PAPER

Check for updates

Efficient vibrationally correlated calculations using *n*-mode expansion-based kinetic energy operators†

Frederik Bader, [iD] *[a] David Lauvergnat [iD] [b] and Ove Christiansen [iD] [c]

Due to its efficiency and flexibility, the *n*-mode expansion is a frequently used tool for representing molecular potential energy surfaces in quantum chemical simulations. In this work, we investigate the performance of *n*-mode expansion-based models of kinetic energy operators in general polyspherical coordinate systems. In particular, we assess the operators with respect to accuracy in vibrationally correlated calculations and their effect on potential energy surface construction with the adaptive density guided approach. Our results show that the *n*-mode expansion-based operator variants are reliable and systematically improvable approximations of the full kinetic energy operator. Moreover, we introduce a workflow to generate the *n*-mode expanded kinetic energy operators on-the-fly within the adaptive density guided approach. This scheme can be applied in studies of species and coordinate systems, for which an analytical form of the kinetic energy operator is not available.

1 Introduction

Solving the nuclear Schrödinger equation for studying stationary vibrational states or time-dependent dynamics invariably involves dealing with the curse of dimensionality for both the nuclear wave functions and the potential energy surfaces (PES). Both, wave function and PES, are formally a function of all coordinates of the system. This is a significant challenge for representing and computing them, since the number of coordinates increases with system size. Focusing on the PES, numerous approaches have been taken, which are often based on systematic expansions and/or fitting procedures. One purely expansion-based approach is to represent the PES as a Taylor expansion around an expansion point.[1,2] However, it is well known that this expansion displays a limited radius of convergence, and limited accuracy as well as a restricted range of applicability in practice.[1–5] An alternative approach is *n*-mode expansions,[6–10] also known as mode-coupling expansion,[11] cut high dimensional model representation (HDMR),[12] or cluster expansion.[13] In the *n*-mode approximation the full-dimensional

PES is written as a sum of many small sub-unit PESs up to a given mode-coupling level of *n*. Such *n*-mode representations have in many cases proven to be accurate, fast converging and compatible with a variety of computational methods.[7,9,14–18] One factor that influences the convergence of the expansions is the choice of coordinate system. For molecules with a well-defined semi-rigid reference structure, rectilinear normal coordinates are a commonly used and highly effective coordinate system. They are strongly decoupled for configurations close to the reference and the corresponding deformational kinetic energy operator (KEO), the well-known Watson KEO,[19] is generally applicable due to its easily transferable functional form. However, for large amplitude motion, such as highly excited vibrations or torsions,[20–22] other – curvilinear – coordinate systems can represent the nuclear motion more efficiently. The application of the latter type of coordinates can facilitate the treatment of the PES and the vibrational wave functions, but typically comes at the cost of more complex KEO forms, which are not straightforwardly transferable to other systems.[23,24] Consequently, *n*-mode expansions of PESs and property surfaces, *e.g.* dipole moments, have been used in combination with recti- and curvilinear coordinate systems.[9,10,25–28] This holds true for the KEO as well. When working with the Watson KEO, *n*-mode expansions have been used to treat the elements of the inverse effective moment-of-inertia tensor (commonly denoted $\mu_{\alpha\beta}$).[11,29,30] Regarding curvilinear coordinates, Strobusch and Scheurer demonstrated vibrational self-consistent field (VSCF) and vibrational configuration interaction (VCI) calculations for H_2O_2 in primitive internal coordinates with *n*-mode expansion-based

[a] Department of Chemistry, Aarhus University, DK-8000 Aarhus C, Denmark. E-mail: f.bader@chem.au.dk

[b] Université Paris-Saclay, CNRS, Institut de Chimie Physique UMR8000, Orsay 91405, France. E-mail: david.lauvergnat@universite-paris-saclay.fr

[c] Department of Chemistry, Aarhus University, DK-8000 Aarhus C, Denmark. E-mail: ove@chem.au.dk

† Electronic supplementary information (ESI) available: Tables with all vibrational state energies analyzed in Step 1 and Step 3 of our study. See DOI: **https://doi.org/10.1039/d4cp00423j**

This journal is © the Owner Societies 2024

Phys. Chem. Chem. Phys., 2024, **26**, 11469–11481 | **11469**

KEOs.[31,32] To this end, they started from the most general form of deformational KEOs in curvilinear coordinates (eqn (1) below) and examined various n-mode representations of the G matrix and the extrapotential V_{ep}. On this basis, Strobusch and Scheurer ensuingly developed an adaptive scheme to increase the efficiency of their n-mode KEO models.[33,34] Aside from this, Yurchenko and cooperators have used general KEO models based on generalized power series expansions of G_{ij} and V_{ep},[35,36] and many previous works have employed general coordinate systems combined with different low-order KEO approximations that have been useful in practice.[37–39] Typically, the latter lead to a KEO with a restricted mode-coupling, but without being defined or characterized like this.

The current work can be understood as part of a series of studies on vibrationally correlated calculations in general curvilinear coordinate systems. In a first study,[24] Klinting, Lauvergnat and Christiansen investigated the performance of full kinetic energy operators parametrized by polyspherical coordinates in full vibrational configuration interaction (FVCI) and vibrational coupled cluster calculations (VCC). This was facilitated by a new interface between the software suites Tnum/Tana and MidasCpp.[40,41] The former is able to construct different representations of kinetic energy operators in general coordinates for a variety of coordinate systems,[38,42–45] while the latter produces potential energy surfaces as well as solutions to the vibrational Schrödinger equation through different approaches. The polyspherical coordinates were a natural choice as a starting point, since an analytical and exact kinetic energy operator can be automatically obtained by Tana and the kinetic energy operators derived from these coordinates always have a sum-of-products (SOP) form. A SOP form is a requirement for using operators in MidasCpp, which is not met by all curvilinear coordinate systems. The investigation with the analytical KEOs showed that the application of curvilinear coordinate systems can lead to more efficient representations of the PES and the vibrational wave function. This was deduced from a reduced number of electronic structure calculations in the adaptive density guided PES construction as well as faster convergence behaviour in the vibrationally correlated calculations. Recently, we continued this work and presented corresponding calculations with Taylor expansion-based variants of the KEOs in polyspherical coordinates.[46] Our objective was to introduce KEO approximations that allow for systematic restrictions of the intermode coupling in the KEO and, at the same time, have a sum-of-products form irrespective of the selected general coordinate system. This is important in view of the fact that, while the exact analytical KEO may not always couple all modes in all terms, it formally features quite high mode-coupling levels. The latter could become a hindrance in many cases, since the cost of, e.g., VSCF, VCI and VCC methods grows with the mode coupling in the Hamiltonian. As a key finding, we reported that several of the proposed Taylor expansion-based schemes reproduce reference FVCI and VCC calculations with the analytical polyspherical KEOs to an adequate accuracy. Thus, the new approximations paved the way for using the computational schemes implemented in MidasCpp with curvilinear coordinate systems other than

the polyspherical ones. In the current study, we further extend the interface between Tnum/Tana and MidasCpp to investigate n-mode expansion-based KEOs in polyspherical coordinates. Our work follows the spirit of Strobusch and Scheurer in generating hierarchies of kinetic energy operators based on the n-mode expansion.[31,32] There are, however, a few methodological differences between their approach and ours. Instead of grid-based representations of G_{ij} and V_{ep}, we construct computationally convenient continuous forms of these quantities. Moreover, in contrast to Strobusch and Scheurer, we do not explicitly generate derivatives of the G matrix numerically through interpolation. Finally, our kinetic energy operators are straightforwardly applicable not just for VSCF and VCI calculations, but also for VCC parametrizations of the vibrational wave functions. Thus, they provide additional attractive options for high accuracy computations, also for larger systems. Just like the n-mode PESs, the n-mode KEO models can, through grid construction and fitting procedures, be brought into a sum-of-products form and can thus be used in MidasCpp for any general coordinate system. As n-mode expansions are more flexible than Taylor expansions, we hope to find the KEOs based on the former to be more efficient approximations of the full KEOs. To evaluate this, we perform vibrationally correlated calculations with the n-mode KEO variants and compare them to analogous calculations with the full KEO as well as selected Taylor expansion-based KEO models. In addition to that, we probe in which way the choice of the KEO influences the potential energy surfaces produced by the adaptive density guided approach in MidasCpp. Finally, we study measures to reduce the computational effort associated with applying the n-mode KEO variants.

2 Theory

2.1 Kinetic energy operators in general coordinates

The starting point of our study is the deformational kinetic energy operator $\hat{T}_{def}$ in general coordinates with the volume element $d\tau_{def}^\rho$,

$$\hat{T}_{def} = -\frac{1}{2} \sum_{i,j=1}^{3N-6} \frac{1}{\rho(\mathbf{Q})} \frac{\partial}{\partial Q_i} \rho(\mathbf{Q}) G_{ij}(\mathbf{Q}) \frac{\partial}{\partial Q_j} + V_{ep}(\mathbf{Q}), \quad (1)$$

$$d\tau_{def}^\rho = \rho(Q)dQ_1 dQ_2 \ldots dQ_M, \quad (2)$$

with the $M = 3N - 6$ internal coordinates $\mathbf{Q} = (Q_1, \ldots Q_M)$, the contravariant components of the metric tensor associated with the deformational molecular motion G_{ij} and the extrapotential V_{ep}. The latter arises from the selected normalization function $\rho(\mathbf{Q})$, resulting in a non-Euclidean volume element. In this study, we always choose $\rho = 1$. More detailed derivations of this operator form are presented in other works.[35,42,46–48] To employ this KEO in efficient computations of wave functions and energies we require a convenient representation of its action. As it stands, the KEO is exact and we do in fact have access to computing function values of G_{ij} and V_{ep}. However, the latter, like the PES, are in principle full-dimensional, i.e. they formally couple all degrees of freedom, even if the direct coupling of many modes by the KEO is often quite small. This makes these forms expensive to use in

11470 | *Phys. Chem. Chem. Phys.*, 2024, **26**, 11469–11481

This journal is © the Owner Societies 2024

many contexts for larger systems. To be able to obtain representations that can bridge both small and large systems as well as intermediate and high accuracy, we choose to describe these quantities in terms of n-mode expansions. Using the operator form of eqn (1) in combination with n-mode expansions of G_{ij} ensures that the resulting KEO is Hermitian.

2.2 n-Mode expansions

The n-mode expansion is a well-established scheme to express the full functional dependence of a given quantity, $F(\mathbf{Q})$, in terms of a sum of contributions, each of which depends only on a subset of the variables. For instance, the expansion of the property $F(\mathbf{Q})$ depending on M modes around a reference point $\mathbf{Q}_{\mathrm{ref}}$,

$$F(\mathbf{Q}) = \bar{F}^{(0)} + \bar{F}^{(1)}(\mathbf{Q}) + \bar{F}^{(2)}(\mathbf{Q}) + \cdots + \bar{F}^{(M)}(\mathbf{Q}), \quad (3)$$

comprises the property value at the reference point, $\bar{F}^{(0)} = F^0$, a contribution of terms depending solely on a single mode, the one-mode (1M) terms summarized in $\bar{F}^{(1)}(\mathbf{Q})$, a contribution by terms coupling exactly two modes, $\bar{F}^{(2)}(\mathbf{Q})$, and analogous higher-order contributions coupling triplets, quartets and more modes. If all coupling terms up to Mth order are included in the expansion, the latter is exact. For practical purposes, however, the expansion is often truncated, facilitated by a fast convergence of the expansion with respect to increasing mode-coupling levels. The individual terms summarized in the different contributions $\bar{F}^{(i)}(\mathbf{Q})$ can be generally expressed in terms of the "bar"-properties as

$$\bar{F}^{\mathbf{m}_j}\left(\mathbf{Q}_{\mathbf{m}_j}\right) = S^{\mathbf{m}_j} \sum_{n'=0}^{j} (-1)^{j-n'} \binom{j}{n'} F^{\mathbf{m}_{n'}}\left(\mathbf{Q}_{\mathbf{m}_{n'}}\right). \quad (4)$$

In eqn (4), $\mathbf{m}_k$ is a vector containing k mode labels and $S^{\mathbf{m}_k}$ is an operator that symmetrizes the ensuing expression with respect to the mode labels in $\mathbf{m}_k$. The term $F^{\mathbf{m}_k}(\mathbf{Q}_{\mathbf{m}_k})$ is a k-dimensional cut of the function $F(\mathbf{Q})$, where only the modes contained in $\mathbf{m}_k$ are allowed to vary while the other Q are frozen to the reference point. For example, a 2-dimensional cut F^{m_1,m_2} is given as

$$F^{m_1,m_2}(Q_{m_1}, Q_{m_2}) = F(0, \ldots, 0, Q_{m_1}, 0, \ldots, 0, Q_{m_2}, 0, \ldots, 0), \quad (5)$$

where the reference point is set to $\mathbf{Q}_{\mathrm{ref}} = (0, 0, \ldots, 0, 0)$. In order to illustrate the form of the bar-properties, we provide an explicit example for a 2-mode term:

$$\bar{F}^{\mathbf{m}_2}(\mathbf{Q}_{\mathbf{m}_2}) = S^{\mathbf{m}_2} \sum_{n'=0}^{2} (-1)^{2-n'} \binom{2}{n'} F^{\mathbf{m}_{n'}}\left(\mathbf{Q}_{\mathbf{m}_{n'}}\right) \quad (6)$$

$$= S^{\mathbf{m}_2}\big(F^{\mathbf{m}_0} - 2{\cdot}F^{\mathbf{m}_1}(\mathbf{Q}_{\mathbf{m}_1}) + F^{\mathbf{m}_2}(\mathbf{Q}_{\mathbf{m}_2})\big) \quad (7)$$

$$= S^{m_1,m_2}\big(F^0 - 2{\cdot}F^{m_1}(Q_{m_1}) + F^{m_1,m_2}(Q_{m_1}, Q_{m_2})\big) \quad (8)$$

$$\begin{aligned} = \frac{1}{2} \cdot \big(&F^0 - 2 \cdot F^{m_1}(Q_{m_1}) + F^{m_1,m_2}(Q_{m_1}, Q_{m_2}) \\ &+ F^0 - 2 \cdot F^{m_2}(Q_{m_2}) + F^{m_2,m_1}(Q_{m_2}, Q_{m_1})\big) \end{aligned} \quad (9)$$

$$= F^{m_1,m_2}(Q_{m_1}, Q_{m_2}) - F^{m_1}(Q_{m_1}) - F^{m_2}(Q_{m_2}) + F^0 \quad (10)$$

In the last step of the above set of equations we have exploited that the property function is symmetric with respect to permutations of modes in the mode index vector. Eqn (10) shows that the "bar" property for a mode combination $\mathbf{m}_j$ is actually cleaned of all terms of lower order, avoiding any over-counting of contributions.

In general, any property can be written as a mode-coupling expansion

$$F(\mathbf{Q}) = \sum_{\mathbf{m}\in\mathrm{MCR}} \bar{F}^{\mathbf{m}}(\mathbf{Q}_{\mathbf{m}}) \quad (11)$$

where MCR is the mode combination range, *i.e.* the set of mode combinations (MC) included in the expansion. If all MCs are included, the exact function is obtained. We can truncate the expansion to the standard n-mode case by regarding all MCs up to a cardinal number n, or choose MCRs tailored to various physical cases. While in this paper we focus on the standard n-mode case for the G_{ij} matrix elements and V_{ep}, the principles apply generally. Having formally represented any property in terms of low-dimensional bar-properties, the question now becomes how to construct these functions. There are numerous approaches to this. For example, one could set up a static grid and either represent the function as values on that grid, or in a continuous way by subsequent fitting procedures. The term "static-grid" will in this work indicate that the n-mode expansions are built from fitting a pre-defined number of function values for each MC on a regular grid in pre-selected coordinate domains to analytical functions. How to choose an adequate grid is a non-trivial problem, and an automatic setup may save significant time and computational resources.

2.3 Adaptive density guided approach

In order to automatically construct n-mode expansions, the adaptive density guided approach (ADGA) can be employed.[49–53] This scheme has been designed to generate n-mode expansion-based forms of potential energy surfaces without prior knowledge of the relevant parts of the examined system's configuration space.

To this end, an ADGA run consists of iterations of systematically expanding a grid of electronic energies and producing intermediate continuous representations of the PES by fitting predefined general functions to these energies. In accordance with the n-mode expansions, the ADGA proceeds in a hierarchical way, producing, subsequently, surfaces from 1M- up to kM-coupling. The exploration of the chemical space in the course of the ADGA is driven by the vibrational wave function density obtained from intermediate VSCF calculations. In particular, the convergence of the ADGA is probed in every iteration for every unconverged mode combination by the ADGA integral,

$$T_{l^{\mathbf{m}_n}}^{\mathbf{m}_n} = \int_{I_{l^{m_1}}^{m_1}} \int_{I_{l^{m_2}}^{m_2}} \cdots \int_{I_{l^{m_n}}^{m_n}} \rho_{\mathrm{ave}}^{\mathbf{m}_n}(\mathbf{Q}) \bar{V}^{\mathbf{m}_n}(\mathbf{Q}) \mathrm{d}Q_{m_1} \mathrm{d}Q_{m_2} \cdots \mathrm{d}Q_{m_n}, \quad (12)$$

where $\mathbf{m}_n$ is a given mode combination of n modes, and $I_{l^{m_k}}^{m_k}$ represents the integration bounds for the interval l^{m_k} on the domain of Q_{m_k}. The intervals are defined in terms of

consecutive electronic energies on the grid, and the ADGA integrals are evaluated for every interval. Moreover, in eqn (12), $\bar{V}^{\mathbf{m}_n}(\mathbf{Q})$ is the bar potential of the given mode combination and $\rho_{\text{ave}}^{\mathbf{m}_n}(\mathbf{Q})$ is the corresponding vibrational density. The latter is determined from the modals optimized in the VSCF calculation with the current iteration of the PES. By default, one-mode densities for a given mode m are calculated as an average over a selected number of modals $\phi_{j_m}^m(Q_m)$,

$$\rho_{\text{ave}}^m(Q_m) = \frac{1}{N_{\text{modal}}} \sum_{j_m=1}^{N_{\text{modal}}} \left| \phi_{j_m}^m(Q_m) \right|^2, \qquad (13)$$

and then used to construct a full density for the mode combination as

$$\rho_{\text{ave}}^{\mathbf{m}_n}(\mathbf{Q}) = \prod_{m \in \mathbf{m}_n} \rho_{\text{ave}}^m(Q_m). \qquad (14)$$

As a result of this default, VSCF calculations are typically only run until the one-mode surfaces are converged. There are three parameters that govern the convergence of the ADGA: ε_{rel}, ε_{abs} and ε_ρ, which control, respectively, the relative and absolute errors of the ADGA integrals, as well as the fraction of the vibrational density outside of the ADGA's integration grid.

The ADGA is also able to generate n-mode expansion-based molecular property surfaces analogous to the PES. These are constructed alongside the latter, meaning that the molecular properties of interest are sampled at the same grid points as the electronic energies and are used to obtain continuous representations in the same way as for the PES. However, the molecular property surfaces do not affect the convergence of the ADGA. As a consequence, the property surfaces produced by ADGA will only ever be valid in the same domain as the PES.

3 Computational methods

In this work, we study the performance of kinetic energy operator approximations based on different n-mode expansions of G_{ij} and V_{ep} in time-independent vibrationally correlated calculations. To this end, we expand the previously described interface between the software suites MidasCpp and Tnum/Tana. Our approach consists of the following three steps:

1. First, we examine KEOs generated from static-grid n-mode expansions of the G matrix elements and V_{ep} in vibrationally correlated calculations with preexisting PESs and known integration domains. (**Step 1**)
2. Second, we investigate the on-the-fly construction of n-mode expanded KEO variants in the course of an ADGA run to generate a potential energy surface. (**Step 2**)
3. Finally, we test the validity of two schemes to reduce the computational cost associated with the n-mode expansion-based KEO models, namely decreasing the size of the fitting basis in the construction of the expansions as well as screening of the expansions' coefficients. (**Step 3**)

3.1 Molecular property surfaces for the KEO

We use MidasCpp's facilities for constructing general property surfaces to obtain n-mode expansions of the G_{ij} matrix elements and the extrapotential V_{ep}. The G matrix has M^2 entries, but is symmetric. Thus, only the $M(M + 1)/2$ expansions of, e.g., the upper triangle of the G matrix as well as the single expansion of V_{ep} are required to set up a n-mode representation of the KEO in eqn (1). Importantly, exploiting this symmetry in the construction of the KEO eliminates another source of Hermiticity issues in the KEO. We started our tests, however, from n-mode KEOs based on expansions of all M^2 G_{ij} matrix elements. For the calculations of **Step 1** and **Step 3**, this does not cause any problems, speaking for a robust and accurate fitting procedure in MidasCpp. In **Step 2** we encountered small Hermiticity issues in the VSCF part of the ADGA, such that ultimately we switched to n-mode KEOs exploiting the G matrix symmetry for all calculations of this step of the study. This emphasizes the importance of using explicitly Hermitian approaches to avoid numerical subtleties.

Moreover, in **Step 2**, we use the ADGA to construct n-mode expansions of the G_{ij} matrix elements and the extrapotential V_{ep} alongside a PES. From the former, we generate KEO variants to use together with the PES in subsequent vibrationally correlated calculations. Since a KEO for the VSCF calculations has to be provided at the ADGA's input stage, this appears circular. However, in situations where a SOP form of the full KEO is not available and the important part of the coordinate domain is uncertain, the ADGA is a convenient way to construct a reliable KEO representation from minimal prior knowledge. Our approach to this end is the following: starting from a KEO variant not requiring insight into coordinate domains, such as one obtained from a Taylor expanded G_{ij}, we construct one-mode coupled (1M) representations of G_{ij} with the ADGA. Then, we start a new ADGA run that targets surfaces with the next highest maximum mode coupling, i.e. 2M coupling, and uses the preceding step's n-mode expansions of G_{ij} to construct a starting KEO. This sequence could be expanded until the maximum mode coupling level of the PES is reached, if required. Thus, by concatenating ADGA runs with a systematically increasing maximum mode coupling level, we aim to produce increasingly accurate n-mode expansion-based KEO variants to use in ADGA runs and vibrationally correlated calculations.

3.2 Test systems and coordinate systems

We re-use the test systems from our previous study on the application of Taylor expansion-based KEO approximations in vibrationally correlated calculations.[46] Accordingly, the test set consists of the five triatomic and two tetra-atomic molecules listed in Table 1 as well as *trans*-formic acid. Moreover, we

Table 1 Examined molecules in this study

N_{atoms}	3	4	5
Molecules	H_2O, H_2S F_2O, HOF, SO_2	CH_2O H_2O_2	*trans*-HCOOH

11472 | *Phys. Chem. Chem. Phys.*, 2024, **26**, 11469–11481

This journal is © the Owner Societies 2024

describe the test systems with the same polyspherical coordinate systems (see Fig. 1 in our previous work).[46]

3.3 Step 1: static-grid n-mode expansions

In this step, we aim to show that n-mode expansions of G_{ij} and V_{ep} lead to adequate KEO representations. To achieve this, we run full vibrational configuration interaction and vibrational coupled cluster calculations in polyspherical coordinates with preexisting potential energy surfaces and new KEO variants based on n-mode expansions of G_{ij} and V_{ep}.

3.3.1 Potential energy surfaces. For the vibrationally correlated calculations we use only precedingly published potential energy surfaces. We describe the triatomic molecules with the 3M-coupled PESs presented in our previous study on Taylor expansion-based KEO variants.[46] Likewise, for the tetra-atomic molecules, we use the 4M-coupled PESs generated by Klinting, Christiansen and Lauvergnat.[24] To model *trans*-formic acid we apply the 5M-coupled PES of Richter and Carbonnière.[54]

3.3.2 Kinetic energy operators. We construct the KEOs for the vibrationally correlated calculations through static grid n-mode expansions of each G_{ij} matrix element and the extrapotential V_{ep}. To this end, we employ the static grid n-mode expansion functionalities of MidasCpp,[11,41] the curvilinear KEO construction capabilities of Tnum/Tana and the previously described interface between the two software suites.[24,40,46] For all molecules we investigate n-mode expansions of 1M-, 2M- and 3M-coupling level. The reference structures for the KEO expansions are those presented in the ESI of our previous work.[46] The coordinate domains for the n-mode expansions are selected to match the valid domains of the potential energy surfaces (ESI in our preceding work, Table S2).[46] The individual one-, two- and three-mode cuts in the expansions are sampled by 64, 32 and 16 points, respectively, for each coordinate. The fitting bases for the construction of the continuous representations of G_{ij} and V_{ep} comprise simple polynomials up to order 12 (1M-coupling level) and products of these polynomials with a sum of the exponents less or equal 12 (2M- and 3M-coupling levels). From our experience, these settings for the grid and fitting bases produce quite accurate results. Thus, errors are expected to be primarily due to the approximation of low order mode coupling itself, as corroborated in our results. We call the n-mode expansion-based KEOs XM[G,V] (X = 1, 2, 3), where the "X" denotes X-mode expansions of G_{ij} and V_{ep}. To obtain reference values, we perform calculations with the analytical vibrational polyspherical KEOs and the Taylor expansion-based

KEO variant G2,V2 of our previous work.[46] In this study, we refer to the latter by 2M[G,V]2T (and 2M[G]2T when V_{ep} is not included), where the "2T" denotes the 2nd order Taylor expansions. An overview of all KEOs applied in the current work is given in Table 2.

3.3.3 Vibrationally correlated calculations. For the triatomic test cases, we perform full vibrational interaction calculations targeting the vibrational ground state as well as the fundamental and first overtone excited states.[55–57] The vibrational states are described by a product basis of 15 modal functions per mode. These modals are obtained from a VSCF calculation with 100 primitive B-spline basis functions of 10th order.[58]

The vibrational states of formaldehyde, hydrogen peroxide and *trans*-formic acid are obtained from calculations with the vibrational coupled cluster method VCC[3] and corresponding response calculations.[59–62] For all three molecules, the states are represented by a direct product basis of 12 modals per mode, which are obtained from an earlier VSCF calculation with 180 primitive B-spline functions of order 12. We calculate the 121 lowest vibrational states of H_2O_2, which include all fundamentals and first overtone excitations below 4000 cm^{-1}. For formaldehyde and formic acid we target the ground state as well as all fundamental and first overtone excitations below 4000 cm^{-1}.

3.4 Step 2: on-the-fly KEO construction within the ADGA

In the second step of our investigation we study how the choice of the kinetic energy operator affects the potential energy surface construction with the adaptive density guided approach. Moreover, we probe a workflow that – starting from a Taylor expansion-based KEO – allows us to sequentially improve the KEO in the course of generating a PES by running through a sequence of ADGAs of lower mode coupling level. The ADGA potential energy surfaces are validated and compared through vibrational state energies calculated with the VCC[3] method, the respective PESs and different KEO models. We examine ADGA runs on formaldehyde and *trans*-formic acid.

3.4.1 Potential energy and property surfaces. First, we perform structure optimizations and harmonic frequency analyses for both molecules. These are based on electronic structure calculations using the RI-MP2 method and a cc-pVDZ basis set.[63–68] In the self-consistent field step we use the resolution-of-identity approximation to treat the Coulomb and exchange integrals with the auxiliary basis set for cc-pVDZ.[69] In the

Table 2 Kinetic energy operator variants used in this study

$\hat{T}$ label	Description	Applications	Max. coupl.
Full	Full vibrational polyspherical KEO	PES construction with the ADGA, vibrationally correlated calculations	n_{Atoms}: 3/4/5 MC_{max}: 3/5/6
2M[G,V]2T, 2M[G]2T	KEOs based on 2nd order Taylor expansions of G_{ij} (and V_{ep})	PES construction with the ADGA, vibrationally correlated calculations	4
XM[G,V] (X: 1, 2, 3)	KEO based on n-mode expansions of G_{ij} and V_{ep} up to order X	PES construction with the ADGA, vibrationally correlated calculations	$X + 2$
XM[G]-SQYT with (X,Y): (1,2), (2,2)	KEO of XM-expanded G_{ij} from sequential ADGA starting with YM[G]YT KEO	PES construction with the ADGA	$X + 2$

This journal is © the Owner Societies 2024

Phys. Chem. Chem. Phys., 2024, **26**, 11469–11481 | **11473**

correlation part of the calculations we use the cc-pVDZ auxiliary basis set for the resolution-of-identity scheme as well.[70,71] The convergence criteria for the electronic structure calculations are changes of the total energy and the one-electron density less than $10^{-9}E_H$ and $10^{-9}r_B^{-3}$, respectively. We consider the structure optimizations converged when the structure's energy changes by less than $10^{-9}E_H$ and the change of the Cartesian gradient's norm is below 10^{-5}. The ensuing frequency analysis is checked for imaginary frequencies to ensure convergence to a minimum structure. The electronic structure calculations, structure optimizations and frequency analyses are run with TURBOMOLE V7.5.[72]

After that, the potential energy surfaces are constructed using the adaptive density guided approach implemented in MidasCpp. The ADGA calculations are started from the previously optimized molecular structures and produce 3M-coupled PESs. The electronic structure method used by the ADGA is the same as described above. The parameters that control the convergence of the ADGA are set to $\varepsilon_{rel} = 10^{-3}$, $\varepsilon_{abs} = 10^{-6}$ and $\varepsilon_\rho = 10^{-4}$. A continuous representation of the PES is obtained by fitting products of polynomials to the electronic structure data. The sum of the exponents of the individual polynomials in every product is restricted to not exceed 12. We use the dynamic ADGA grid extension scheme.[73] For the VSCF calculations inside the ADGA we employ a primitive basis of 180 B-spline functions of order twelve. Four VSCF-optimized modals are included in the calculation of the averaged one-mode densities.

Aside from the potential energy surface, we use the ADGA to produce further property surfaces, namely 3M-expansions of the G_{ij} matrix elements and the extrapotential V_{ep}. To this end, at every molecular configuration sampled during the construction of the PES, Tnum/Tana is called to provide the corresponding G_{ij} and V_{ep}. The output of Tnum/Tana is parsed into MidasCpp, which ultimately creates the n-mode expansions in the same way as for the PES. These additional property surfaces do not affect the convergence of the ADGA procedure, but can be used in subsequent vibrationally correlated calculations. For each of the two molecules four PESs are generated, each characterized by a different KEO.

3.4.2 Kinetic energy operators. For the construction of the potential energy surfaces with ADGA we employ multiple kinetic energy operator variants, in particular the full vibrational polyspherical KEO and the Taylor expansion-based 2M[G]2T KEO. Furthermore, we run ADGA calculations with two other KEO schemes, where we systematically build up and apply n-mode expansion-based KEO variants starting from a Taylor expansion-based KEO approximation. The workflow for these latter two ADGA runs is the following: we start from the 2M[G]2T KEO and generate 1M-coupled n-mode representations of the electronic ground state potential energy surface as well as G_{ij} matrix elements and V_{ep}. Then, we discard the 1M-PES and the extrapotential term and start a new ADGA calculation with the KEO based on the 1M expansions of G_{ij}. At this point, we can run the ADGA with the target maximum mode-coupling level using the updated KEO, which in our case yields a 3M-coupled PES and 3M representations of G_{ij} and V_{ep}.

We refer to the KEO used in the VSCF calculations of this last ADGA run as 1M[G]-SQ2T, indicating a sequentially improved KEO based on a 1M-expansion of G_{ij} that is obtained starting from a second order Taylor expansion of G_{ij}. Alternatively, another ADGA run of a mode-coupling level lower than the target can be added to the workflow. This run will use the 1M[G]-SQ2T KEO to generate 2M-coupled representations of the PES, the G_{ij} matrix elements and V_{ep}. Again, we discard the 2M-PES and the 2M extrapotential and use the remaining expansions to construct the two-step 2M[G]-SQ2T KEO. The latter we use in the VSCF calculations of the final ADGA run to obtain a 3M-coupled PES and corresponding G_{ij} and V_{ep} expansions. The described procedure of starting a completely new ADGA run with each KEO update is fairly wasteful of computational resources, given that the PES construction is typically one of the most costly parts of studying vibrational properties. We point out that the electronic energies sampled in preceding ADGA runs in this sequential workflow can be reused in the current run with the latest KEO to avoid the repetition of potentially expensive electronic structure calculations. There may be some computational overhead in repeating the fitting procedures, however, we note that the properties being fitted several times are not at the final mode coupling level. Thus, the resulting overhead is very far from dominating.

For each of the four PESs, which we generate for CH_2O and HCOOH, we run three VCC[3] calculations with different KEOs. These are the full polyspherical vibrational KEO, the 2M[G,V]2T KEO and a KEO based on the n-mode expansions of G_{ij} and V_{ep} yielded by the final ADGA step of each PES construction. For CH_2O the latter KEO is based on the full 3M-expansions, while for HCOOH we remove the 3M-coupled terms from the n-mode expansions before assembling the KEO.

3.4.3 Vibrational correlation method. For the vibrationally correlated calculations used to compare the PESs constructed by ADGA runs with different KEOs we employ the same computational procedure as in **Step 1**, that is we perform VCC[3] (response) calculations of the ground state, fundamental and first overtone excitation energies below 4000 cm^{-1} with the same primitive and modal basis sets.

3.5 Step 3: reducing the computational cost

We evaluate the effect of smaller fitting bases and coefficient screening in the n-mode expanded KEO approximations on the vibrational state energies of *trans*-formic acid. The computational parameters are the same as those described in **Step 1**. In order to assess the possibility of reducing the computational cost in the calculations with n-mode expanded KEO variants by coefficient screening, we repeat the calculations on formic acid of **Step 1**. However, instead of using the full n-mode expansions of G_{ij} and V_{ep} with a polynomial fitting basis of maximum order 12, we screen the individual terms in the expansions, dropping those with coefficients larger than a given threshold. To study the fitting bases, we generate a new set of n-mode expansion-based KEO variants of 1M- and 2M-coupling level with the restriction that the maximum total polynomial order of the functions in the fit basis (FBMO) must not exceed 4, 6, 8 and 10,

11474 | *Phys. Chem. Chem. Phys.*, 2024, **26**, 11469–11481

This journal is © the Owner Societies 2024

respectively. Here, total polynomial order refers to the sum of the exponents in the polynomial products. With these new KEO models we repeat the vibrationally correlated calculations on *trans*-formic acid outlined in **Step 1**.

4 Results and discussion

4.1 Step 1: static-grid *n*-mode expansions

The objective of **Step 1** of our work is to establish the *n*-mode expansion-based KEOs as viable KEO representations for accurate calculations of vibrational properties. We demonstrate this through the analysis of the vibrational state energies from vibrationally correlated calculations.

4.1.1 Tri-atomic molecules. To start with the calculated vibrational energies of the tri-atomic test molecules, the energies, maximum (MAX) and mean (MAD) absolute deviations are summarized in Fig. 1 and listed in Table S1 in the ESI.† Judging from these, the KEO approximations based on the static-grid *n*-mode expansion of G_{ij} and V_{ep} appear to reproduce the reference calculations with the full vibrational polyspherical KEO very accurately. Using one-mode expansions (and correspondingly the 3M-coupled 1M[G,V] KEO variant) the MADs of the ground state energy as well as the fundamental and first overtone excitation energies do not exceed 0.12 cm^{-1}. The maximum absolute deviation across all test systems is 0.44 cm^{-1} for H$_2$O. For the 2M- and 3M-coupled expansions, the deviations from the reference typically become less than 0.01 cm^{-1}, *i.e.* they vanish. In Fig. 1 and Table S1 (ESI†) we also show vibrational energies calculated using the 2M[G,V]2T KEO variant introduced in our preceding work on Taylor expansion-based KEO approximations.[46] This KEO variant is based on a second order Taylor expansion of G_{ij} and V_{ep} and features up to 4M-coupling. The deviations for the 2M[G,V]2T KEO are typically larger than for the other KEOs with peak values for MAX and MAD of 5.86 cm^{-1} (HOF: $2\nu_{OH}$) and 1.78 cm^{-1} (H$_2$O), respectively. Thus, it is clear that the *n*-mode expanded KEO models perform better for the majority of these test cases even if only a 1M coupled *n*-mode expansion is used. The exceptions are F$_2$O and SO$_2$, where the performance of the 2M[G,V]2T and 1M[G,V] KEOs is comparable

and very good already. As discussed in our previous work on the Taylor expansion-based KEOs this is mostly due to the inverse mass dependence of G_{ij}.

4.1.2 Larger molecules. Next, we discuss the vibrationally correlated calculations of formaldehyde, hydrogen peroxide and *trans*-formic acid. The results of the calculations are displayed in Fig. 2 and Tables S2–S4 in the ESI,† which show the individual vibrational energies for the different KEO approximations.

In Fig. 2 the mean (MAD) and maximum (MAX) absolute deviations of the vibrational energies calculated with the KEO variants relative to the full vibrational KEO are shown for the tetra-atomic molecules and *trans*-formic acid. The MADs of the vibrational state energies are found to decrease when going from the Taylor expansion-based 2M[G,V]2T KEO variant to any of the *n*-mode expansion-based KEO models. Likewise, with increasing maximum mode coupling level in the *n*-mode expansions, the MADs decrease consistently. In particular, the largest MAD among the three test molecules amounts to 1.08 cm^{-1} (CH$_2$O) and 0.08 cm^{-1} (HCOOH) for the 1M[G,V] and 2M[G,V] KEO models, respectively. For the tetra-atomic molecules, the *n*-mode expansions up to the 3M coupling level yield KEO models that reproduce the calculations with the full KEO with MADs below 0.01 cm^{-1}. We note that for CH$_2$O and H$_2$O$_2$, small deviations from the reference calculation with the analytical KEO are still expected for the 3M[G,V] KEO variant, even though both KEOs feature operator terms with up to 5M-coupling. That is because the former KEO contains several terms with four-mode contributions in the non-derivative part, which are not caught by a 3M-expansion of G_{ij}. Furthermore, the *n*-mode expanded KEOs are obtained from fitting function values obtained on a numerical grid. Because of this, we would expect errors, though very small, even if the *n*-mode expansions covered all of the mode coupling in the full KEO. For *trans*-formic acid the VCC[3] calculations with the 3M expansions of G_{ij} with the full polynomial fitting bases of up to 12th order are
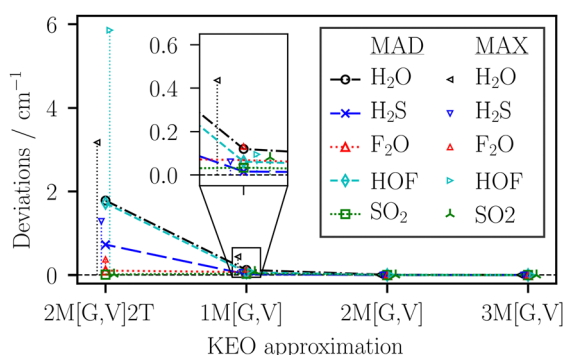

Fig. 1 Mean (MAD) and maximum (MAX) absolute deviations of the calculated vibrational state energies for the tri-atomic molecules and the different KEO variants. Deviations are calculated with respect to the FVCI calculations with the full KEO.
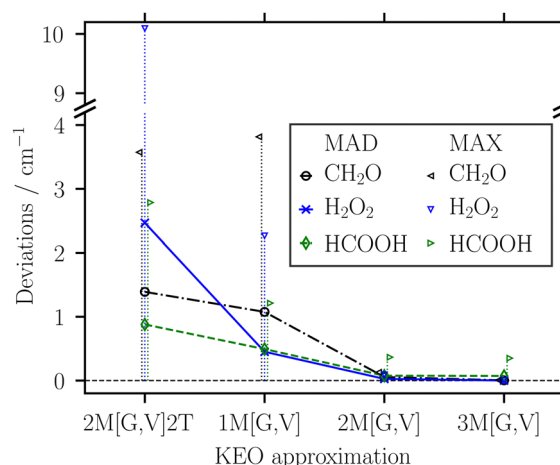

Fig. 2 Mean (MAD) and maximum (MAX) absolute deviations of the calculated vibrational state energies for the tetra-atomic molecules and *trans*-formic acid as well as the different KEO variants. Deviations are calculated with respect to the VCC[3] calculations with the full KEO.

This journal is © the Owner Societies 2024

Phys. Chem. Chem. Phys., 2024, **26**, 11469–11481 | 11475

not practical due to the large number of terms of the KEO (see Table 3 below). Despite this, we have performed such a calculation, which produces a MAD of the vibrational state energies of 0.07 cm^{-1}. Analogously to the tetra-atomic species, the 3M[G,V] KEO is not able to reproduce the reference calculations exactly, as the analytical KEO of HCOOH contains 6M-coupling terms as well as terms with non-derivative contributions with more than 3M-coupling.

Concerning the maximum absolute deviations of the vibrational state energies for the larger test molecules, Fig. 2 shows that for the most part they behave analogously to the MADs. The one exception is the 1M[G,V] KEO in CH$_2$O, which produces a MAX of 3.82 cm^{-1}, slightly larger than the corresponding 2M[G,V]2T MAX (3.57 cm^{-1}). Aside from this, the MAXs of the 1M[G,V] and 2M[G,V] KEO variants are within 1–4 cm^{-1} and below 0.4 cm^{-1}, respectively. The MAXs of the 3M[G,V] KEO models are below 0.02 cm^{-1} for the tetra-atomic molecules and 0.35 cm^{-1} for formic acid.

As a final note on the validity of our results we superficially compare our calculations on H$_2$O$_2$ to the work of Strobusch and Scheurer.[32] The latter constructed KEOs in primitive internal coordinates based on n-mode expansions of G_{ij} and V_{ep} up to the 1M- and 2M-coupling level and applied them to determine the ground state and fundamental excitation energies of hydrogen peroxide from VCI calculations. Due to differing vibrational correlation methods, PESs, numerical procedures, and reference values, these calculations are not directly comparable to ours. Nonetheless, we think it is an encouraging sign that the deviations from the respective references are of a similar magnitude. To be precise, for their KEOs denoted "T(1,1, –, –)" (1M-expanded G_{ij}) and "T(2,2,2,2)" (expansion comparable to our 2M[G,V] approach), Strobusch and Scheurer report MADs of 1.3 cm^{-1} and 0.7 cm^{-1}, as well as MAXs of 2.1 cm^{-1} and 1.9 cm^{-1}, respectively,[32] values that are slightly larger but comparable to our results.

To sum up the first step of our investigation, the analysis of the calculations with KEOs based on static-grid n-mode expansions of G_{ij} and V_{ep} clearly demonstrates that such expansions produce effective KEO representations. When stopped at the 1M coupling level, the expansions give rise to three mode-coupled KEOs that perform equally well or better than the four mode-coupled Taylor expansion-based 2M[G,V]2T scheme across all test systems. This indicates that more functional flexibility than granted by the second order Taylor expansion can be crucial for accurately capturing the 1M part of the G_{ij}, which seems more important than including a low-order description of couplings in G and V_{ep}. The KEOs based on expansions with a higher maximum mode coupling level consistently improve upon the 1M[G,V] model. The 3M[G,V] KEOs (five-mode coupled at a maximum), in particular, reproduce the reference values obtained with the full vibrational polyspherical KEO with maximum deviations of 0.01 cm^{-1} or less for the tri- and tetra-atomic molecules. However, also the results obtained with the 2M[G,V] models are close to converged, which is in agreement with the observations by Strobusch and Scheurer.[32] Our tests show that, on one hand, the n-mode approach gives

better accuracy with lower mode coupling than the Taylor expansion-based schemes and therefore should be computationally attractive. On the other hand, they also show that applying high mode coupling level variants like $\hat{T}$[3M[G,V]] with the large grids and many-term fittings used here is not practical for larger molecular systems, unless further measures are taken to reduce the computational cost, as the number of terms in the KEO quickly becomes prohibitive. We therefore return to the subject of an economical representation in a later section, where more details are given on this subject.

4.2 Step 2: on-the-fly KEO construction within the ADGA

In the next step of our study, we examine to what extent ADGA-PESs are affected by the choice of the KEO for the VSCF calculations within the ADGA. We construct PESs with the ADGA and several KEO variants (KEO$_{PES}$ in Fig. 3) and subsequently run VCC[3] calculations with the resulting PESs and different KEOs (KEO$_{VCC}$ in Fig. 3). Fig. 3 shows the mean and maximum absolute deviations of vibrational state energies of formaldehyde and formic acid relative to energies obtained from VCC[3] calculations using the full KEO and the ADGA-PES generated with the full KEO. The first general observation is that for both molecules the scale of the deviations among the vibrational state energies calculated for the different KEOs and PESs is within the range of 3.5 cm^{-1}, *i.e.* relatively small. This indicates that the PES constructed by the ADGA is not overly sensitive to the employed KEO variant, provided that the latter is not based on a very crude approximation. In addition to that, for both systems and each potential energy surface, *i.e.* the different KEO$_{PES}$, the MADs and MAXs of the correlated calculations are almost equal for the xM[G,V] and the full KEO, while the VCC[3] calculations with the 2M[G,V]2T KEO produce larger deviations. This is in agreement with the results presented in the preceding **Step 1**.

To go into further detail, the MADs obtained for formaldehyde are slightly higher for the PES determined with the 2M[G,V]2T KEO compared to those obtained with the sequential or full KEOs, which produce nearly matching MADs for any KEO$_{VCC}$. The MAXs obtained for VCC[3] calculations with the 2M[G,V]2T KEO are stable at 3.4 cm^{-1} for all PESs. For corresponding calculations with the 3M[G,V] as well as the full KEO the MAX for the PES constructed with the 2M[G,V]2T KEO is at roughly 2 cm^{-1}, and almost vanishes when calculated for the remaining PESs. For *trans*-HCOOH the deviations of the vibrational state energies across the different PESs (with different KEO$_{PES}$) are even less pronounced than for CH$_2$O. The MADs and MAXs remain at basically 0.8 cm^{-1} and 2.8 cm^{-1}, respectively, across the different PESs when using the 2M[G,V]2T KEO in the correlated calculations. For the other two KEO$_{VCC}$ the MADs are below 0.1 cm^{-1} for all PESs. We note, however, that there is a slight improvement of the MAD for the PESs based on the 2M[G]-SQ2T and the full KEOs compared to the 2M[G,V]2T and 1M[G]-SQ2T KEOs, as is depicted in the inset of the lower panel in Fig. 3. The MAXs for PESs generated with the 2M[G,V]2T and the 1M[G]-SQ2T KEO scheme are
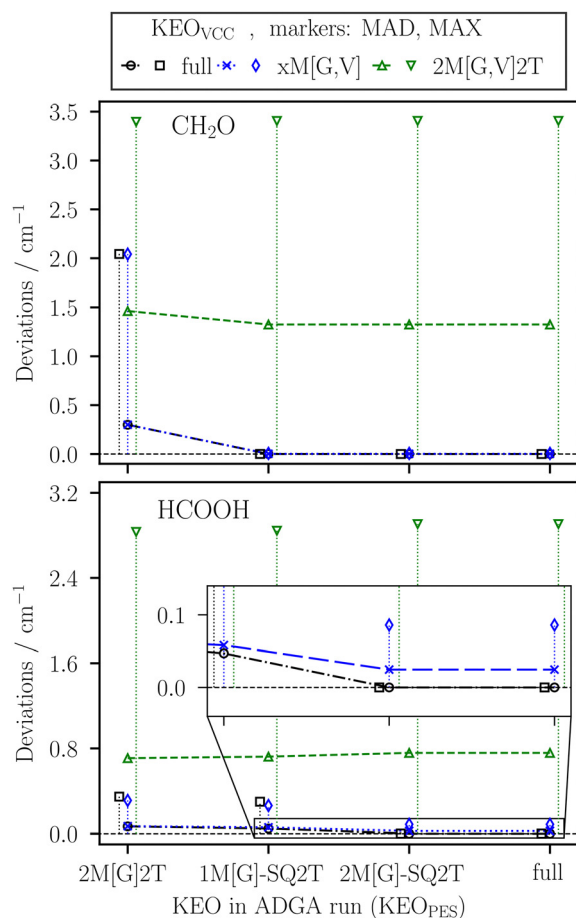
11476 | *Phys. Chem. Chem. Phys.*, 2024, **26**, 11469–11481

This journal is © the Owner Societies 2024

Fig. 3 Mean and maximum deviation across the VCC[3] ground state, fundamental and first overtone excitation energies below 4000 cm^{-1} for CH$_2$O (top) and *trans*-HCOOH (bottom). Calculated with different KEOs (KEO$_{VCC}$) as well as for several ADGA–PESs generated with different KEO models (KEO$_{PES}$). For CH$_2$O "*x*" in "*x*M[G,V]" is 3, for HCOOH it is 2. Deviations are calculated relative to the VCC[3] calculations with the full KEO and the ADGA PES generated with the full KEO.

slightly larger (around 0.3 cm^{-1}) than for the other two PESs (just below 0.1 cm^{-1}).

We can draw a few important conclusions from these results. First, the relatively small range of the absolute deviations across all PESs indicates that in many use cases where an analytical representation of the full KEO is not available, any of the investigated KEO approximations could be a sensible choice for the KEO in an ADGA run. Second, even in situations where neither the full KEO nor a Taylor expansion-based variant is applicable, concatenating ADGA runs of step-wisely increasing mode-coupling level presents a systematically improvable workflow to construct a reliable *n*-mode expansion-based general-coordinate KEO alongside the PES.

4.3 Step 3: reducing the computational cost

In the preceding Section 4.1 we have shown that the general KEO variants based on *n*-mode expansions of G_{ij} and V_{ep} generally produce very accurate vibrational energies across all molecules in our test set. As expected, the accuracy of the

results appears to be controllable through the maximum mode coupling level allowed in the *n*-mode expansions. With respect to the applicability of the KEO schemes introduced here, these are very encouraging results, as high and tunable accuracy is a characteristic of great importance to new approximations. However, another important factor is the numerical cost associated with applying the KEO approximations. With respect to that, the *n*-mode expansion based KEOs display three possible bottlenecks, namely the maximum level of mode coupling, the number of *n*-mode expansions for the elements of the G matrix and the number of terms in the resulting operators. Concerning the first obstacle, a high mode coupling level in the Hamiltonian will result in a steep scaling of the cost with respect to the number of modes, especially if the operator is combined with a more accurate vibrational correlation method. Hence, the maximum mode coupling level in the PES and KEO should be kept as low as possible. In view of that it is reassuring that even with the fairly low 3M-coupling of the 1M[G,V] KEO adequate results can be obtained. With regards to the second obstacle, we note that, other than constructing the KEO from only the upper triangle form of the symmetric G matrix, not much can be done to avoid the increasing number of G_{ij} expansions with increasing number of modes. Yet, we do not necessarily expect this part to be a bottleneck compared to, *e.g.*, accurate electronic structure calculations of the PES, as the calculation and the fitting of the G_{ij} matrix elements is relatively fast. Beyond this, with increasing system size the number of terms in the individual *n*-mode expansions of G_{ij} grows drastically. Of course, this is also tied into the choice of maximum mode coupling level in the *n*-mode expansions for the KEO. For instance, if a KEO variant based on 1M expansions is used, the size of the fitting basis will impact the overall computational cost less compared to expansions of higher mode coupling.

To address the third obstacle in particular, we investigate two approaches to reducing the numerical cost of applying the *n*-mode expansion based KEO approximations: first, we assess the possibility of using coefficient screening to restrict the number of terms in the *n*-mode expansions. Second, we examine to what extent the calculated vibrational energies are affected by the reduction of the number of polynomial functions in the continuous representations of G_{ij} and V_{ep}.

As a final note, we point out that while the calculation of the extrapotential is characterized by steep computational scaling,[74] it will most likely not be the limiting factor. That is because it is only a single scalar term (unlike the numerous G_{ij} matrix elements giving rise to derivative operators), which in MidasCpp, is by default absorbed into the PES if both are represented by the same fitting basis. Moreover, its construction is not part of any expensive iterative steps in the vibrationally correlated calculation and since its effect on the calculated vibrational properties is typically fairly small, it could even be fully disregarded in many applications.

4.3.1 Coefficient screening. One way of reducing the number of operator terms is the screening of coefficients in the *n*-mode expansion when constructing the KEOs in MidasCpp.

This journal is © the Owner Societies 2024

Phys. Chem. Chem. Phys., 2024, **26**, 11469–11481 | **11477**

Table 3 Number of operator terms in the KEO and MADs of the ground state energy as well as fundamental and first overtone excitation energies below 4000 cm^{-1} for *trans*-HCOOH. Determined for different thresholds ε_{KEO}, which were used to screen the operator terms in the KEO. Calculated for KEO approximations based on different static-grid n-mode expansions of the G matrix and V_{ep}, relative to the full vibrational polyspherical KEO. Calculated with VCC[3]. Wave numbers are reported in cm^{-1}

	3M[G,V]		2M[G,V]		1M[G,V]	
ε_{KEO}	N_{terms}	MAD	N_{terms}	MAD	N_{terms}	MAD
0.0	356 306	0.07	36646	0.08	2062	0.49
10^{-9}	60 811	0.07	12665	0.08	1106	0.49
10^{-8}	44 542	0.07	9701	0.08	948	0.49
10^{-7}	27 265	0.07	6695	0.08	755	0.49
10^{-6}	12 345	0.09	3912	0.09	535	0.51
10^{-5}	3381	2.85	1711	2.85	275	2.70
1.0	46	12.55	46	12.55	46	12.55

Table 3 shows the number of operator terms and the MADs across the calculated vibrational state energies of *trans*-formic acid and for different maximum mode-coupling levels after screening the KEO terms with different thresholds. In Tables S5–S7 in the ESI,† the corresponding individual vibrational state energies are listed for the 1M[G,V], 2M[G,V] and 3M[G,V] KEO variants.

If the screening threshold is set to 0.0, we recover the energies presented in **Step 1** of our investigation (Table S4 in the ESI†). On the other hand, if the screening threshold is set to 1.0, the resulting KEO corresponds to that obtained from a zero-order n-mode expansion of G_{ij} and V_{ep} and also the Taylor expansion-based G0 KEO variant from our previous work.[46] For thresholds between these limiting cases an expected trend is discernible, namely that with decreasing threshold the number of operator terms in the KEO consistently grows while the MADs steadily drop before plateauing at the expected values (see Table S4 in the ESI†). Importantly, a relatively large screening threshold appears to be sufficient to capture most of the KEO's effect. For our test case a threshold of 10^{-6} resulted in MADs very close to the ones obtained with no screening at all, while regarding only a fraction of the corresponding KEO terms. In particular, this threshold resulted in KEO representations with approximately 25%, 11% and 3% of the operator terms expected for the full 1M[G,V], 2M[G,V] and 3M[G,V] KEOs with a FBMO of 12, respectively. Screening is conceptually simple and straightforward to apply. However, it can of course be considered wasteful to first construct many operator terms only to disregard most of them later. To balance this, we also consider reducing the size of the numerical representations earlier on.

4.3.2 Fitting basis of the *n*-mode expansions. A corresponding approach to reducing the number of operator terms is to restrict the number of functions used in the fitting procedure for the n-mode expansions of G_{ij} and V_{ep}. Fig. 4 and Table S8 in the ESI,† provide insights on vibrational state energies of *trans*-formic acid obtained for KEOs based on 1M- and 2M-expansions of G_{ij} and V_{ep} as well as different maximum total polynomial orders of the fitting basis (FBMO).

In the upper part of Fig. 4 the number of terms in the kinetic energy operators for different FBMOs are displayed. The numbers
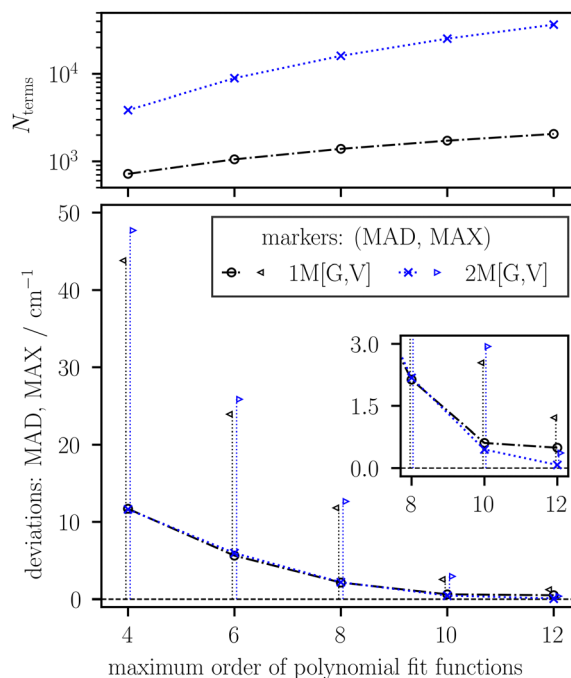

Fig. 4 Number of KEO terms and mean (MAD) and maximum (MAX) absolute deviations of the calculated vibrational energies for *trans*-HCOOH and different maximum total polynomial orders of the functions in the fitting basis (FBMO) for the construction of the KEO n-mode expansions. Calculated relative to the full polyspherical vibrational KEO.

comprise the terms stemming from the n-mode expansions of G_{ij} and V_{ep} after MidasCpp has internally summarized equivalent terms as much as possible. It is obvious that the number of terms increases drastically with increasing maximum polynomial order of the fitting basis. For the 1M[G,V] KEO model the number of terms in the KEO is in the range of roughly 700 (FBMO = 4) to 2000 (FBMO = 12), which is a sizable increase from the 378 terms in the analytical full KEO. The 1M[G,V] KEO with FBMO = 12 has, however, about 500 terms less than the corresponding 2M[G,V]2T KEO,[46] while also having a lower maximum mode coupling level and producing more accurate results. When going to 2M-expansions of G_{ij} and V_{ep}, the number of KEO terms increases steeply again. Even with an FBMO of only 4, there are almost twice as many terms as for the 1M[G,V] variant with FBMO = 12. For the largest fitting basis, the 2M[G,V] KEO of HCOOH contains approximately 36 000 operator terms. In view of molecular systems with more than five atoms, it is thus clear that any reduction of the fitting basis would correspond to a significant saving of operator terms, especially for the n-mode expansions with mode-coupling levels higher than one.

To estimate if such restrictions of the fitting basis are sensible with respect to accuracy, we regard the maximum and mean absolute deviations across the calculated vibrational energies of HCOOH shown in the lower part of Fig. 4. They indicate that polynomial fitting functions of high order are crucial for highly accurate results. To start with the general trend, the MADs and MAXs decrease consistently with increasing size of the fitting basis for both, the 1M[G,V] and 2M[G,V]

11478 | *Phys. Chem. Chem. Phys.*, 2024, **26**, 11469–11481

This journal is © the Owner Societies 2024

KEO variants. When the sum of the polynomial orders in each term of the n-mode expansions is restricted to 4, the MAD and MAX are relatively large, amounting to just above 10 cm^{-1} and around 45 cm^{-1} for both, the 1M[G,V] and 2M[G,V] KEOs, respectively. Upon increasing the FBMO from 4 to 6 and from 6 to 8, the MAD and MAX are approximately halved compared to the preceding lower order. For FBMO = 8 the MAXs and MADs are below 13 cm^{-1} and 3 cm^{-1}, respectively, for the KEOs based on 1M- as well as 2M-expansions of G_{ij} and V_{ep}. When going to FBMO = 10 there is another significant decrease of the absolute deviations compared to a maximum order of 8, especially for the maximum deviations. At that point the MAXs and MADs are below 3 cm^{-1} and 0.7 cm^{-1}, respectively, for both of the examined KEOs. The data points with a maximum overall polynomial order of FBMO = 12 correspond to those presented in **Step 1** of our investigation (Fig. 2, Table S4 in the ESI†). Here, the MADs are 0.49 cm^{-1} and 0.08 cm^{-1} for the 1M[G,V] and 2M[G,V] KEO variants, respectively, while the corresponding MAXs amount to 1.21 cm^{-1} and 0.37 cm^{-1}. If very accurate results are targeted, a fitting basis of FBMO = 12 appears to be necessary, giving rise to KEOs with a relatively large amount of terms.

In view of these results, we assess the coefficient screening to be the more effective measure to reduce the computational cost for calculations with the n-mode expansion-based KEO variants. However, the two approaches are not mutually exclusive, such that a reduction of the fitting basis' size is readily combined with coefficient screening. To illustrate this, we perform a final set of test calculations on *trans*-HCOOH with the KEO variants based on a 2M-expansion of G_{ij} and V_{ep}. Herein, we use an FBMO of twelve to fit the 1M part of the expansions, while fitting the 2M part with polynomial functions contained in FBMOs from four to ten. The resulting KEOs we use in VCC[3] calculations with coefficient screening with a threshold of $\varepsilon_{KEO} = 10^{-7}$. The results in Table 4 (and Table S9 in the ESI†) indicate that both schemes can be well combined and that the largest fitting bases do not necessarily always produce the best results.

Moreover, they show that the large deviations reported in Fig. 4 for the lower FBMOs are mainly driven by insufficient accuracy in the 1M part of the expansion. With a FBMO = 4 in the 1M- and 2M-parts of the expansions, for example, the MAD and MAX are around 10 cm^{-1} and 45 cm^{-1}, whereas they are only around 1 cm^{-1} and 5 cm^{-1} for FBMOs of 12 and 4 for the 1M- and 2M-parts, respectively. On the basis of this last test, it appears that combining coefficient screening with tailored fitting bases for the terms at each mode coupling level is the most sensible and efficient approach.

5 Conclusions

We have presented our work on n-mode expansion-based approximations of kinetic energy operators in (curvilinear) polyspherical coordinates. The focal point of the study has been to evaluate the performance of these KEO variants in accurate vibrationally correlated calculations and in PES constructions with the ADGA. Moreover, we have examined a pair of approaches to reduce the computational cost of applying the n-mode KEOs.

In order to facilitate the construction and application of the KEO variants, the capabilities of the Tnum/Tana-MidasCpp interface of our preceding works have been further expanded. To be precise, Tnum/Tana has been extended to include output tailored to MidasCpp's workflow of constructing n-mode expansions, while the latter has been enhanced with the possibility to assemble the KEOs from the n-mode expansions. We have shown that the n-mode KEO variants approximate the full KEO accurately and in a systematically improvable way. They are highly flexible and always display the numerically advantageous sum-of-products form. Compared to the Taylor expansion-based schemes, they are generally more costly, but appear to be more effective as well. For instance, the 3M-coupled KEOs built from 1M-expansions of G_{ij} and V_{ep} perform consistently better than the 4M-coupled Taylor expanded 2M[G,V]2T KEOs introduced in our preceding work. Moreover, the increase of numerical cost can be alleviated by coefficient screening and fine-tuning the fitting bases in the n-mode expansion. In addition to that, we demonstrated the construction, usage and systematic improvement of reliable n-mode expansion-based KEO variants in the course of the PES generation with the adaptive density guided approach. This is integral to prospective calculations with curvilinear coordinate systems other than the polyspherical ones. Based on our findings, we propose the following procedure for an efficient and accurate investigation of a molecule's vibrational properties: starting from solely a reference structure and a suitably selected set of coordinates, faithful n-mode representations of the PES and KEO are generated with the proposed workflow of concatenating ADGA runs. The resulting operators are employed in efficient vibrational coupled cluster calculations to access the vibrational properties. To boost the procedure's efficiency, the fitting bases for the KEO n-mode expansions are restricted at the higher mode coupling levels and coefficient screening is applied to every KEO action.

In the future, we aim to probe the n-mode expansion-based KEOs in PES constructions and vibrationally correlated calculations for larger molecular systems, and with general coordinate systems other than the polyspherical ones, especially

Table 4 Number of KEO terms and averaged deviations of the vibrational ground state energy as well as fundamental and first overtone excitation energies below 4000 cm^{-1} for *trans*-HCOOH and a 2M[G,V] KEO. In the KEO the 1M part of the expansion is generated with FBMO = 12, while the 2M part is constructed with different FBMOs (FBMO-2M). Coefficient screening with a threshold of $\varepsilon_{KEO} = 10^{-7}$ is applied. Deviations are given relative to the full KEO. Calculated with VCC[3]. Wave numbers are reported in cm^{-1}

FBMO-2M	12	10	8	6	4
MAX	0.37	0.08	1.34	2.00	4.71
MAD	0.08	0.03	0.26	0.44	1.08
N_{terms}	6695	5779	4828	3804	2252

This journal is © the Owner Societies 2024

Phys. Chem. Chem. Phys., 2024, **26**, 11469–11481 | **11479**

curvilinear normal modes. Moreover, we plan to explore the limits of the n-mode expansion-based KEOs by examining systems for which we expect them to perform less well, *i.e.* those displaying large amplitude motion. For such systems, we can select more complex coordinates that are adapted to the large amplitude motion, *e.g.* linear combinations or flexible coordinates, by adding coordinate transformations available in Tnum. Also, in some cases, for instance H_2O_2, this problem can be mostly avoided by selecting the coordinates and the reference structure for the expansions smartly. Generally though, expansions around a single reference may become inaccurate if the nuclear motion spans a vast portion of the configuration space. Other interesting paths of this strand of research are the application of the n-mode KEO schemes in combination with the Gaussian process regression ADGA or in simulations of molecular dynamics with the recently introduced time-dependent ADGA.[51,53,75] These methods are already implemented in MidasCpp.

Data availability statement

The vibrational state energies analyzed in this work are provided as a part of the ESI.† Exemplary calculations employing the n-mode KEO models are made available in the MidasCpp example suite. The molecular property surfaces generated in the course of this work are available from the authors upon request.

Author contributions

F. Bader: conceptualization, data curation, funding acquisition, investigation, methodology, software, validation, visualization, writing – original draft. **D. Lauvergnat**: formal analysis, methodology, software, writing – original draft. **O. Christiansen**: conceptualization, formal analysis, funding acquisition, supervision, writing – original draft.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 A. G. Császár, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2012, **2**, 273–289.
2 R. C. Fortenberry, X. Huang, A. Yachmenev, W. Thiel and T. J. Lee, *Chem. Phys. Lett.*, 2013, **574**, 1–12.
3 G. Simons, R. G. Parr and J. M. Finlan, *J. Chem. Phys.*, 1973, **59**, 3229–3234.
4 R. Burcl, S. Carter and N. C. Handy, *Chem. Phys. Lett.*, 2003, **373**, 357–365.
5 M. Keçeli, T. Shiozaki, K. Yagi and S. Hirata, *Mol. Phys.*, 2009, **107**, 1283–1301.
6 J. O. Jung and R. B. Gerber, *J. Chem. Phys.*, 1996, **105**, 10332–10348.
7 S. Carter, S. J. Culik and J. M. Bowman, *J. Chem. Phys.*, 1997, **107**, 10458–10469.
8 J. M. Bowman, S. Carter and X. Huang, *Int. Rev. Phys. Chem.*, 2003, **22**, 533–549.
9 G. Rauhut, *J. Chem. Phys.*, 2004, **121**, 9313–9322.
10 J. Kongsted and O. Christiansen, *J. Chem. Phys.*, 2006, **125**, 124108.
11 D. Toffoli, J. Kongsted and O. Christiansen, *J. Chem. Phys.*, 2007, **127**, 204106.
12 H. Rabitz and Ö. F. Alis, *J. Math. Chem.*, 1999, **25**, 197–233.
13 H.-D. Meyer, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2012, **2**, 351–374.
14 T.-S. Ho and H. Rabitz, *J. Chem. Phys.*, 2003, **119**, 6433–6442.
15 J. Rheinecker and J. M. Bowman, *J. Chem. Phys.*, 2006, **125**, 133206.
16 S. Manzhos and T. Carrington, *J. Chem. Phys.*, 2006, **125**, 084109.
17 K. Yagi, S. Hirata and K. Hirao, *Theor. Chem. Acc.*, 2007, **118**, 681–691.
18 F. Richter, P. Carbonnière, A. Dargelos and C. Pouchan, *J. Chem. Phys.*, 2012, **136**, 224105.
19 J. K. Watson, *Mol. Phys.*, 1968, **15**, 479–490.
20 J. M. Bowman, T. Carrington and H.-D. Meyer, *Mol. Phys.*, 2008, **106**, 2145–2182.
21 S. Carter and J. M. Bowman, *J. Chem. Phys.*, 1998, **108**, 4397–4404.
22 M. J. Bramley, W. H. Green Jr and N. C. Handy, *Mol. Phys.*, 1991, **73**, 1183–1208.
23 E. L. Sibert III, J. T. Hynes and W. P. Reinhardt, *J. Phys. Chem.*, 1983, **87**, 2032–2037.
24 E. L. Klinting, D. Lauvergnat and O. Christiansen, *J. Chem. Theory Comput.*, 2020, **16**, 4505–4520.
25 D. Madsen, O. Christiansen and C. König, *Phys. Chem. Chem. Phys.*, 2018, **20**, 3445–3456.
26 M. Schneider and G. Rauhut, *J. Comput. Chem.*, 2023, **44**, 298–306.
27 S. Carter, J. M. Bowman and L. B. Harding, *Spectrochim. Acta, Part A*, 1997, **53**, 1179–1188.
28 Y. Wang, S. Carter and J. M. Bowman, *J. Phys. Chem. A*, 2013, **117**, 9343–9352.
29 M. Neff, T. Hrenar, D. Oschetzki and G. Rauhut, *J. Chem. Phys.*, 2011, **134**, 064105.

11480 | *Phys. Chem. Chem. Phys.*, 2024, **26**, 11469–11481

This journal is © the Owner Societies 2024

30 M. Tschöpe and G. Rauhut, *J. Chem. Phys.*, 2022, **157**, 234105.

31 D. Strobusch and C. Scheurer, *J. Chem. Phys.*, 2011, **135**, 124102.

32 D. Strobusch and C. Scheurer, *J. Chem. Phys.*, 2011, **135**, 144101.

33 D. Strobusch, M. Nest and C. Scheurer, *J. Comput. Chem.*, 2013, **34**, 1210–1217.

34 D. Strobusch and C. Scheurer, *J. Chem. Phys.*, 2014, **140**, 074111.

35 S. N. Yurchenko, W. Thiel and P. Jensen, *J. Mol. Spectrosc.*, 2007, **245**, 126–140.

36 A. Yachmenev and S. N. Yurchenko, *J. Chem. Phys.*, 2015, **143**, 014105.

37 M. Bounouar and C. Scheurer, *Chem. Phys.*, 2008, **347**, 194–207.

38 Y. Scribano, D. M. Lauvergnat and D. M. Benoit, *J. Chem. Phys.*, 2010, **133**, 094103.

39 I. Suwan and R. Gerber, *Chem. Phys.*, 2010, **373**, 267–273.

40 D. Lauvergnat, Tnum-Tana, a Fortran code, which deals with curvilinear coordinates and kinetic energy operators, available from https://github.com/lauvergn/Tnum-Tana, 2002.

41 O. Christiansen, D. G. Artiukhin, F. Bader, I. H. Godtliebsen, E. M. Gras, W. Györffy, M. B. Hansen, M. B. Hansen, M. G. Højlund, N. M. Høyer, R. B. Jensen, A. B. Jensen, E. L. Klinting, J. Kongsted, C. König, D. Madsen, N. K. Madsen, K. Monrad, G. Schmitz, P. Seidler, K. Sneskov, M. Sparta, B. Thomsen, D. Toffoli and A. Zoccante, *MidasCpp 2022.10.0*, available from https://gitlab.com/midascpp/midascpp.

42 D. Lauvergnat and A. Nauts, *J. Chem. Phys.*, 2002, **116**, 8560–8570.

43 M. Ndong, L. Joubert-Doriol, H.-D. Meyer, A. Nauts, F. Gatti and D. Lauvergnat, *J. Chem. Phys.*, 2012, **136**, 034107.

44 M. Ndong, A. Nauts, L. Joubert-Doriol, H.-D. Meyer, F. Gatti and D. Lauvergnat, *J. Chem. Phys.*, 2013, **139**, 204107.

45 E. Marsili, F. Agostini, A. Nauts and D. Lauvergnat, *Philos. Trans. R. Soc., A*, 2022, **380**, 20200388.

46 F. Bader, D. Lauvergnat and O. Christiansen, *J. Chem. Phys.*, 2023, **159**, 214107.

47 A. Nauts and X. Chapuisat, *Mol. Phys.*, 1985, **55**, 1287–1318.

48 E. Mátyus, G. Czakó and A. G. Császár, *J. Chem. Phys.*, 2009, **130**, 134112.

49 M. Sparta, D. Toffoli and O. Christiansen, *Theor. Chem. Acc.*, 2009, **123**, 413–429.

50 M. Sparta, I.-M. Høyvik, D. Toffoli and O. Christiansen, *J. Phys. Chem. A*, 2009, **113**, 8712–8723.

51 G. Schmitz, D. G. Artiukhin and O. Christiansen, *J. Chem. Phys.*, 2019, **150**, 131102.

52 D. G. Artiukhin, E. L. Klinting, C. König and O. Christiansen, *J. Chem. Phys.*, 2020, **152**, 194105.

53 N. M. Høyer and O. Christiansen, *J. Chem. Theory Comput.*, 2024, **20**, 558–579.

54 F. Richter and P. Carbonnière, *J. Chem. Phys.*, 2018, **148**, 064303.

55 J. M. Bowman, K. Christoffel and F. Tobin, *J. Phys. Chem.*, 1979, **83**, 905–912.

56 O. Christiansen and J. M. Luis, *Int. J. Quantum Chem.*, 2005, **104**, 667–680.

57 B. Schröder and G. Rauhut, in *Vibrational Dynamics of Molecules*, ed. J. M. Bowman, World Scientific, Singapore, 2022.

58 D. Toffoli, M. Sparta and O. Christiansen, *Mol. Phys.*, 2011, **109**, 673–685.

59 P. Seidler and O. Christiansen, *J. Chem. Phys.*, 2009, **131**, 234109.

60 O. Christiansen, *J. Chem. Phys.*, 2005, **122**, 194105.

61 P. Seidler and O. Christiansen, *J. Chem. Phys.*, 2007, **126**, 204101.

62 P. Seidler, M. Sparta and O. Christiansen, *J. Chem. Phys.*, 2011, **134**, 054119.

63 M. Häser and R. Ahlrichs, *J. Comput. Chem.*, 1989, **10**, 104–111.

64 F. Weigend and M. Häser, *Theor. Chem. Acc.*, 1997, **97**, 331–340.

65 C. Hättig, A. Hellweg and A. Köhn, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1159–1169.

66 T. H. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007–1023.

67 K. L. Schuchardt, B. T. Didier, T. Elsethagen, L. Sun, V. Gurumoorthi, J. Chase, J. Li and T. L. Windus, *J. Chem. Inf. Model.*, 2007, **47**, 1045–1052.

68 B. P. Pritchard, D. Altarawy, B. Didier, T. D. Gibsom and T. L. Windus, *J. Chem. Inf. Model.*, 2019, **59**, 4814–4820.

69 F. Weigend, *Phys. Chem. Chem. Phys.*, 2002, **4**, 4285–4291.

70 F. Weigend, M. Häser, H. Patzelt and R. Ahlrichs, *Chem. Phys. Lett.*, 1998, **294**, 143–152.

71 F. Weigend, A. Köhn and C. Hättig, *J. Chem. Phys.*, 2002, **116**, 3175–3183.

72 TURBOMOLE V7.5 2020, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from https://www.turbomole.org.

73 E. L. Klinting, B. Thomsen, I. H. Godtliebsen and O. Christiansen, *J. Chem. Phys.*, 2018, **148**, 064113.

74 A. Nauts and D. Lauvergnat, *Mol. Phys.*, 2018, **116**, 3701–3709.

75 D. G. Artiukhin, I. H. Godtliebsen, G. Schmitz and O. Christiansen, *J. Chem. Phys.*, 2023, **159**, 024102.