



**HAL**  
open science

# Rationale for the Extrapolation Procedure in Selected Configuration Interaction

Hugh G. A. Burton, Pierre-François Loos

► **To cite this version:**

Hugh G. A. Burton, Pierre-François Loos. Rationale for the Extrapolation Procedure in Selected Configuration Interaction. *Journal of Chemical Physics*, 2024, 160, pp.104102. 10.1063/5.0192458 . hal-04357241

**HAL Id: hal-04357241**

**<https://hal.science/hal-04357241>**

Submitted on 21 Dec 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Rationale for the Extrapolation Procedure in Selected Configuration Interaction

Hugh G. A. Burton<sup>1, a)</sup> and Pierre-François Loos<sup>2, b)</sup>

<sup>1)</sup> Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, U.K.

<sup>2)</sup> Laboratoire de Chimie et Physique Quantiques (UMR 5626), Université de Toulouse, CNRS, UPS, France.

Selected configuration interaction (SCI) methods have emerged as state-of-the-art methodologies for achieving high accuracy and generating benchmark reference data for ground and excited states in small molecular systems. However, their precision relies heavily on extrapolation procedures to produce a final estimate of the exact result. Using the structure of the exact electronic energy landscape, we provide a rationale for the common linear extrapolation of the variational energy as a function of the second-order perturbative correction. In particular, we demonstrate that the energy gap and the coupling between the so-called internal and external spaces are the key factors determining the rate at which the linear regime is reached. Starting from first principles, we also derive a new non-linear extrapolation formula that improves the post-processing of data generated from SCI methods and can be applied to both ground- and excited-state energies.

## I. INTRODUCTION

Selected configuration interaction (SCI),<sup>1–4</sup> and related methods (such as density-matrix renormalisation group approaches<sup>5–7</sup> and others<sup>8–17</sup>), have taken a prominent role in modern electronic structure theory.<sup>18–20</sup> Their primary purpose is to calculate reference correlation and excitation energies in small molecular systems,<sup>19,21–28</sup> for which they have demonstrated a remarkable ability to yield highly accurate estimates of full configuration interaction (FCI) results. The numerous variations of SCI all perform a sparse exploration of the Hilbert space by selecting only the most energetically relevant determinants. This natural philosophy emerges from the observation that, among the incredibly large number of determinants in the FCI space, only a tiny fraction of them significantly contribute to the energy. Modern versions of SCI include CIPSI (CI using a Perturbative Selection made Iteratively)<sup>3,18,21,29–36</sup> adaptive sampling CI (ASCI),<sup>37–40</sup> semistochastic heatbath CI (SHCI),<sup>23,24,41–45</sup> iterative CI (iCI),<sup>46–50</sup> Monte Carlo CI (MCCI),<sup>51,52</sup> and FCI quantum Monte Carlo (FCIQMC).<sup>53–58</sup>

The SCI wave function corresponds to a truncated CI expansion constructed from determinants in some internal (or model) space  $\mathcal{I}$

$$|\Psi_{\text{var}}\rangle = \sum_{I \in \mathcal{I}} c_I |I\rangle, \quad (1)$$

with the associated variational energy  $E_{\text{var}} = \langle \Psi_{\text{var}} | \hat{H} | \Psi_{\text{var}} \rangle$ , where we assume the normalisation of the variational wave function, i.e.,  $\langle \Psi_{\text{var}} | \Psi_{\text{var}} \rangle = 1$ . The accuracy of  $|\Psi_{\text{var}}\rangle$  can be assessed using the second-order Epstein–Nesbet perturbation correction, computed using the determinants  $\{\alpha\}$  that lie outside the model space

(i.e. in the external space  $\mathcal{A}$ ) as

$$E_{\text{PT}2} = - \sum_{\alpha \in \mathcal{A}} \frac{|\langle \Psi_{\text{var}} | \hat{H} | \alpha \rangle|^2}{H_{\alpha\alpha} - E_{\text{var}}}, \quad (2)$$

where  $H_{\alpha\alpha} = \langle \alpha | \hat{H} | \alpha \rangle$ . The exact FCI wave function and energy are indicated by the limit  $E_{\text{PT}2} \rightarrow 0^-$ .

Despite the sparse exploration of the Hilbert space, these state-of-the-art methods still rely heavily on extrapolation procedures to produce final FCI estimates.<sup>19,20,23,34</sup> In particular, it is widely observed that  $E_{\text{var}}$  becomes approximately proportional to  $E_{\text{PT}2}$  for small  $E_{\text{PT}2}$ , and thus a linear or quadratic extrapolation of  $E_{\text{var}}$  for  $E_{\text{PT}2} \rightarrow 0$  is generally used to estimate the exact energy for an unconverged SCI calculation.<sup>23</sup> The precision and reliability of this post-processing extrapolation procedure are critical in order to produce meaningful estimates. However, to the best of our knowledge, no theoretical justification for the linear (or otherwise) relationship between  $E_{\text{var}}$  and  $E_{\text{PT}2}$  has been proposed.

To illustrate this extrapolation procedure, Fig. 1 shows the evolution of the variational correlation energy of benzene as a function of  $E_{\text{PT}2}$ , computed in the cc-pVDZ basis and within the frozen-core approximation. These SCI calculations were performed with QUANTUM PACKAGE using the CIPSI algorithm<sup>34</sup> and the data are extracted from Ref. 35. The FCI estimate of the correlation energy (solid black line in Fig. 1) was estimated to be  $-862.890 mE_h$  and was obtained by performing a five-point linear fit (dashed black line in Fig. 1) of the CIPSI data. This estimate carries an error of the order of  $1 mE_h$  and the fitting error was estimated to be  $0.266 mE_h$ . From Fig. 1, it is clear that, for sufficiently small  $E_{\text{PT}2}$ , the variational quantity is linear with respect to  $E_{\text{PT}2}$ . However, the SCI data deviate significantly from linearity for larger values of  $E_{\text{PT}2}$ , which we shall address in detail later on.

In this Communication, we provide a rationale to justify the linear extrapolation of the (zeroth-order) variational energy as a function of the second-order perturbative energy. We adopt a geometric approach that considers the variational wave function as a point on the exact

<sup>a)</sup> Electronic mail: [hgaburton@gmail.com](mailto:hgaburton@gmail.com)

<sup>b)</sup> Electronic mail: [loos@irsamc.ups-tlse.fr](mailto:loos@irsamc.ups-tlse.fr)

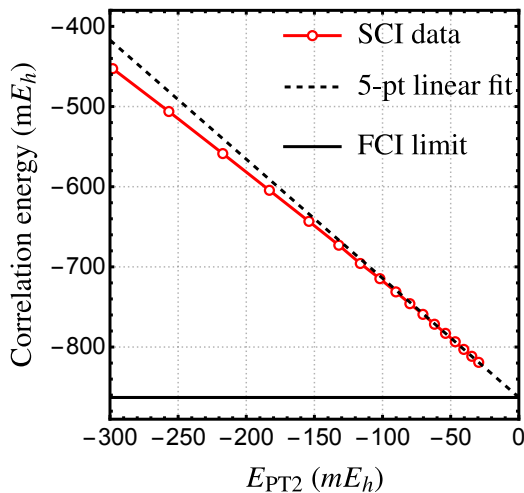


FIG. 1. Frozen-core (variational) correlation energy of benzene as a function of  $E_{PT2}$  computed in the cc-pVDZ basis, as described in Ref. 35.

electronic energy landscape,<sup>59</sup> allowing the second-order perturbative correction to be derived from the local gradient and curvature of this energy landscape. Moreover, we investigate a two-state model in which an analytic relationship between  $E_{\text{var}}$  and  $E_{PT2}$  can be derived, leading to a novel parametrised non-linear formula that facilitates a more robust extrapolation procedure.

## II. RATIONALE FOR THE LINEAR EXTRAPOLATION

The linear relationship between  $E_{\text{var}}$  and  $E_{PT2}$  can be derived from first principles by considering the structure of the electronic energy landscape. While Ref. 59 describes this energy landscape perspective in detail, the salient points are summarised here. Any normalised wave function in the full  $N$ -dimensional Hilbert space

$$|\Psi\rangle = \sum_{I=1}^N v_I |I\rangle \quad (3)$$

can be represented by a vector  $\mathbf{v}$  subject to the normalisation constraint  $\mathbf{v}^\dagger \cdot \mathbf{v} = 1$ , which constrains the wave function to the surface of a hypersphere. The energy is given by the quadratic form

$$E = \mathbf{v}^\dagger \cdot \mathbf{H} \cdot \mathbf{v} \quad (4)$$

and exact eigenstates of the Hamiltonian correspond to stationary points of  $E$  constrained to the surface of the hypersphere, as illustrated in Fig. 2. At any point on the hypersphere, the tangent space  $\mathcal{T}$  contains the vectors that are orthogonal to  $\mathbf{v}$ , which can be collected as the columns of an  $N \times (N - 1)$  matrix  $\mathbf{v}_\perp$ . These tangent vectors correspond to the states  $|T\rangle$  that are orthogonal to  $|\Psi\rangle$  and satisfy

$$\hat{1} = |\Psi\rangle\langle\Psi| + \sum_{T \in \mathcal{T}} |T\rangle\langle T|, \quad (5)$$

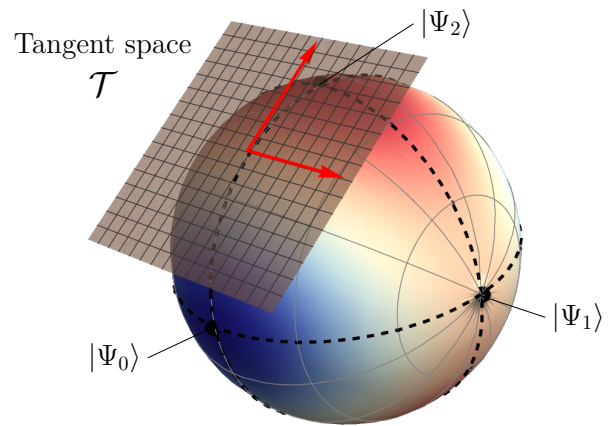


FIG. 2. Sketch of the exact electronic energy landscape in a three-dimensional Hilbert space. Eigenstates correspond to stationary points constrained to the surface of the unit sphere. At any point, the tangent space  $\mathcal{T}$  is spanned by the two vectors (red) that are orthogonal to the position vector.

where  $\hat{1}$  is the identity operator. A constrained step  $\mathbf{s}$  on this landscape is parametrised using a unitary transformation as

$$|\Psi(\mathbf{s})\rangle = \exp\left(\sum_{T \in \mathcal{T}} s_T \left(|T\rangle\langle\Psi| - |\Psi\rangle\langle T|\right)\right) |\Psi\rangle. \quad (6)$$

Assuming real wave functions, the components of the constrained energy gradient are then

$$g_T = \left. \frac{\partial E}{\partial s_T} \right|_{\mathbf{s}=\mathbf{0}} = 2 \langle T | \hat{H} | \Psi \rangle, \quad (7)$$

while the elements of the Hessian matrix of constrained second-derivatives become

$$Q_{TT'} = \left. \frac{\partial^2 E}{\partial s_T \partial s_{T'}} \right|_{\mathbf{s}=\mathbf{0}} = 2 \langle T | \hat{H} - E | T' \rangle. \quad (8)$$

So far, we have only considered the structure of the electronic energy landscape for an arbitrary wave function in the full Hilbert space. For a SCI variational wave function, the only non-zero elements of  $\mathbf{v}$  correspond to determinants included in the internal space, with coefficients  $c_I$ , as defined in Eq. (1). The tangent vectors can then be split into two disjoint sets corresponding to the eigenstates within the internal space that are orthogonal to  $|\Psi_{\text{var}}\rangle$ , denoted  $\mathcal{J} = \{|J\rangle\}_{J \neq \text{var}}$ , and the determinants in the external space, giving  $\mathcal{T} = \mathcal{J} \cup \mathcal{A}$ , as illustrated in Fig. 3. Since  $|\Psi_{\text{var}}\rangle$  is a CI solution within the internal space, such that  $\langle J | \hat{H} | \Psi_{\text{var}} \rangle = 0$ , the gradient given by Eq. (7) is only non-zero in the direction of the tangent vectors in  $\mathcal{A}$ .

The local structure of the energy landscape around the variational wave function  $|\Psi_{\text{var}}\rangle$  is given by a second-order Taylor series expansion as

$$E(\mathbf{s}) = E_{\text{var}} + \mathbf{s}^\dagger \cdot \mathbf{g} + \frac{1}{2} \mathbf{s}^\dagger \cdot \mathbf{Q} \cdot \mathbf{s}. \quad (9)$$

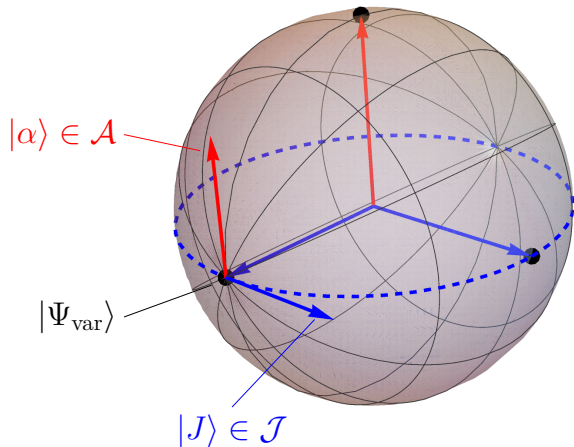


FIG. 3. Sketch of the tangent space construction  $\mathcal{T} = \mathcal{J} \cup \mathcal{A}$ . The variational space  $\mathcal{I}$  (dashed blue circle) is built from two configurations (blue vectors) and the external space  $\mathcal{A}$  contains one configuration (red vector). At  $|\Psi_{\text{var}}\rangle$ , the tangent space is spanned by one tangent direction  $|J\rangle$ , which is locally parallel to the variational space  $\mathcal{I}$ , and one orthogonal tangent direction in the external space  $|\alpha\rangle$ .

Optimising this quadratic form with the Newton–Raphson step  $\mathbf{s} = -\mathbf{Q}^{-1} \cdot \mathbf{g}$  gives an estimate of the difference between the exact energy  $E_{\text{exact}}$  and  $E_{\text{var}}$  as

$$\Delta E = E_{\text{exact}} - E_{\text{var}} \approx -\frac{1}{2} \mathbf{g}^\dagger \cdot \mathbf{Q}^{-1} \cdot \mathbf{g}, \quad (10)$$

or, equivalently,

$$\Delta E \approx - \sum_{\alpha\alpha' \in \mathcal{A}} \langle \Psi | \hat{H} | \alpha \rangle \langle \alpha | (\hat{H} - E_{\text{var}} \hat{1})^{-1} | \alpha' \rangle \langle \alpha' | \hat{H} | \Psi \rangle, \quad (11)$$

where we have exploited the fact that the gradient is zero in the direction of the tangent vectors within  $\mathcal{J}$ . The exact energy landscape is quadratic near an eigenstate,<sup>59</sup> and thus Eq. (10) will become the exact energy correction when  $|\Psi_{\text{var}}\rangle$  is sufficiently close to an eigenstate. Since the matrix elements  $\langle \alpha | (\hat{H} - E_{\text{var}} \hat{1})^{-1} | \alpha' \rangle$  are expensive to compute, we can assume that the Hamiltonian is diagonally dominant in  $\mathcal{A}$  and take the leading order approximation

$$\hat{H}^{(0)} = \sum_{I'I'' \in \mathcal{I}} |I\rangle H_{II'} \langle I'| + \sum_{\alpha \in \mathcal{A}} |\alpha\rangle H_{\alpha\alpha} \langle \alpha|, \quad (12)$$

where  $\hat{H}^{(0)}$  is the zeroth-order Hamiltonian within the Epstein–Nesbet partitioning. The energy correction then reduces to the second-order Epstein–Nesbet expression obtained in Eq. (2) to give

$$\Delta E \approx - \sum_{\alpha \in \mathcal{A}} \frac{|\langle \Psi_{\text{var}} | \hat{H} | \alpha \rangle|^2}{H_{\alpha\alpha} - E_{\text{var}}} = E_{\text{PT2}}. \quad (13)$$

Assuming that  $|\Psi\rangle$  is sufficiently close to an eigenstate, such that the quadratic approximation to the exact energy

landscape is valid, we can combine Eqs. (10), (11), and (13) to obtain  $E_{\text{var}} \approx E_{\text{exact}} - E_{\text{PT2}}$ , thus rationalising the linear extrapolation of  $E_{\text{var}}$  as a function of  $E_{\text{PT2}}$ .

### III. INSIGHTS FROM A TWO-STATE MODEL

In practice, linear or quadratic extrapolation procedures only work using a limited number of points for well-converged calculations (see Fig. 1). As mentioned above, the variational energy generally appears to deviate away from linearity for larger values of  $E_{\text{PT2}}$ . The cause of these deviations can be studied using a two-state model that represents the separation of the internal and external spaces in a SCI calculation, and for which the relationship between  $E_{\text{var}}$  and  $E_{\text{PT2}}$  can be analytically derived.

Our model contains individual states  $|\Psi_{\mathcal{I}}\rangle$  and  $|\Psi_{\mathcal{A}}\rangle$  representing wave functions in the internal and external spaces, respectively, with characteristic energies  $E_{\mathcal{I}}$  and  $E_{\mathcal{A}}$ . The Hamiltonian matrix in this basis is then

$$\mathbf{H} = \begin{pmatrix} E_{\mathcal{I}} & t \\ t & E_{\mathcal{A}} \end{pmatrix}, \quad (14)$$

where  $t = \langle \Psi_{\mathcal{I}} | \hat{H} | \Psi_{\mathcal{A}} \rangle$  represents the strength of the coupling between the internal and external spaces, and  $\delta E = E_{\mathcal{A}} - E_{\mathcal{I}}$  provides a measure of their energetic separation. The exact ground-state energy is

$$E_{\text{exact}} = E_{\mathcal{I}} + \frac{\delta E}{2} - \sqrt{\left(\frac{\delta E}{2}\right)^2 + t^2}. \quad (15)$$

The improvement of  $|\Psi_{\text{var}}\rangle$  during the course of a SCI calculation can be modelled by mixing  $|\Psi_{\mathcal{I}}\rangle$  and  $|\Psi_{\mathcal{A}}\rangle$  to give the parametrisation

$$|\Psi_{\text{var}}(\theta)\rangle = \cos \theta |\Psi_{\mathcal{I}}\rangle + \sin \theta |\Psi_{\mathcal{A}}\rangle, \quad (16)$$

with  $0 \leq \theta < 2\pi$ . The corresponding energy is

$$E_{\text{var}}(\theta) = E_{\mathcal{I}} + \frac{\delta E}{2} (1 - \cos 2\theta) + t \sin 2\theta. \quad (17)$$

Following a Taylor series expansion, the second-order correction is

$$E_{\text{PT2}}(\theta) = -\frac{1}{4} \frac{(2t \cos 2\theta + \delta E \sin 2\theta)^2}{\delta E \cos 2\theta - 2t \sin 2\theta}. \quad (18)$$

By solving Eq. (18) for  $\theta$ , we can invert these equations to express  $E_{\text{var}}$  in terms of  $E_{\text{PT2}}$  as

$$E_{\text{var}} = E_{\mathcal{I}} + \frac{\delta E}{2} - E_{\text{PT2}} - \sqrt{\left(\frac{\delta E}{2}\right)^2 + t^2 + (E_{\text{PT2}})^2}. \quad (19)$$

which naturally reduces to Eq. (15) for  $E_{\text{PT2}} = 0$ . This expression reveals that the more general form of  $E_{\text{var}}$  as a function of  $E_{\text{PT2}}$  involves a square-root term that deviates away from linearity, and that this departure

from the linear regime is directly related to the energetic separation ( $\delta E$ ) and coupling strength ( $t$ ) between the internal and external spaces. For  $(E_{\text{PT}2})^2 \ll (\frac{\delta E}{2})^2 + t^2$ , we recover the linear behaviour  $E_{\text{var}} \approx E_{\text{exact}} - E_{\text{PT}2}$ . In a real SCI calculation, we expect  $|t| \ll |\frac{\delta E}{2}|$ . Therefore, the larger the energy separation, the sooner the linear regime is reached, while a large coupling between  $\mathcal{I}$  and  $\mathcal{A}$  also leads more rapidly to the linear regime. These features are further demonstrated through the series expansion of  $E_{\text{var}}$  at small  $E_{\text{PT}2}$ ,

$$E_{\text{var}} = E_{\text{exact}} - E_{\text{PT}2} - \frac{E_{\text{PT}2}^2}{2\sqrt{(\frac{\delta E}{2})^2 + t^2}} + \mathcal{O}(E_{\text{PT}2}^3), \quad (20)$$

which shows that the quadratic behaviour is minimal when  $|\delta E|$  or  $|t|$  is large.

This functional relationship can be illustrated by evaluating  $E_{\text{var}}$  for various values of  $E_{\text{PT}2}$  in the limit of interest, as shown in Fig. 4 for  $E_{\mathcal{I}} = -1$ ,  $\delta E = 1$  and  $t = 1$ . As  $E_{\text{PT}2}$  gets larger,  $E_{\text{var}}$  strongly deviates from linearity and bears a close similarity to previous SCI data.<sup>19,21,23–27,34</sup> Figure 1 nicely illustrates this square-root behaviour for a realistic system, and the similarities between Figs. 1 and 4 are striking.

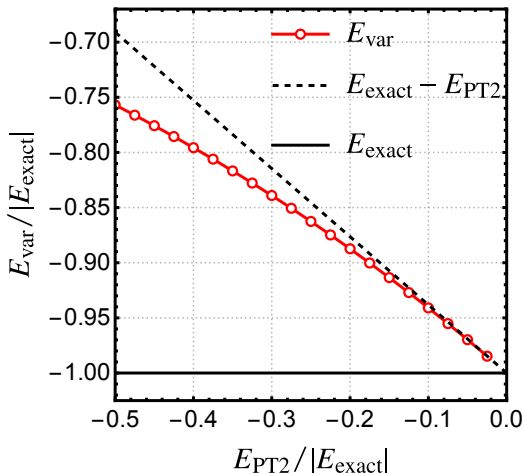


FIG. 4.  $E_{\text{var}}$  (red markers) as a function of  $E_{\text{PT}2}$  for the two-state system with  $E_{\mathcal{I}} = -1$ ,  $\delta E = 1$  and  $t = 1$ . These data deviate from the linear approximation (black dashed line) due to the square-root term in Eq. (19).

For real systems, this two-state scenario can always be engineered by using a singular value decomposition of the gradient vector to transform the tangent vectors such that only one direction has a non-zero gradient. In other words, we can transform the external space  $\mathcal{A}$  such that only a single state  $|\Psi_{\mathcal{A}}\rangle = \sum_{\alpha \in \mathcal{A}} c_{\alpha} |\alpha\rangle$  couples to  $|\Psi_{\text{var}}\rangle$  through the Hamiltonian. Furthermore, since the states within  $\mathcal{I}$  are decoupled by solving the CI problem, this two-state model can be constructed for both ground and low-lying excited states, the only difference being the precise values of  $E_{\mathcal{I}}$ ,  $\delta E$ , and  $t$ . While this transformation

is not feasible in practice, it provides a conceptual link between real SCI data and the present two-state model.

#### IV. A NON-LINEAR EXTRAPOLATION FORMULA

The insights from our two-state model suggest that a non-linear functional form is more suitable for extrapolating SCI data. The separation of the model and perturbation space in SCI means that the relationship between  $E_{\text{var}}$  and  $E_{\text{PT}2}$  is not as straightforward as the model system. However, as more determinants are added to the model space, the variational wave function follows a path towards the exact ground state that is likely to resemble Eq. (19). The concave form of Eq. (19) suggests that a linear extrapolation procedure will generally underestimate the exact correlation energy. Therefore, we propose a new non-linear extrapolation formula

$$E_{\text{var}}(E_{\text{PT}2}; a, b, c) = a + \frac{|c|}{2} - bE_{\text{PT}2} - \sqrt{\left(\frac{c}{2}\right)^2 + (bE_{\text{PT}2})^2}, \quad (21)$$

where  $a$ ,  $b$ , and  $c$  are fitting parameters. This expression corresponds to a form of quadratic approximant, which has been previously used in the resummation of divergent perturbation expansions.<sup>60–63</sup> Crucially, Eq. (21) reduces to a linear fit for  $|E_{\text{PT}2}| \ll |\frac{c}{2b}|$  and can reproduce the observed non-linearity for larger  $E_{\text{PT}2}$ . The fitted value of  $a$  provides the estimate for the FCI result.

The variation of the FCI estimate of the correlation energy of benzene computed in the cc-pVDZ basis using the linear or non-linear extrapolation procedure is compared in Fig. 5. These data show that the non-linear formula can accurately fit the SCI data at larger  $E_{\text{PT}2}$  values than the linear procedure. In the left panel of Fig. 6, we study the influence of the number of points included in the linear

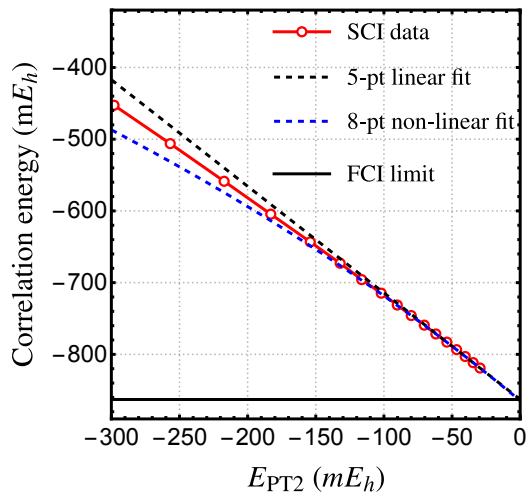


FIG. 5. Comparison of the linear and non-linear extrapolation procedure for the correlation energy of benzene computed in the cc-pVDZ basis, using the SCI data from Ref. 35.

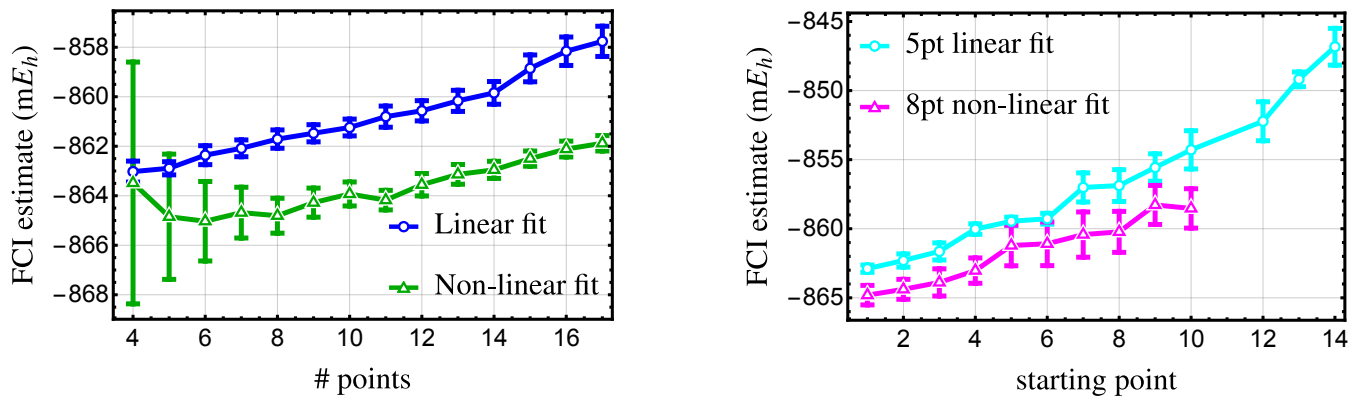


FIG. 6. Evolution of the FCI estimate of the correlation energy of benzene computed in the cc-pVDZ basis. Left: A variable number of fitting points is considered in the linear (blue) or non-linear (green) fits, with each fit starting from the point associated with the smallest value of  $E_{PT2}$ . Right: A fixed-length window of consecutive points is used for the linear (cyan) or non-linear (magenta) fits. The point of index 1 corresponds to the smallest value of  $E_{PT2}$ , and a higher index for the starting point corresponds to a larger  $E_{PT2}$  correction. The error bars indicate the standard errors associated with the fitting procedure.

or non-linear fits, each of them starting from the point associated with the smallest value of  $E_{PT2}$ . The error bars indicate the standard errors associated with the fitting procedure. Although the FCI estimates obtained from the non-linear formula [see Eq. (21)] have larger fitting errors for a small number of points, they quickly stabilise and remain relatively consistent compared to those obtained from the linear procedure, which rise much faster. Consequently, a larger number of points can, and should, be employed for the non-linear extrapolation formula, while the linear extrapolation procedure becomes systematically worse when more points are used.

In the right panel of Fig. 6, we consider a fixed-length window of consecutive points for the linear and non-linear fits, but we vary the index of the starting point, with the point at index 1 corresponding to the smallest value of  $E_{PT2}$  (i.e., the higher the index of the starting point, the larger the  $E_{PT2}$  correction). Using 5 and 8 points for the linear and non-linear fits, respectively, we again show that the non-linear procedure is slightly more stable than its linear counterpart with respect to the starting point, although both extrapolation schemes eventually underestimate the correlation energy for larger  $E_{PT2}$  values. These results indicate that, compared to the linear approach, the non-linear extrapolation procedure can provide a more accurate estimate of the FCI result for SCI calculations that are less well converged.

## V. CONCLUDING REMARKS

In this Communication, we have proposed a theoretical rationale for the linear extrapolation procedure of the variational zeroth-order energy,  $E_{\text{var}}$ , as a function of the second-order perturbative correction,  $E_{PT2}$ , commonly employed in SCI methods. Our derivation is based on connecting the SCI variational wave function to the properties of the underlying electronic energy landscape. The

accuracy of the extrapolation of  $E_{\text{var}}$  as  $E_{PT2}$  approaches 0 is critical for the accurate determination of the final FCI estimate.

Our investigations led us to the discovery of a novel non-linear extrapolation formula that more effectively captures the behaviour of  $E_{\text{var}}$  for larger  $E_{PT2}$  values, thereby enhancing the robustness of extrapolations toward the FCI limit. Based on a two-state model, we derived the analytic form of this non-linear extrapolation formula and examined its mathematical properties. Specifically, we illustrated that the rate at which the linear regime is attained is primarily determined by the energetic gap and the coupling between the internal and external spaces. As a concrete example, we studied the ground-state correlation energy of benzene, which illustrates the versatility and reliability of this new extrapolation procedure.

We anticipate that this study will facilitate SCI calculations on larger Hilbert spaces, while providing confidence in the validity of extrapolation procedures. Furthermore, we expect that our two-state model and our framing of the SCI approach within the electronic energy landscape framework will allow other intriguing mathematical aspects of these methods to be explored in the future.

## SUPPLEMENTARY MATERIAL

See the [supplementary material](#) for the SCI data of Fig. 1 and the raw data associated with Fig. 6.

## ACKNOWLEDGEMENTS

HGAB thanks Downing College, Cambridge for support through the Kim and Julianna Silverman Research Fellowship. PFL thanks financial support from the European Research Council (ERC) under the European Union’s

Horizon 2020 research and innovation programme (Grant agreement No. 863481).

## DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available within the article and its supplementary material.

## REFERENCES

- 1 C. F. Bender and E. R. Davidson, *Phys. Rev.* **183**, 23 (1969).
- 2 J. L. Whitten and M. Hackmeyer, *J. Chem. Phys.* **51**, 5584 (1969).
- 3 B. Huron, J. P. Malrieu, and P. Rancurel, *J. Chem. Phys.* **58**, 5745 (1973).
- 4 R. J. Buenker and S. D. Peyerimhoff, *Theor. Chim. Acta* **35**, 33 (1974).
- 5 S. R. White, *Phys. Rev. Lett.* **69**, 2863 (1992).
- 6 S. R. White, *Phys. Rev. B* **48**, 10345 (1993).
- 7 G. K.-L. Chan and S. Sharma, *Annu. Rev. Phys. Chem.* **62**, 465 (2011).
- 8 J. J. Eriksen, F. Lipparini, and J. Gauss, *J. Phys. Chem. Lett.* **8**, 4633 (2017).
- 9 J. J. Eriksen and J. Gauss, *J. Chem. Theory Comput.* **14**, 5180 (2018).
- 10 J. J. Eriksen and J. Gauss, *J. Chem. Theory Comput.* **15**, 4873 (2019).
- 11 J. J. Eriksen and J. Gauss, *J. Phys. Chem. Lett.* **27**, 7910 (2019).
- 12 M. Motta and S. Zhang, *WIREs Comput. Mol. Sci.* **8**, e1364 (2018).
- 13 J. Lee, F. D. Malone, and D. R. Reichman, *J. Chem. Phys.* **153**, 126101 (2020).
- 14 E. Xu, M. Uejima, and S. L. Ten-no, *Phys. Rev. Lett.* **121**, 113001 (2018).
- 15 E. Xu, M. Uejima, and S. L. Ten-no, *J. Phys. Chem. Lett.* **11**, 9775 (2020).
- 16 I. Magoulas, K. Gururangan, P. Piecuch, J. E. Deustua, and J. Shen, *J. Chem. Theory Comput.* **17**, 4006 (2021).
- 17 G. Karthik, D. J. Emiliano, S. Jun, and P. Piotr, *J. Chem. Phys.* **155**, 174114 (2021).
- 18 P.-F. Loos, A. Scemama, and D. Jacquemin, *J. Phys. Chem. Lett.* **11**, 2374 (2020).
- 19 J. J. Eriksen, T. A. Anderson, J. E. Deustua, K. Ghanem, D. Hait, M. R. Hoffmann, S. Lee, D. S. Levine, I. Magoulas, J. Shen, N. M. Tubman, K. B. Whaley, E. Xu, Y. Yao, N. Zhang, A. Alavi, G. K.-L. Chan, M. Head-Gordon, W. Liu, P. Piecuch, S. Sharma, S. L. Ten-no, C. J. Umrigar, and J. Gauss, *J. Phys. Chem. Lett.* **11**, 8922 (2020).
- 20 J. J. Eriksen, *J. Phys. Chem. Lett.* **12**, 418 (2021).
- 21 P.-F. Loos, Y. Damour, and A. Scemama, *J. Chem. Phys.* **153**, 176101 (2020).
- 22 M. Caffarel, T. Applencourt, E. Giner, and A. Scemama, *J. Chem. Phys.* **144**, 151103 (2016).
- 23 A. A. Holmes, C. J. Umrigar, and S. Sharma, *J. Chem. Phys.* **147**, 164111 (2017).
- 24 A. D. Chien, A. A. Holmes, M. Otten, C. J. Umrigar, S. Sharma, and P. M. Zimmerman, *J. Phys. Chem. A* **122**, 2714 (2018).
- 25 P. F. Loos, A. Scemama, A. Blondel, Y. Garniron, M. Caffarel, and D. Jacquemin, *J. Chem. Theory Comput.* **14**, 4360 (2018).
- 26 P.-F. Loos, M. Boggio-Pasqua, A. Scemama, M. Caffarel, and D. Jacquemin, *J. Chem. Theory Comput.* **15**, 1939 (2019).
- 27 P. F. Loos, F. Lipparini, M. Boggio-Pasqua, A. Scemama, and D. Jacquemin, *J. Chem. Theory Comput.* **16**, 1711 (2020).
- 28 M. Véril, A. Scemama, M. Caffarel, F. Lipparini, M. Boggio-Pasqua, D. Jacquemin, and P.-F. Loos, *WIREs Comput. Mol. Sci.* **11**, e1517.
- 29 E. Giner, A. Scemama, and M. Caffarel, *Can. J. Chem.* **91**, 879 (2013).
- 30 E. Giner, A. Scemama, and M. Caffarel, *J. Chem. Phys.* **142**, 044115 (2015).
- 31 M. Caffarel, T. Applencourt, E. Giner, and A. Scemama, “Using cipsi nodes in diffusion monte carlo,” in *Recent Progress in Quantum Monte Carlo* (2016) Chap. 2, pp. 15–46.
- 32 Y. Garniron, A. Scemama, P.-F. Loos, and M. Caffarel, *J. Chem. Phys.* **147**, 034101 (2017).
- 33 Y. Garniron, A. Scemama, E. Giner, M. Caffarel, and P. F. Loos, *J. Chem. Phys.* **149**, 064103 (2018).
- 34 Y. Garniron, K. Gasperich, T. Applencourt, A. Benali, A. Ferté, J. Paquier, B. Pradines, R. Assaraf, P. Reinhardt, J. Toulouse, P. Barbaresco, N. Renon, G. David, J. P. Malrieu, M. Véril, M. Caffarel, P. F. Loos, E. Giner, and A. Scemama, *J. Chem. Theory Comput.* **15**, 3591 (2019).
- 35 Y. Damour, M. Véril, F. Kossoski, M. Caffarel, D. Jacquemin, A. Scemama, and P.-F. Loos, *J. Chem. Phys.* **155**, 134104 (2021).
- 36 Y. Damour, R. Quintero-Monsebaiz, M. Caffarel, D. Jacquemin, F. Kossoski, A. Scemama, and P.-F. Loos, *J. Chem. Theory Comput.* **19**, 221 (2023).
- 37 J. B. Schriber and F. A. Evangelista, *J. Chem. Phys.* **144**, 161106 (2016).
- 38 N. M. Tubman, J. Lee, T. Y. Takeshita, M. Head-Gordon, and K. B. Whaley, *J. Chem. Phys.* **145**, 044112 (2016).
- 39 N. M. Tubman, D. S. Levine, D. Hait, M. Head-Gordon, and K. B. Whaley, *arXiv* (2018), 10.48550/arXiv.1808.02049, 1808.02049.
- 40 N. M. Tubman, C. D. Freeman, D. S. Levine, D. Hait, M. Head-Gordon, and K. B. Whaley, *J. Chem. Theory Comput.* **16**, 2139 (2020).
- 41 A. A. Holmes, N. M. Tubman, and C. J. Umrigar, *J. Chem. Theory Comput.* **12**, 3674 (2016).
- 42 S. Sharma, A. A. Holmes, G. Jeanmairet, A. Alavi, and C. J. Umrigar, *J. Chem. Theory Comput.* **13**, 1595 (2017).
- 43 Y. Yao, E. Giner, J. Li, J. Toulouse, and C. J. Umrigar, *J. Chem. Phys.* **153**, 124117 (2020).
- 44 Y. Yao and C. J. Umrigar, *J. Chem. Theory Comput.* **17**, 4183 (2021).
- 45 H. R. Larsson, H. Zhai, C. J. Umrigar, and G. K.-L. Chan, *J. Am. Chem. Soc.* **144**, 15932 (2022).
- 46 W. Liu and M. Hoffmann, *Theor. Chem. Acc.* **133**, 1481 (2014).
- 47 W. Liu and M. R. Hoffmann, *J. Chem. Theory Comput.* **12**, 1169 (2016).
- 48 Y. Lei, W. Liu, and M. R. Hoffmann, *Mol. Phys.* **115**, 2696 (2017).
- 49 N. Zhang, W. Liu, and M. R. Hoffmann, *J. Chem. Theory Comput.* **16**, 2296 (2020).
- 50 N. Zhang, W. Liu, and M. R. Hoffmann, *J. Chem. Theory Comput.* **17**, 949 (2021).
- 51 J. P. Coe, *J. Chem. Theory Comput.* **14**, 5739 (2018).
- 52 J. P. Coe, A. Moreno Carrascosa, M. Simmermacher, A. Kirrander, and M. J. Paterson, *J. Chem. Theory Comput.* **18**, 6690 (2022).
- 53 G. H. Booth, A. J. W. Thom, and A. Alavi, *J. Chem. Phys.* **131**, 054106 (2009).
- 54 D. Cleland, G. H. Booth, and A. Alavi, *J. Chem. Phys.* **132**, 041103 (2010).
- 55 N. S. Blunt, S. D. Smart, G. H. Booth, and A. Alavi, *J. Chem. Phys.* **143**, 134117 (2015).
- 56 K. Ghanem, A. Y. Lozovoi, and A. Alavi, *J. Chem. Phys.* **151**, 224108 (2019).
- 57 J. E. Deustua, J. Shen, and P. Piecuch, *Phys. Rev. Lett.* **119**, 223003 (2017).
- 58 J. E. Deustua, I. Magoulas, J. Shen, and P. Piecuch, *J. Chem. Phys.* **149**, 151101 (2018).
- 59 H. G. A. Burton, *J. Chem. Theory Comput.* **18**, 1512 (2022).
- 60 A. Marie, H. G. A. Burton, and P.-F. Loos, *J. Phys.: Condens. Matter* **33**, 283001 (2021).

<sup>61</sup>D. Z. Goodson, *WIREs Comput. Mol. Sci.* **2**, 743 (2012).

<sup>62</sup>D. Z. Goodson, in *Mathematical Physics in Theoretical Chemistry*, Developments in Physical & Theoretical Chemistry, edited by S. Blinder and J. House (Elsevier, 2019) p. 295.

<sup>63</sup>I. L. Mayer and B. Y. Tong, **18**, 3297 (1985).