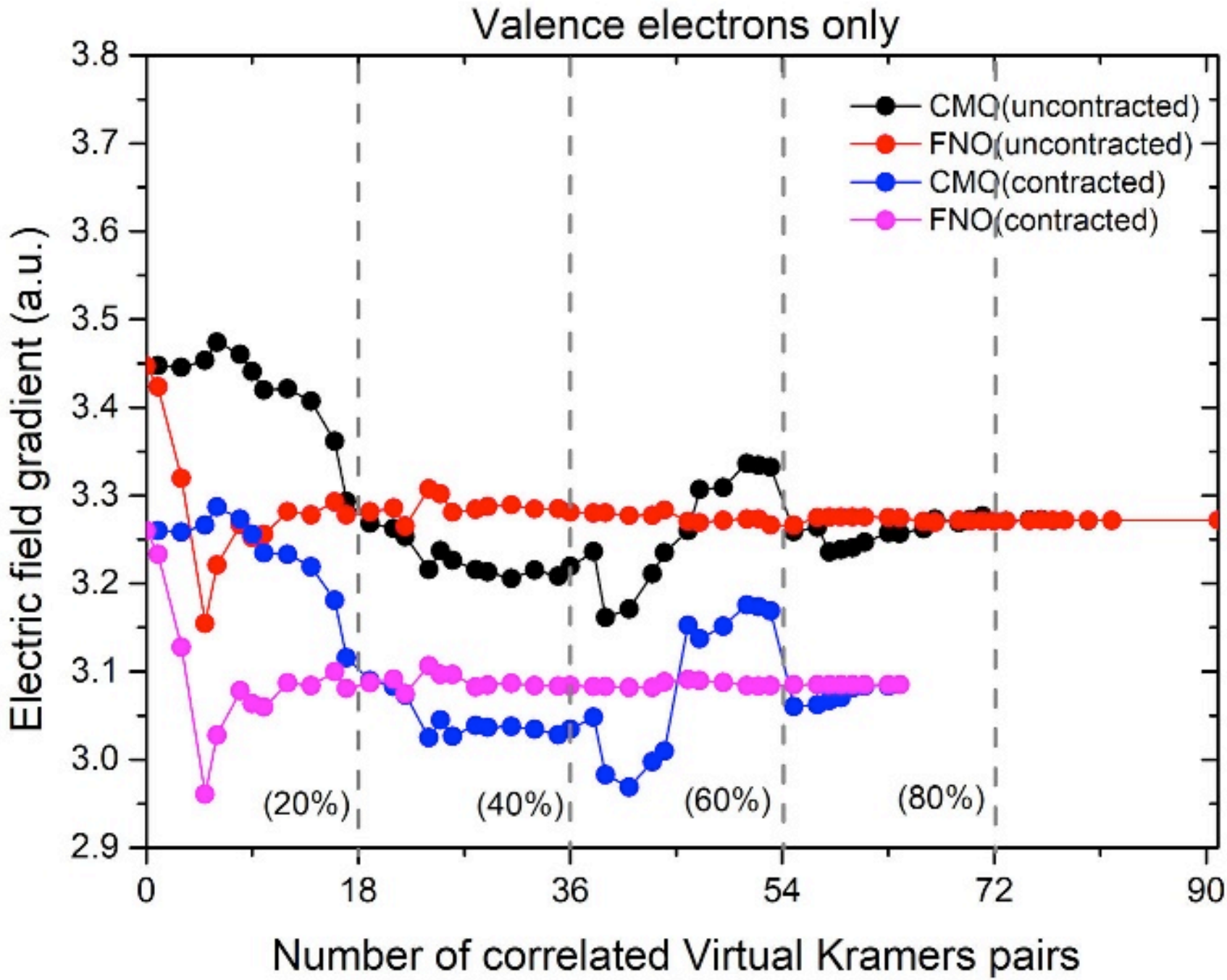

All electrons
CMO(uncontracted)
FNO(uncontracted)
CMO(contracted)
FNO(contracted)
Dipole Moment(a.u.)
-41.30
-41.32
-41.34
-41.36
-41.38
-41.40
-41.42
-41.44
0
18
36
54
72
90
(20%)
(40%)
(60%)
(80%)
Number of correlated Virtual Kramers pairs

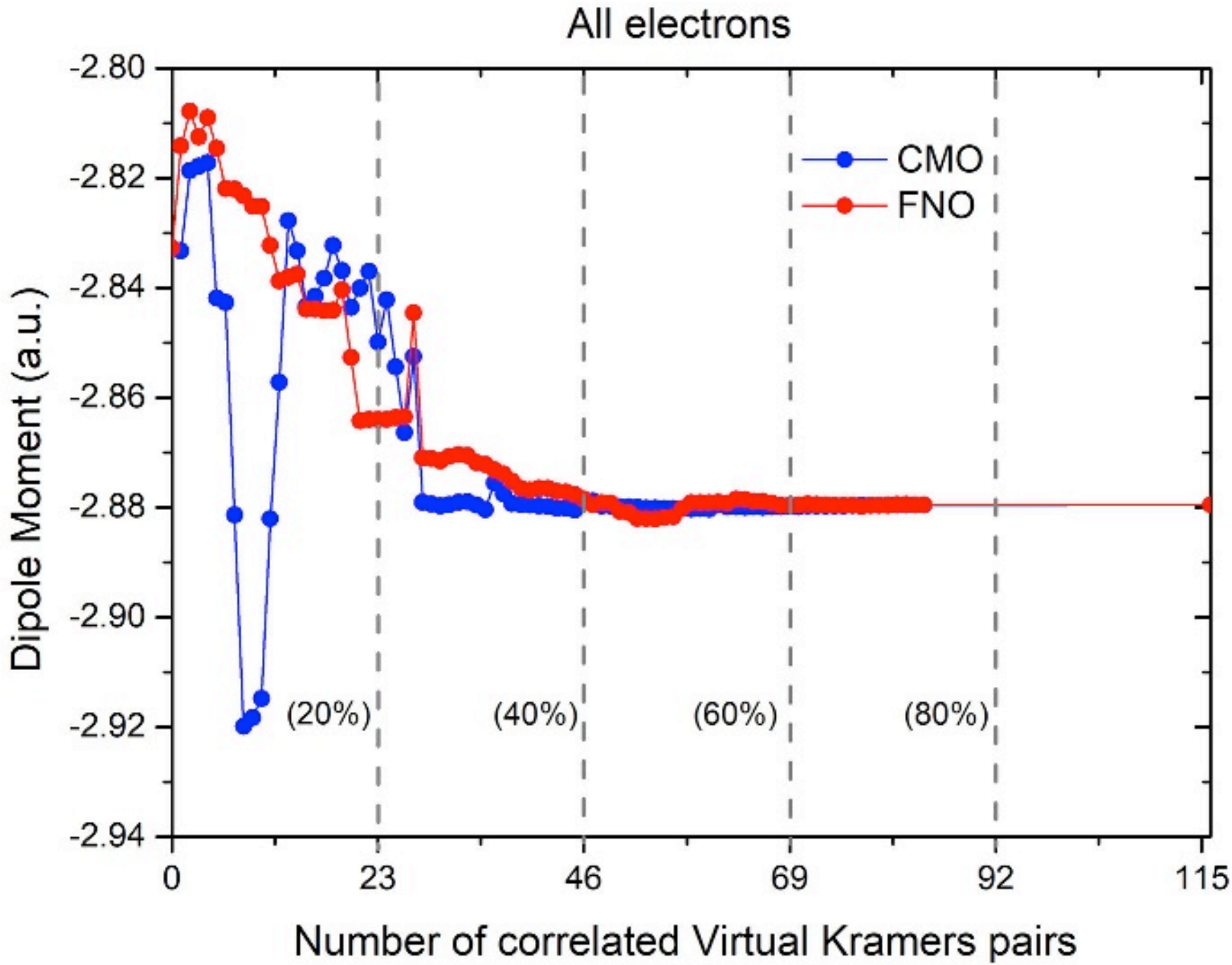

Valence electrons only
CMO(uncontracted)
FNO(uncontracted)
CMO(contracted)
FNO(contracted)
Dipole Moment(a.u.)
-41.30
-41.32
-41.34
-41.36
-41.38
-41.40
-41.42
-41.44
0
18
36
54
72
90
(20%)
(40%)
(60%)
(80%)
Number of correlated Virtual Kramers pairs

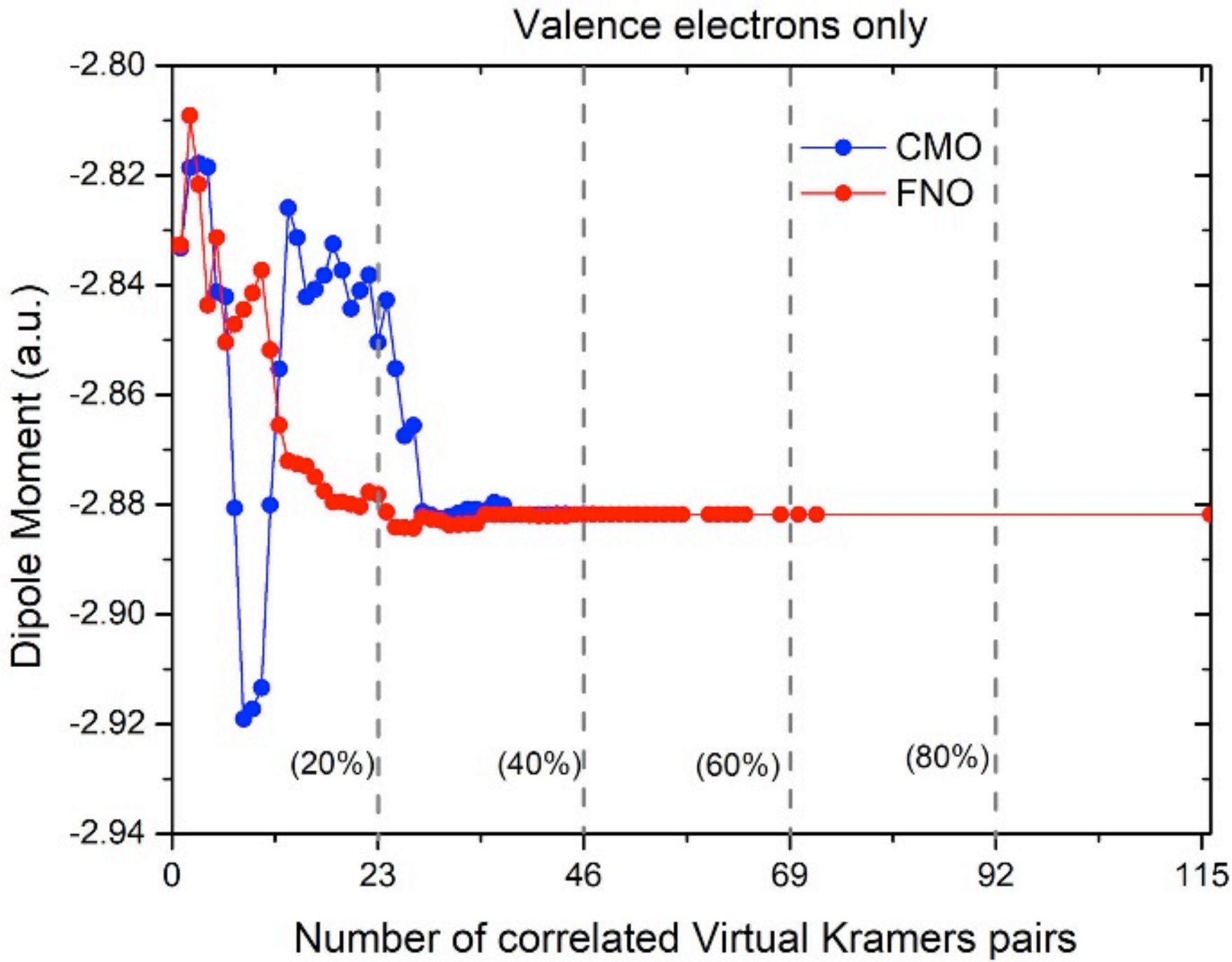

All electrons
98.1
98.0
97.9
97.8
97.7
97.6
Quadrupole Moment(a.u.)
0
18
36
54
72
90
Number of correlated Virtual Kramers pairs
(20%)
(40%)
(60%)
(80%)
CMO(uncontracted)
FNO(uncontracted)
CMO(contracted)
FNO(contracted)

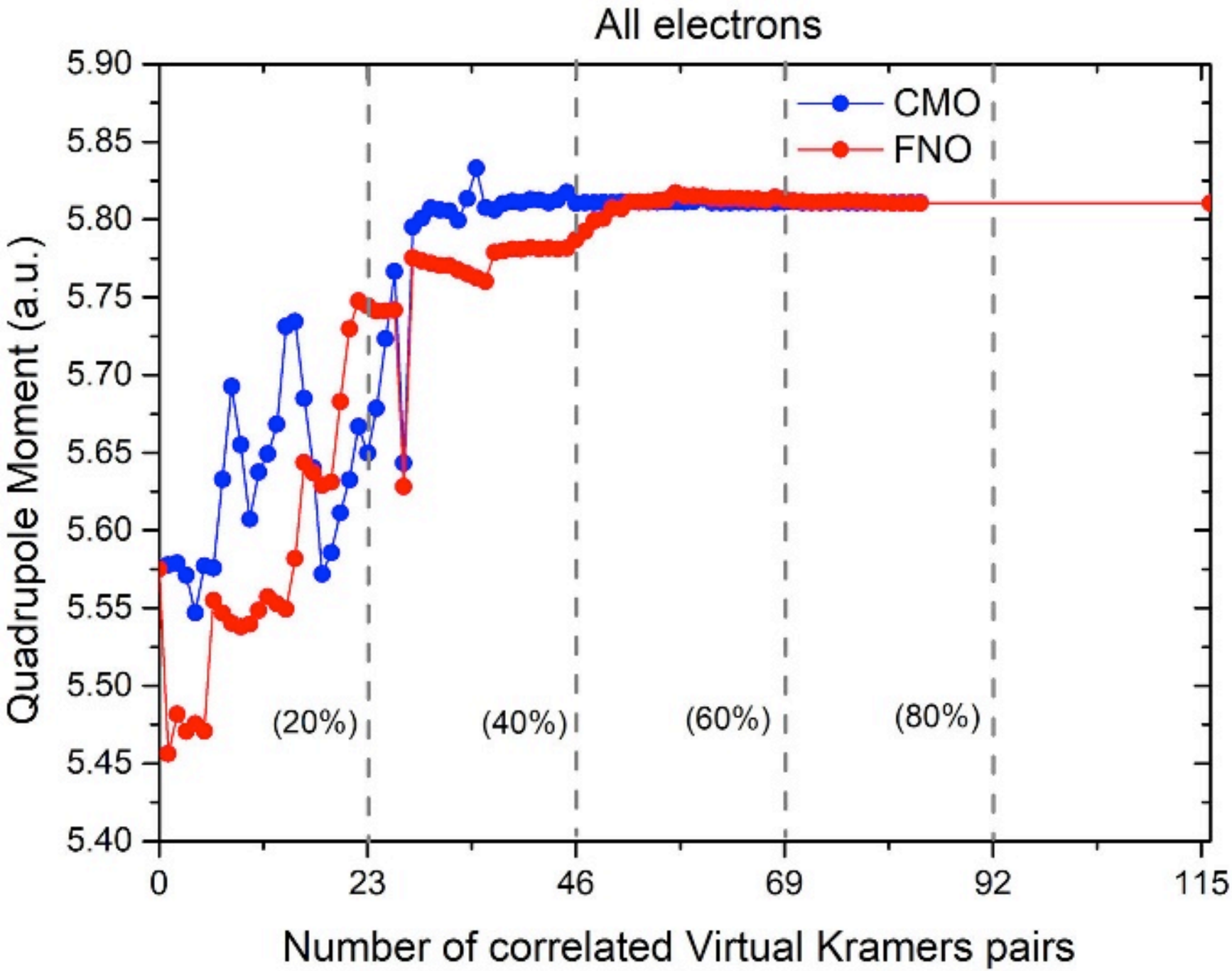

Valence electrons only
98.1
98.0
97.9
97.8
97.7
97.6
Quadrupole Moment(a.u.)
CMO(uncontracted)
FNO(uncontracted)
CMO(contracted)
FNO(contracted)
(20%)
(40%)
(60%)
(80%)
0
18
36
54
72
90
Number of correlated Virtual Kramers pairs

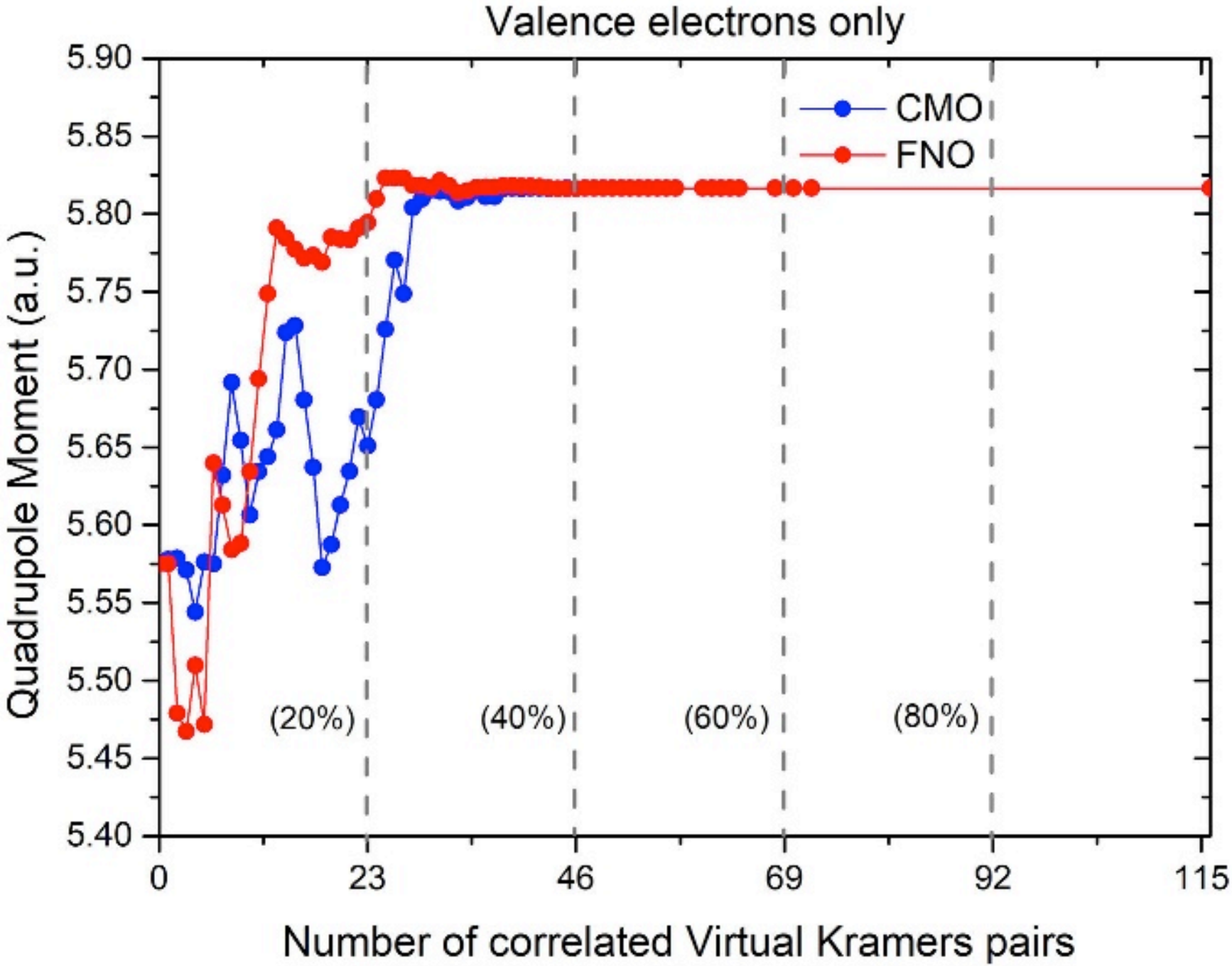

All electrons
CMO(uncontracted)
FNO(uncontracted)
CMO(contracted)
FNO(contracted)
Electric field gradient (a.u.)
3.8
3.7
3.6
3.5
3.4
3.3
3.2
3.1
3.0
2.9
(20%)
(40%)
(60%)
(80%)
0
18
36
54
72
90
Number of correlated Virtual Kramers pairs

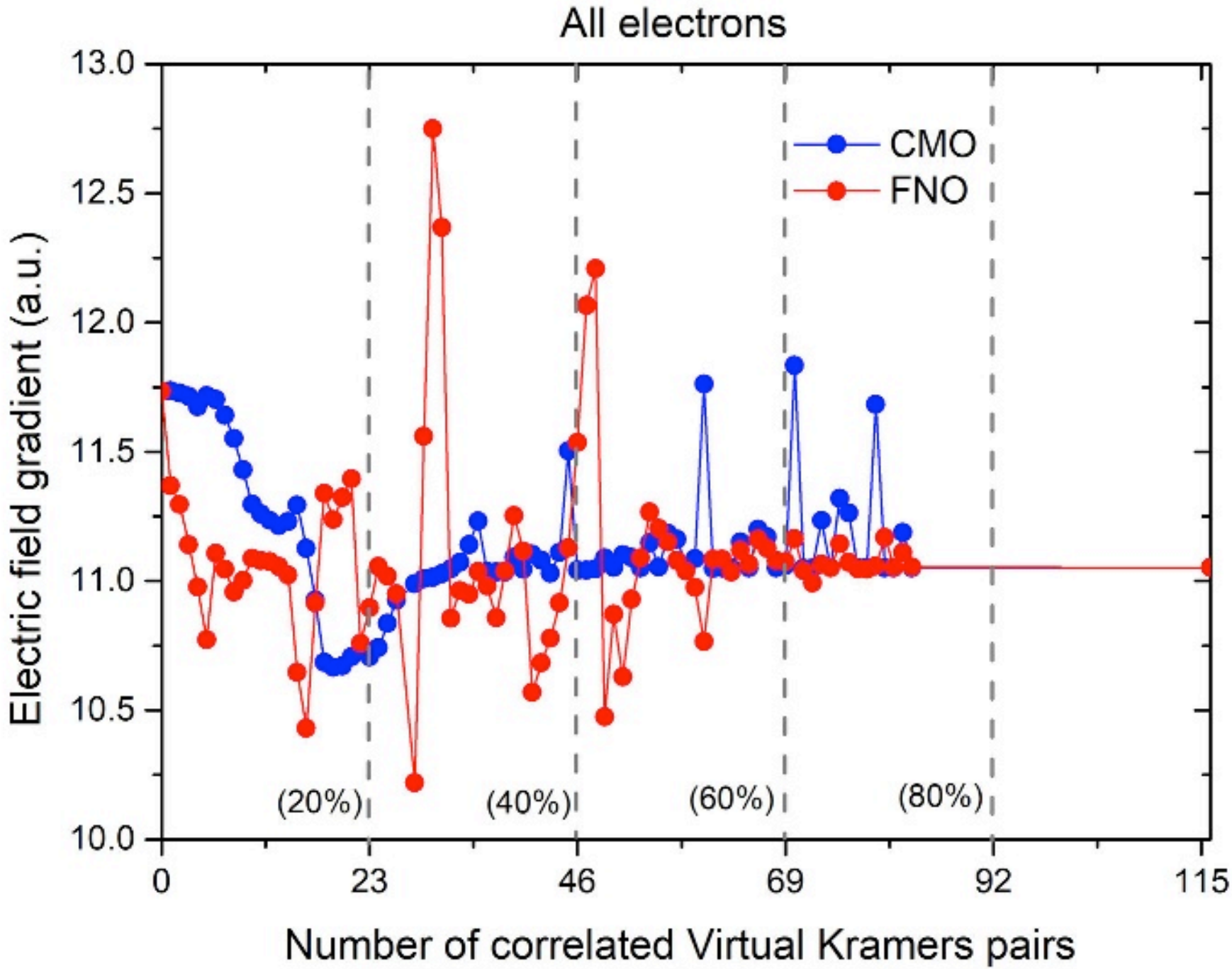

Valence electrons only
CMQ(uncontracted)
FNO(uncontracted)
CMQ(contracted)
FNO(contracted)
Electric field gradient (a.u.)
3.8
3.7
3.6
3.5
3.4
3.3
3.2
3.1
3.0
2.9
(20%)
(40%)
(60%)
(80%)
0
18
36
54
72
90
Number of correlated Virtual Kramers pairs

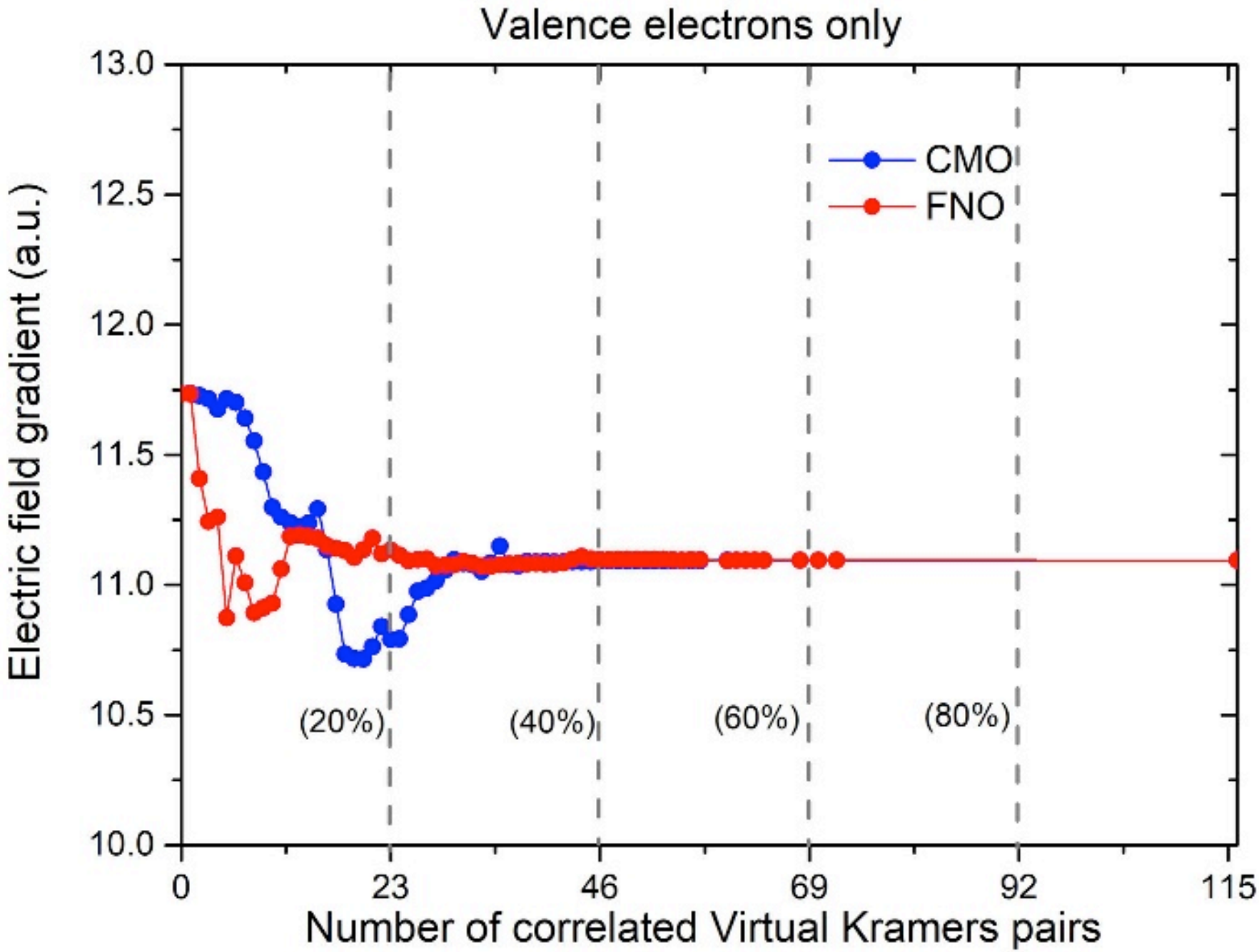

All electrons
CMO
FNO
Dipole Moment (a.u.)
-2.80
-2.82
-2.84
-2.86
-2.88
-2.90
-2.92
-2.94
0
23
46
69
92
115
(20%)
(40%)
(60%)
(80%)
Number of correlated Virtual Kramers pairs


Valence electrons only
CMO
FNO
Dipole Moment (a.u.)
-2.80
-2.82
-2.84
-2.86
-2.88
-2.90
-2.92
-2.94
(20%)
(40%)
(60%)
(80%)
0
23
46
69
92
115
Number of correlated Virtual Kramers pairs


All electrons
CMO
FNO
Quadrupole Moment (a.u.)
5.90
5.85
5.80
5.75
5.70
5.65
5.60
5.55
5.50
5.45
5.40
(20%)
(40%)
(60%)
(80%)
0
23
46
69
92
115
Number of correlated Virtual Kramers pairs


Valence electrons only
CMO
FNO
Quadrupole Moment (a.u.)
5.90
5.85
5.80
5.75
5.70
5.65
5.60
5.55
5.50
5.45
5.40
(20%)
(40%)
(60%)
(80%)
0
23
46
69
92
115
Number of correlated Virtual Kramers pairs


All electrons
CMO
FNO
Electric field gradient (a.u.)
13.0
12.5
12.0
11.5
11.0
10.5
10.0
0
23
46
69
92
115
(20%)
(40%)
(60%)
(80%)
Number of correlated Virtual Kramers pairs


Valence electrons only
Electric field gradient (a.u.)
13.0
12.5
12.0
11.5
11.0
10.5
10.0
CMO
FNO
(20%)
(40%)
(60%)
(80%)
0
23
46
69
92
115
Number of correlated Virtual Kramers pairs

Fraction of the PV CCSD value

Fraction of virtual space retained