A Tale of Two Vectors

A Lanczos Algorithm For Calculating RPA Mean Excitation Energies

Zamok, Luna; Coriani, Sonia; Sauer, Stephan P. A.

Published in:
Journal of Chemical Physics

Link to article, DOI:
10.1063/5.0071144

Publication date:
2022

Document Version
Peer reviewed version

A Tale of Two Vectors: A Lanczos Algorithm For Calculating RPA Mean Excitation Energies

Luna Zamok,[*,†,‡] Sonia Coriani,[*,‡] and Stephan P. A. Sauer[*,†]

†*Department of Chemistry, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark*

‡*Department of Chemistry, Technical University of Denmark, Kemitorvet Bldg 207, 2800 Kongens Lyngby, Denmark*

E-mail: luza@kemi.dtu.dk; soco@kemi.dtu.dk; sauer@chem.ku.dk

Abstract

The experimental and theoretical determination of the mean excitation energy, $I(0)$, and the stopping power, $S(v)$, of a material is of great interest in particle and material physics, as well as radiation therapy. For calculations of $I(0)$, the complete set of electronic transitions in a given basis set is required, effectively limiting such calculations to systems with a small number of electrons, even at the random-phase approximation (RPA)/time-dependent Hartree-Fock (TDHF) or time-dependent density-functional theory (TDDFT) level. To overcome such limitations, we present here the implementation of a Lanczos algorithm adapted for the paired RPA/TDHF eigenvalue problem in the Dalton program and show that it provides good approximations of the entire RPA eigenspectra in a reduced space. We observe rapid convergence of $I(0)$ with the number of Lanczos vectors as the algorithm favors the transitions with large contributions. In most cases, the algorithm recovers RPA $I(0)$ values of up to 0.5 % accuracy at less than a quarter of the full space size. The algorithm not only exploits the RPA paired structure to save computational resources, but it is also preserves certain

sum-over-states properties, as first demonstrated by Johnson *et al.* [Comput. Phys. Commun. 1999, 120, 155]. The block Lanczos RPA solver, as presented here, thus shows promise for computing mean excitation energies for systems larger than what was computationally feasible before.

1 Introduction

Interaction of fast ions and matter leads to various processes in the target material that are of interest in multiple fields in science, including nuclear and material physics and radiation therapy. These processes, such as ionisation, electronic and vibrational excitations or charge transfer, can be described by Bethe's quantum mechanical formula for calculating the stopping power.[1,2] Stopping power can be understood as a retarding force that acts on the charged particle moving through matter, which then, due to different interactions with said target matter, suffers loss of energy. Bethe's formula consists of the speed and charge of the projectile and the mean excitation energy of the target material, where the latter is the only non-trivial quantity in the formula. For a given material, it describes the mean loss of energy the charged particle experiences when traversing through the target material.

The mean excitation energy can be determined from experiment or *ab initio* calculations.[3,4] It belongs to the group of molecular properties in quantum chemistry called sum-over-states properties, whose computation requires calculating the complete set of excitation energies and transition moments to both bound states as well as the continuum, since they cannot be recast in closed form. So, for instance, in contrast to the dipole polarizability, the mean excitation energy cannot be calculated as a linear response function. This makes the determination of the *ab initio* mean excitation energy for larger atoms and molecules rather difficult, if not, in many cases, unfeasible. As a result, the random phase approximation (RPA)[5] level of theory is usually used, while higher-level electron correlated calculations have only been performed for a few atoms and the H_2 molecule.[6] Since in this case electron correlation effects were found to be small,[6] there are grounds to believe that the RPA

approach is not only computationally feasible, but also delivers a reasonable description. Nonetheless, even within RPA, computing mean excitation energies for systems with a large number of electrons is still either impractical or unattainable. This is due to the roughly $\mathcal{O}(N_e^6)$ scaling, for a given number of electrons N_e, of the full diagonalization of the RPA matrix. Therefore, good approximations to the full eigenspectrum are highly desirable. In the case of the mean excitation energy, approximating the full eigenspectrum is particularly challenging, as it is well known[7] that the largest contribution to the mean excitation energy comes from the numerical integration over the continuum. A good representation of the continuum is therefore a prerequisite for accurated values of the mean excitation energy. To attain such a balanced approximation, we propose to use the family of Lanczos[8] algorithms. These algorithms exhibit simultaneous convergence from both extreme ends of the eigenspectrum[9] and can therefore be expected to give a balanced description of the eigenspectrum at a reasonably sized subspace. To investigate this, we have, in the present study, implemented a Lanczos algorithm as a solver for the paired RPA eigenvalue problem in the Dalton[10,11] program, and tested it for several small systems in order to obtain and analyze its convergence behaviour when computing sum-over-states properties in practice. Besides the work of Brabec *et al.*,[12] we are unaware of recent implementations of a Lanczos algorithm explicitly designed to be used to calculate sum-over-states properties at the RPA level. Brabec *et al.* have computed absorption spectra obtained from the sum-over-states imaginary polarizability at TD-DFT level using Casida[13] formulation. Versions of the general algorithm have previously been implemented and used to compute linear response properties and spectra at various levels of theory, including RPA.[14–22]

In more recent implementations, targeting linear response, nonsymmetric Lanczos solvers were used to calculate vibrational coupled cluster response functions[23] as well as regular and damped electronic coupled-cluster linear response functions.[16,24] The latter were observed to perform well for computing absorption spectra over the entire frequency range, including the X-ray region, which could be accessed thanks to the convergence properties of the algorithm.

The same solver was used by Sauer *et al.* to compute the mean excitation energies for a few atoms and the H_2 molecule,[6] however, it was not computationally feasible for systems with a large number of electrons. Also, the solver has been extensively used to generate pseudospectra in connection with the computational simulation of total photoionisation cross-section using Stieltjes Imaging and analytic continuation approaches.[25–27]

Several implementations that use Lanczos-type algorithms for computing the lowest excitations at the RPA level were developed and published between 2000 and 2009.[18–21] These authors[18–21] cast the RPA eigenvalue problem (or the equivalent time-dependent density functional theory[28] (TDDFT) eigenvalue problem) in the form of Harmonic Oscillator representation. This representation results in two coupled eigenvalue problems that obey the Thouless variational[29] principle, which states that the lowest frequency can be obtained as a minimum of the total energy, and thus has a defined minimum and is suitable for computing the lowest eigenvalues. Tsiper *et al.*[18] presented a coupled Lanczos algorithm that is run several times to find the lowest eigenvalues one by one, while Chernyak *et al.*[19] employed the same procedure but used a nonsymmetric Lanczos algorithm. More recently, Tretiak *et al.*[20] used the coupled Lanczos algorithm analogous to Tsiper *et al.*[18] and compared its convergence to the performance of the Davidson[30] and Arnoldi[31] algorithms when solving the RPA eigenvalue problem for the lowest eigenmodes and found it to be slower to converge the first excited states than the latter two. In 2008, Rocca *et al.*[21] implemented a nonsymmetric Lanczos solver for TDDFT, where a biconstant extrapolation procedure for the coefficients in Lanczos matrix was also presented. The authors could in this way extrapolate to a large Lanczos matrix from a much smaller one, and their procedure was shown to perform well when computing absorption spectra for a couple of large molecules.

The implementation we report here is based on a Lanczos algorithm for the RPA eigenvalue problem presented in 1999 by Johnson *et al.*[32] As anticipated, to the best of our knowledge, this is the only implementation that explicitly uses the algorithm to compute sum-over-states properties that cannot be straightforwardly formulated as response func-

tions. Alternatively, one could have employed the algorithm of Brabec *et al.,*[12] that is based on the Casida formulation. However, this would have required different linear transformations than those available in the Dalton program we build upon. Johnson *et al.*[32] used the excitation/de-excitation symmetry in the RPA eigenvalue problem to devise Lanczos algorithm that preserves this symmetry, while simultaneously halving the number of iterations needed to build a subspace of desired size. Johnson *et al.* also showed that the algorithm preserves certain sum rules for the sum-over-states properties, for which we include here a more detailed proof. Our implementation of this algorithm includes the computation of the dipole oscillator strengths in the Lanczos basis, as well as the sums, $S(0)$ and $L(0)$, needed to calculate the mean excitation energy. The resulting block Lanczos RPA solver is tested by using it to compute mean excitation energies for a set of molecules. The convergence properties and basis set performance are discussed, and the RPA mean excitation energies are briefly compared with data from experiments.

2 Theory

For a given material, the stopping power describes the slowing down of a charged particle traversing it due to the inelastic collisions between the charged particle and the bound electrons in the material. It is defined by the Bethe formula,[1] as derived via first-order nonrelativistic perturbation theory:

$$S(v) = \frac{4\pi Z^2 N_e}{v^2} \ln\left(\frac{2v^2}{I(0)}\right) \quad . \tag{1}$$

Here, N_e is the number of electrons in the material, while Z and v are the charge and speed of the particle traversing said material, all expressed in Hartree atomic units. The only nontrivial quantity in Eq. (1) is the natural logarithm of the mean excitation energy $I(0)$:

$$\ln I(0) = \frac{L(0)}{S(0)} = \frac{\sum_{n\neq 0} \ln(\omega_n) f_{n0}}{\sum_{n\neq 0} f_{n0}} \tag{2}$$

which is defined through the sums $S(0)$ and $L(0)$, both involving the dipole oscillator strengths f_{n0}.

The summations in Eq. (2) imply both a summation over the discrete and bound excited states of a molecule and an integration over the continuum. In practical RPA calculations with a finite basis set, the latter is replaced by a summation over "pseudostates" which can be regarded as a discretization of the continuum. $S(0)$ fulfills the Thomas-Reiche-Kuhn sum rule,[33–35] that states that, in the complete basis limit, $S(0) = N_e$, the total number of electrons in the system. The dipole oscillator strength, f_{n0}, in Eq. (2) is defined as:

$$f_{n0} = \frac{2}{3}\omega_n |\langle n|\hat{\boldsymbol{\mu}}|0\rangle|^2 \tag{3}$$

where 0 is the ground state, n is the (unperturbed) wavefunction[36] of the n^{th} excited state with excitation energy ω_n (in atomic units), and $\hat{\boldsymbol{\mu}}$ is the electric dipole moment operator.

Computing $I(0)$ for a molecule or atom thus requires computing all its excitation energies and dipole oscillator strengths. Both quantities can be obtained from (the poles and residues of) the linear response function (LRF) or polarization propagator.[37] In the presence of a general time-dependent electromagnetic field, the expectation value of an operator representing some property is expanded in perturbation series, where the LRF is the first order expansion coefficient. The time-dependent Hartree-Fock[38] (TDHF) LRF, also known as RPA[5] LRF, can be expressed as:

$$\langle\langle \hat{P}; \hat{O} \rangle\rangle_\omega = (\mathbf{P}^L(\hat{P}))^T (\mathbf{E} - \omega\mathbf{1})^{-1} \mathbf{P}^R(\hat{O}) = \tag{4}$$
$$\begin{pmatrix} {}^e\mathbf{P}^T(\hat{P}) & -{}^e\mathbf{P}^T(\hat{P}) \end{pmatrix} \left[\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ -\mathbf{B} & -\mathbf{A} \end{pmatrix} - \omega \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{pmatrix} \right]^{-1} \begin{pmatrix} {}^e\mathbf{P}(\hat{O}) \\ {}^e\mathbf{P}(\hat{O}) \end{pmatrix}$$

where ω is the frequency of the perturbing field, $\mathbf{P}^L(\hat{P})^T$ and $\mathbf{P}^R(\hat{O})$ are property gradient vectors and $(\mathbf{E} - \omega\mathbf{1})$ is the principal propagator matrix consisting of the RPA electronic Hessian, $\mathbf{E}$. The electronic Hessian is here split in the conventional $\mathbf{A}$ and $\mathbf{B}$ blocks[39] of

dimensions $N \times N$, where N is the number of single excitations in a given basis. The elements in the $\mathbf{A}$ and $\mathbf{B}$ blocks are differences in HF MO-energies (diagonal elements of $\mathbf{A}$) and two-electron integrals.[40] The elements of the ${}^{e}\mathbf{P}^{R}(\hat{O})$ and ${}^{e}\mathbf{P}^{L}(\hat{P})$ are defined as:

$$ {}^{e}\mathbf{P}(\hat{O})_{\tilde{n}} = \left\langle \tilde{0} \middle| \hat{O} \middle| \tilde{n} \right\rangle \qquad {}^{e}\mathbf{P}(\hat{P})_{\tilde{n}} = \left\langle \tilde{n} \middle| \hat{P} \middle| \tilde{0} \right\rangle \tag{5} $$

where $|\tilde{0}\rangle$ and $|\tilde{n}\rangle$ are, respectively, the Hartree-Fock Slater determinant and a Slater determinant where orbitals in $|\tilde{0}\rangle$ are swapped by a virtual orbital, i.e. a singly excited determinant. Note, however, that in Dalton, $\tilde{n}$ are spin-adapted linear combinations of singly excited HF determinants.

The values of ω for which the expression in Eq. (4) becomes singular are interpreted as vertical excitation energies.[40] They can be found by examining when the resolvent matrix $(\mathbf{E} - \omega\mathbf{1})$ becomes singular, i.e. solving the eigenvalue problem:

$$ \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ -\mathbf{B} & -\mathbf{A} \end{pmatrix} \begin{pmatrix} {}^{e}\tilde{\mathbf{X}}_n \\ {}^{d}\tilde{\mathbf{X}}_n \end{pmatrix} = \omega_n \begin{pmatrix} {}^{e}\tilde{\mathbf{X}}_n \\ {}^{d}\tilde{\mathbf{X}}_n \end{pmatrix} \tag{6} $$

Here, ω_n and $\tilde{\mathbf{X}}_n^R = \begin{pmatrix} {}^{e}\tilde{\mathbf{X}}_n & {}^{d}\tilde{\mathbf{X}}_n \end{pmatrix}^T$ are, respectively, the n^{th} eigenvalue and right eigenvector. The latter consists of the upper and lower parts, ${}^{e}\tilde{\mathbf{X}}_n$ and ${}^{d}\tilde{\mathbf{X}}_n$. The corresponding left eigenvector, $\tilde{\mathbf{X}}_n^L$, has a minus sign on the lower part. Swapping the two components of the eigenvector $\tilde{\mathbf{X}}_n^R$ (or $\tilde{\mathbf{X}}_n^L$) results in the eigenvalue of the opposite sign, i.e. $-\omega_n$. As a result of this symmetry, the eigenvectors corresponding to the excitation energies, ω_n, can be used to build the eigenvectors corresponding to the de-excitation energies, $-\omega_n$.[39] Since the eigenvalues and eigenvectors come in such pairs, the RPA eigenvalue equations are said to have a paired structure.[41] This can be utilized when carrying out linear transformations with $\mathbf{E}$ in the iterative solution of the equations, thus reducing the computational costs. Exploiting the paired structure in iterative solvers was shown to lead to faster convergence[41] in Davidson style solvers.

Computing the RPA mean excitation energy, $I(0)$, Eq. (2), involves calculating the dipole oscillator strengths, Eq. (3), which in turn requires computing the RPA excitation energies and electric dipole transition moments. The square of the norm of the n^{th} RPA dipole transition moment, $|\langle n|\hat{\boldsymbol{\mu}}|0\rangle|^2$, can be computed as a residue of the RPA LRF:

$$|\langle n|\hat{\boldsymbol{\mu}}|0\rangle|^2 = \lim_{\omega\to\omega_n}(\omega-\omega_n)(\mathbf{P}^L(\hat{\boldsymbol{\mu}}))^T(\mathbf{E}-\omega\mathbf{1})^{-1}\mathbf{P}^R(\hat{\boldsymbol{\mu}}) \tag{7}$$

here, n and 0 are the RPA ground and n^{th} excited states. Expressing $\mathbf{E}$ in spectral form, this expression simplifies to:

$$|\langle n|\hat{\boldsymbol{\mu}}|0\rangle|^2 = \lim_{\omega\to\omega_n}(\omega-\omega_n)(\mathbf{P}^L(\hat{\boldsymbol{\mu}}))^T\tilde{\mathbf{R}}(\boldsymbol{\Omega}-\omega\mathbf{1})^{-1}\tilde{\mathbf{L}}\mathbf{P}^R(\hat{\boldsymbol{\mu}}) = (\mathbf{P}^L(\hat{\boldsymbol{\mu}}))^T\tilde{\mathbf{X}}_n^R\tilde{\mathbf{X}}_n^L\mathbf{P}^R(\hat{\boldsymbol{\mu}}) \tag{8}$$

where the columns of $\tilde{\mathbf{R}}$ and the rows of $\tilde{\mathbf{L}}$ are the right and left eigenvectors of $\mathbf{E}$, i.e. $\tilde{\mathbf{R}} = \left[\tilde{\mathbf{X}}_1^R,..,\tilde{\mathbf{X}}_{2N}^R\right]$ and $\tilde{\mathbf{L}} = \left[(\tilde{\mathbf{X}}_1^L)^T,..,(\tilde{\mathbf{X}}_{2N}^L)^T\right]$, and $\boldsymbol{\Omega}$ is a diagonal matrix with the eigenvalues of $\mathbf{E}$ on the diagonal. Thus the squares of the dipole transition moments, and, in turn, the dipole oscillator strengths can be obtained by computing the eigenvectors and eigenvalues of $\mathbf{E}$, as well as the electronic dipole moment property gradient vectors. In other words, computing $I(0)$ at the RPA level requires solving the RPA matrix eigenvalue problem.

2.1 Lanczos for the paired RPA Eigenvalue Problem

Matrix eigenvalue problems in quantum chemistry are usually solved with the help of iterative subspace methods, since the matrices involved are usually too large to be explicitly diagonalized, or even stored. A family of such iterative methods, based on power iteration,[42] called Lanczos[8] methods, are known to perform well if a balanced approximation of the entire eigenspectrum is sought. The Lanczos algorithm(-s) successively builds Krylov subspaces[43,44] in such a manner that the extrema eigenpairs (i.e. the eigenvalues and their corresponding eigenvectors) converge fastest. When calculating sum rules, e.g. approximate $I(0)$ and $S(0)$ values, and therefore approximating the full eigenspectra, this convergence

pattern can be expected to translate into a faster convergence of the said properties. The algorithm builds a subspace incrementally by optimizing Rayleigh quotients[44] of the eigenvalues it tries to solve for. In the basic Lanczos method, a subspace is built by creating one Lanczos vector at each iteration, thus building an orthonormal basis that tridiagonalizes a real square symmetric matrix. Truncating the iterations before the full-space size is reached, and finding the eigenvalues and eigenvectors of such tridiagonal matrix, gives an approximation to the full-space ones.

Various generalizations of the method for nonhermitian matrices or matrices with degenerate eigenvalues have been developed. In floating point arithmetic, the algorithm suffers from round-off errors and so all Lanczos vectors need to be stored and re-orthogonalized to each other in order to avoid the appearance of "ghost" eigenvalues.[45–47] In the nonsymmetric case, the iterations also break down if either the inner product of the new (left and right) Lanczos vectors or the norm of one of them is zero. The latter case indicates that an invariant subspace for the matrix in question has been found, while the former means that the desired oblique projection to the reduced space cannot be found. This is called a "serious break-down" (see more on this in Ref. 48). This break-down as well as near-break-downs can be circumvented by the look-ahead Lanczos methods (see, e.g., Ref. 43) or restarting the iterations; neither is implemented in our current solver yet.

Both symmetric and nonsymmetric Lanczos algorithms have their block analogues (see Refs. 9 and 43), where the former seems to have been developed by Golub and Underwood.[49] The algorithms are useful in cases of degenerate or clustered eigenvalues.[9] Johnson *et al.*[32] adapted a version of the nonsymmetric block Lanczos algorithm for the RPA eigenvalue problem, where the eigenvalues are known to come in numerically degenerate pairs ($\pm\omega_n$). This algorithm preserves the paired structure of the RPA eigenvalue problem by augmenting the Lanczos basis in each iteration by pairs of right and left Lanczos vectors that have the same symmetry as the (right and left) RPA eigenvectors. We have therefore chosen to implement and use a version of this algorithm, and present here a more detailed derivation

of it.

The full block-tridiagonalization (where N is the dimension of the RPA $\mathbf{A}$ and $\mathbf{B}$ matrices) facilitated by this flavor of block Lanczos method can be expressed as:

$$\tilde{\mathbf{Q}}^{(N)T}\mathbf{E}\mathbf{Q}^{(N)} = \mathbf{T}^{(N)} = \begin{pmatrix} \mathbf{M}_1 & \tilde{\mathbf{C}}_1^T & & \cdots & 0 \\ \mathbf{C}_1 & \mathbf{M}_2 & \ddots & & \vdots \\ & \ddots & \ddots & \ddots & \\ \vdots & & \ddots & \ddots & \tilde{\mathbf{C}}_{N-1}^T \\ 0 & \cdots & & \mathbf{C}_{N-1} & \mathbf{M}_N \end{pmatrix} \tag{9}$$

$$\mathbf{Q}^{(N)} = [\mathbf{Q}_1|...|\mathbf{Q}_N] \qquad \tilde{\mathbf{Q}}^{(N)} = [\tilde{\mathbf{Q}}_1|...|\tilde{\mathbf{Q}}_N] \qquad \mathbf{Q}_i, \tilde{\mathbf{Q}}_i \in \mathbb{R}^{2N\times 2}$$

$$\mathbf{Q}^{(N)}, \tilde{\mathbf{Q}}^{(N)}, \mathbf{E} \in \mathbb{R}^{2N\times 2N} \qquad \mathbf{M}_i, \mathbf{C}_i, \tilde{\mathbf{C}}_i \in \mathbb{R}^{2\times 2}$$

The vectors in the left and right Lanczos blocks of vectors, $\tilde{\mathbf{Q}}_i$ and $\mathbf{Q}_i$, are constructed to have the same symmetry as the left and right RPA (excitation and de-excitation) eigenvectors, this way enforcing the RPA paired structure in the Lanczos basis:

$$\mathbf{Q}_i \equiv \begin{pmatrix} \mathbf{X}_i & \mathbf{Y}_i \\ \mathbf{Y}_i & \mathbf{X}_i \end{pmatrix} \qquad \tilde{\mathbf{Q}}_i \equiv \begin{pmatrix} \mathbf{X}_i & -\mathbf{Y}_i \\ -\mathbf{Y}_i & \mathbf{X}_i \end{pmatrix} \qquad \mathbf{X}_i, \mathbf{Y}_i \in \mathbb{R}^{N\times 1} \tag{10}$$

Where, since the initial Lanczos blocks, $\mathbf{Q}_1$ and $\tilde{\mathbf{Q}}_1$, must be bi-orthonormal i.e. $\tilde{\mathbf{Q}}_1^T\mathbf{Q}_1 = \mathbf{I}^{(2)}$, the normalization of the first (and each of the following) Lanczos vectors is defined as $\mathbf{X}_1^T\mathbf{X}_1 - \mathbf{Y}_1^T\mathbf{Y}_1 = 1$. Stopping after k iterations yields a block-tridiagonal matrix $\mathbf{T}^{(k)}$ which will have an eigenspectrum that is an approximation of that of $\mathbf{E}$. As a result of the symmetry present in the $\mathbf{E}$ and the Lanczos blocks of vectors, there is some redundancy in the recursion relations for the linear transformations carried out in the block Lanczos algorithm. Due to the bi-orthonormality requirement between Lanczos blocks, $\tilde{\mathbf{Q}}_i^T\mathbf{Q}_j = \delta_{ij}\mathbf{I}^{(2)}$, the recursion

for creating the new right and left Lanczos blocks, $\mathbf{R}_k$ and $\mathbf{S}_k$, at iteration k is:

$$\mathbf{R}_k = \mathbf{E}\mathbf{Q}_k = \begin{pmatrix} \mathbf{A}\mathbf{X}_k + \mathbf{B}\mathbf{Y}_k & \mathbf{A}\mathbf{Y}_k + \mathbf{B}\mathbf{X}_k \\ -\mathbf{B}\mathbf{X}_k - \mathbf{A}\mathbf{Y}_k & -\mathbf{B}\mathbf{Y}_k - \mathbf{A}\mathbf{X}_k \end{pmatrix} \equiv \begin{pmatrix} \mathbf{X}'_{k+1} & -\mathbf{Y}'_{k+1} \\ \mathbf{Y}'_{k+1} & -\mathbf{X}'_{k+1} \end{pmatrix} \tag{11}$$

$$\mathbf{S}_k = \mathbf{E}^T\tilde{\mathbf{Q}}_k = \begin{pmatrix} \mathbf{A}\mathbf{X}_k + \mathbf{B}\mathbf{Y}_k & -\mathbf{A}\mathbf{Y}_k - \mathbf{B}\mathbf{X}_k \\ \mathbf{B}\mathbf{X}_k + \mathbf{A}\mathbf{Y}_k & -\mathbf{B}\mathbf{Y}_k - \mathbf{A}\mathbf{X}_k \end{pmatrix} \equiv \begin{pmatrix} \mathbf{X}'_{k+1} & \mathbf{Y}'_{k+1} \\ -\mathbf{Y}'_{k+1} & -\mathbf{X}'_{k+1} \end{pmatrix} \tag{12}$$

Notice that the Lanczos vectors in $\mathbf{R}_k$ and $\mathbf{S}_k$ are linearly dependent. Next, the new diagonal block of the block-tridiagonal matrix is defined as:

$$\mathbf{M}_k = \tilde{\mathbf{Q}}_k^T\mathbf{E}\mathbf{Q}_k = \tilde{\mathbf{Q}}_k^T\mathbf{R}_k \equiv \begin{pmatrix} e_k & d_k \\ -d_k & -e_k \end{pmatrix} \tag{13}$$

The components of the previous two Lanczos blocks are then projected out from the new right and left Lanczos block:

$$\mathbf{Q}_{k+1}\mathbf{C}_k = \mathbf{R}'_k = \mathbf{R}_k - \mathbf{Q}_k\mathbf{M}_k - \mathbf{Q}_{k-1}\tilde{\mathbf{C}}_{k-1}^T \tag{14}$$

$$\tilde{\mathbf{Q}}_{k+1}\tilde{\mathbf{C}}_k = \mathbf{S}'_k = \mathbf{S}_k - \tilde{\mathbf{Q}}_k\mathbf{M}_k^T - \tilde{\mathbf{Q}}_{k-1}\mathbf{C}_{k-1}^T \tag{15}$$

Knowing the banded form of $\mathbf{T}^{(N)}$, the super- and sub-diagonal blocks of coefficients, $\tilde{\mathbf{C}}_{k-1}^T$ and $\mathbf{C}_{k-1}$, can be shown to be equal to each other and are defined as:

$$\tilde{\mathbf{C}}_{k-1}^T = \tilde{\mathbf{Q}}_{k-1}^T\mathbf{E}\mathbf{Q}_k = \tilde{\mathbf{Q}}_{k-1}^T\mathbf{R}_k = \tag{16}$$

$$\begin{pmatrix} \mathbf{X}_{k-1}^T\mathbf{X}'_{k+1} - \mathbf{Y}_{k-1}^T\mathbf{Y}'_{k+1} & -\mathbf{X}_{k-1}^T\mathbf{Y}'_{k+1} + \mathbf{Y}_{k-1}^T\mathbf{X}'_{k+1} \\ -\mathbf{Y}_{k-1}^T\mathbf{X}'_{k+1} + \mathbf{X}_{k-1}^T\mathbf{Y}'_{k+1} & \mathbf{Y}_{k-1}^T\mathbf{Y}'_{k+1} - \mathbf{X}_{k-1}^T\mathbf{X}'_{k+1} \end{pmatrix} \equiv \begin{pmatrix} a_{k-1} & b_{k-1} \\ -b_{k-1} & -a_{k-1} \end{pmatrix}$$

$$\mathbf{C}_{k-1} = (\mathbf{Q}_{k-1}^T\mathbf{E}^T\tilde{\mathbf{Q}}_k)^T = (\mathbf{Q}_{k-1}^T\mathbf{S}_k)^T = \tilde{\mathbf{C}}_{k-1}^T \tag{17}$$

Explicitly writing out Eqs. (14) and (15), it is easy to confirm that the symmetry between

Lanczos vectors in $\mathbf{R}_k$ and $\mathbf{S}_k$ in Eqs. (11)–(12) is left unchanged. The new unnormalized Lanczos blocks, $\mathbf{R}'_k$ and $\mathbf{S}'_k$, thus contain only one linearly independent Lanczos vector. It is therefore sufficient to only execute the recursion for one of them, say the first vector in the new unnormalized $\mathbf{Q}_{k+1}$:

$$(\mathbf{R}'_k)_1 \equiv \begin{pmatrix} \mathbf{X}''_{k+1} \\ \mathbf{Y}''_{k-1} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'_{k+1} - e_k\mathbf{X}_k + d_k\mathbf{Y}_k - a_{k-1}\mathbf{X}_{k-1} + b_{k-1}\mathbf{Y}_{k-1} \\ \mathbf{Y}'_{k+1} - e_k\mathbf{Y}_k + d_k\mathbf{X}_k - a_{k-1}\mathbf{Y}_{k-1} + b_{k-1}\mathbf{X}_{k-1} \end{pmatrix} \tag{18}$$

so that the three remaining Lanczos vectors are handled implicitly, while still retaining the benefits of the block Lanczos approach. To normalize a new Lanczos vector, the inverses of $\tilde{\mathbf{C}}_k$ and $\mathbf{C}_k$ need to be found. This is done by examining their product prior to normalization, i.e. product of $\mathbf{S}'^T_k$ and $\mathbf{R}'_k$:

$$\mathbf{S}'^T_k\mathbf{R}'_k = \tilde{\mathbf{C}}^T_k\mathbf{C}_k = \begin{pmatrix} a_k^2 - b_k^2 & 0 \\ 0 & a_k^2 - b_k^2 \end{pmatrix} = \begin{pmatrix} \mathcal{N}_{k+1} & 0 \\ 0 & \mathcal{N}_{k+1} \end{pmatrix} \tag{19}$$

As the values of a_k and b_k are not uniquely defined, the easiest choice is to always set one of them to zero. Then, looking at the product of coefficient matrices $\tilde{\mathbf{C}}^T_k\mathbf{C}_k$ in Eq. (19), it is obvious that if a_k is set to zero, $\mathcal{N}_{k+1}$ will be negative, and if b_k is set to zero, one has a positive $\mathcal{N}_{k+1}$. The scheme for setting the coefficients in the coefficient block matrices is therefore:

$$\begin{aligned} \mathcal{N}_{k+1} > 0: &\quad a_k = \sqrt{\mathcal{N}_{k+1}}\,, \qquad b_k = 0 \\ \mathcal{N}_{k+1} < 0: &\quad b_k = \sqrt{-\mathcal{N}_{k+1}}\,, \quad a_k = 0 \end{aligned} \tag{20}$$

The $\mathcal{N}_{k+1}$ is the square of normalization constant for the $(k+1)^{th}$ Lanczos blocks. Since the recursion in this implicit algorithm is performed for only one Lanczos vector, $(\mathbf{R}'_k)_1$, a negative $\mathcal{N}_{k+1}$ means that a new right Lanczos block, $\mathbf{Q}_{k+1}$, is normalized with an antidiagonal matrix $\mathbf{C}_k^{-1}$. This in turn means that the upper and lower components of each Lanczos

Algorithm 1: Lanczos for RPA

Data: $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{N\times N}$, the excitation part of the property gradient vector as a starting vector ${}^{e}\mathbf{P}(\hat{O}) \in \mathbb{R}^{N\times 1}$, k_{max} Lanczos chain size (user specified)

Result: $\mathbf{A}'^{(k_{max})}$, $\mathbf{B}'^{(k_{max})} \in \mathbb{R}^{k_{max}\times k_{max}}$, $\mathbf{U}^{(k_{max})} = [\mathbf{X}_1|..|\mathbf{X}_{k_{max}}]$, $\mathbf{V}^{(k_{max})} = [\mathbf{Y}_1|..|\mathbf{Y}_{k_{max}}] \in \mathbb{R}^{N\times k_{max}}$

begin

$\mathbf{X}_1 = {}^{e}\mathbf{P}(\hat{O})/||{}^{e}\mathbf{P}(\hat{O})||_2$

$\mathbf{Y}_1 = \mathbf{0}$

$k = 1$

while $k < k_{max}$ **do**

$\mathbf{X}'_{k+1} = \mathbf{A}\mathbf{X}_k + \mathbf{B}\mathbf{Y}_k$

$\mathbf{Y}'_{k+1} = -\mathbf{B}\mathbf{X}_k - \mathbf{A}\mathbf{Y}_k$

$e_k = \mathbf{X}^T_{k+1}\mathbf{X}_k - \mathbf{Y}^T_{k+1}\mathbf{Y}_k$

$d_k = \mathbf{X}_{k+1}\mathbf{Y}_k - \mathbf{Y}_{k+1}\mathbf{X}_k$

$\mathbf{X}'_{k+1} = \mathbf{X}'_{k+1} - e_k\mathbf{X}_k + d_k\mathbf{Y}_k - a_{k-1}\mathbf{X}_{k-1} + b_{k-1}\mathbf{Y}_{k-1}$

$\mathbf{Y}'_{k+1} = \mathbf{Y}'_{k+1} - e_k\mathbf{Y}_k + d_k\mathbf{X}_k - a_{k-1}\mathbf{Y}_{k-1} + b_{k-1}\mathbf{X}_{k-1}$

Bi-orthogonalize $\mathbf{X}_{k+1}$, $\mathbf{Y}_{k+1}$ against all previous $\mathbf{X}_k$, $\mathbf{Y}_k$

$\mathcal{N}_{k+1} = \mathbf{X}'^T_{k+1}\mathbf{X}'_{k+1} - \mathbf{Y}'^T_{k+1}{}'\mathbf{Y}'_{k+1}$

if $|\mathcal{N}_{k+1}| < tol$ **then**

stop

if $\mathcal{N}_{k+1} > 0$ **then**

$a_k = \sqrt{\mathcal{N}_{k+1}}$

$b_k = 0$

else

$a_k = 0$

$b_k = \sqrt{|\mathcal{N}_{k+1}|}$

$\mathbf{X}'_{k+1} = -\mathbf{Y}'_{k+1}$

$\mathbf{Y}'_{k+1} = -\mathbf{X}'_{k+1}$

$\mathbf{X}_{k+1} = \mathbf{X}'_{k+1}/\sqrt{|\mathcal{N}_{k+1}|}$

$\mathbf{Y}_{k+1} = \mathbf{Y}'_{k+1}/\sqrt{|\mathcal{N}_{k+1}|}$

$k = k + 1$

vector are swapped and $\mathbf{Q}_{k+1}$ has an overall minus sign as compared to the definition of all $\mathbf{Q}_i$ in Eq. (10). Therefore, during the iterations, when a negative $\mathcal{N}_{k+1}$ is found, we can change the sign and flip the Lanczos vector components, this way only accumulating $(\mathbf{Q}_k)_1$, from which the rest of the Lanczos vectors can be easily computed if needed. Such implicit handling of the pairs of Lanczos vectors effectively reduces the computational costs of one iteration by 75 %, as compared with general nonsymmetric or nonsymmetric block Lanczos algorithms that are not adapted for RPA. The full Lanczos for RPA algorithm is presented in Algorithm 1.

Finally, stopping after k iterations, the banded $\bar{\mathbf{T}}^{(k)}$ matrix can be reordered to have the same symmetry as the RPA $\mathbf{E}$ matrix (see Section S2).

$$\begin{pmatrix} \mathbf{U}^{(k)T} & -\mathbf{V}^{(k)T} \\ -\mathbf{V}^{(k)T} & \mathbf{U}^{(k)T} \end{pmatrix} \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ -\mathbf{B} & -\mathbf{A} \end{pmatrix} \begin{pmatrix} \mathbf{U}^{(k)} & \mathbf{V}^{(k)} \\ \mathbf{V}^{(k)} & \mathbf{U}^{(k)} \end{pmatrix} = \begin{pmatrix} \mathbf{A}'^{(k)} & \mathbf{B}'^{(k)} \\ -\mathbf{B}'^{(k)} & -\mathbf{A}'^{(k)} \end{pmatrix} = \bar{\mathbf{T}}^{(k)} \qquad (21)$$

$$\mathbf{U}^{(k)} = [\mathbf{X}_1, .., \mathbf{X}_k] \qquad \mathbf{V}^{(k)} = [\mathbf{Y}_1, .., \mathbf{Y}_k]$$
$$\mathbf{A}, \mathbf{B} \in \mathbb{R}^{N\times N} \qquad \mathbf{U}^{(k)}, \mathbf{V}^{(k)} \in \mathbb{R}^{N\times k} \qquad \mathbf{A}'^{(k)}, \mathbf{B}'^{(k)} \in \mathbb{R}^{k\times k}$$

where each i^{th} column in the matrices $\mathbf{U}^{(k)}$ and $\mathbf{V}^{(k)}$ is, respectively, the upper and lower parts of the first vector in the i^{th} right Lanczos block as seen in Eq. (10). Alike all nonsymmetric Lanczos algorithms, this one will also suffer a "serious break-down" if the product $\mathbf{S}_k'^T\mathbf{R}'_k$ is singular or near-singular. The iterations can also break down if the norm of the new Lanczos vector is zero, which just means that an invariant subspace for the $\mathbf{E}$ matrix has been found. In our study we only observed it to occur when using the smallest basis sets. The algorithm is prone to loss of (bi-)orthogonality of Lanczos vectors due to round-off errors in the floating-point arithmetic, which can result in appearance of the ghost eigenvalues.[45–47] Each new Lanczos vector is therefore explicitly re-orthogonalized to all the previous right

and left Lanczos vectors (of excitation and de-excitation symmetry):

$$\begin{pmatrix} \mathbf{X}''_{k+1} \\ \mathbf{Y}''_{k+1} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'_{k+1} \\ \mathbf{Y}'_{k+1} \end{pmatrix} - \mathbf{Q}^{(k)} \tilde{\mathbf{Q}}^{(k)T} \begin{pmatrix} \mathbf{X}'_{k+1} \\ \mathbf{Y}'_{k+1} \end{pmatrix} \tag{22}$$

2.1.1 Sum-over-states Properties in Lanczos Basis

One can compute molecular properties directly in the Lanczos basis, when the starting vector contains the property gradient vector of the property in question. This was shown to hold for electronic properties by Coriani *et al.,*[16,26] and for vibrational properties by Christiansen and coworkers,[17,50] under the assumption that the non-linear contributions can be neglected in the coupled cluster (CC) case. They would otherwise require a back-transformation to full space. In Refs. 16,26, expressions were derived for the (asymmetric Lanczos-based) Coupled Cluster (CC) linear response function and transition strengths, and for the EOM-CCSD transition strengths, respectively. The same approach can be easily taken in the RPA case. Moreover, using the property gradient vector as a starting vector can be expected to better converge the eigenmodes corresponding to the transitions that give the largest contributions to the mean excitation energy, as these transitions have the largest overlap with the gradient property vector. As the oscillator strength sums require square norms of the electronic dipole transition moments, as expressed in Eq. (7), the iterations are started with Lanczos blocks containing the dipole moment gradient property vectors, $\mathbf{P}^R(\hat{\boldsymbol{\mu}})$ and $\mathbf{P}^L(\hat{\boldsymbol{\mu}})^T$, as defined in Eqs. (4) and (5):

$$\mathbf{Q}_1 = \frac{1}{n_\mu} \begin{pmatrix} {}^e\mathbf{P}(\hat{\boldsymbol{\mu}}) & \mathbf{0} \\ \mathbf{0} & {}^e\mathbf{P}(\hat{\boldsymbol{\mu}}) \end{pmatrix} \qquad \tilde{\mathbf{Q}}_1^T = \frac{1}{n_\mu} \begin{pmatrix} {}^e\mathbf{P}^T(\hat{\boldsymbol{\mu}}) & \mathbf{0} \\ \mathbf{0} & {}^e\mathbf{P}^T(\hat{\boldsymbol{\mu}}) \end{pmatrix} \qquad {}^e\mathbf{P}(\hat{\boldsymbol{\mu}}) \in \mathbb{R}^{N\times 1} \tag{23}$$

Here, n_μ is the Euclidean norm of the excitation part of the dipole moment gradient property vector. Since the Lanczos blocks are saved in the manner as expressed in Eq. (21), the two vectors from each of the first Lanczos blocks, $\mathbf{Q}_1$ and $\tilde{\mathbf{Q}}_1^T$, will be the 1^{st} and $(k+1)^{th}$

columns and rows in the $\mathbf{Q}^{(k)}$ and $\tilde{\mathbf{Q}}^{T(k)}$ matrices, respectively. Thus the (right and left) dipole moment gradient property vector can be expressed through the matrices of Lanczos blocks:

$$\mathbf{P}^R(\hat{\boldsymbol{\mu}}) = n_\mu \mathbf{Q}^{(k)}(\mathbf{e}_1^{(k)} + \mathbf{e}_{k+1}^{(k)}) \qquad \mathbf{P}^L(\hat{\boldsymbol{\mu}})^T = n_\mu (\mathbf{e}_1^{(k)} - \mathbf{e}_{k+1}^{(k)})^T \tilde{\mathbf{Q}}^{(k)T} \tag{24}$$
$$\mathbf{e}_1^{(k)}, \mathbf{e}_{k+1}^{(k)} \in \mathbb{R}^{2k\times 1}$$

Here, $\mathbf{e}_1^{(k)}$ and $\mathbf{e}_{k+1}^{(k)}$ are unit basis vectors where, respectively, i.e. the 1^{st} or the $(k+1)^{th}$ elements are set to 1, and all other elements are zero. It can be shown that the propagator matrix, $(\mathbf{E} - \omega\mathbf{1})^{-1}$, can be expressed in the diagonal Lanczos basis (assuming that the truncation error in similarity transformation of $\mathbf{E}$ to $\bar{\mathbf{T}}^{(k)}$ is small enough):

$$(\mathbf{E} - \omega\mathbf{1})^{-1} \overset{(21)}{=} (\mathbf{Q}^{(k)}\bar{\mathbf{T}}^{(k)}\tilde{\mathbf{Q}}^{(k)T} - \omega\mathbf{1})^{-1} = (\mathbf{Q}^{(k)}\mathbf{R}\boldsymbol{\Lambda}\mathbf{L}\tilde{\mathbf{Q}}^{(k)T} - \omega\mathbf{1})^{-1} =$$
$$\mathbf{Q}^{(k)}\mathbf{R}(\boldsymbol{\Lambda} - \omega\mathbf{1})^{-1}\mathbf{L}\tilde{\mathbf{Q}}^{(k)T} \tag{25}$$

$\boldsymbol{\Lambda}$ is a diagonal matrix with the eigenvalues of $\bar{\mathbf{T}}^{(k)}$ on the diagonal, while the rows of $\mathbf{L}$ and the columns of $\mathbf{R}$ matrices are the left and right eigenvectors of $\bar{\mathbf{T}}^{(k)}$, and the fact that Lanczos vectors and eigenvectors make up bi-orthonormal bases, i.e. $\tilde{\mathbf{Q}}^{(k)T}\mathbf{Q}^{(k)} = \mathbf{1}^{(k)}$ and $\mathbf{LR} = \mathbf{1}$, was used. Inserting Eqs. (24) and (25) into Eq. (7) yields:

$$|\langle n|\hat{\boldsymbol{\mu}}|0\rangle|^2 = \lim_{\omega\to\omega_n}(\omega - \omega_n)(\mathbf{P}^L(\hat{\boldsymbol{\mu}}))^T(\mathbf{E} - \omega\mathbf{1})^{-1}\mathbf{P}^R(\hat{\boldsymbol{\mu}})$$
$$= \lim_{\omega\to\omega_n}(\omega - \omega_n)\left[n_\mu^2(\mathbf{e}_1^{(k)} - \mathbf{e}_{k+1}^{(k)})^T\tilde{\mathbf{Q}}^{(k)T}\mathbf{Q}^{(k)}\mathbf{R}(\boldsymbol{\Lambda} - \omega\mathbf{1})^{-1}\mathbf{L}\tilde{\mathbf{Q}}^{(k)T}\mathbf{Q}^{(k)}(\mathbf{e}_1^{(k)} + \mathbf{e}_{k+1}^{(k)})\right]$$
$$\overset{\bar{\omega}_n\approx\omega_n}{=} n_\mu^2(\mathbf{e}_1^{(k)} - \mathbf{e}_{k+1}^{(k)})^T\mathbf{R}_n\mathbf{L}_n(\mathbf{e}_1^{(k)} + \mathbf{e}_{k+1}^{(k)}) = n_\mu^2(R_{1,n} - R_{k+1,n})(L_{n,1} + L_{n,k+1}) \tag{26}$$

Here, $\mathbf{R}_n$ and $\mathbf{L}_n$ are the right and left eigenvectors of $\bar{\mathbf{T}}^{(k)}$ corresponding to the n^{th} excitation, and it is assumed that the n^{th} eigenvalue of $\mathbf{E}$ is approximated well enough by $\bar{\omega}_n$ (eigenvalue of $\bar{\mathbf{T}}^{(k)}$). As it is now clear, only the dipole moment gradient property vector and the 1^{st}

and $(k+1)^{th}$ elements of the corresponding left and right eigenvectors of the Lanczos $\bar{\mathbf{T}}^{(k)}$ matrix are needed in order to compute $|\langle n|\hat{\boldsymbol{\mu}}|0\rangle|^2$.

2.1.2 Preservation of Sum Rules

Johnson *et al.*[32] have proven that the Lanczos algorithm for the RPA also preserves certain energy-weighted sum-rules of the linear response transitions. In Section S1, we present a detailed derivation of this property, generalized for all versions of nonsymmetric Lanczos algorithm as applied to the RPA matrix $\mathbf{E}$. A brief overview is given in the following.

As the full set of the RPA eigenvectors form a complete basis, $\mathbf{E}$ can be expanded in it:

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ -\mathbf{B} & -\mathbf{A} \end{pmatrix} = \sum_n^N \omega_n \left[\begin{pmatrix} {}^e\tilde{\mathbf{X}}_n \\ {}^d\tilde{\mathbf{X}}_n \end{pmatrix} \begin{pmatrix} {}^e\tilde{\mathbf{X}}_n & -{}^d\tilde{\mathbf{X}}_n \end{pmatrix} - \begin{pmatrix} {}^d\tilde{\mathbf{X}}_n \\ {}^e\tilde{\mathbf{X}}_n \end{pmatrix} \begin{pmatrix} -{}^d\tilde{\mathbf{X}}_n & {}^e\tilde{\mathbf{X}}_n \end{pmatrix} \right] \tag{27}$$

where the sum is over the positive eigenvalues only. It can be easily seen that multiplying the RPA matrix in this expression with itself would change the sign inside the square brackets. Generalizing for the RPA matrix to some power m then yields:

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ -\mathbf{B} & -\mathbf{A} \end{pmatrix}^m = \sum_n^N \omega_n^m \left[\begin{pmatrix} {}^e\tilde{\mathbf{X}}_n \\ {}^d\tilde{\mathbf{X}}_n \end{pmatrix} \begin{pmatrix} {}^e\tilde{\mathbf{X}}_n & -{}^d\tilde{\mathbf{X}}_n \end{pmatrix} + (-1)^m \begin{pmatrix} {}^d\tilde{\mathbf{X}}_n \\ {}^e\tilde{\mathbf{X}}_n \end{pmatrix} \begin{pmatrix} -{}^d\tilde{\mathbf{X}}_n & {}^e\tilde{\mathbf{X}}_n \end{pmatrix} \right] \tag{28}$$

and so for odd powers of m the term in the square brackets will have a minus sign. Computing some sum-over-states property, $M(m)$, using the appropriate property operator $\hat{O}$, the expression usually takes such form:

$$M(m) = \sum_n \omega_n^m |\langle n|\hat{O}|0\rangle|^2 \tag{29}$$

For RPA, inserting Eq. (8) into Eq. (29), yields:

$$M(m) = \sum_{n}^{N} \omega_n^m (\mathbf{P}^L(\hat{O}))^T ({}^e\tilde{\mathbf{X}}_n^R)({}^e\tilde{\mathbf{X}}_n^L)\mathbf{P}^R(\hat{O}) \tag{30}$$
$$= \sum_{n}^{N} (\mathbf{P}^L(\hat{O}))^T \begin{pmatrix} {}^e\tilde{\mathbf{X}}_n \\ {}^d\tilde{\mathbf{X}}_n \end{pmatrix} \omega_n^m \begin{pmatrix} {}^e\tilde{\mathbf{X}}_n & -{}^d\tilde{\mathbf{X}}_n \end{pmatrix} \mathbf{P}^R(\hat{O})$$

As the product of the gradient property vector and the RPA de-excitation eigenvector can be expressed through the analogous product with the RPA excitation eigenvector, Eq. (30) can be rewritten as an expansion in a linear combination of the RPA excitation and de-excitation eigenvectors:

$$M(m) = \tag{31}$$
$$\frac{1}{2}(\mathbf{P}^L(\hat{O}))^T \sum_{n}^{N} \omega_n^m \left[\begin{pmatrix} {}^e\tilde{\mathbf{X}}_n \\ {}^d\tilde{\mathbf{X}}_n \end{pmatrix} \begin{pmatrix} {}^e\tilde{\mathbf{X}}_n & -{}^d\tilde{\mathbf{X}}_n \end{pmatrix} - \begin{pmatrix} {}^d\tilde{\mathbf{X}}_n \\ {}^e\tilde{\mathbf{X}}_n \end{pmatrix} \begin{pmatrix} -{}^d\tilde{\mathbf{X}}_n & {}^e\tilde{\mathbf{X}}_n \end{pmatrix} \right] \mathbf{P}^R(\hat{O})$$

Recognizing Eq. (28) here for odd m, a final expression for the sum-over-states properties energy-weighted to an odd power m can be written as:

$$M(m) = \frac{1}{2}(\mathbf{P}^L(\hat{O}))^T \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ -\mathbf{B} & -\mathbf{A} \end{pmatrix}^m \mathbf{P}^R(\hat{O}) = \frac{1}{2}(\mathbf{P}^L(\hat{O}))^T \mathbf{E}^m \mathbf{P}^R(\hat{O}) \tag{32}$$

The sum of the oscillator strengths, $S(0)$, needed to compute the mean excitation energy, fulfills the odd power requirement[40] and can be written as[40] $S(0) = \frac{2}{3}M(1)$. It can then be expressed as:

$$S(0) = \frac{1}{3}(\mathbf{P}^L(\hat{O}))^T \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ -\mathbf{B} & -\mathbf{A} \end{pmatrix} \mathbf{P}^R(\hat{O}) = \frac{1}{3}(\mathbf{P}^L(\hat{O}))^T \mathbf{E}\mathbf{P}^R(\hat{O}) \tag{33}$$

If the Lanczos algorithm is initiated with (normalized) property gradient vectors (or a set of

vectors that are linear combinations of them) as the first Lanczos vector or block, and if the RPA matrix $\mathbf{E}$ is expressed through a general projection into the Lanczos subspace of size i, which leaves the right and left gradient property vectors unchanged, we observe that the Lanczos transformation does not change the expression of $S(0)$ in Eq. (33):

$$S(0) \approx \frac{1}{3}(\mathbf{P}^L(\hat{O}))^T \mathbf{Q}^{(i)} \tilde{\mathbf{Q}}^{(i)T} \mathbf{E} \mathbf{Q}^{(i)} \tilde{\mathbf{Q}}^{(i)T} \mathbf{P}^R(\hat{O}) = \frac{1}{3}(\mathbf{P}^L(\hat{O}))^T \mathbf{E} \mathbf{P}^R(\hat{O}) \tag{34}$$

This means that, for $m = 1$, the RPA sum-over-states properties will be preserved by the algorithm already after the first iteration. Using the aforementioned projection and remembering the span of Krylov subspaces, this can be generalized to all odd m. The nonsymmetric Lanczos algorithms, when initiated with vector(-s) containing the relevant gradient property vector, will preserve the odd-m-power energy-weighted sum-over-states properties after $\frac{m+1}{2}$. Note that in quantum chemistry the first power of the excitation energy is usually absorbed in the oscillator strengths, thus in this notation the preservation of RPA sum-over-states properties by Lanczos algorithms will be valid for the even-m-weighted sums.

3 Implementation Details

Based on the algorithm presented in this work, a block Lanczos RPA solver was implemented in a development version of Dalton,[10,11] as a part of the atomic integral direct (AO) SOPPA module.[51] The solver computes $S(0)$, $L(0)$ and $I(0)$ for a user-chosen subspace size. The main driver of this solver generates the arrays needed to store these, and loops over the components of the dipole moment operator. Here, a routine is called to compute the initial vector, where the upper part is the excitation part of the electric dipole moment property gradient vector and the lower part is zeros, as well as the norm of the said gradient property vector, $||^e\mathbf{P}(\hat{\mu}^{label})||_2$. The matrix–vector products, $\mathbf{AX}_k$, $\mathbf{AY}_k$, $\mathbf{BX}_k$ and $\mathbf{BY}_k$, are computed with the existing (atomic integral driven) routine. The Lanczos recursion is then performed in a routine, which also returns the elements of the Lanczos matrix $\bar{\mathbf{T}}$, as well

as a newly computed and normalized Lanczos vector, which has also been bi-orthogonalized to all the previous ones at least once. A check for break-down is executed after each iteration by comparing the absolute of the "square norm", $|\mathcal{N}_{k+1}|$, of a new Lanczos vector to the threshold (currently set to `1.0e-12`). The vector is written to disk if $|\mathcal{N}_{k+1}|$ is larger than the threshold. After the desired (user specified) number of Lanczos vectors has been generated or a break-down has been encountered, the Lanczos matrix $\bar{\mathbf{T}}$ is diagonalized. Its eigenvalues are then sorted in ascending order, and the eigenvectors are bi-orthonormalized. The eigenvectors are afterwards used to compute the electric dipole moment transition strengths in the Lanczos basis, which in turn are used to compute $S(0)$, $L(0)$ and $I(0)$ (eV) for the current component. The solver can perform intermediate diagonalizations/property calculations at user chosen intervals (say, every 100 iterations). In lieu of the full space values to compare to, this feature could be used to give a clue of whether the $L(0)$ and $I(0)$ values have converged at the chosen subspace size, k_{max} (see Algorithm 1).

4 Computational Details

The Implicit Block Lanczos RPA solver was used to compute sum of dipole oscillator strengths, $S(0)$, and mean excitation energies, $I(0)$, for a set of 18 small molecules, namely AlH, BF, BH, BH_3, C_2H_2, CO, CO_2, F_2, H_2CO, H_2O, HCl, HCN, HF, LiH, N_2, N_2O, NaH, PH_3, and the Neon atom. The geometries were previously optimized using the correlation-consistent cc-pVTZ basis[52,53] at the CCSD(T) level of theory.[7,54–56] Since both core-valence and augmentation functions are needed for this property to converge with respect to the basis set,[6] correlation-consistent core-valence double zeta to pentuple zeta quality basis sets, (aug-)cc-pCVnZ, where n = D, T, Q, 5,[52,53,57,58] both augmented and not, were used to simultaneously evaluate the performance of the basis sets. The values at some iteration k, $S(0)^{(k)}$ and $I(0)^{(k)}$, were compared to full space values, $S_{diag}(0)$ and $I_{diag}(0)$ (or $S_{2N}(0)$ and $I_{2N}(0)$, where $N \times N$ are the dimensions of the RPA $\mathbf{A}$ and $\mathbf{B}$ matrices), in order to deter-

mine if and to what degree these properties have converged (the full set of data supporting our conclusions can be found at the link given in the Data Availability section).

5 Results and Discussion

Lanczos $I(0)$: BH and PH_3 Molecules

As discussed above, the Lanczos algorithm is mathematically shown to preserve the sum-over-states properties that are weighted by the odd power of excitation energies. In practice, this is confirmed by computing one such property, the sum of oscillator strengths, $S(0)$, where convergence is observed after the first iteration in all computations, and the differences between the subspace and full space values are within the limits of numerical noise (see data found at the link given in the Data Availability section). According to the Thomas-Reiche-Kuhn sum rule,[33–35] if a complete basis is used, the calculated $S(0)$ must reproduce the number of electrons in a system. Let us consider e.g. the BH and PH_3 molecules, i.e systems containing, respectively, 6 and 18 electrons. As seen in Tables 1–2, the aug-cc-pCVnZ basis set series converges within 4 significant figures of the correct value, i.e. to $S_{2N}(0) = 6.00$ and $S_{2N}(0) = 18.00$, at aug-cc-pCV5Z. The corresponding cc-pCVnZ basis set series never converges to more than 2 significant digits, and thus never reaches the complete basis limit, and so we discourage the use of these basis sets. The RPA mean excitation energies computed in this study appear to decrease with the cardinal number of the basis set (i.e. D, T, Q, 5). Adding diffuse functions to a basis set seems to result in monotonic decrease in the $I(0)$ as

Table 1: The full space $S_{2N}(0)$ and $I_{2N}(0)$ (in eV) values computed for BH molecule (containing 6 electrons) with different basis sets.

Basis	$S_{2N}(0)$	$I_{2N}(0)$	Basis	$S_{2N}(0)$	$I_{2N}(0)$
cc-pCVDZ	6.77	71.72	aug-cc-pCVDZ	6.47	64.91
cc-pCVTZ	6.15	48.00	aug-cc-pCVTZ	6.05	47.27
cc-pCVQZ	6.07	46.67	aug-cc-pCVQZ	6.01	46.36
cc-pCV5Z	6.04	46.50	aug-cc-pCV5Z	6.00	46.25

Table 2: The full space $S_{2N}(0)$ and $I_{2N}(0)$ (in eV) values computed for PH_3 molecule (containing 18 electrons) with different basis sets.

Basis	$S_{2N}(0)$	$I_{2N}(0)$	Basis	$S_{2N}(0)$	$I_{2N}(0)$
cc-pCVDZ	19.24	117.16	aug-cc-pCVDZ	18.37	111.10
cc-pCVTZ	18.12	99.44	aug-cc-pCVTZ	17.74	100.14
cc-pCVQZ	18.18	107.63	aug-cc-pCVQZ	17.94	107.45
cc-pCV5Z	18.20	109.66	aug-cc-pCV5Z	18.00	109.68

well, a trend that was previously observed by Sauer *et al.*[6] The outliers for these trends were the molecules that contain a period-3 atom (AlH, NaH, HCl, PH_3), as it is well know that for such systems the Dunning basis set series tend not to converge smoothly.[59,60] Dunning *et al.* have shown that smooth convergence for some properties can be achieved by re-optimizing and expanding the set of d-functions in the basis set. In future work, it could be useful to investigate if this approach would help to recover the basis series here as well.

The full space values, $I_{2N}(0)$, for most of the molecules in this study converged at quadruple-zeta quality basis, as can be seen in Table 1. The $I_{2N}(0)$ computed for the BH molecule using aug-cc-pCVQZ is within 0.2 % of the $I_{2N}(0)$ value at aug-cc-pCV5Z. Therefore, aug-cc-pCVQZ can be used for such calculations, if use of a larger basis set is not computationally feasible.

Even though the expression for $I(0)$ does not fulfill the requirement for preservation of the sum rules discussed earlier, fast convergence of its values was observed in most calculations. This can be appreciated in Figures 1–2, that show the values of $I^{xx}_{diag}(0)$ and $I^{zz}_{diag}(0)$ computed at every 10 new Lanczos vectors with several basis sets for the BH and PH_3 molecules. Almost instantly, i.e. for subspaces corresponding to only a couple percentage points of the full space size, the Lanczos $I(0)$ values were within $\sim$5–10 % of the full space values.

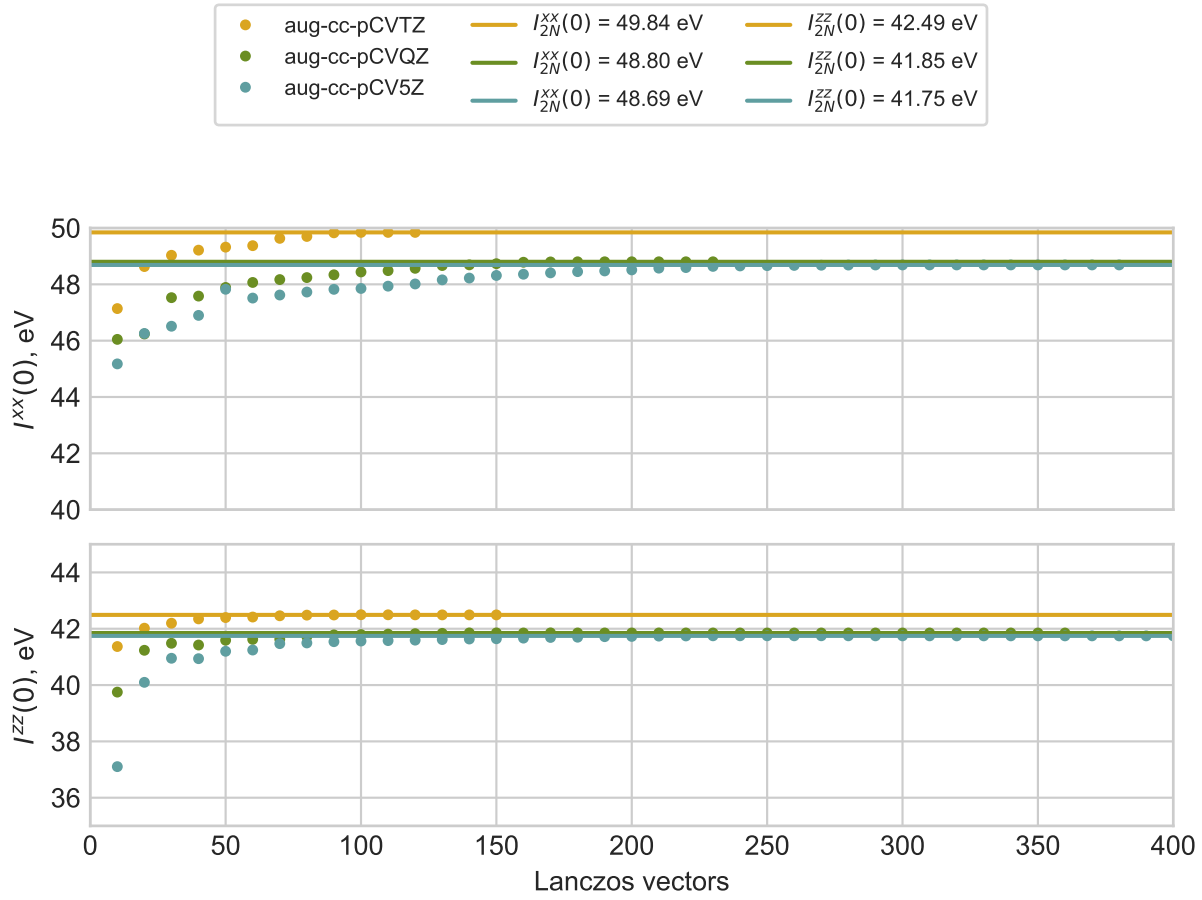

Figure 1: BH. Convergence of $I^{xx}(0)$ and $I^{zz}(0)$ computed with different basis, using the block Lanczos RPA solver in Dalton. The $I^{yy}(0)$ results are identical to those of $I^{xx}(0)$, since x- and y-components transform as the same irreducible representation of the full point group, and are therefore not depicted here. The values compared to, $I^{xx}_{2N}(0)$, $I^{yy}_{2N}(0)$ and $I^{zz}_{2N}(0)$, are computed after diagonalizing the full space electronic Hessian ($2N$ = 378, 378 and 570 for aug-cc-pCV5Z basis and x-, y- and z-components, respectively). Note that 1 Lanczos iteration implicitly generates 2 new Lanczos vectors.

The number of Lanczos vectors needed to converge to the full-space value of the property grows with the size of the basis set. However, there is little, or, in some cases, no increase in the number of Lanczos vectors needed if diffuse basis functions are added (see Tables 3–4 for PH_3 and BH, and Tables S1–S54 in Section S4.1, Supplementary Material, for the rest of the molecules). Qualitatively, this can be expected, since the augmented basis sets include the diffuse functions that can better describe the electronic excitations, i.e. allow the electrons to move further away from the nucleus, as demonstrated in numerous previous

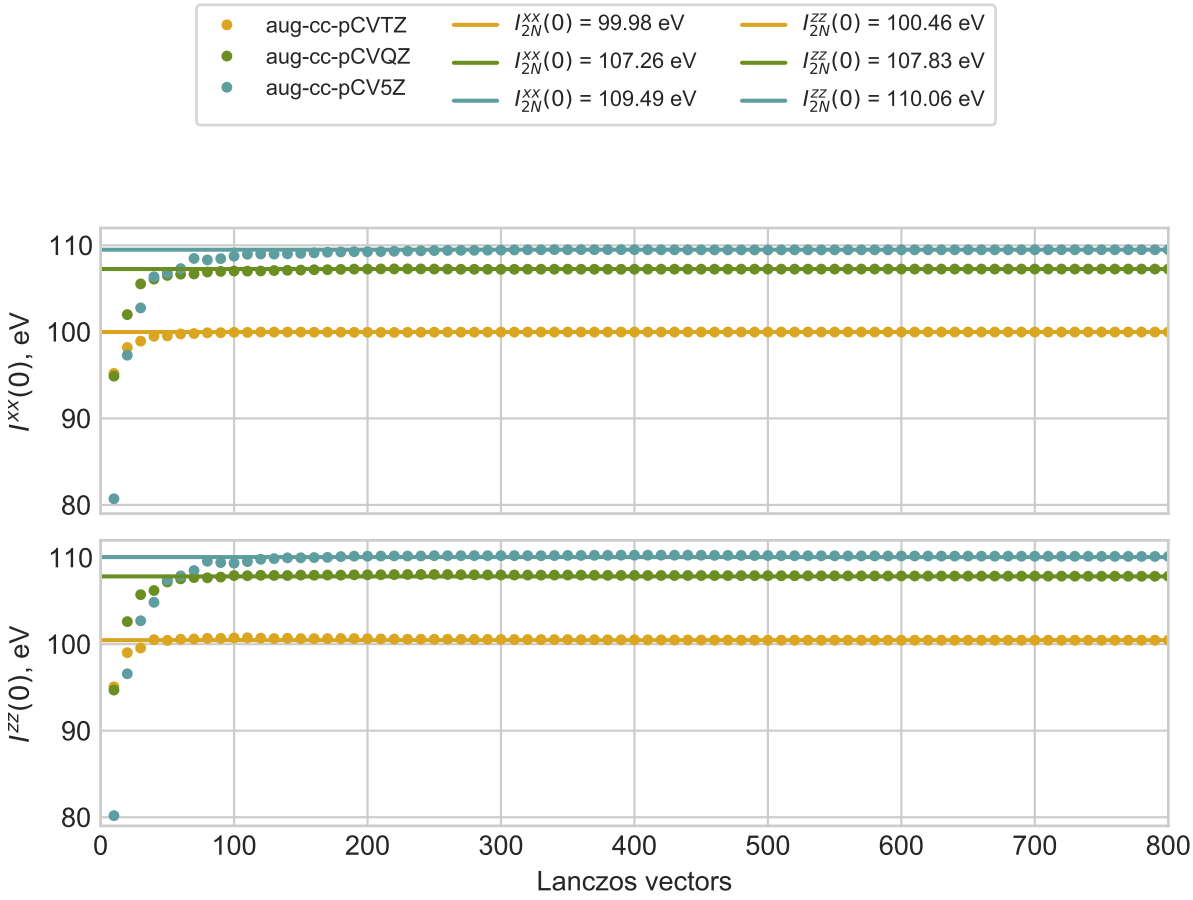

Figure 2: PH_3. Convergence of $I^{xx}(0)$ and $I^{zz}(0)$ computed with different basis, using the block Lanczos RPA solver in Dalton. The $I^{yy}(0)$ results are identical to those of $I^{xx}(0)$, since x- and y-components transform as the same irreducible representation of the full point group, and are therefore not depicted here. The values compared to, $I^{xx}_{2N}(0)$, $I^{yy}_{2N}(0)$ and $I^{zz}_{2N}(0)$, are computed after diagonalizing the full space electronic Hessian ($2N = 3722$, 4324 and 4324 for aug-cc-pCV5Z basis and x-, y- and z-components, respectively). Note that 1 Lanczos iteration implicitly creates 2 new Lanczos vectors.

studies.[6,61,62] Unsurprisingly, the augmented basis sets thus give a more correct description of the electronic excitations, which enter the definition of the $S(0)$ and $I(0)$ expressions (see Eq. (2)). Based on the results obtained here, the aug-cc-pCVQZ basis set emerges as the best compromise between accuracy and computational cost.

The fraction of the full space needed to converge Lanczos $I(0)$ decreases with the augmentation of the basis set and, in most cases, also with the increase in cardinal number, as can be seen in Tables 3–4. This implies that the increase in computational cost when

Table 3: PH_3. Convergence of Lanczos iteration, when computing the x-, y- and z-components of $I(0)$ (in eV) with different basis sets. The results for the x- and y-component are identical since they transform as the same irreducible representation of the full point group. The reference value is the $I(0)$ computed after diagonalizing the full RPA electronic Hessian. Note that the algorithm used here creates 2 Lanczos vectors in one iteration.

Basis	$I^{xx}_{2N}(0)$ $/I^{yy}_{2N}(0)$	Full space size, $2N$	Convergence within 1.0 %, No. Lanczos vectors, (% of full space)	Convergence within 0.5 %, No. Lanczos vectors, (% of full space)
cc-pCVDZ	116.68	252	20 (7.94 %)	30 (11.90 %)
cc-pCVTZ	99.31	728	50 (6.87 %)	80 (10.99 %)
cc-pCVQZ	107.35	1540	50 (3.25 %)	90 (5.84 %)
cc-pCV5Z	109.45	2778	100 (3.60 %)	150 (5.40 %)
aug-cc-pCVDZ	110.86	416	10 (2.40 %)	20 (4.81 %)
aug-cc-pCVTZ	99.98	1080	40 (3.70 %)	40 (3.70 %)
aug-cc-pCVQZ	107.26	2152	50 (2.32 %)	80 (3.72 %)
aug-cc-pCV5Z	109.49	3722	90 (2.42 %)	110 (2.96 %)
	$I^{zz}_{2N}(0)$			
cc-pCVDZ	118.09	342	20 (5.85 %)	20 (5.85 %)
cc-pCVTZ	99.71	928	40 (4.31 %)	60 (6.47 %)
cc-pCVQZ	108.19	1880	50 (2.66 %)	60 (3.19 %)
cc-pCV5Z	110.07	3288	80 (2.43 %)	130 (3.95 %)
aug-cc-pCVDZ	111.59	556	20 (3.60 %)	20 (3.60 %)
aug-cc-pCVTZ	100.46	1350	30 (2.22 %)	40 (2.96 %)
aug-cc-pCVQZ	107.83	2582	50 (1.94 %)	60 (2.32 %)
aug-cc-pCV5Z	110.06	4342	80 (1.84 %)	110 (2.53 %)

enlarging the size of the basis set (and thus the number of excitations and the length of each Lanczos vector) is partly offset again by the fact that a smaller percentage or number of Lanczos vectors is needed for a converged result. This looks promising when it comes to using a larger basis set, e.g. aug-cc-pCVQZ, for larger systems. In addition, we note that the Lanczos $I(0)$ values decreased monotonically with the cardinal number and always converged from below (with the aforementioned exception of molecules with a period-3 atom), see, e.g. Figure 1. This means that a truncation error in the Lanczos algorithm might be partly offset by the basis set error, which might be exploited when computing Lanczos $I(0)$ for systems where the use of a basis set of recommended size is unfeasible. The convergence of Lanczos

Table 4: BH. Convergence of Lanczos iteration, when computing the x-, y- and z–component of $I(0)$ (in eV) with different basis sets. The results for the x- and y-component are identical since they transform as the same irreducible representation of the full point group. The reference value is the $I(0)$ computed after diagonalizing the full RPA electronic Hessian. Note that the algorithm used here creates 2 Lanczos vectors in one iteration.

Basis	$I^{xx}_{2N}(0)$ $/I^{yy}_{2N}(0)$	Full space size, $2N$	Convergence within 1.0 %, No. Lanczos vectors, (% of full space)	Convergence within 0.5 %, No. Lanczos vectors, (% of full space)
cc-pCVDZ	79.08	30	30 (100.00 %)	30 (100.00 %)
cc-pCVTZ	50.81	78	40 (51.28 %)	60 (76.92 %)
cc-pCVQZ	49.24	162	90 (55.56 %)	110 (67.90 %)
cc-pCV5Z	49.04	288	140 (48.61 %)	170 (59.03 %)
aug-cc-pCVDZ	71.15	48	30 (62.50 %)	40 (83.33 %)
aug-cc-pCVTZ	49.84	114	60 (52.63 %)	70 (61.40 %)
aug-cc-pCVQZ	48.80	222	90 (40.54 %)	120 (54.05 %)
aug-cc-pCV5Z	48.69	378	140 (37.04 %)	180 (47.62 %)
	$I^{zz}_{2N}(0)$			
cc-pCVDZ	58.12	54	10 (18.52 %)	10 (18.52 %)
cc-pCVTZ	42.77	138	30 (21.74 %)	40 (28.99 %)
cc-pCVQZ	41.90	264	20 (7.58 %)	60 (22.73 %)
cc-pCV5Z	41.78	444	70 (15.77 %)	90 (20.27 %)
aug-cc-pCVDZ	53.68	90	10 (11.11 %)	30 (33.33 %)
aug-cc-pCVTZ	42.49	198	30 (15.15 %)	40 (20.20 %)
aug-cc-pCVQZ	41.85	354	50 (14.12 %)	80 (22.60 %)
aug-cc-pCV5Z	41.75	570	70 (12.28 %)	90 (15.79 %)

$I(0)$ was usually faster for the nonlinear molecules (BH_3, H_2CO, H_2O and PH_3), as can be seen in Figure 3 and Table 3 versus Table 4. The number of Lanczos vectors needed to achieve the value of the limiting component of $I_{2N}(0)$ up to 0.5 % of the full space value is either approximately the same or just slightly larger than the number of Lanczos vectors needed to converge the other component(-s) of $I(0)$. Taking PH_3 and the aug-cc-pCVQZ basis set as example, see Table 3, where the z-component is the limiting one, the number of Lanczos vectors needed to achieve the latter convergence criteria is 50 for both the x and z components. The faster convergence for non-linear molecules, especially that of the limiting component of $I(0)$, would allow the computation of $I(0)$ for larger molecules, which

are usually nonlinear.

As can be seen in Figures 3–4 and Table 4, the (smaller) x- and y-components of $I(0)$ converge very slowly in the case of BH and AlH, even when the larger (and augmented) basis sets are used. Both BH and AlH are well known to have low-lying electronic excited states, with RPA values of 2.65 eV and 3.00 eV, respectively. However, neither the speed of the convergence of these states nor the size of their contributions to the overall $I(0)$ were found to be the culprit. The slow convergence might be explained if the gaps between the interior eigenvalues, i.e. the numerically smallest in the RPA case, of these molecules are small, as the block Lanczos algorithm is postulated[63,64] (see also Ref. 65 and 9) to converge the clusters of interior eigenvalues slowest.

Convergence Analysis: All Molecules

Figure 3 depicts the normal distribution of the number of Lanczos vectors needed to converge the largest component of $I(0)$, which for most of the molecules here is $I^{zz}(0)$, to within 0.5 % of the full space value. All the values were computed with the aug-cc-pCVQZ basis set. In this case, a number of Lanczos vectors corresponding to 10–25% of the full space was needed. This convergence criterion should give the values that fall within the RPA error (as compared to CCSD numbers[6]) even for the smallest systems. The corresponding numbers for the aug-cc-pCV5Z basis set are almost exactly the same (see Figure 4 and S20 in Section S3.3, Supplementary Material). For $I^{xx}(0)$ and $I^{yy}(0)$, which for the majority of these molecules are the components with the smallest number of excitations, the corresponding number of Lanczos vectors needed to achieve the same precision is 10–30 % when using both aug-cc-pCVQZ and aug-cc-pCV5Z basis sets. Although there is greater variation in how fast the smaller components of $I(0)$ converge, the mean is ≈17–19 % of the full space for all of them. The standard deviation from the mean tends to be smaller for the larger basis sets (see Figure 4), though it is not the case for $I^{zz}(0)$ where, e.g., the standard deviation from the average aug-cc-pCVTZ value is smaller than that of the aug-cc-pCV5Z value.

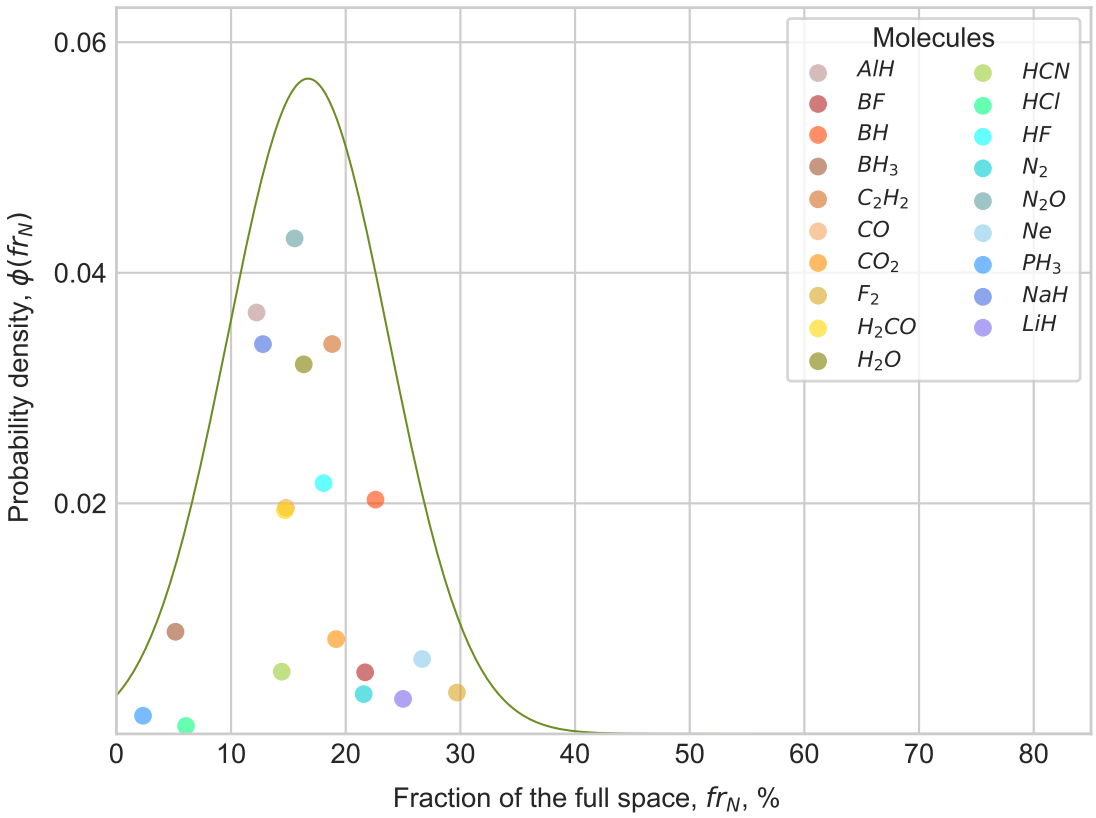

Figure 3: Normal distribution of the convergence of $I^{zz}(0)$ computed for 18 different molecules/atoms with aug-cc-pCVQZ basis set, using our block Lanczos RPA solver. The value compared to, $I^{zz}_{diag}(0)$, is computed after diagonalizing the full space electronic Hessian and the convergence threshold set to 0.5% of said value.

We have also briefly investigated how the $I(0)$ converges when accumulating (one by one) all the states computed after explicit diagonalization. We confirmed that almost all the states are needed to converge $I(0)$ (see Figures S25–S30 in Section S3.4), as previously observed by Sauer *et al.*[54] This well illustrates why Lanczos is the right choice of algorithm for computing such properties, as it simultaneously converges both ends of the eigenspectrum, and thus gives a good approximation of the higher states early on, as the numerically largest positive and negative RPA eigenvalues converge first. Additionally, when the gradient property vector is contained in the initial Lanczos vector, the excitations that have large overlap with it will be favoured.

The algorithm was sometimes found to exhibit signs of "transitory" instabilities. This

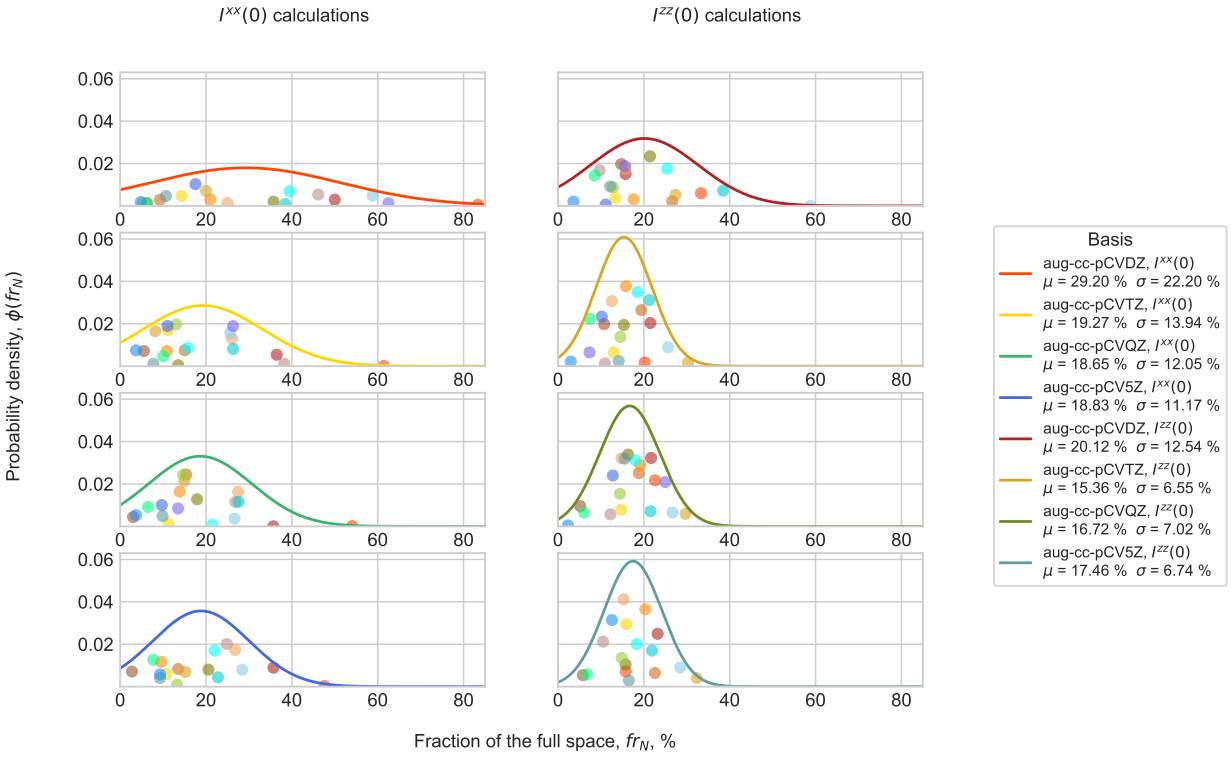

Figure 4: Normal distribution of the convergence of $I^{xx}(0)$ (left) and $I^{zz}(0)$ (right) to up to 0.5% accuracy, as computed for 19 different atomic and molecular systems with 8 different basis sets using our block Lanczos RPA solver. The value compared to, $I^{zz}_{diag}(0)$, is computed after diagonalizing the full space electronic Hessian.

means that the error in the largest eigenvalue, ω_N, could "blow up" for a certain value of k_{max} smaller than the full dimension, but the same did not happen for a calculation with a Lanczos chain length one larger or smaller than this k_{max}. This is likely due to a large norm of the newly computed Lanczos vector at this specific iteration. As the appearance of such an instability is easily detectable, one could either restart the algorithm and add one more Lanczos vector or alternatively go one iteration back and diagonalize the $\mathbf{T}^{(N)}$ matrix without the last two rows and columns.

Quality of RPA $I(0)$

Finally, we should try and assess the quality of our RPA results for $I(0)$ by comparing them to the (semi-) experimental as well as the higher level of theory *ab initio* values in

Table 5. The challenge here was that the latter are not easy to come by, especially for the molecules, while the (semi-) experimental methods have some considerable faults, not to mention the comparison being complicated by lack of inclusion of proper description of electron correlation, vibrational effects or higher-multipole and relativistic effects in the computed data where necessary. Experimentally, $I(0)$ can be estimated using Bethe formula (see Eq. (1)) by making a fit when analyzing the results from experimental stopping power measurements. Even when measuring at high enough energies, using a well-known source of particles and well-characterized description of energy loss, the use of different approaches when including the corrections to Bethe formula[66] can result in different $I(0)$ values from the same set of data.

Table 5: Comparison of $I(0)$ (in eV) computed with Lanczos for RPA (full space) and literature values.

System	Present work	CCSD [a]	Average values (OSS)[b]	$S(v)$ experiment
Ne	137.24^{I} 137.34^{II}	135.4	137	133.8^{c}
H_2O	72.97^{I} 73.00^{II}		72.0	70.9^{e} 60.0^{d}
C_2H_2	57.93^{I} 58.00^{II}		57.8	
N_2	83.51^{I} 83.56^{II}		83.0	85.6^{f} 97.8^{c}
CO	85.74^{I} 85.82^{II}		85.9	95.419^{c}
HCl	155.13^{I} 152.26^{II} 142.23^{III}		146	
CO_2	89.96^{I} 90.03^{II}		89.6	90.0^{g} 102.35^{c}
N_2O	86.65^{I} 86.71^{II}		90.5	103.96^{c}

[a]Sauer *et al.*[6] ; [b]Kamakura *et al.*[67] ; [c]Janni *et al.*[68] ; [d]Porter and Thwaites[69] ; [e]Haque *et al.*[70] [f]Shiomi-Tsuda *et al.*[71] ; [g]Porter *et al.*[72] . The values from present work are computed with [I]aug-cc-pCV5Z, [II]aug-cc-pCVQZ and [III]aug-cc-pCVTZ basis sets.

The semiexperimental values of $I(0)$ estimated by computing the sums from oscillator strength spectra (OSS) in the entire energy range strongly depend on the energy range covered. As discussed earlier, it is important to include weak transitions (as usually most of the spectrum is needed) when calculating $I(0)$, and there is a risk here that these will not show up in photoabsorption cross section. Compared to the average OSS $I(0)$ values

as estimated from a set of spectra by Kamakura *et al.*,[67] the RPA $I(0)$ values, computed using either the aug-cc-pCV5Z or the aug-cc-pCVQZ basis set, were within 0.2–4.3 %. The same RPA $I(0)$ values were within 0.05–21.6 % of $I(0)$ reported from different stopping-power measurements. This is, though, not surprising, since the $I(0)$ reported for the same molecule from different experiments can vary just as greatly, sometimes in the range of 10 eV. As $I(0)$ is usually only feasible to compute at a higher level of theory for small systems, so far, only our RPA $I(0)$ for Ne atom can be compared to a corresponding CCSD value. The RPA $I(0)$ for Ne is only 1.3 % higher than that calculated at the CCSD level,[6] which puts the electron correlation effects at the same size as observed previously by Sauer *et al.*

6 Conclusions

We have implemented a Lanczos algorithm for solving the paired RPA eigenvalue problem, as first introduced by Johnson *et al.*,[32] in the Dalton software package[10,11] and used it to calculate $S(0)$ and $I(0)$. The algorithm gives a balanced description of the electronic spectrum in an iterative subspace, and was therefore ideal for computing sum-over-states properties where excitations from the entire eigenspectrum are needed for convergence, e.g. as we here confirmed in case of $I(0)$. A simple way to compute the transition strengths directly in the Lanczos basis, as introduced by Coriani *et al.*,[16] was adapted for this algorithm and implemented. The Lanczos algorithm for RPA was shown to preserve the odd-m-power energy-weighted sums after $\frac{m+1}{2}$ iterations, as first proposed by Johnson *et al.* and generalized here. The performance of the block Lanczos RPA solver was tested by computing $S(0)$ and $I(0)$ at every 5 iterations (i.e. 10 Lanczos vectors, until the full space was reached) for a set of 18 small molecules and the Ne atom. In all cases, Lanczos $S(0)$ converged after the first iteration, as expected due to the preservation of sum rules. The correct numbers (up to 3-4 significant figures), though, were only reproduced when using the basis sets aug-cc-pCVQZ and aug-cc-pCV5Z, which are sufficiently large to reproduce the TRK sum rules.[33–35] The

$I(0)$ values for the majority of the systems converge well and decrease with the cardinal number of the Dunning basis set (i.e. D, T, Q, 5).[52,53,57,58] This trend is not seen for some systems that include period-3 elements, and it is suggested that the inclusion of extra d-functions might be able to restore the aforementioned trend. The difference between the $I(0)$ values computed with the aug-cc-pCVQZ and aug-cc-pCV5Z bases is negligible, and therefore the former could be used for this type of calculation. When using the aug-cc-pCVQZ (or aug-cc-pCV5Z) basis set, a number corresponding to only 10–30% of the full space is needed to obtain a Lanczos $I^{zz}(0)$ within 0.5 % of the full space RPA value in most cases. In general, the fraction of the full space needed to converge the subspace $I(0)$ values is relatively stable with respect to the basis-set size (starting from the triple-ζ quality basis). For nonlinear systems, the largest component of $I(0)$ not only converged faster than the smaller components of $I(0)$, it also required the same or a smaller number of Lanczos vectors. This looks promising when aiming at computing Lanczos $S(0)$ and $I(0)$ for larger molecules, as these are nonlinear. In future updates, we plan to include several functionalities, such as a check for large gaps between eigenvalues, some inexpensive check for convergence monitoring and a re-starting scheme for cases of (near-) "serious break-downs". The evolution of the diagonal and subdiagonal coefficients of the Lanczos matrix during the Lanczos iteration should be investigated to see if there are any trends that could point to devising a reasonable and useful extrapolation scheme, *á la* Rocca *et al.*[21] Devising an efficient diagonalization algorithm that takes the sparsity of a block Lanczos matrix (re-ordered or nonsymmetric banded form) into account would also save computational resources, and thus be critically important for computing eigenspectrum approximations for larger systems. The current solver can be easily extended to compute other sum-over-states properties. In perspective, it could also be interesting to formulate and adapt the block Lanczos RPA solver for higher level linear response methods that have equivalent block structure, e.g., the second order polarization propagator approximation (SOPPA).[51,73,74]

Supplementary Material

The detailed proof for preservation of sum rules, the re-ordering scheme for the $\bar{\mathbf{T}}^{(k)}$ matrix as well as additional Tables and Figures (depicting normal distribution of the number of Lanczos vectors needed to converge $I(0)$, as well as cummulative convergence of $I(0)$ for BH and BH_3 molecules) are available in the Supplementary Material.

Acknowledgement

S.C. acknowledges financial support from the Independent Research Fund Denmark–Natural Sciences, Research Project 2, Grant No. 7014-00258B.

Author Declarations

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

S.P.A.S. and S.C. conceptualized and supervised the project. L.Z. implemented the methodology, carried out all calculations, and prepared the first draft of the paper. All authors discussed the science in the paper and contributed to its revision into its final version.

Data Availability

All data from this study, including output files, additional figures and tables, can be found online:

`https://sid.erda.dk/wsgi-bin/ls.py?share_id=FFv0I5CK5x`

References

(1) Bethe, H. Zur Theorie des Durchgangs schneller Korpuskularstrahlen durch Materie. *Ann. Phys.* **1930**, *5*, 325.

(2) Bethe, H. Bremsformel für Elektronen Relativistischer Geschwindigkeit. *Z. Phys.* **1932**, *76*, 293–299.

(3) Zeiss, G.; Meath, W.; MacDonald, J.; Dawson, D. Accurate Evaluation of Stopping and Straggling Mean Excitation Energies for N, O, H_2 , N_2 , O_2 , NO, NH_3 , H_2O, and N_2O Using Dipole Oscillator Strength Distributions: A Test of the Validity of Bragg's Rule. *Radiation Research* **1977**, *70*.

(4) Sabin, J. R.; Oddershede, J.; Cabrera-Trujillo, R.; Sauer, S. P. A.; Deumens, E.; Öhrn, Y. Stopping power of molecules for fast ions. *Mol. Phys.* **2010**, *108*, 2891–2897.

(5) Rowe, D. J. Equations-of-Motion Method and the Extended Shell Model. *Rev. Mod. Phys.* **1968**, *40*, 153–166.

(6) Sauer, S. P. A.; Haq, I. U.; Sabin, J. R.; Oddershede, J.; Christiansen, O.; Coriani, S. Coupled cluster calculations of mean excitation energies of the noble gas atoms He, Ne and Ar and of the H_2 molecule. *Mol. Phys.* **2014**, *112*, 751–761.

(7) Jensen, P. W. K.; Sauer, S. P. A.; Oddershede, J.; Sabin, J. Mean excitation energies for molecular ions. *Nucl. Instrum. Methods Phys. Res., B* **2017**, *394*, 73–80.

(8) Lanczos, C. An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators. *J. Res. Natl. Bur. Stand.* **1950**, *45*, 255–282.

(9) Cullum, J. K.; Willoughby, R. A. *Lanczos Algorithms For Large Symmetric Eigenvalue Computations, Vol. I: Theory*; IBM T. J. Watson Research Center, 1984.

(10) Aidas, K.; Angeli, C.; Bak, K. L.; Bakken, V.; Bast, R.; Boman, L.; Christiansen, O.; Cimiraglia, R.; Coriani, S.; Dahle, P.; Dalskov, E. K.; Ekström, U.; Enevoldsen, T.; Eriksen, J. J.; Ettenhuber, P.; Fernández, B.; Ferrighi, L.; Fliegl, H.; Frediani, L.; Hald, K.; Halkier, A.; Hättig, C.; Heiberg, H.; Helgaker, T.; Hennum, A. C.; Hettema, H.; Hjertenæs, E.; Høst, S.; Høyvik, I.-M.; Iozzi, M. F.; Jansík, B.; Jensen, H. J. Aa.; Jonsson, D.; Jørgensen, P.; Kauczor, J.; Kirpekar, S.; Kjærgaard, T.; Klopper, W.; Knecht, S.; Kobayashi, R.; Koch, H.; Kongsted, J.; Krapp, A.; Kristensen, K.; Ligabue, A.; Lutnæs, O. B.; Melo, J. I.; Mikkelsen, K. V.; Myhre, R. H.; Neiss, C.; Nielsen, C. B.; Norman, P.; Olsen, J.; Olsen, J. M. H.; Osted, A.; Packer, M. J.; Pawlowski, F.; Pedersen, T. B.; Provasi, P. F.; Reine, S.; Rinkevicius, Z.; Ruden, T. A.; Ruud, K.; Rybkin, V. V.; Sałek, P.; Samson, C. C. M.; de Merás, A. S.; Saue, T.; Sauer, S. P. A.; Schimmelpfennig, B.; Sneskov, K.; Steindal, A. H.; Sylvester-Hvid, K. O.; Taylor, P. R.; Teale, A. M.; Tellgren, E. I.; Tew, D. P.; Thorvaldsen, A. J.; Thøgersen, L.; Vahtras, O.; Watson, M. A.; Wilson, D. J. D.; Ziolkowski, M.; Ågren, H. The Dalton quantum chemistry program system. *WIREs Comput. Mol. Sci.* **2014**, *4*, 269–284.

(11) Dalton, a molecular electronic structure program, Release Dalton2020.alpha (2020), see http://daltonprogram.org.

(12) Brabec, J.; Lin, L.; Shao, M.; Govind, N.; Yang, C.; Saad, Y.; Ng, E. G. Efficient Algorithms for Estimating the Absorption Spectrum within Linear Response TDDFT. *J. Chem. Theory Comput.* **2015**, *11*, 5197–5208.

(13) Casida, M. E. Time-Dependent Density Functional Response Theory for Molecules. In *In Recent Advances in Density Functional Methods*; Chong, D. E., Ed.; World Scientific, Singapore, 1995; pp 155–192.

(14) Trofimov, A.; Krivdina, I.; Weller, J.; Schirmer, J. Algebraic-diagrammatic construction

propagator approach to molecular response properties. *Chem. Phys.* **2006**, *329*, 1–10, Electron Correlation and Multimode Dynamics in Molecules.

(15) Koch, H.; Jørgensen, P. Coupled cluster response functions. *J. Chem. Phys.* **1990**, *93*, 3333–3344.

(16) Coriani, S.; Fransson, T.; Christiansen, O.; Norman, P. Asymmetric-Lanczos-Chain-Driven Implementation of Electronic Resonance Convergent Coupled-Cluster Linear Response Theory. *J. Chem. Theory Comput.* **2012**, *8*, 1616–1628.

(17) Hansen, M. B.; Seidler, P.; Győrffy, W.; Christiansen, O. A Lanczos-chain driven approach for calculating damped vibrational configuration interaction response functions. *J. Chem. Phys.* **2010**, *133*, 114102.

(18) Tsiper, E. V. A classical mechanics technique for quantum linear response. *J. Phys. B: At. Mol. Opt. Phys.* **2001**, *34*, L401–L407.

(19) Chernyak, V.; Schulz, M. F.; Mukamel, S.; Tretiak, S.; Tsiper, E. V. Krylov-space algorithms for time-dependent Hartree–Fock and density functional computations. *J. Chem. Phys.* **2000**, *113*, 36–43.

(20) Tretiak, S.; Isborn, C. M.; Niklasson, A. M. N.; Challacombe, M. Representation independent algorithms for molecular response calculations in time-dependent self-consistent field theories. *J. Chem. Phys.* **2009**, *130*, 054111.

(21) Rocca, D.; Gebauer, R.; Saad, Y.; Baroni, S. Turbo charging time-dependent density-functional theory with Lanczos chains. *J. Chem. Phys.* **2008**, *128*, 154105.

(22) Shee, A.; Zgid, D. Coupled Cluster as an Impurity Solver for Green's Function Embedding Methods. *J. Chem. Theory Comput.* **2019**, *15*, 6010–6024.

(23) Thomsen, B.; Hansen, M. B.; Seidler, P.; Christiansen, O. Vibrational absorption spectra from vibrational coupled cluster damped linear response functions calculated using an asymmetric Lanczos algorithm. *The Journal of Chemical Physics* **2012**, *136*, 124101.

(24) Coriani, S.; Christiansen, O.; Fransson, T.; Norman, P. Coupled-cluster response theory for near-edge x-ray-absorption fine structure of atoms and molecules. *Phys. Rev. A* **2012**, *85*, 022507.

(25) Cukras, J.; Coriani, S.; Decleva, P.; Christiansen, O.; Norman, P. Photoionization cross section by Stieltjes imaging applied to coupled cluster Lanczos pseudo-spectra. *J. Chem. Phys.* **2013**, *139*, 094103.

(26) Cabral Tenorio, B. N.; Chaer Nascimento, M. A.; Rocha, A. B.; Coriani, S. Lanczos-based equation-of-motion coupled-cluster singles-and-doubles approach to the total photoionization cross section of valence excited states. *J. Chem. Phys.* **2019**, *151*, 184106.

(27) Tenorio, B. N. C.; Moitra, T.; Nascimento, M. A. C.; Rocha, A. B.; Coriani, S. Molecular inner-shell photoabsorption/photoionization cross sections at core-valence-separated coupled cluster level: Theory and examples. *J. Chem. Phys.* **2019**, *150*, 224104.

(28) Runge, E.; Gross, E. K. U. Density-Functional Theory for Time-Dependent Systems. *Phys. Rev. Lett.* **1984**, *52*, 997–1000.

(29) Thouless, D. J. Vibrational States of Nuclei in the Random Phase Approximation. *Nucl. Phys.* **1961**, *22*, 78.

(30) Davidson, E. R. The Iterative Calculation of a Few of the Lowest Eigenvalues and Corresponding Eigenvectors of Large Real-Symmetric Matrices. *J. Comput. Phys.* **1975**, *17*, 87–94.

(31) Arnoldi, W. E. The principle of minimized iterations in the solution of the matrix eigenvalue problem. *Quart. Appl. Math.* **1951**, *9*, 17–29.

(32) Johnson, C. W.; Bertsch, G. F.; Hazelton, W. D. Lanczos Algorithm and Energy-weighted Sum Rules for Linear Response. *Comput. Phys. Commun.* **1999**, *120*, 155–161.

(33) Thomas, W. Über die Zahl der Dispersionselektronen, die einem stationären Zustande zugeordnet sind. (Vorläufige Mitteilung). *Naturwissenschaften* **1925**, *13*, 627.

(34) Reiche, F.; Thomas, W. Über die Zahl der Dispersionselektronen, die einem stationären Zustand zugeordnet sind. *Z. Phys.* **1925**, *34*, 510.

(35) Kuhn, W. Über die Gesamtstärke der von einem Zustande ausgehenden Absorptionslinien. *Z. Phys.* **1925**, *33*, 408.

(36) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618–622.

(37) Olsen, J.; Jørgensen, P. Linear and nonlinear response functions for an exact state and for an MCSCF state. *J. Chem. Phys.* **1985**, *82*, 3235–3264.

(38) Hartree, D. R. The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part II. Some Results and Discussion. *Math. Proc. Camb. Philos. Soc.* **1928**, *24*, 111–132.

(39) Jørgensen, P.; Simons, J. *Second Quantization Based Methods in Quantum Chemistry*; Academic Press, New York, 1981; pp 146–148.

(40) Sauer, S. P. A. *Molecular Electromagnetism: A Computational Chemistry Approach*; Oxford University Press, Oxford, 2011.

(41) Olsen, J.; Jensen, H. J. A.; Jørgensen, P. Solution of the large matrix equations which occur in response theory. *J. Comput. Phys.* **1988**, *74*, 265 – 282.

(42) Mises, R. V.; Pollaczek-Geiringer, H. Praktische Verfahren der Gleichungsauflösung. *J. Appl. Math. Mech.* **1929**, *9*, 152–164.

(43) Golub, G. H.; Loan, C. F. V. *Matrix Computations 4th Edition*; The Johns Hopkins University Press, 2013.

(44) Saad, Y. *Iterative Methods for Sparse Linear Systems*; The Society for Industrial and Applied Mathematics, 2003.

(45) Cullum, J.; Willoughby, R. A. Lanczos and the Computation in Specified Intervals of the Spectrum of Large, Sparse, Real Symmetric Matrices. *Sparse Matrix Proc.* **1978**, *1*, 220–255.

(46) Parlett, B.; Reid, J. Tracking the Progress of the Lanczos Algorithm for Large Symmetric Eigenproblems. *IMA J. Numer. Anal.* **1981**, *1*, 135–155.

(47) Scott, D. S. The Advantages of Inverted Operators in Rayleigh-Ritz Approximations. *SIAM J. Sci. Comput.* **1982**, *3*, 68–8.

(48) Wilkinson, J. H. *The Algebraic Eigenvalue Problem*; Oxford University Press, 1965.

(49) Cullum, J.; Donath, W. E. A block Lanczos algorithm for computing the q algebraically largest eigenvalues and a corresponding eigenspace of large, sparse, real symmetric matrices. 1974 IEEE Conference on Decision and Control including the 13th Symposium on Adaptive Processes. IEEE, 1974. 1974; pp 505–509.

(50) Godtliebsen, I. H.; Christiansen, O. A band Lanczos approach for calculation of vibrational coupled cluster response functions: simultaneous calculation of IR and Raman anharmonic spectra for the complex of pyridine and a silver cation. *Phys. Chem. Chem. Phys.* **2013**, *15*, 10035–10048.

(51) Bak, K. L.; Koch, H.; Oddershede, J.; Christiansen, O.; Sauer, S. P. A. Atomic integral

driven second order polarization propagator calculations of the excitation spectra of naphthalene and anthracene. *J. Chem. Phys.* **2000**, *112*, 4173–4185.

(52) Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(53) Peterson, K. A.; Dunning, T. H. Accurate correlation consistent basis sets for molecular core–valence correlation effects: The second row atoms Al–Ar, and the first row atoms B–Ne revisited. *J. Chem. Phys.* **2002**, *117*, 10548–10560.

(54) Sauer, S. P. A.; Sabin, J. R.; Oddershede, J. Calculated molecular mean excitation energies for some small molecules. *Nucl. Instr. and Meth. B* **1995**, *100*, 458–463.

(55) Sauer, S. P. A.; Sabin, J. R.; Oddershede, J. Test of the Validity of Bragg's Rule for Mean Excitation Energies of Small Molecules and Ions. *Nucl. Instr. and Meth. B* **2019**, *444*, 112–116.

(56) Sauer, S. P. A.; Sabin, J. R.; Oddershede, J. Bond correction factors and their applications to the calculation of molecular mean excitation energies. *Nucl. Instr. and Meth. B* **2020**, *468*, 28–36.

(57) Woon, D. E.; Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. V. Core-valence basis sets for boron through neon. *J. Chem. Phys.* **1995**, *103*, 4572–4585.

(58) Prascher, B. P.; Woon, D. E.; Peterson, K. A.; Dunning, T. H.; Wilson, A. K. Gaussian basis sets for use in correlated molecular calculations. VII. Valence, core-valence, and scalar relativistic basis sets for Li, Be, Na, and Mg. *Theor. Chem. Acc.* **2011**, *128*, 69–82.

(59) Dunning, T. H.; Peterson, K. A.; Wilson, A. K. Gaussian basis sets for use in correlated

molecular calculations. X. The atoms aluminum through argon revisited. *J. Chem. Phys.* **2001**, *114*, 9244–9253.

(60) Sauer, S. P. A.; Raynes, W. T. Unexpected differential sensitivity of nuclear spin-spin coupling constants to bond stretching in BH_4^-, NH_4^+ and SiH_4. *J. Chem. Phys.* **2000**, *113*, 3121–3129.

(61) Sauer, S. P. A.; Oddershede, J.; Sabin, J. R. Directional Dependence of the Mean Excitation Energy and Spectral Moments of the Dipole Oscillator Strength Distribution of Glycine and Its Zwitterion. *J. Phys. Chem. A* **2006**, *110*, 8811–8817.

(62) Silva-Junior, M. R.; Sauer, S. P. A.; Schreiber, M.; Thiel, W. Basis set effects on coupled cluster benchmarks of electronically excited states: CC3, CCSDR(3) and CC2. *Mol. Phys.* **2010**, *108*, 453–465.

(63) Bai, Z.; Day, D.; Ye, Q. ABLE: An Adaptive Block Lanczos Method for Non-Hermitian Eigenvalue Problems. *SIAM J. Matrix Anal. Appl.* **1999**, *20*, 1060–1082.

(64) Li, R.-C.; Zhang, L.-H. Convergence of the Block Lanczos Method for Eigenvalue Clusters. *Numer. Math.* **2015**, *131*, 83–113.

(65) Li, R.-C. Sharpness in rates of convergence for the symmetric Lanczos method. *Math. Comput.* **2010**, *79*, 419–435.

(66) Deasy, J. ICRU Report 49, Stopping Powers and Ranges for Protons and Alph Particles. *Med. Phys.* **1994**, *21*, 709–710.

(67) Kamakura, S.; Sakamoto, N.; Ogawa, H.; Tsuchida, H.; Inokuti, M. Mean excitation energies for the stopping power of atoms and molecules evaluated from oscillator-strength spectra. *J. Appl. Phys.* **2006**, *100*, 064905.

(68) Janni, J. F. Energy loss, range, path length, time-of-flight, straggling, multiple scattering, and nuclear interaction probability: In two parts. Part 1. For 63 compounds Part 2. For elements $1 \leqslant Z \leqslant 92$. *At. Data Nucl. Data Tables* **1982**, *27*, 147 – 339.

(69) Porter, L. E.; Thwaites, D. I. Physical state effects on the mean excitation energy of water as determined from α particle stopping-power measurements. *Phys. Rev. A* **1982**, *25*, 3407–3410.

(70) Haque, A.; Nikjoo, H. Stopping power for alpha particles in organic liquids and vapours. *Nucl. Instrum. Methods Phys. Res., B* **1991**, *53*, 15 – 23.

(71) Shiomi-Tsuda, N.; Sakamoto, N.; Ogawa, H.; Tanaka, M.; Saitoh, M.; Kitoba, U. Stopping powers of N_2 and O_2 for protons from 4.0 to 13.0 MeV. *Nucl. Instrum. Methods Phys. Res., B* **1999**, *149*, 17 – 24.

(72) Porter, L. Additivity and effective charge studies with light projectiles and target gases of CH_4, CO_2, He, Ne, Ar, Kr, Xe, Cl_2 and Br_2. *Nucl. Instrum. Methods Phys. Res., B* **1994**, *93*, 203 – 209.

(73) Nielsen, E. S.; Jørgensen, P.; Oddershede, J. Transition moments and dynamic polarizabilities in a second order polarization propagator approach. *J. Chem. Phys.* **1980**, *73*, 6238–6246.

(74) Schnack-Petersen, A. K.; Simmermacher, M.; Fasshauer, E.; Jensen, H. J. A.; Sauer, S. P. A. The Second-Order-Polarization-Propagator-Approximation (SOPPA) in a four-component spinor basis. *J. Chem. Phys.* **2020**, *152*, 134113.