Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Tensor Decomposition in post-Hartree Fock
Methods.

by

*Udo Benedikt, Alexander Auer, Mike Espig, and Wolfgang
Hackbusch*

Tensor Decomposition in post-Hartree Fock Methods.

I. Two-Electron Integrals and MP2

Udo Benedikt,[1] Alexander A. Auer,[1, a)] Mike Espig,[2, b)] and Wolfgang Hackbusch[2]

[1)]*Max-Planck-Institut for Iron Research GmbH, Max-Planck-Strasse 1,*

D-40237 Düsseldorf, Germany

[2)]*Max-Planck-Institute for Mathematics in the Sciences, Inselstraße 22,*

D-04103 Leipzig, Germany

(Dated: 22 October 2010)

A new approximation for post-Hartree Fock (HF) methods is presented applying tensor decomposition techniques in the canonical product tensor format. In this ansatz, multidimensional tensors like integrals or wavefunction parameters are processed as an expansion of one-dimensional representing vectors. This approach has the potential to decrease the computational effort and the storage requirements of conventional algorithms drastically while allowing for rigorous truncation and error estimation. For post-HF ab initio methods for example, storage is reduced to $\mathcal{O}(d \cdot R \cdot n)$ with d being the number of dimensions of the full tensor, R being the expansion length (rank) of the tensor decomposition and n being the number of entries in each dimension (i.e. the orbital index). If all tensors are expressed in the canonical format, the computational effort for any subsequent tensor contraction can be reduced to $\mathcal{O}(R^2 \cdot n)$. We discuss details of the implementation, especially the decomposition of the two electron integrals, the AO-MO transformation, the MP2 energy expression and the perspective for Coupled Cluster methods. An algorithm for rank-reduction is presented that parallelizes trivially. For a set of representative examples the scaling of the decomposition rank with system and basis set size is found to be $\mathcal{O}(N^{1.8})$ for the AO integrals, $\mathcal{O}(N^{1.4})$ for the MO integrals and $\mathcal{O}(N^{1.2})$ for the MP2 t_2-amplitudes (N denotes a measure of system size) if the upper bound of the error in the ℓ^2-norm is chosen as $\varepsilon = 10^{-2}$. This leads to an error in the MP2 energy in the order of mHartree.

[a)]Electronic mail: auer@mpie.de

[b)]Electronic mail: Mike.Espig@mis.mpg.de

I. INTRODUCTION

Tensor decomposition techniques originate from the early works of Hitchcock[1] and its first applications with the work by Tucker[2], Carroll and Chang[3] and Harshman[4] in the field of psychometrics. Later, these decomposition techniques were used also in the fields of chemometrics, where they have become very popular[5–8]. Over the last ten years tensor decomposition techniques found their way into numerical linear algebra[9,10], signal processing[11,12], computer vision[13–15], data mining[16–18], neuroscience[19,20] and many more. The advantage of using decomposed tensors is mainly due to the fact of reduced data handling and decreased complexity of mathematical operations dealing with decomposed high dimensional tensors. Decomposition techniques often allow to treat large amount of data very efficiently, which would not be possible with standard procedures due to the "curse of dimensionality"[21] of higher-order tensors.

A prominent example for "curse of dimensionality" problems are Coupled Cluster (CC), Configuration Interaction (CI) or perturbative post-Hartree Fock (HF) methods. For example, the computational effort for the CCSDT method scales as $\mathcal{O}(N^8)$, for CCSDTQ it scales as $\mathcal{O}(N^{10})$, where N is a parameter of system size. Due to the steep scaling of the number of parameters to store and manipulate, applications of these exquisite methods are limited to small systems.

In recent years a lot of effort has been devoted to overcome this steep scaling by reducing the number of wavefunction parameters. One ansatz is to use tensor decomposition techniques in order to compress the amount of data. Methods like the Cholesky decomposition (CD) can be used to decompose four dimensional arrays like the two-electron integrals into two dimensional quantities[22–33]. Also other approximations like density fitting (DF)[34–37] / resolution of identity (RI) techniques[38–41] are used for approximating the two-electron integrals. These approaches allow to reduce the computational effort typically by one order of magnitude and have been commonly used for a long time in the framework of density functional theory (DFT). In explicitly correlated methods[42–47] such approaches are essential in order to overcome problems associated with the arising additional multi-centre integrals. If decomposition techniques are applied to the the energy denominator like in the in Laplace transformed Møller-Plesset perturbation theory (MP2)[33,48–54], the complexity of such methods can, in combination with further approximations, be reduced to linear scaling[51,55–57].

A further possibility to reduce the number of parameters in post-HF calculations in this spirit is the construction of an optimal valence space representation in which the number of orbitals that have to be correlated is significantly decreased[58–68]. In the framework of higher order CC methods also Singular-Value-Decomposition (SVD) techniques have been studied[69,70].

In this work we present an adaption of a numerical scheme from applied mathematics for post-HF electronic structure methods. In contrast to most existing decomposition schemes like RI/DF or CD this approach is not aimed at reducing the effort to calculate the two-electron integrals but rather to reduce the steep scaling of post-HF methods. In this scheme, multidimensional tensors are represented by using an expansion in one-dimensional quantities. Rather than constructing a low rank approximation of a highly dimensional tensor a priori, which is for example difficult for wavefunction parameters, the detour via a trivial decomposition with high rank and a subsequent rank reduction is taken. In the optimal case, the rank reduction procedure should then yield parameters that have been compressed to their most compact form. For the practicability of the resulting approximation the decisive quality is of cause the scaling of the expansion length (rank) with system and basis set size, which is investigated in detail for the two-electron integrals and an estimate of the CC amplitudes as obtained from MP2.

This paper is organized in the following way: In Sec. II A a short introduction into the tensor decomposition techniques applied in this work is given. The tensor format and the rank reduction algorithm are discussed in detail. In the following sections II B - II D the application of the tensor decomposition and low rank approximation are demonstrated for different tensors in post-HF methods and the implications for the amount of storage and the complexity of tensor manipulations are discussed. In Section IV the results of a series of benchmark calculations are presented as a proof of principle. Finally, in Section V the implications of the results on the amount of storage and the complexity of tensor manipulations are summarized followed by a short outlook for high-level post-HF ab initio methods.

II. THEORY

A. Tensor product approximation and rank reduction

With the help of decomposition methods a general d-dimensional tensor can be factorized into a sum of representing vectors[1,71–75]. For this purpose different tensor formats exist, like the TUCKER model[2,9,76] or the canonical product format (CP)[1,77–81] that is often obtained by a parallel factor decomposition (PARAFAC), also termed canonical decomposition (CANDECOMP)[3,4,9,82–84]. In the following, we focus on the CP format because it offers certain advantages for post-HF methods and yields a representing format of minimal dimensionality. In the CP format a tensor $A(w, x, y, z) \in \mathbb{R}^{n \times n \times n \times n}$ in 4-dimensions (for example the two-electron integrals) is expanded as

$$\mathbf{A} = \sum_{r=1}^{R} \mathbf{a}_r^{(w)} \otimes \mathbf{a}_r^{(x)} \otimes \mathbf{a}_r^{(y)} \otimes \mathbf{a}_r^{(z)} \tag{1}$$

with summation length R called rank of the representation and n describing the number of entries in each dimension[85]. The four representing vectors $\mathbf{a}_r^{(w)}, \mathbf{a}_r^{(x)}, \mathbf{a}_r^{(y)}$ and $\mathbf{a}_r^{(z)}$ are connected by application of the Kronecker product. Here and in the following, it should be noted the superscripts $(w), (x), (y)$ and (z) in the representing vectors of a tensor in the CP format are, strictly speaking, not indices (in sense of indices to denote matrix elements) as such but are rather used to distinguish the representing vectors for the different dimensions of the original tensor. E.g. in Eqn. 1 the vectors $\mathbf{a}_r^{(w)}$ are the r representing vectors for the first dimension (w) of the tensor $A(w, x, y, z)$ and each has a length of n.

While shown for the example of four dimensions above, this type of decomposition can be extended to any dimensionality. In contrast to the full 4-dimensional tensor that requires to store n^4 entries, the memory requirement is reduced to $4 \cdot n \cdot R$. In this case, the complexity of algebraic operations can also be decreased to only linear scaling with respect to the dimension so that the "curse of dimensionality" can be overcome[77,78]. However, this ansatz is only beneficial if the 4-dimensional tensor in Eqn. 1 can be represented by a low-rank approximation, where the rank R is significantly smaller than n^3.

This decomposition is related to Cholesky decomposition and RI/density fitting, and while the latter differ in the procedure to obtain the representing matrices, they share a

similar form of decomposition :

$$A(w, x, y, z) = \sum_{r=1}^{R} a_r^{wx} \cdot a_r^{yz}. \tag{2}$$

The difference to the CP format (Eqn. 1) is that rather than having one representing matrix (a_r^{wx} and a_r^{yz} in Eqn. 2 denote different elements of the same matrix) with for example $n^2 \cdot R$ elements, in the CP format different representing vectors are used that contain $n \cdot R$ elements. The disadvantage is, that the procedure to construct the low rank representation is an iterative scheme that (currently) requires the computation, and partly storage, of all integrals. Consequently, the application of this scheme is computationally more demanding than the calculation of the integrals by itself and is thus not likely to be of advantage for DFT or HF algorithms. However, in contrast to RI/DF or CD this ansatz can be applied to any kind of tensor like wavefunction parameters such as higher order amplitudes or CI coefficients, arbitrary integrals etc. Therefore, the technique presented here are intended for the usage in high-level post-HF methods. As also lambda and perturbed amplitudes that occur in analytical derivatives of CC theory[86,87] can be decomposed, this approach can also directly be applied in the calculation of molecular properties via analytical derivatives.

While for two dimensions, there exists a unique procedure to obtain an optimal low-rank approximation, namely the SVD, this is not the case for higher dimensions. In principle, the best way to obtain a representation in the CP format is to derive it a priori from the equations that define the high dimensional tensors, such that these do not have to be calculated explicitly[79,80,88–90]. However, if this is not possible, one can always find a trivial decomposition of this tensor with very high rank, that can be reduced with a reduction algorithm (c.f. Eqn. 3)[84,91]. In Fig. 1 the trivial decomposition is illustrated for a small example in two dimensions. For the trivial decomposition, the first representing vector codes the number of the row as a unit vector. The last representing vector is initialized with all entries in the specific row described by the first vector. To decompose the full matrix shown in Fig. 1 three pairs of representing vectors are necessary leading to an initial rank of three. For a 4-dimensional quantity four vectors are required, three unit vectors holding the first three indices in vector format and the last representing vector containing all values for this special multi-index. In this trivial decomposition the initial rank for a four dimensional tensor is always n^3. To obtain a low rank approximation, this initial quantity can be reduced

in rank within a given accuracy ε:

$$J(\tilde{\mathbf{A}}) := ||\mathbf{A} - \tilde{\mathbf{A}}|| \leq \varepsilon \quad \text{with} \quad \tilde{\mathbf{A}} = \sum_{r=1}^{\tilde{R}} \tilde{\mathbf{a}}_r^{(w)} \otimes \tilde{\mathbf{a}}_r^{(x)} \otimes \tilde{\mathbf{a}}_r^{(y)} \otimes \tilde{\mathbf{a}}_r^{(z)}, \tag{3}$$

where $\tilde{\mathbf{A}}$ is the new approximated tensor with the reduced rank $\tilde{R}$. This is done by a minimization procedure where the error ε in the ℓ^2-norm, to which the low-rank approximation will be converged, is given as input to the reduction algorithm. For solving this minimization problem there are different choices like an alternating least square (ALS) scheme[77,78], a modified Newton method[80,90] or an accelerated gradient (AG)[84] algorithm.

In this work the AG algorithm (see Fig. 2) has been used, which has certain advantages over the other methods, such as a better convergence and a complexity comparable to the ALS method. For a detailed analysis of the AG methods see reference 84. The crucial part of the AG algorithm is the computation of the exact line search parameter $\alpha_k \in \mathbb{R}_{\geq 0}$. Given a direction $\mathbf{D}^k$, a solution of the one-dimensional nonlinear equation

$$p(\alpha_k) = \left\langle J'(\tilde{\mathbf{A}}^k + \alpha \mathbf{D}^k), \mathbf{D}^k \right\rangle = 0.$$

has to be found. Normally the exact line search is avoided and an Armijo type inexact line search[92] is applied. Generally, for CP tensors of order d, the function p is a polynomial of degree at most $2d - 1$. Hence, a third order derivative-free procedure (3-PG) for finding zeros of a function[93] is applied. The 3-PG method is globally linear convergent for a function $f \in C^2[x, y]$, where $x, y \in \mathbb{R}$ with $f(x)f(y) < 0$. The order of convergence is defined by the real root of the polynomial $\tau \mapsto \tau^3 - \tau^2 - \tau - 1$ (≈ 1.8393). Moreover, the 3-PG method is equivalent to the Newton method for polynomials of degree three.

The complexity of the computation of the gradient J' is $\mathcal{O}(d \cdot \tilde{R} \cdot n \cdot (\tilde{R} + R))$, where d is the dimension of the tensor. Since the most expensive part in the AG method is the calculation of the gradient, the overall complexity of the AG method is

$$k_{\max} \mathcal{O}(d \cdot \tilde{R} \cdot n \cdot (\tilde{R} + R)), \tag{4}$$

where $k_{\max}$ denotes the maximal number of iterations in the minimization procedure (see Fig. 2).

The current rank-reduction scheme shown in Fig. 2 is organised as follows : First, a pivoting routine is used to find a large entry in the original tensor on a special cross over all

dimensions, which is used as a new rank in the representing vectors for the approximation. This new rank is then optimized to improve the **A**-residual value and is added to the set ranks in $\tilde{\mathbf{A}}$. An iterative procedure is then used to improve all ranks in the representing vectors to further lower the difference between $\tilde{\mathbf{A}}$ and $\mathbf{A}$. If this difference is lower than the given threshold parameter ε, the procedure is stopped and the final number of ranks has been obtained. Otherwise, the cycle starts again with the pivoting routine and adds another rank to the approximated $\tilde{\mathbf{A}}$. This way, the rank grows one by one during the iterations until the given accuracy is reached. A detailed description of the full algorithm can be found in references 90 and 84.

It should be pointed out that this scheme has not been modified in any way in order to optimise it for the problem of approximating two-electron integrals or wavefunction parameters. As the work presented here should be understood as a proof of principle, there is extensive room for improvement from a viewpoint of computational efficiency. Current efforts are directed at making the procedure computationally efficient and will be reported in a further publication.

Furthermore, it should be noted that in the scheme applied in this work, no use of permutational symmetry for the indices of the tensors are made. While this results in a numerical "permutational symmetry violation" in principle, this does not pose a problem in practice (see also section IV A).

B. Decomposition of the two-electron integrals and integral transformation

As the amount of integrals to be computed increases rapidly with system and basis set size, several approaches for low rank approximation of the two-electron integral tensor already exist, as already mentioned in the introduction. For a representation in the CP tensor format the two-electron integrals can be cast into a decomposed form by trivial decomposition, so that the initial rank R scales with $\mathcal{O}(N^3)$, where N is the number of basis functions

$$\langle \mu\nu|\sigma\rho \rangle = \sum_{r=1}^{R} \boldsymbol{\chi}_r^{(\mu)} \otimes \boldsymbol{\chi}_r^{(\nu)} \otimes \boldsymbol{\chi}_r^{(\sigma)} \otimes \boldsymbol{\chi}_r^{(\rho)}. \tag{5}$$

Here and in the following we use the convention that μ, ν, σ, ρ denote AO-indices, a, b, c, d denote virtual and i, j, k, l occupied indices, e, f denote virtual and m, n denote occupied summation indices. It should be noted, that superscripts in parenthesis do not denote indices

in the sense of denoting matrix elements, but are used to distinguish the representing vectors for the different dimensions of the tensor.

From Eqn. 5 it can be seen that for the trivial decomposition the number of parameters to store even increases from N^4 for the full tensor to $4 \cdot R \cdot N = 4 \cdot N^4$ for the representation in the CP format. Thus it is only beneficial to use integrals in the CP decomposed format for post-HF approximations if the rank can be reduced to less than $\frac{1}{4}N^3$. Only then the memory requirements for the representation in the CP format are lower than the amount of storage for the full two-electron integrals. Thus, the decisive quantity is the final rank after reduction and especially the scaling of the reduced rank with system and basis set size. Together with the prefactor of the rank reduction algorithm, this scaling behaviour will determine the cross over point to conventional algorithms. Yet, already from the success of schemes like RI/DF and CD for lowering the amount of storage for the integrals one can estimate that it should be possible to find a low rank representation in the CP tensor format that exhibits low scaling with system and basis set size. Furthermore, as for CD techniques linear scaling algorithms have been devised[30], similar approaches should also be applicable to the tensor decomposition methods presented in the following.

Having cast the two-electron integrals in the AO-basis into the CP format, the AO-MO integral transformation (here given for the four-virtual index integrals)

$$\langle ab|cd \rangle = \sum_{\mu\nu\sigma\rho} C_\mu^a C_\nu^b C_\sigma^c C_\rho^d \langle \mu\nu|\sigma\rho \rangle \tag{6}$$

can be written as the separate transformation of the four representing vectors $\boldsymbol{\chi}$

$$v_{cd}^{ab} = \sum_{\mu\nu\sigma\rho} C_\mu^a C_\nu^b C_\sigma^c C_\rho^d \sum_{r=1}^{R} \boldsymbol{\chi}_r^{(\mu)} \otimes \boldsymbol{\chi}_r^{(\nu)} \otimes \boldsymbol{\chi}_r^{(\sigma)} \otimes \boldsymbol{\chi}_r^{(\rho)} \tag{7}$$

$$= \sum_{r=1}^{R} \left(\sum_\mu C_\mu^a \boldsymbol{\chi}_r^{(\mu)} \right) \otimes \left(\sum_\nu C_\nu^b \boldsymbol{\chi}_r^{(\nu)} \right) \otimes \left(\sum_\sigma C_\sigma^c \boldsymbol{\chi}_r^{(\sigma)} \right) \otimes \left(\sum_\rho C_\rho^d \boldsymbol{\chi}_r^{(\rho)} \right) \tag{8}$$

$$v_{cd}^{ab} = \sum_{r=1}^{R'=R} \boldsymbol{v}_r^{(a)} \otimes \boldsymbol{v}_r^{(b)} \otimes \boldsymbol{v}_r^{(c)} \otimes \boldsymbol{v}_r^{(d)}. \tag{9}$$

Thus, the integral transformation is carried out by simple matrix-vector multiplications of the MO coefficient matrices C with the corresponding representing vectors of the decomposed two-electron integrals in the AO-basis. The complexity of this transformation can be reduced from $\mathcal{O}(N^4 \cdot virt + N^3 \cdot virt^2 + N^2 \cdot virt^3 + N \cdot virt^4)$ for the conventional algorithm to a scaling of $\mathcal{O}(N \cdot virt \cdot R)$ in the CP tensor format, were $virt$ denotes the number

of virtual orbitals. One important fact is that in principle the rank does not change during the transformation ($R' = R$). This means that the compression in the CP tensor format is independent of the basis chosen to represent the two-electron integrals. If a low rank approximation is found starting from the canonical orbitals it should have the same rank as for localized or natural orbitals. Furthermore, after a low rank approximation has been found in the AO basis, a low rank representation of the integrals in the MO basis can be obtained.

All other types of MO integrals, as they for example occur in CC calculations, can be obtained in the same way, but the CP format allows also to construct other types of MO integrals in a very convenient way. If only the MO integrals with four occupied and four virtual indices are transformed, all other types of integrals can then be composed from the representing vectors of these two objects. In Eqn. 10 this is demonstrated for the two occupied, two virtual index integrals.

$$
\left.
\begin{array}{l}
v_{cd}^{ab} = \sum_{r=1}^{R'} \boldsymbol{v}_r^{(a)} \otimes \boldsymbol{v}_r^{(b)} \otimes \boldsymbol{v}_r^{(c)} \otimes \boldsymbol{v}_r^{(d)} \\
v_{ij}^{kl} = \sum_{r=1}^{R'} \boldsymbol{v}_r^{(k)} \otimes \boldsymbol{v}_r^{(l)} \otimes \boldsymbol{v}_r^{(i)} \otimes \boldsymbol{v}_r^{(j)}
\end{array}
\right\}
\Rightarrow v_{ij}^{ab} = \sum_{r=1}^{R'} \boldsymbol{v}_r^{(a)} \otimes \boldsymbol{v}_r^{(b)} \otimes \boldsymbol{v}_r^{(i)} \otimes \boldsymbol{v}_r^{(j)}
\qquad (10)
$$

While the rank of the integral representation does not change during transformation, the occupied and the virtual space are only subspaces of the AO-basis, so that the physical information the integrals carry is reduced upon transformation. Due to this fact, it should be possible to reduce the ranks of the different MO-integrals even further.

It should be noted that an alternative could also be to first transform the AO-integrals after trivial decomposition to the MO-basis and only reduce the rank of the MO-integrals or to perform the trivial decomposition and subsequent reduction only for the MO-integrals after they have been obtained from any quantum chemistry software package. Furthermore, it might be advantageous to derive schemes in which the starting point for the decomposition and rank reduction are representations of the two-electron integrals as obtained from more efficient schemes like RI or DF techniques.

C.   MP2 algorithm based on decomposed integrals and denominator

While we discuss an MP2 algorithm in the following, this should be regarded as a first step towards the application of tensor decomposition techniques in CC theory. The greatest benefit of low-dimensional, low-rank approximations is expected for higher order CC methods, for which the computational effort is prohibitive for larger applications due to the manipulation of tensors with six, eight or more dimensions.

In the framework of CC theory MP2 can be considered to yield a first-order estimate of the t_2 amplitudes[94]:

$$t_{ij}^{ab} = v_{ij}^{ab} \cdot D_{ij}^{ab} \tag{11}$$

where v_{ij}^{ab} are antisymmetrized two-electron integrals and D_{ij}^{ab} is the energy denominator defined in Eq. 12.

$$D_{ij}^{ab} = \frac{1}{-\epsilon_a - \epsilon_b + \epsilon_i + \epsilon_j} \tag{12}$$

This ansatz is not only convenient from a formal point of view, but also allows to estimate the performance of the CP tensor format for post-HF methods like CC.

If Eqn. 11 is based on integrals represented using a low rank CP tensor approximation, the complete amplitude expression can be decomposed if also the energy denominator D_{ij}^{ab} is cast into a product format. For this purpose the approximation of $1/x$ by exponential sums[48,52,54] is used. One element of the denominator can be written as

$$\frac{1}{-\epsilon_a - \epsilon_b + \epsilon_i + \epsilon_j} \approx \sum_{s=1}^{S} \omega_s \cdot exp(-\alpha_s(-\epsilon_a - \epsilon_b + \epsilon_i + \epsilon_j)) \tag{13}$$

$$= \sum_{s=1}^{S} \omega_s \cdot exp(\alpha_s\epsilon_a) \cdot exp(\alpha_s\epsilon_b) \cdot exp(-\alpha_s\epsilon_i) \cdot exp(-\alpha_s\epsilon_j), \tag{14}$$

so that the full tensor D_{ij}^{ab} can be obtained in the CP format as

$$D_{ij}^{ab} \approx \sum_{s=1}^{S} \boldsymbol{\epsilon}_s^{(a)} \otimes \boldsymbol{\epsilon}_s^{(b)} \otimes \boldsymbol{\epsilon}_s^{(i)} \otimes \boldsymbol{\epsilon}_s^{(j)}. \tag{15}$$

The accuracy of this approximation, like in the Laplace MP2 method[33,48–52,54], can be adjusted by choosing an appropriate threshold which results in a fixed decomposition rank for the denominator expression[52]. If decomposed integrals and the denominator in the CP format are combined, the amplitudes from Eqn. 11 can be obtained as

$$t_{ij}^{ab} \approx \left(\sum_{r=1}^{R'} \boldsymbol{v}_r^{(a)} \otimes \boldsymbol{v}_r^{(b)} \otimes \boldsymbol{v}_r^{(i)} \otimes \boldsymbol{v}_r^{(j)} \right) \cdot \left(\sum_{s=1}^{S} \boldsymbol{\epsilon}_s^{(a)} \otimes \boldsymbol{\epsilon}_s^{(b)} \otimes \boldsymbol{\epsilon}_s^{(i)} \otimes \boldsymbol{\epsilon}_s^{(j)} \right) \tag{16}$$

$$= \sum_{r=1}^{R'} \sum_{s=1}^{S} \left(\boldsymbol{v}_r^{(a)} \cdot \boldsymbol{\epsilon}_s^{(a)} \right) \otimes \left(\boldsymbol{v}_r^{(b)} \cdot \boldsymbol{\epsilon}_s^{(b)} \right) \otimes \left(\boldsymbol{v}_r^{(i)} \cdot \boldsymbol{\epsilon}_s^{(i)} \right) \otimes \left(\boldsymbol{v}_r^{(j)} \cdot \boldsymbol{\epsilon}_s^{(j)} \right) \tag{17}$$

$$= \sum_{q=1}^{Q=R' \cdot S} \boldsymbol{t}_q^{(a)} \otimes \boldsymbol{t}_q^{(b)} \otimes \boldsymbol{t}_q^{(i)} \otimes \boldsymbol{t}_q^{(j)}. \tag{18}$$

The initial rank Q of the amplitudes is determined by the rank of the MO integrals R' and the rank of the energy denominator S as $Q = R' \cdot S$. In consequence, the initial rank for the amplitudes is always larger than the rank of the two-electron integrals. However, due to the fact that the t_2 amplitudes are obtained from the MO integrals by weighting them with the energy denominator, the optimal rank of t_2 should be approximately the same or even smaller than the rank of the MO integrals. Therefore, it should be possible to reduce the initial rank of t_2 by applying the rank-reduction algorithm presented in Sec. II A.

Finally, the MP2 energy using t_2 and the MO-integrals in the new CP tensor format can simply be calculated as inner products of t_{ij}^{ab} and v_{ab}^{ij}:

$$E_{MP2} = \frac{1}{4} \sum_{efmn} t_{mn}^{ef} v_{ef}^{mn} \tag{19}$$

$$= \frac{1}{4} \sum_{efmn} \left(\sum_{q=1}^{Q} \boldsymbol{t}_q^{(e)} \otimes \boldsymbol{t}_q^{(f)} \otimes \boldsymbol{t}_q^{(m)} \otimes \boldsymbol{t}_q^{(n)} \right) \cdot \left(\sum_{r=1}^{R} \boldsymbol{v}_r^{(m)} \otimes \boldsymbol{v}_r^{(n)} \otimes \boldsymbol{v}_r^{(e)} \otimes \boldsymbol{v}_r^{(f)} \right)$$

$$= \frac{1}{4} \sum_{q=1}^{Q} \sum_{r=1}^{R} \left(\sum_{e=1}^{virt} \boldsymbol{t}_q^{(e)} \cdot \boldsymbol{v}_r^{(e)} \right) \otimes \left(\sum_{f=1}^{virt} \boldsymbol{t}_q^{(f)} \cdot \boldsymbol{v}_r^{(f)} \right) \otimes \left(\sum_{m=1}^{occ} \boldsymbol{t}_q^{(m)} \cdot \boldsymbol{v}_r^{(m)} \right) \otimes \left(\sum_{n=1}^{occ} \boldsymbol{t}_q^{(n)} \cdot \boldsymbol{v}_r^{(n)} \right)$$

$$E_{MP2} = \frac{1}{4} \sum_{q=1}^{Q} \sum_{r=1}^{R} \left\langle \boldsymbol{t}_q^{(e)}, \boldsymbol{v}_r^{(e)} \right\rangle \otimes \left\langle \boldsymbol{t}_q^{(f)}, \boldsymbol{v}_r^{(f)} \right\rangle \otimes \left\langle \boldsymbol{t}_q^{(m)}, \boldsymbol{v}_r^{(m)} \right\rangle \otimes \left\langle \boldsymbol{t}_q^{(n)}, \boldsymbol{v}_r^{(n)} \right\rangle. \tag{20}$$

Thus, only scalar products of the different representing vectors of t_{ij}^{ab} and v_{ab}^{ij} have to be evaluated. The calculation of the MP2 energy expression in this format scales as $\mathcal{O}(Q \cdot R' \cdot (virt + virt + occ + occ)) \approx \mathcal{O}(Q \cdot R' \cdot N)$. Overall, the complexity of the outlined MP2 algorithm in the CP format scales linear with system size and the rank of the tensors. As a consequence, the scaling of the rank with system and basis set size and the scaling of the rank reduction algorithm itself determine the actual scaling of the algorithm.

D.   A note on CC methods using tensor decomposition techniques

The computationally most demanding step in CC calculations are tensor contractions of amplitudes with different integrals that occur in the CC amplitude equations. If all integrals and amplitudes are expressed as low-rank representations in decomposed format, any tensor contraction can be formulated in a way similar to the evaluation of the MP2 energy discussed above (see Sect. II C). There, all contractions of internal indices are evaluated by simple scalar products, while the remaining external indices are obtained by simple copy operations of the representing vectors of integrals and amplitudes. This way, all contributions to the residual are immediately obtained in the CP format. The scaling of each tensor contraction in the CP format is $\mathcal{O}(N \cdot Q \cdot R)$, were Q denotes the rank of the amplitudes and R denotes the rank of the Hamilton-matrix element. However, the rank of the resulting object is now determined by the rank of the t_2 amplitudes times the rank of the corresponding integral and has then to be reduced by application of the rank reduction algorithm. The overall scaling will result as $\mathcal{O}(N)$ if Q and R are independent of system size, $\mathcal{O}(N^3)$ if Q and R scale linear, $\mathcal{O}(N^5)$ if Q and R scale quadratic with system size and so on. The implications for CC methods and benchmark calculations from a first pilot implementation, which is currently being developed in our group, will be presented in a forthcoming paper.

III.   COMPUTATIONAL DETAILS

The current implementation of the rank reduction procedure uses input data obtained with a development version of the CFOUR program package[95]. The two-electron integrals[96] and the results from the Hartree Fock calculation together with the ε-value for the upper bound of the error in the ℓ^2-norm are taken as input for the new algorithm.[97]

The rank reduction algorithm presented in Sec. II A scales with initial rank times final rank and the number of iterations, so that for large initial ranks the rank reduction is a very time consuming step. Due to the trivial decomposition that is currently used to cast the two-electron integrals into the CP tensor format, the initial rank always scales as $\mathcal{O}(N^3)$, where N is the number of basis functions. However, the canonical format allows for an efficient parallel procedure by treating large quantities in a different way: The large initial tensor in CP format is split up into slices including a fixed number of ranks. The rank

reduction is then applied to the individual slices, so that reduced slices are obtained, which can then again be merged (see Fig. 3). In order to obtain a low rank approximation this procedure can be repeated until the rank does not change any more or a full rank reduction can be carried out for the full merged tensor representation. This procedure can trivially be distributed to multiple processes and also lends itself to a distributed integral direct algorithm that is currently being developed in our group. This would also eliminate the need for storing the two-electron integrals as done in the current pilot implementation and as it is often customary in CC codes.

A typical full rank reduction of AO-two-electron integrals with an initial rank of 5000 to $\varepsilon = 10^{-4}$ takes 8.6 hours of CPU time on one core of a workstation with an AMD Opteron 2218 Stepping 2 processor with 2.6 GHz. The parallelized sliced reduction with slices of 1000 and the same threshold takes $5 \cdot 0.4$ CPU hours on 5 CPUs. A subsequent full reduction to $\varepsilon = 10^{-4}$ of the prereduced slices takes 3.7 hours on a single CPU, so the sliced reduction followed by a full reduction leads to good rank reduction and affordable CPU time as will further be discussed in section IV.[98]

IV. RESULTS

A. Decomposition of the two-electron integrals

In this section the decomposition of AO integrals and the rank reduction to obtain a low rank approximation for a small test-set of molecules (see Tab. I) are discussed. Here, different schemes for the rank reduction were tested: A single sliced rank reduction (sr) as described in section III, a single sliced reduction followed by a full rank reduction (sfr) and a full rank reduction (fr) of the AO integral tensor. From Table I it can be seen that the single sliced rank reduction does not lead to a good reduction for the AO integrals. The compression of ranks[99] in CH_4 for example is between 53 % for the highest accuracy ($\varepsilon = 10^{-6}$) and 92 % for the lowest accuracy ($\varepsilon = 10^{-2}$). If the remaining representing vectors are combined to one slice and reduced (sfr) the compression can be improved further. The final rank for $\varepsilon = 10^{-6}$ can be reduced by 82 % while the rank for $\varepsilon = 10^{-2}$ can be decreased by 96 %. Furthermore, even from these examples it can be seen that the compression for larger examples is better.

13

If a full reduction (fr) is carried out for the AO integrals the results are very similar to those obtained by the sliced reduction followed by a full reduction (sfr). Here, the final ranks are almost identical and deviate at most about 0.5 %. It should be noted that the deviation decreases for smaller values of ε. Therefore, the best balance between computational time and reduction of rank is obtained with a sliced reduction followed by a full reduction (sfr). This scheme is used for all larger examples (initial rank > 5000) in the following for which a full rank reduction is not feasible due to the limitations of the current implementation of the rank reduction algorithm[100]. In order to assess the scaling of the final rank after reduction of the two-electron integrals, a LiH chain in a 6-31G basis[101] and a H_2O molecule using different basis sets have been studied. The results obtained from single sliced rank reductions (sr) are presented in Figure 4 and show a scaling $\mathcal{O}(n^{2.5})$ with system size and $\mathcal{O}(N^{2.8})$ with basis set size which is almost independent of the ε value used. This means that the scaling of the initial rank ($\mathcal{O}(n^{2.6})$ and $\mathcal{O}(N^3)$ respectively)[102] cannot be reduced drastically by a single sliced rank reduction (sr). However, the prefactor can be decreased by more than one order of magnitude especially for larger examples and larger values of ε. For a value of $\varepsilon = 10^{-2}$ also a subsequent full reduction of the prereduced AO integrals (sfr) from the LiH chain has been performed (see Fig. 4). There it can be seen, that a more rigorous reduction reduces the scaling further to $\mathcal{O}(n^{1.8})$, while the prefactor does not change very much with respect to the sliced reduction.

The dependence of the final rank on the initial slice sizes for sliced decomposition has been tested for the H_2O example with different basis sets. Here, slices with 500, 1000 and 2000 ranks have been reduced and the scaling with basis set size can be compared (see Table II). Generally, larger slice sizes lead to better reduction of ranks and also the scaling with basis set size can be reduced by choosing larger slice sizes. It should be noted that a subsequent full rank reduction would yield practically the rank of the fr-procedure, no matter what slice size has been chosen in the initial sliced prereduction (c.f. Tab. I). In conclusion, the optimal performance of the rank reduction algorithm is achieved with slice sizes between 1000 and 2000.

In order to assess the error that arises due to the fact that in the decomposed format no permutational symmetry is used, the values of integrals that should be identical by permutational symmetry have been compared after reassembly. This has been done for the AO integrals as well as for the MO integrals for several test cases and different values of ε.

14

In all cases the difference of two equivalent integrals were smaller than 10^{-7} which is at least two order of magnitude smaller than the error in the ℓ^2-norm introduced by the low-rank approximation and can thus safely be neglected.

B.   AO-MO transformation with decomposed integrals

The AO-MO transformation can be carried out as described in Sec. II B. While the ranks do not change due to the transformation it should be possible to reduce the rank of the MO integrals further. Thus, different types of MO integrals are calculated in the CP format (c.f. Eqn. 9 and 10) and the rank reduction algorithm is applied to the resulting MO integrals of the LiH chain example. The same accuracy parameters that have been used for the AO integrals are applied. The scaling with system size is shown in Figure 5.   Here it can be seen, that approximate low rank representations for different kind of MO integrals exhibit almost linear scaling of reduced ranks with system size, especially for v_{ij}^{ab}. Other types of integrals show a scaling between $\mathcal{O}(n^{2.6})$ for v_{cd}^{ab} with $\varepsilon = 10^{-6}$ and $\mathcal{O}(n^{1.4})$ for v_{ij}^{ab} with $\varepsilon = 10^{-2}$. To check for consistency, also a series of alkyl-chains have been calculated that show practically the same behaviour (see Fig. 6).   From this it can be concluded that the amount of storage for the MO integrals can be decreased from $\mathcal{O}(N^4)$ scaling to optimally $\mathcal{O}(N^{2.4})$ (see also Sec. II B). Therefore, especially the bottleneck in storing the v_{cd}^{ab} can be moderated by using the new CP tensor format in a low rank approximation. It should be noted that the overall scaling of the number of parameters for representing the two-electron integrals with approximately $\mathcal{O}(N^2)$ to $\mathcal{O}(N^3)$ is similar to RI/DF and CD. For these, the representing matrices are of size $N^2 \cdot R$. For CD, for example, it has been found that the rank scales approximately linear with systems size if a stable error is maintained[27]. For the CP format on the other hand, the size of the representing vectors are with $N \cdot R$ of lower dimensionality, while the rank itself exhibits a steeper scaling with system size.

C.   MP2 algorithm with decomposed integrals

If the MO integrals are available in the decomposed format, the MP2 amplitudes t_{ij}^{ab} can also be obtained in the CP format (c.f. Sec. II C). For the approximation of the energy denominator currently a very high accuracy is used - a rank of 42 is chosen, for which an

error in the ℓ^2-norm of the energy denominator of $\varepsilon = 7 \cdot 10^{-12}$ is obtained[52]. As already mentioned in Sec. II C the initial rank of the t_2 amplitudes is always larger than the rank of the corresponding v_{ij}^{ab} integral tensor due to the construction of this object, so it must be possible to reduce the initial rank of t_{ij}^{ab}. In Figure 7, the scaling of the final ranks of MP2 amplitudes with system and basis set size for LiH chains and the H_2O molecule are given.

For the t_2 amplitudes the ranks can be reduced to similar or even better values than the ranks of the MO integrals. For the lowest accuracy ($\varepsilon = 10^{-2}$) the scaling with system and basis set size is almost linear ($\mathcal{O}(n^{1.4})$ and $\mathcal{O}(N^{1.2})$, respectively). This means that the number of parameters for the t_2 tensor can be decreased from a scaling with $\mathcal{O}(N^4)$ to $\mathcal{O}(N^{2.2})$. Especially the low scaling with basis set size should be pointed out, as other approximations, for example local correlation methods, that exhibit low scaling with system size often do not show advantageous scaling with basis set size.

For any application of the tensor decomposition techniques presented here the error in the total energy and its dependence on the value of the ε-threshold is important, also because this parameter has an impact on the overall scaling of the method. Thus, the MP2 energy is calculated with all tensors in the CP format as explained in Sec. II C. Table III shows the absolute errors in the MP2 energy for the test molecules calculated using the 6-31G basis set. For the calculation of the MP2 energy two schemes have been applied: First, quantities obtained from decomposed and reduced AO integrals are used. In the second case, the t_2 amplitudes and the corresponding MO integrals are obtained from a conventional calculation and then transformed into the CP format and reduced to the given accuracy. From Table III it can be seen that for the latter case (amplitudes and integrals obtained conventionally) the accuracy relative to the full MP2 result is better than for successive application of the rank reduction. The errors can be decreased by more than 3 orders of magnitude by changing the ε-parameter of the reduction from 10^{-2} to 10^{-4}. Thus, the error propagation during multiple approximations must be considered carefully and the ε-threshold of the approximation has to be adjusted.

Figure 8 shows the errors in the MP2 energy for a series of alkyl chains calculated using the 6-31G basis set. As the quantities used here are obtained from a conventional calculation and only one rank reduction has to be applied to the amplitudes and integrals also larger values of ε are tested in the approximation. For values of ε between 10^{-1} and 10^{-2} the error in the MP2 energy is in the range of 1 mHartree. If μHartree accuracy is required or

16

multiple rank reductions are applied the value of ε should be decreased to 10^{-3} or 10^{-4}.

V. DISCUSSION AND OUTLOOK

In this paper we have presented the application of a novel tensor decomposition technique to electronic structure methods. Using the canonical tensor format together with a low rank approximation / rank reduction procedure provides the opportunity to reduce the storage requirements and computational effort of post-HF ab initio methods drastically.

For the AO integrals, the scaling for the rank with system and basis set size is found to be $\mathcal{O}(N^{1.8})$ if a sliced reduction followed by a full reduction with $\varepsilon = 10^{-2}$ is applied. In contrast the full AO integral tensor, for which the rank scales as $\mathcal{O}(N^3)$, storage requirements are decreased from N^4 to $4 \cdot R \cdot N = 4 \cdot N^{2.8}$.

As the rank of the decomposed tensors is independent of the basis chosen, it does not change during the AO MO transformation. However, it is possible to compress the rank of the MO integrals further so that the scaling of the rank R' with system size can be reduced to $\mathcal{O}(N^{1.4})$ for v_{ij}^{ab} with $\varepsilon = 10^{-2}$ or slightly higher depending on the type of MO integral and applied ε-parameter. As a consequence, the memory required for storage of the MO integrals is reduced to $4 \cdot R' \cdot N = 4 \cdot N^{2.4}$. Furthermore, the complexity of the transformation operation itself is decreased from the formal $\mathcal{O}(N^5)$ scaling of the canonical transformation to approximately $\mathcal{O}(virt \cdot N \cdot R) \approx \mathcal{O}(N^{3.8})$, if R scales as $\mathcal{O}(N^{1.8})$ as shown for $\varepsilon = 10^{-2}$.

In order to assess the implications for CC theory, an estimate for the t_2 amplitudes as obtained from MP2 has been investigated. In this case, the amplitudes can be obtained by conversion to the CP format if a Laplace-like decomposition is applied to the energy denominator. After the reduction procedure, the rank of the amplitudes is found to be reduced to almost linear scaling with system and basis set size ($\mathcal{O}(n^{1.4})$ or $\mathcal{O}(N^{1.2})$ for $\varepsilon = 10^{-2}$). The storage requirements for the amplitudes are thus reduced from a formal scaling with N^4 to $4 \cdot Q \cdot N \approx 4 \cdot N^{2.2}$ for $\varepsilon = 10^{-2}$.

Once amplitudes and integrals are available in the CP format, the MP2 energy can be calculated by simple scalar products of the corresponding representing vectors. Using amplitudes and integrals that have been reduced with ε values between 10^{-1} to 10^{-2} still yields mHartree accuracy in the MP2 energy. With slightly smaller values of ε for the rank reduction even μHartree accuracy can be achieved. The complexity of the evaluation of the

energy expression in the CP format scales as $\mathcal{O}(R' \cdot Q \cdot N) = \mathcal{O}(N^{3.6})$ if a scaling of R' as $\mathcal{O}(N^{1.4})$ and Q as $\mathcal{O}(N^{1.2})$ as described above is assumed for a value of $\varepsilon = 10^{-2}$. However, as the conventional algorithm based on canonical orbitals scales with $\mathcal{O}(N^4)$, the complexity for the MP2 energy equation can not be reduced much by the introduction of decomposed tensors in the CP format.

The largest benefit can be expected for the evaluation of more complex tensor contractions like in the CC amplitude equations. Contractions between amplitudes and MO integrals can be evaluated by simple scalar products over inner contraction variables and copy operations for the remaining representing vectors. Therefore, these contractions would scale no longer with $\mathcal{O}(N^6)$ but rather with $\mathcal{O}(N^{3.6})$, assuming rank reductions with $\varepsilon = 10^{-2}$ that yield amplitude and integral representations for which the ranks scale as $\mathcal{O}(N^{1.2})$ and $\mathcal{O}(N^{1.4})$, respectively. While this should hold for arbitrary order CC theory the scaling of ranks for higher order amplitudes still has to be investigated.

In conclusion, tensor decomposition techniques and their rigorous application bears the potential to overcome the "curse of dimensionality" for post-HF ab initio methods. If all quantities are expressed in decomposed representing vectors the high scaling of tensor contractions and storage requirements are drastically reduced trough tensor decomposition and low rank approximation. Ultimately, the rank reduction algorithm will become the time determining step. This step scales approximately as $\mathcal{O}(\text{initial rank} \cdot \text{final rank})$. Due to the trivial decomposition technique currently used, the initial rank scales as $\mathcal{O}(N^3)$ while the final rank scales roughly as $\mathcal{O}(N^2)$ depending on the quantity of interest and the ε parameter chosen. In consequence, the overall scaling of the rank reduction would be $\mathcal{O}(N^5)$, if the scaling with the number of iterations of the current algorithm is eliminated, and can be further decreased if the initial tensor is sliced into smaller parts. Treating the representing vectors by slices of ranks also allows for an efficient parallelization of the reduction and contraction procedures. The refinement of the rank reduction algorithm and the implementation of a full CC code are the subject of current work in our group and will be discussed in a forthcoming paper.

REFERENCES

[1] F. Hitchcock, J. Math. Phys. **6**, 164 (1927).

[2] L. R. Tucker, Psychometrika **31**, 279 (1966).

[3] J. D. Carroll and J. J. Chang, Psychometrika **35**, 283 (1970).

[4] R. A. Harshman and M. E. Lundy, Computational Statistics and Data Analysis **18**, 39 (1994).

[5] C. J. Appellof and E. R. Davidson, Anal. Chem. **53**, 2053 (1981).

[6] R. Henrion, J. Chemom. **7**, 477 (1993).

[7] A. K. Smilde, Y. D. Wang, and B. R. Kowalski, J. Chemom. **8**, 21 (1994).

[8] R. Bro, Crit. Rev. Anal. Chem. **36**, 279 (2006).

[9] L. De Lathauwer, B. De Moor, and J. Vandewalle, SIAM J. Matrix Anal. Appl. **21**, 1253 (2000).

[10] A. N. Langville and W. J. Stewart, Numer. Linear Algebra Appl. **11**, 723 (2004).

[11] L. De Lathauwer and J. Vandewalle, Linear Algebra Appl. **391**, 31 (2004).

[12] L. De Lathauwer and J. Castaing, Signal Processing **87**, 322 (2007).

[13] M. A. O. Vasilescu, 16th International Conference On Pattern Recognition, Proceedings, 456(2002).

[14] M. A. O. Vasilescu and D. Terzopoulos, 16th International Conference On Pattern Recognition, Proceedings, 511(2002).

[15] D. Vlasic, M. Brand, H. Pfister, and J. Popovic, ACM Transactions On Graphics **24**, 426 (2005).

[16] E. Acar, S. A. Camtepe, M. S. Krishnamoorthy, and B. Yener, Intelligence and Security Informatics, Proceedings **3495**, 256 (2005).

[17] J. M. Sun, S. Papadimitriou, and P. S. Yu, ICDM 2006: Sixth International Conference

on Data Mining, Proceedings, 1076(2006).

[18] L. Omberg, G. H. Golub, and O. Alter, P. Natl. Acad. Sci. USA **104**, 18371 (2007).

[19] E. Martínez-Montes, P. A. Valdés-Sosa, F. Miwakeichi, R. I. Goldman, and M. S. Cohen, Neuroimage **22**, 1023 (2004).

[20] F. Miwakeichi, E. Martínez-Montes, P. A. Valdés-Sosa, N. Nishiyama, H. Mizuhara, and Y. Yamaguchia, Neuroimage **22**, 1035 (2004).

[21] R. Bellman, *Adaptive control processes: A guided tour.* (Princeton University Press, 1961) p. 255.

[22] S. Wilson, Comput. Phys. Commun. **58**, 71 (1990).

[23] I. Røeggen and E. Wisløff-Nilssen, Chem. Phys. Lett. **132**, 154 (1986).

[24] R. Lindh, U. Ryu, and B. Liu, J. Chem. Phys. **95**, 5889 (1991).

[25] H. Koch, A. S. de Merás, and T. B. Pedersen, J. Chem. Phys. **118**, 9481 (2003).

[26] F. Aquilante, T. B. Pedersen, and R. Lindh, J. Chem. Phys. **126**, 194106 (2007).

[27] F. Aquilante and T. B. Pedersen, Chem. Phys. Lett. **449**, 354 (2007).

[28] L. Boman, H. Koch, and A. S. de Merás, J. Chem. Phys. **129**, 134107 (2008).

[29] I. Røeggen and T. Johansen, J. Chem. Phys. **128**, 194107 (2008).

[30] S. Schweizer, J. Kussmann, B. Doser, and C. Ochsenfeld, J. Comput. Chem. **29**, 1004 (2008).

[31] F. Aquilante, L. Gagliardi, T. B. Pedersen, and R. Lindh, J. Chem. Phys. **130**, 154107 (2009).

[32] F. Weigend, M. Kattannek, and R. Ahlrichs, J. Chem. Phys. **130**, 164106 (2009).

[33] J. Zienau, L. Clin, B. Doser, and C. Ochsenfeld, J. Chem. Phys. **130**, 204112 (2009).

[34] J. L. Whitten, J. Chem. Phys. **58**, 4496 (1973).

[35] F. R. Manby, J. Chem. Phys. **119**, 4607 (2003).

[36] H. J. Werner, F. R. Manby, and P. J. Knowles, J. Chem. Phys. **118**, 8149 (2003).

[37] M. Schütz and F. R. Manby, Phys. Chem. Chem. Phys. **5**, 3349 (2003).

[38] R. Kendall and H. Früchtl, Theor. Chem. Acc. **97**, 158 (1997).

[39] O. Vahtras, J. Almlöf, and M. Feyereisen, Chem. Phys. Lett. **213**, 514 (1993).

[40] M. Feyereisen, G. Fitzgerald, and A. Komornicki, Chem. Phys. Lett. **208**, 359 (1993).

[41] R. Ahlrichs, Phys. Chem. Chem. Phys. **6**, 5119 (2004).

[42] W. Kutzelnigg, Theor. Chim. Acta. **68**, 445 (1985).

[43] E. F. Valeev and H. F. Schaefer, J. Chem. Phys. **113**, 3990 (2000).

[44] S. Ten-No, Chem. Phys. Lett. **398**, 56 (2004).

[45] W. Klopper, F. R. Manby, S. Ten-No, and E. F. Valeev, Int. Rev. Phys. Chem. **25**, 427 (2006).

[46] T. B. Adler, H. J. Werner, and F. R. Manby, J. Chem. Phys. **130**, 054106 (2009).

[47] T. B. Adler and H. J. Werner, J. Chem. Phys. **130**, 241101 (2009).

[48] M. Häser and J. Almlöf, J. Chem. Phys. **96**, 489 (1992).

[49] M. Häser, Theor. Chim. Acta. **87**, 147 (1993).

[50] A. K. Wilson and J. Almlöf, Theor. Chim. Acta. **95**, 49 (1997).

[51] P. Y. Ayala and G. E. Scuseria, J. Chem. Phys. **110**, 3660 (1999).

[52] D. Braess and W. Hackbusch, IMA J. Numer. Anal. **25**, 685 (2005).

[53] D. S. Lambrecht, B. Doser, and C. Ochsenfeld, J. Chem. Phys. **123**, 184102 (2005).

[54] A. Takatsuka, S. Ten-No, and W. Hackbusch, J. Chem. Phys. **129**, 044112 (2008).

[55] G. E. Scuseria and P. Y. Ayala, J. Chem. Phys. **111**, 8330 (1999).

[56] P. Y. Ayala, K. N. Kudin, and G. E. Scuseria, J. Chem. Phys. **115**, 9698 (2001).

[57] B. Doser, D. S. Lambrecht, J. Kussmann, and C. Ochsenfeld, J. Chem. Phys. **130**, 064107 (2009).

[58] W. Meyer, Int. J. Quantum Chem. **5**, 341 (1971).

[59] W. Meyer, J. Chem. Phys. **58**, 1017 (1973).

[60] R. Ahlrichs and F. Driessler, Theor. Chim. Acta. **36**, 275 (1975).

[61] R. Ahlrichs, F. Driessler, H. Lischka, V. Staemmler, and W. Kutzelnigg, J. Chem. Phys. **62**, 1235 (1975).

[62] L. Adamowicz and R. J. Bartlett, J. Chem. Phys. **86**, 6314 (1987).

[63] L. Adamowicz, R. J. Bartlett, and A. J. Sadlej, J. Chem. Phys. **88**, 5749 (1988).

[64] P. Neogrády, M. Pitoňák, and M. Urban, Mol. Phys. **103**, 2141 (2005).

[65] M. Pitoňák, P. Neogrády, V. Kellö, and M. Urban, Mol. Phys. **104**, 2277 (2006).

[66] M. Pitoňák, F. Holka, P. Neogrády, and M. Urban, J. Mol. Struct. Theochem **768**, 79 (2006).

[67] F. Neese, F. Wennmohs, and A. Hansen, J. Chem. Phys. **130**, 114108 (2009).

[68] F. Neese, A. Hansen, and D. G. Liakos, J. Chem. Phys. **131**, 064103 (2009).

[69] T. Kinoshita, O. Hino, and R. J. Bartlett, J. Chem. Phys. **119**, 7756 (2003).

[70] O. Hino, T. Kinoshita, and R. J. Bartlett, J. Chem. Phys. **121**, 1206 (2004).

[71] T. G. Kolda, SIAM J. Matrix Anal. Appl. **23**, 243 (2001).

[72]T. Zhang and G. H. Golub, SIAM J. Matrix Anal. Appl. **23**, 534 (2001).

[73]M. Ishteva, L. De Lathauwer, P. A. Absil, and S. Van Huffel, Numerical Analysis and Applied Mathematics **1048**, 274 (2008).

[74]M. Ishteva, L. De Lathauwer, P. A. Absil, and S. Van Huffel, Numer. Algor. **51**, 179 (2009).

[75]T. G. Kolda and B. W. Bader, SIAM Review **51**, 455 (2009).

[76]L. De Lathauwer, B. De Moor, and J. Vandewalle, SIAM J. Matrix Anal. Appl. **21**, 1324 (2000).

[77]G. Beylkin and M. J. Mohlenkamp, P. Natl. Acad. Sci. USA **99**, 10246 (2002).

[78]G. Beylkin and M. J. Mohlenkamp, SIAM J. Sci. Comput. **26**, 2133 (2005).

[79]S. R. Chinnamsetty, M. Espig, B. N. Khoromskij, W. Hackbusch, and H. J. Flad, J. Chem. Phys. **127**, 084110 (2007).

[80]M. Espig, *Effiziente Bestapproximation mittels Summen von Elementartensoren in hohen Dimensionen*, Ph.D. thesis, Universität Leipzig (2008).

[81]E. Acar and B. Yener, IEEE T. Knowl. Data. En. **21**, 6 (2009).

[82]R. Bro, Chemom. Intell. Lab. Syst. **38**, 149 (1997), ISSN 0169-7439.

[83]R. Bro and H. A. L. Kiers, J. Chemom. **17**, 274 (2003).

[84]M. Espig and W. Hackbusch, in preparation(2010).

[85]This expansion is exact as long as the rank R is large enough. As has been shown, any tensor in d dimensions can be represented exactly with a rank of $R = n^{d-1}$, where n denotes the number of entries in each dimension.

[86]J. Gauss, J. F. Stanton, and R. J. Bartlett, J. Chem. Phys. **95**, 2623 (1991).

[87]J. F. Stanton and J. Gauss, Int. Rev. Phys. Chem. **19**, 61 (2000).

[88]W. Hackbusch and B. Khoromskij, J. Complex. **23**, 697 (2007).

[89]W. Hackbusch and B. N. Khoromskij, SIAM J. Matrix Anal. Appl. **30**, 1233 (2008).

[90]M. Espig, L. Grasedyck, and W. Hackbusch, Constr. Approx. **30**, 557 (2009).

[91]M. Espig, W. Hackbusch, T. Rohwedder, and R. Schneider, Numer. Math., paper submitted(2009).

[92]L. Armijo, Pac. J. Math. **16**, 1 (1966).

[93]P. Kosmol, *Methoden zur numerischen Behandlung nichtlinearer Gleichungen und Optimierungsaufgaben* (Teubner Verlag; Auflage: 2., 1993).

[94]R. J. Bartlett, Ann. Rev. Phys. Chem. **32**, 359 (1981).

[95]CFOUR, Coupled Cluster techniques for Computational Chemistry, a quantum-chemical program package by J.F. Stanton, J. Gauss, M.E. Harding, P.G. Szalay with contributions from A.A. Auer, R.J. Bartlett, U. Benedikt, C. Berger, D.E. Bernholdt, O. Christiansen, M. Heckert, O. Heun, C. Huber, D. Jonsson, J. Jusélius, K. Klein, W.J. Lauderdale, D. Matthews, T. Metzroth, D.P. O'Neill, D.R. Price, E. Prochnow, K. Ruud, F. Schiffmann, S. Stopkowicz, A. Tajti, M.E. Varner, J. Vázquez, F. Wang, J.D. Watts and the integral packages MOLECULE (J. Almlöf and P.R. Taylor), PROPS (P.R. Taylor), ABACUS (T. Helgaker, H.J. Aa. Jensen, P. Jørgensen, and J. Olsen), and ECP routines by A. V. Mitin and C. van Wüllen. For the current version, see http://www.cfour.de.

[96]The antisymmetrized two electron integrals were used as described in Ref. 103. All integrals smaller than a threshold of 10^{-14} were neglected.

[97]All calculations were carried out on the "Chemnitz High performance Linux Cluster" (CHiC)[104].

[98]It should be noted that this is the current performance using a pilot implementation - an optimized scheme for two-electron integrals and wavefunction parameters is the subject of current work and will be presented in forthcoming publications.

[99]The compression has been calculated as: $\text{compression} = 100 \cdot \left(1 - \left(\frac{\text{reduced rank}}{\text{initial rank}} \right) \right)$.

[100]With the current pilot implementation that is a general-purpose, BLAS-level 2 based C++ development platform, only initial ranks up to 10000 are computationally feasible. For an example of this size, the rank reduction algorithm takes in the order of a few weeks using a single core on a modern workstation computer (AMD Opteron 2218 Stepping 2, 2.6 GHz, 4 GB Memory).

[101]The geometry for LiH monomer was taken from Ref. 105: $R_{LiH} = 159.5$ pm. The LiH chain was build up as a linear chain LiH–LiH–... using a distance of 300 pm between the H-atom of one molecule and the Li-atom of the next.

[102]It should be noted that the initial scaling with system size is actually not $\mathcal{O}(n^3)$ due to the fact that integrals smaller than 10^{-14} are neglected.

[103]J. F. Stanton, J. Gauss, J. D. Watts, and R. J. Bartlett, J. Chem. Phys. **94**, 4334 (1991).

[104]F. Mietke, T. Mehlan, T. Hoefler, and W. Rehm, Proceedings of the 3rd Workshop KiCC (Kommunikation in Clustern und Clusterverbundsystemen)(2007).

[105]K. P. Huber and G. H. Herzberg, *Molecular Spectra and Moleculare Structure IV: Con-*

stants of Diatomic Molecules (Van Nostrand-Reinhold, New York, 1979).

[106] A. R. Hoy, I. M. Mills, and G. Strey, Mol. Phys. **24**, 1265 (1972).

[107] J. L. Duncan and I. M. Mills, Spectrochimica Acta **20**, 523 (1964).

[108] D. L. Gray and A. G. Robiette, Mol. Phys. **37**, 1901 (1979).

FIGURES

Trivial decomposition leading to a representation of rank $= 3$

$$\begin{pmatrix} 1 & 4 & 1 \\ 6 & 4 & 2 \\ 11 & 4 & 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 4 \\ 1 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 6 \\ 4 \\ 2 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 11 \\ 4 \\ 3 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 4 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 6 & 4 & 2 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 11 & 4 & 3 \end{pmatrix}$$

reduced rank representation

$$\begin{pmatrix} 1 & 4 & 1 \\ 6 & 4 & 2 \\ 11 & 4 & 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \otimes \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \otimes \begin{pmatrix} -2 \\ 2 \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} 3 & 2 & 1 \\ 6 & 4 & 2 \\ 9 & 6 & 3 \end{pmatrix} + \begin{pmatrix} -2 & 2 & 0 \\ 0 & 0 & 0 \\ 2 & -2 & 0 \end{pmatrix}$$

FIG. 1. Example for a trivial decomposition and a reduced rank representation in two dimensions.

1: Choose initial $\tilde{\mathbf{A}}^0 \in V^{\tilde{R} \times d}$ and parameter $\varepsilon \in \mathbb{R}_{>0}$. Define iteration count $k := 0$, compute the gradient $\mathbf{G}^0 := \nabla J(\tilde{\mathbf{A}}^0)$ and $\mathbf{D}^0 := -\mathbf{G}^0$.

2: **while** the gradient $\|\mathbf{G}^k\| > \varepsilon$ **do**

3:    Compute the smallest root $\alpha_k \in [0,1]$ of the polynomial $p(\alpha) := \left\langle \nabla J(\tilde{\mathbf{A}}^k + \alpha \mathbf{d}^k), \mathbf{D}^k \right\rangle$, i.e.
$$\alpha_k := \min\left\{ \alpha \in \mathbb{R}_{\geq 0} : p(\alpha) := \left\langle \nabla J(\tilde{\mathbf{A}}^k + \alpha \mathbf{D}^k), \mathbf{D}^k \right\rangle = 0 \right\}.$$

4:    Update the representation system of $\tilde{\mathbf{A}}$, i.e. $\tilde{\mathbf{A}}^{k+1} := \tilde{\mathbf{A}}^k + \alpha_k \mathbf{D}^k$.

5:    Compute the gradient for the updated system, i.e. $\mathbf{G}^{k+1} := \nabla J(\tilde{\mathbf{A}}^{k+1})$.

6:    Compute $\beta_k := \frac{\left\langle \mathbf{G}^{k+1} - \mathbf{G}^k, \mathbf{G}^{k+1} \right\rangle}{\|\mathbf{G}^k\|^2}$, $\gamma_k := \max\{0, \beta_k\}$.

7:    Update the new search direction, i.e. $\mathbf{D}^{k+1} := -\mathbf{G}^{k+1} + \gamma_k \mathbf{D}^k$.

8:    $k \mapsto k+1$.

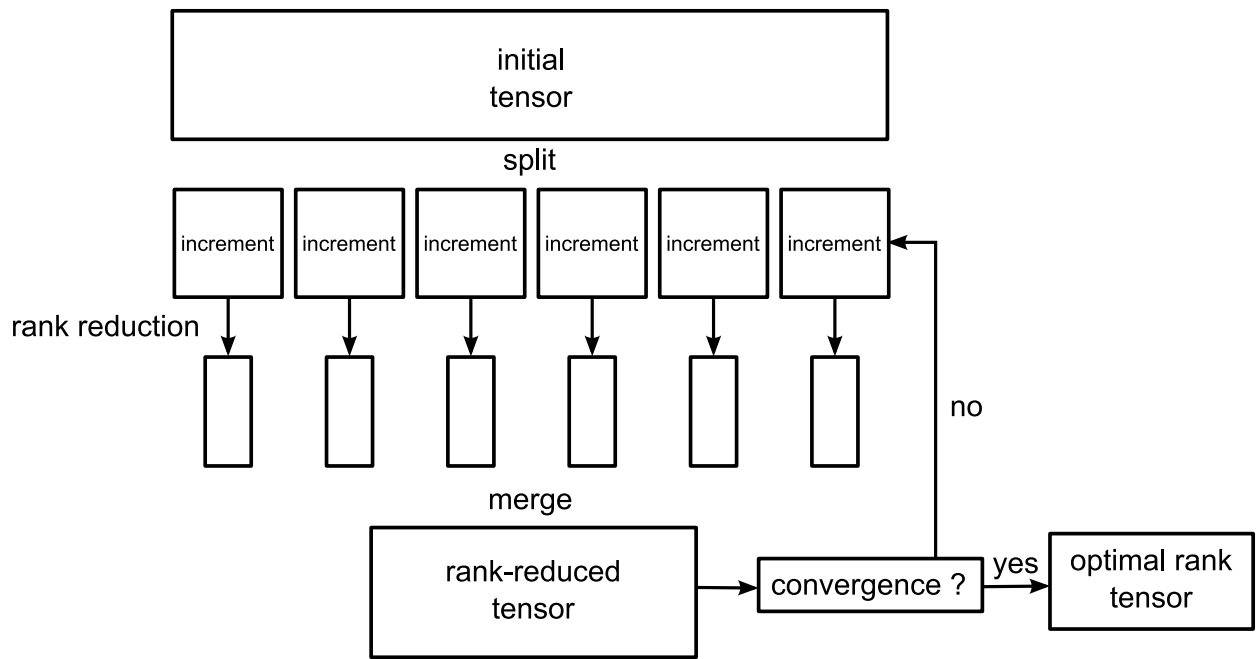FIG. 2. Schematic representation for the Accelerated Gradient (AG) Method.

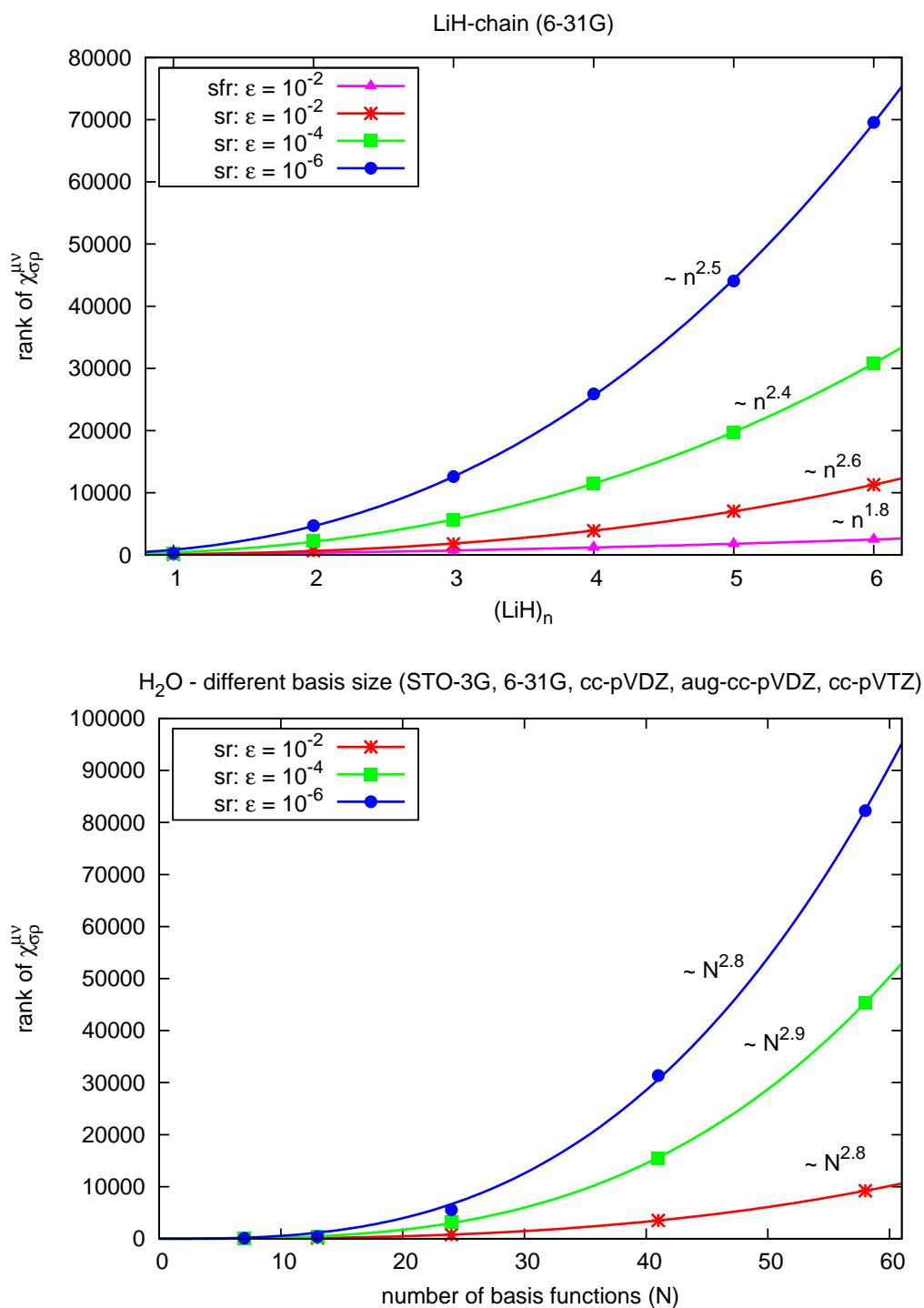FIG. 3. Schematic representation of the parallel rank reduction currently used for large examples.

FIG. 4. Scaling of reduced ranks for AO-integrals with system size for LiH-chains using the 6-31G basis set and with number of basis functions for H_2O using the basis set series: STO-3G, 6-31G, cc-pVDZ, aug-cc-pVDZ, cc-pVTZ. Both plots show results from a single sliced (sr) reduction (c.f. Fig. 3). For LiH slices with rank 1000 and for H_2O slices with rank 2000 are chosen. For the LiH chains also the sfr reduction for $\varepsilon = 10^{-2}$ is shown.
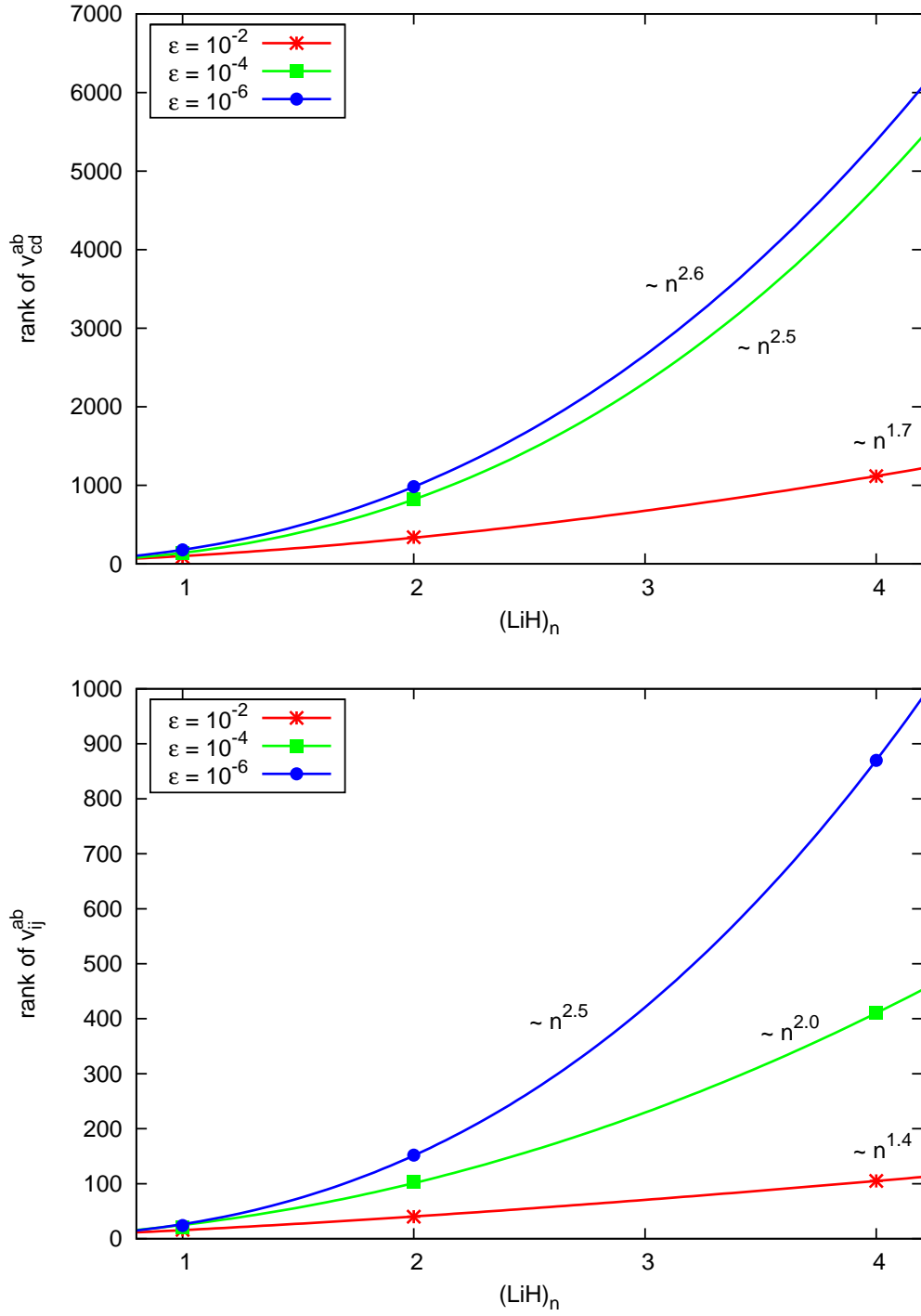
FIG. 5. Scaling of ranks with system size in LiH chains calculated using the 6-31G basis set. For the small examples ($n = 1 - 2$) a single full reduction (fr) is used, while the larger chains are reduced in slices (rank = 2000) followed by a subsequent full reduction (sfr).

FIG. 6. Scaling of ranks with system size in alkyl chains (C_nH_{2n+2}) calculated using the 6-31G basis set. For the small examples ($n = 1 - 2$) a single full reduction (fr) is used, while the larger chains are reduced in slices (rank = 2000) followed by a subsequent full reduction (sfr).
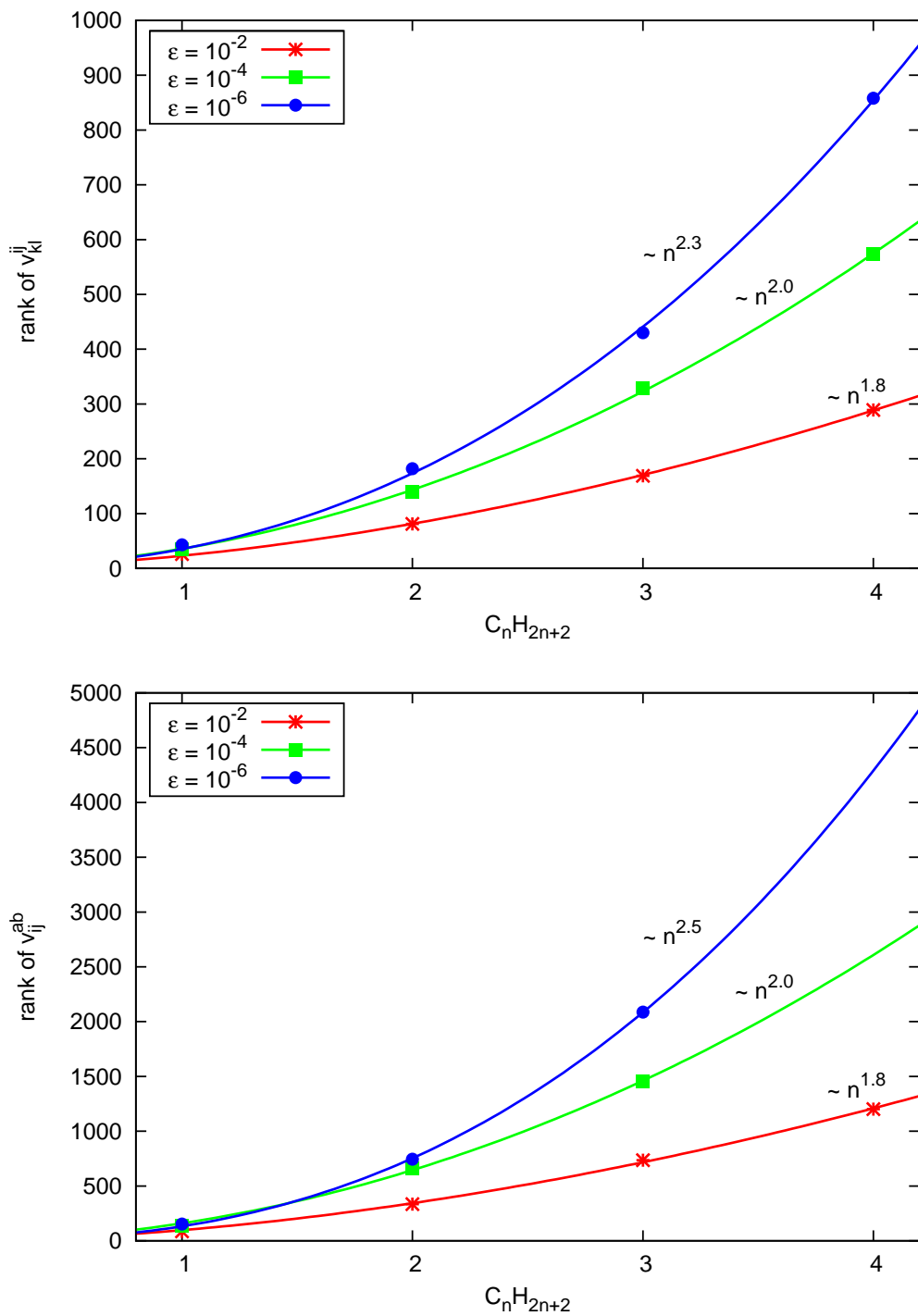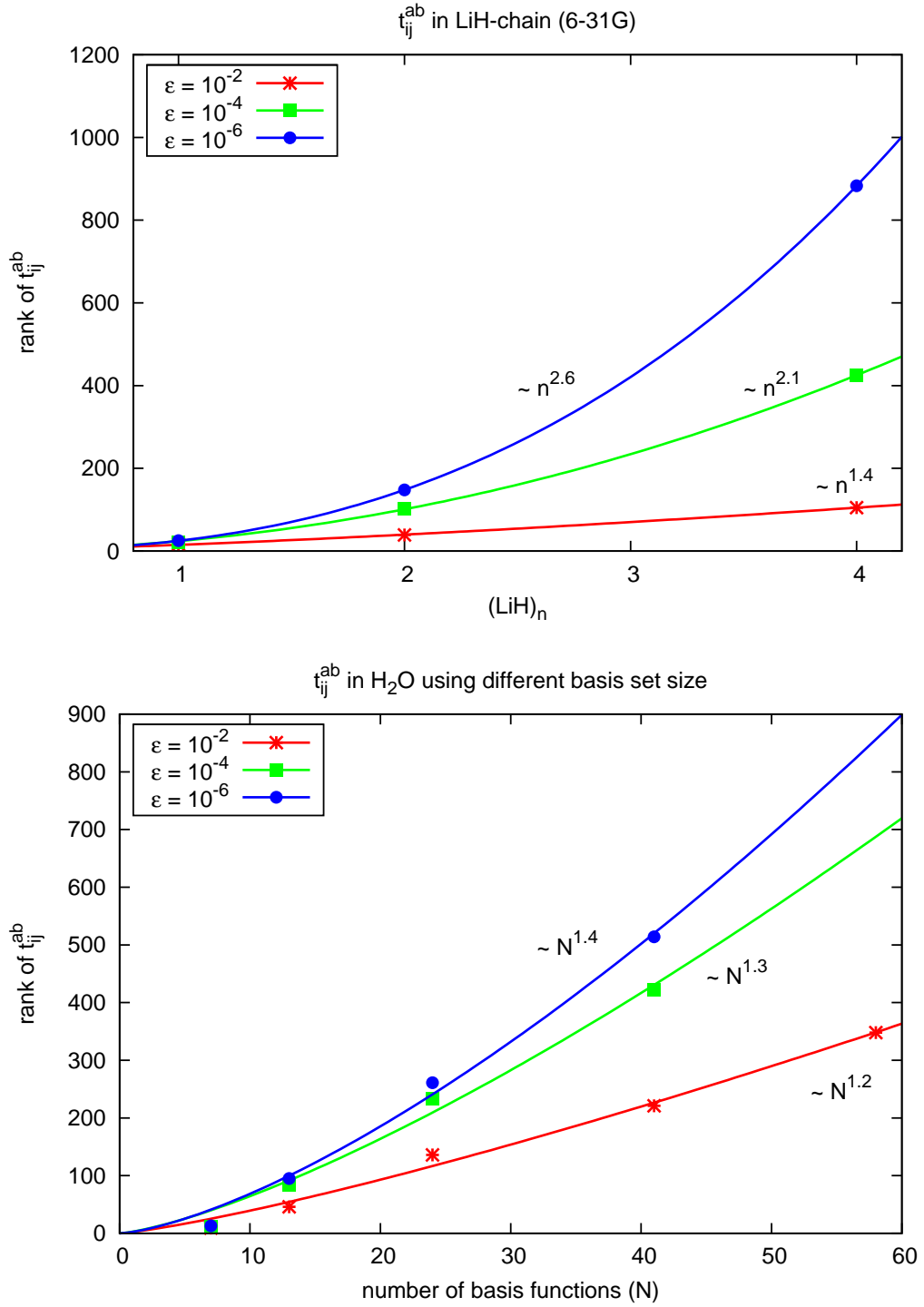
31

FIG. 7. Scaling of ranks for the t_2-amplitudes with system size for $(LiH)_n$ and scaling with basis set size for H_2O. For the small examples (initial rank $<$ 5000) a full reduction (fr) is used, while the larger examples are reduced in slices (rank = 2000) followed by a subsequent full reduction (sfr).
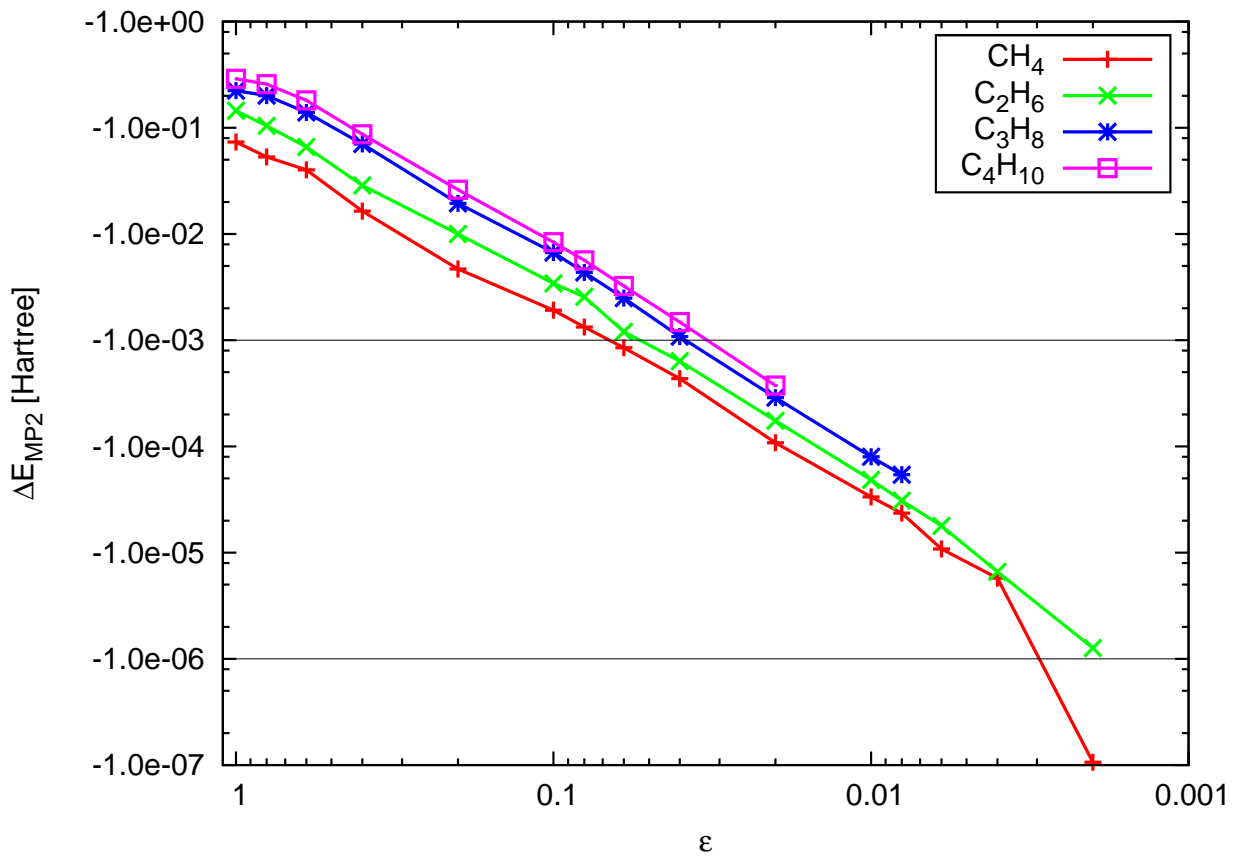
FIG. 8. Error in MP2 energies for different values of ε in a series of alkyl-chains. The amplitudes and MO integrals are obtained from a conventional calculation. The tensors are transformed into the CP format and reduced by a full rank-reduction.

TABLES

TABLE I. Initial and final ranks for different accuracies for the AO-integrals in H_2O, NH_3 and CH_4 using the 6-31G basis set. [a] N denotes the number of elements per dimension that is equivalent to the number of basis functions.

molecule	N	initial rank	sliced reduction (sr) (size=1000) ε			sliced + full (sfr) reduction (size=1000) ε			full reduction (fr) ε		
			10^{-2}	10^{-4}	10^{-6}	10^{-2}	10^{-4}	10^{-6}	10^{-2}	10^{-4}	10^{-6}
H_2O	13	2005	205	598	852	148	337	403	142	337	403
NH_3	15	3375	307	966	1409	183	488	603	172	476	599
CH_4	17	4793	398	1397	2233	210	648	844	199	632	840

[a] The following geometries were used for the molecules (distance R [pm], angle θ [degree]): H_2O: $R_{OH} = 95.72$, $\theta_{HOH} = 104.52$[106]; NH_3: $R_{NH} = 101.1$, $\theta_{HNH} = 106.7$[107]; CH_4: $R_{CH} = 108.58$, $\theta_{HCH} = 109.47$[108]

TABLE II. Reduced ranks for sliced reductions (sr) with different slice sizes for AO integrals in H_2O using different basis sets.

basis set	N	initial rank	slice = 500 ε			slice = 1000 ε			slice = 2000 ε		
			10^{-2}	10^{-4}	10^{-6}	10^{-2}	10^{-4}	10^{-6}	10^{-2}	10^{-4}	10^{-6}
6-31G	13	1283	253	698	1000	205	598	852	149	346	474
cc-pVDZ	24	8553	1610	5207	7066	1107	4180	6914	782	3257	5555
aug-cc-pVDV	41	49054	9029	30436	41108	5380	21466	34449	3501	15493	31350
cc-pVTZ	58	150771	27293	95480	127789	15562	68872	109808	9219	45369	87386

TABLE III. Errors in MP2 energies in mHartree for different values of ε using the 6-31G basis. The results in the first column are obtained with quantities build from a decomposed and reduced AO integral tensor. The results in the second column are calculated with quantities obtained from a conventional calculation which are then transformed to the new format and reduced.

ε	from decomposed AO integrals			from canonical quantities		
	H_2O	NH_3	CH_4	H_2O	NH_3	CH_4
10^{-2}	1.503023	5.308963	5.978639	-0.039275	-0.032095	-0.032298
10^{-4}	0.002606	0.004576	0.002512	0.000017	0.000234	0.000126
10^{-6}	0.000026	-0.000208	0.000016	-0.000001	0.000001	0.000002