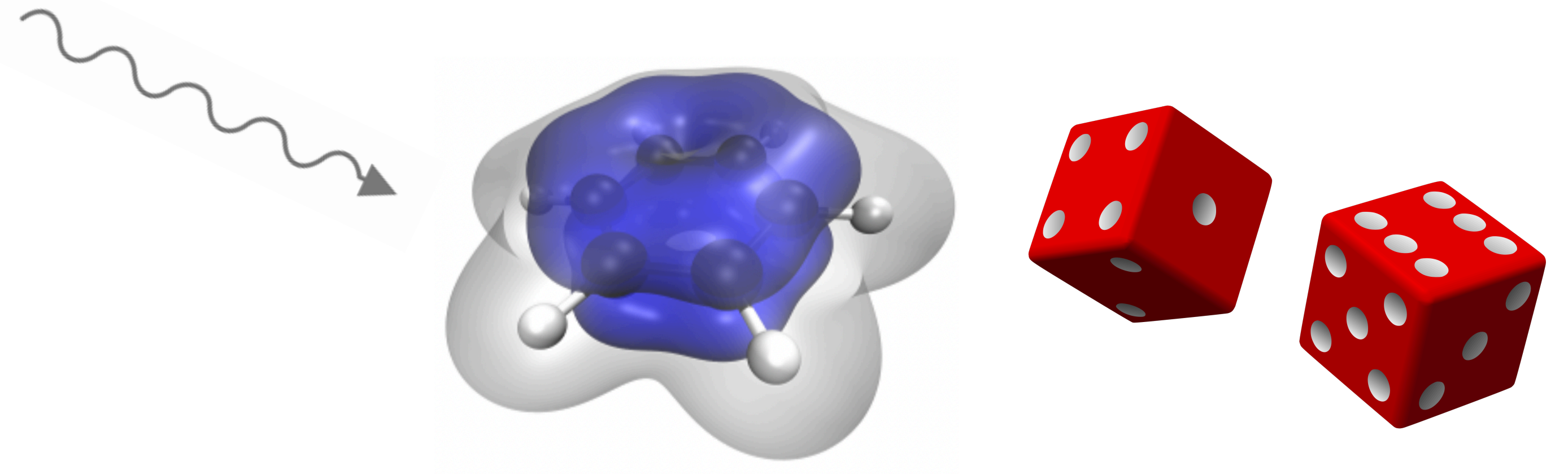


Natural Quantum Monte Carlo Computation of Excited States

ECT* Workshop

Many-Body Quantum Physics With Machine Learning

6 September 2023



David Pfau

Google DeepMind, Imperial College London

pfau@google.com



Simon Axelrod



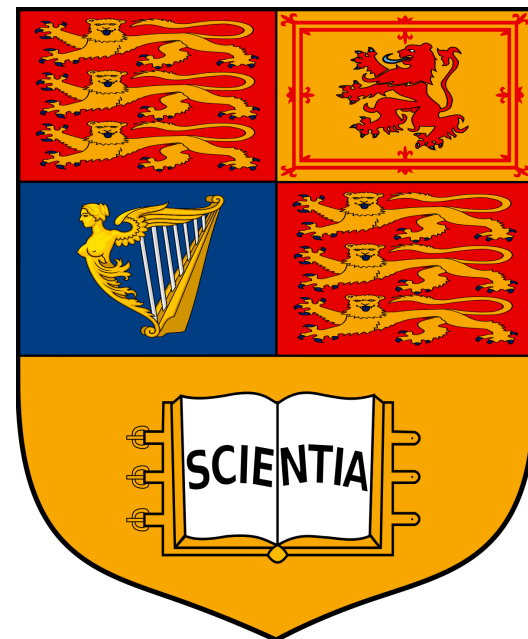
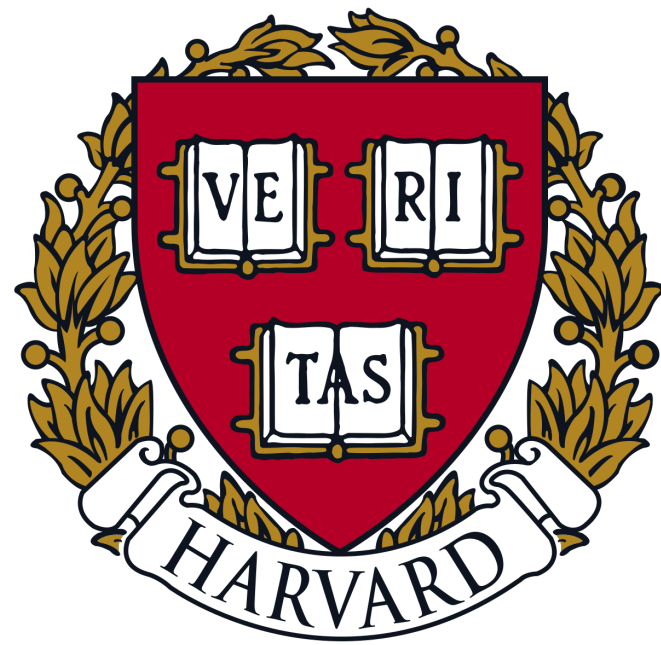
Halvard Sutterud



Ingrid von Glehn



James Spencer



Excited States

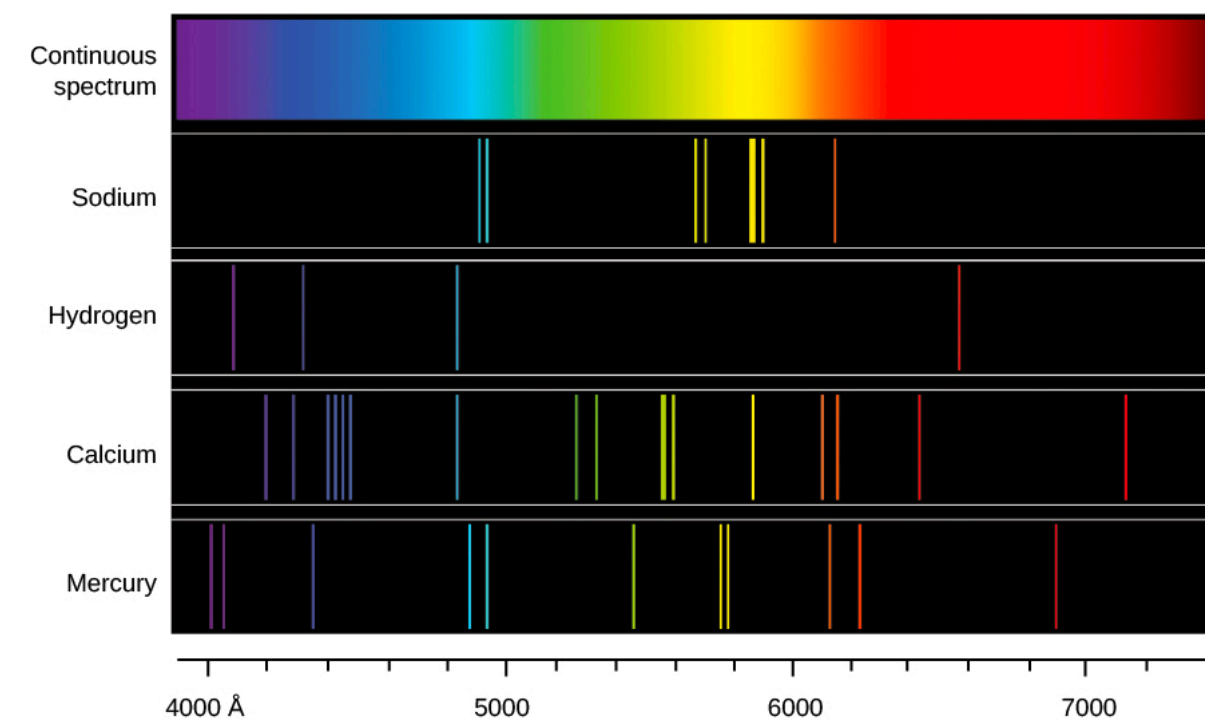
$$\hat{H}\psi = E\psi$$

$$E \neq \min\langle\psi^\dagger \hat{H}\psi\rangle$$

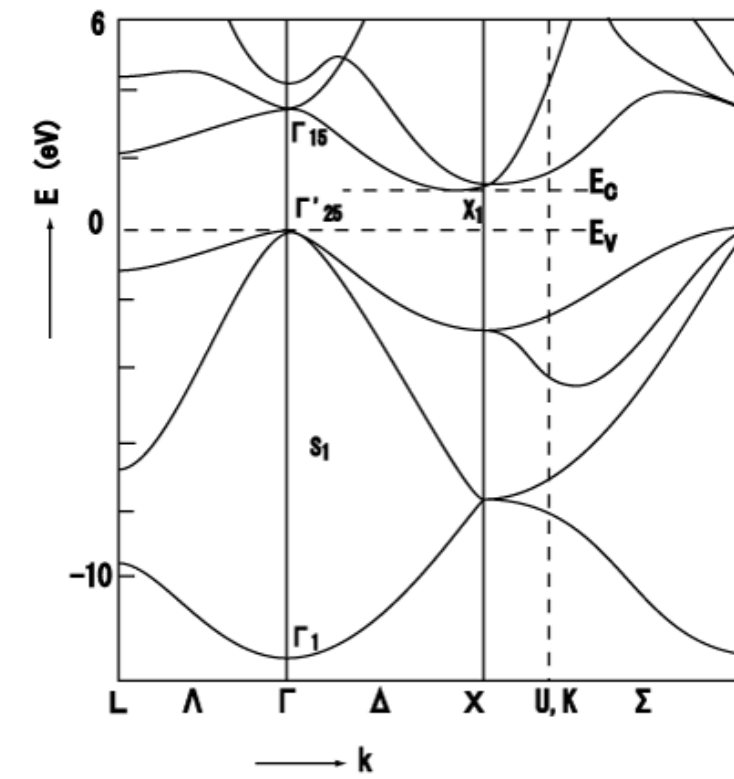
Excited States



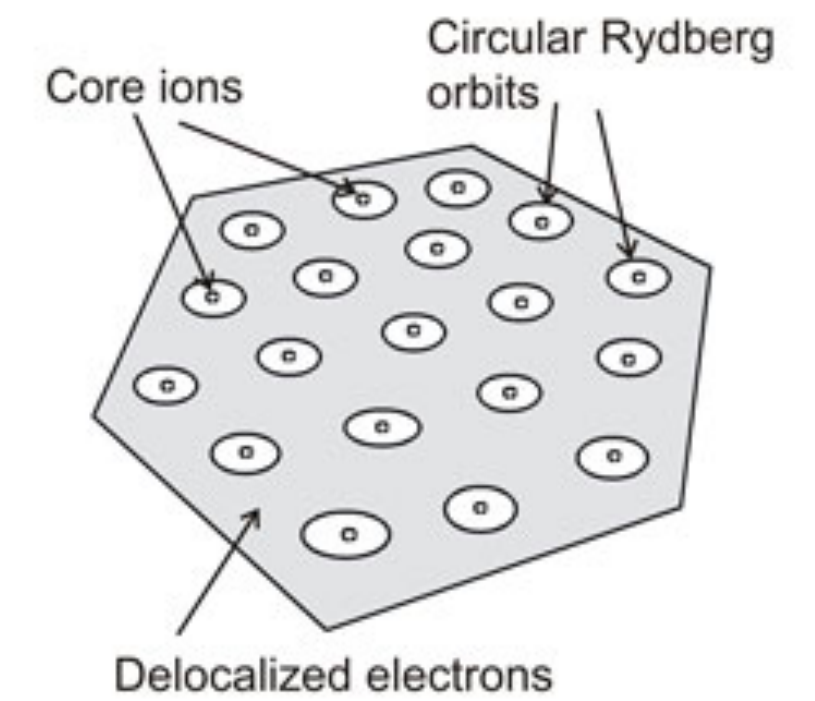
Astrochemistry



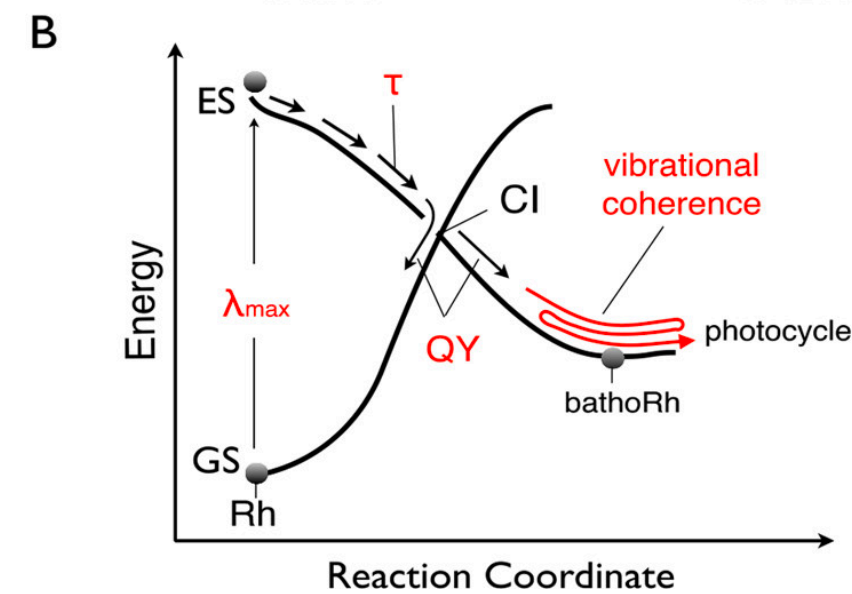
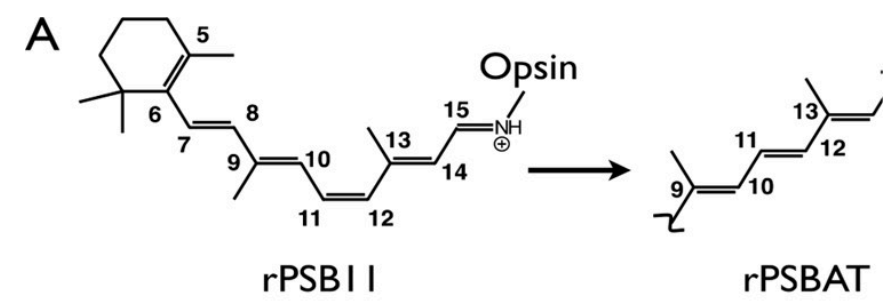
Spectroscopy



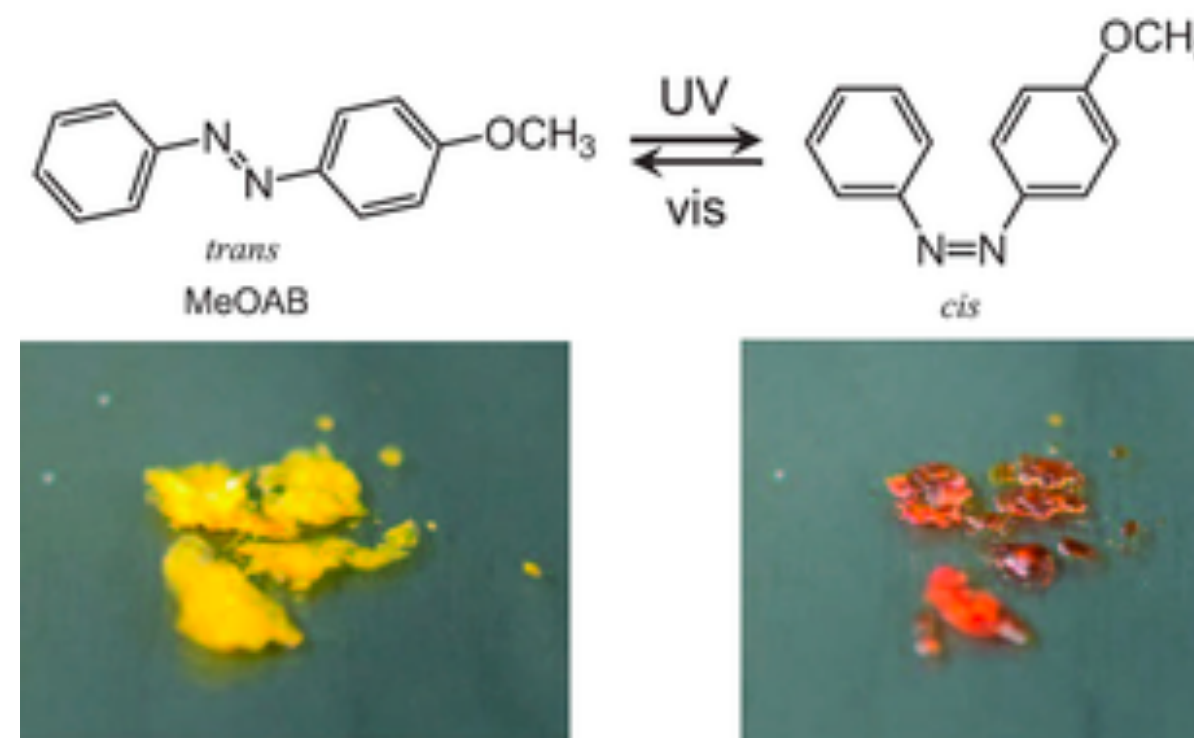
Band Structure



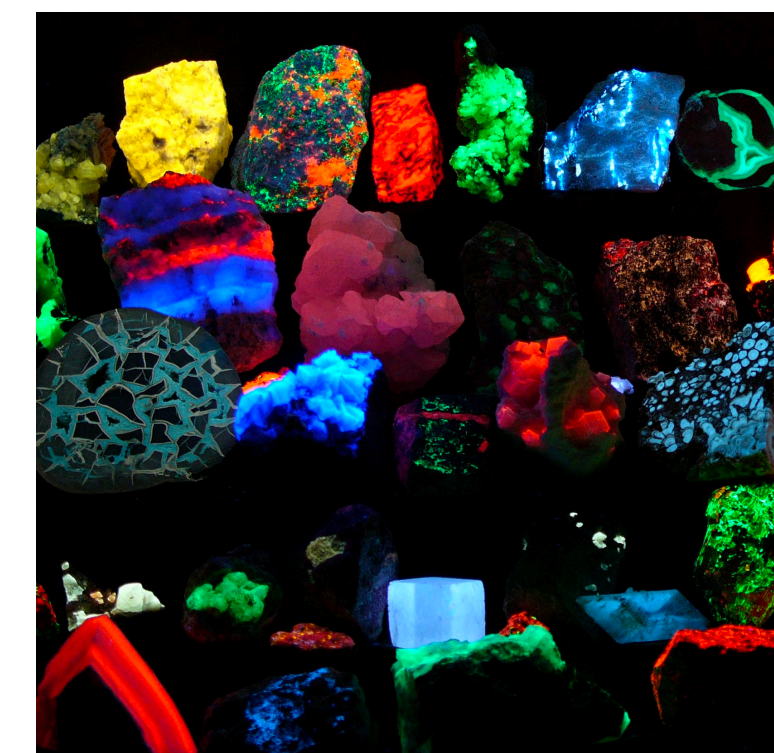
Rydberg Matter



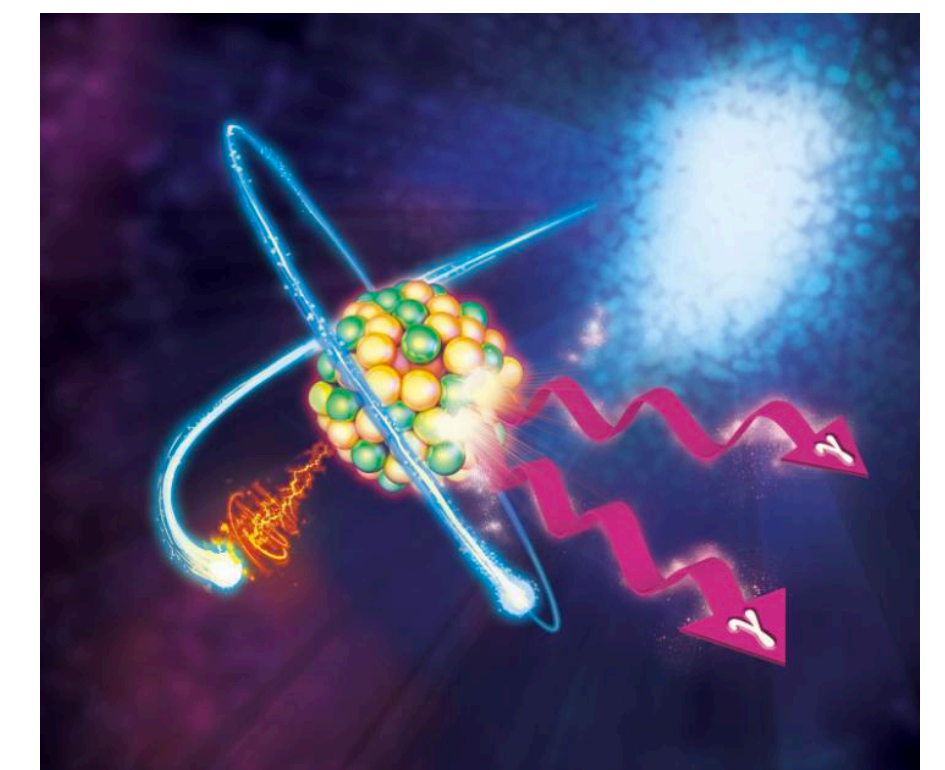
Biochemistry of vision



Photoswitchable Molecules

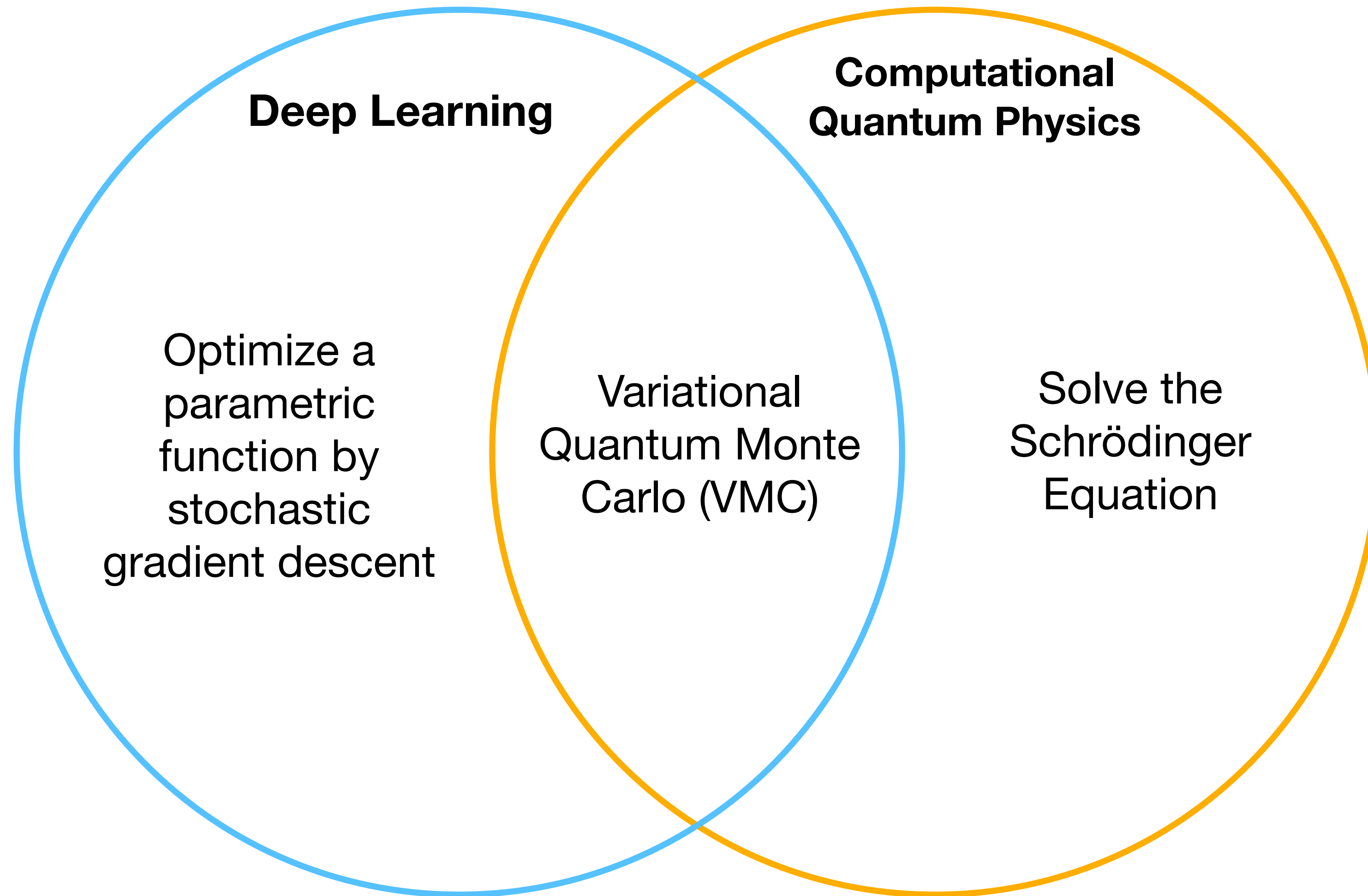


Fluorescence and Phosphorescence



Nuclear Excitations

Deep Learning and Many-Body Quantum Physics



Deep Learning and Many-Body Quantum Physics

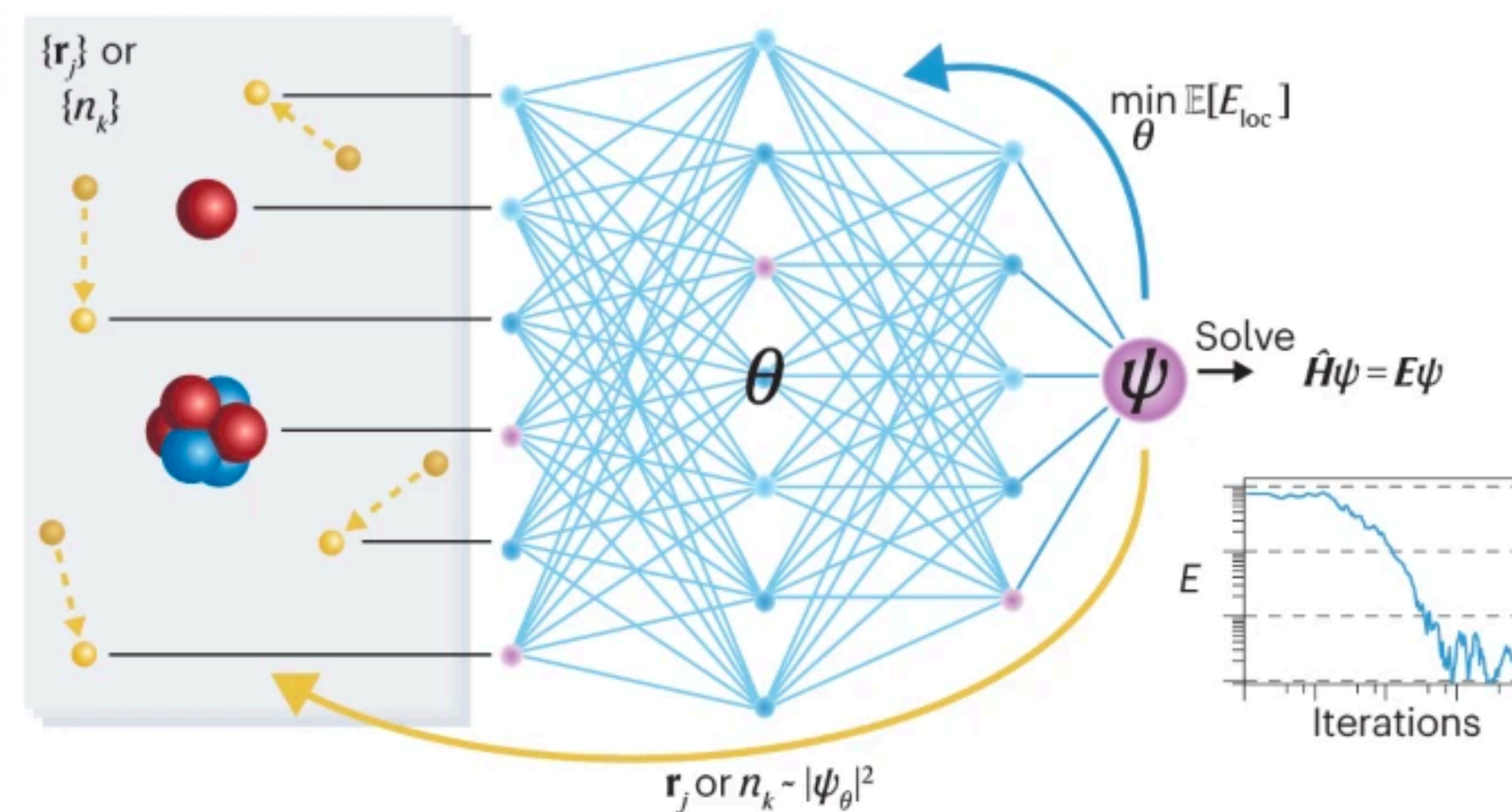
nature reviews chemistry

Review Article | [Published: 09 August 2023](#)

Ab initio quantum chemistry with neural-network wavefunctions

[Jan Hermann](#), [James Spencer](#), [Kenny Choo](#), [Antonio Mezzacapo](#), [W. M. C. Foulkes](#), [David Pfau](#) ,

[Giuseppe Carleo](#)  & [Frank Noé](#) 



Deep Learning and Many-Body Quantum Physics

“Classic” VMC

Long MCMC runs between parameter updates

Small number of parameter updates

Few “nonlinear” parameters, many linear coefficients

Large effective batch size - nearly deterministic

Neural Network VMC

Short MCMC runs between parameter updates

Large number of parameter updates

Many “nonlinear” parameters, no linear coefficients

Small batch size - very stochastic

Part I: Ground State VMC

Energy Minimisation

$$\min_{\psi} \frac{\langle \psi \hat{H} \psi \rangle}{\langle \psi^2 \rangle} =$$

Energy Minimisation

$$\min_{\psi} \frac{\langle \psi \hat{H} \psi \rangle}{\langle \psi^2 \rangle} = \min_{\psi} \frac{\mathbb{E}_q \left[q^{-1}(\mathbf{x}) \psi(\mathbf{x}) \hat{H} \psi(\mathbf{x}) \right]}{\mathbb{E}_q \left[q^{-1}(\mathbf{x}) \psi^2(\mathbf{x}) \right]}$$

← Empirical estimates from finite data will be biased!

$$= \min_{\psi} \mathbb{E}_{\psi^2} \left[\psi^{-1}(\mathbf{x}) \hat{H} \psi(\mathbf{x}) \right]$$

← Bias only goes away if ψ^2 is used as Monte Carlo distribution

ψ^2 is the *natural* distribution to sample from to generate unbiased energy and gradient estimates

Energy Minimisation - Excited States

Rayleigh Quotient

$$\frac{\langle \psi \hat{H} \psi \rangle}{\langle \psi^2 \rangle}$$

Only gives states which *span* the bottom eigenfunctions

Works even if ψ_1, \dots, ψ_K are not orthogonal

What is the natural generalization of ψ^2 for Monte Carlo?

Generalized Rayleigh Quotient

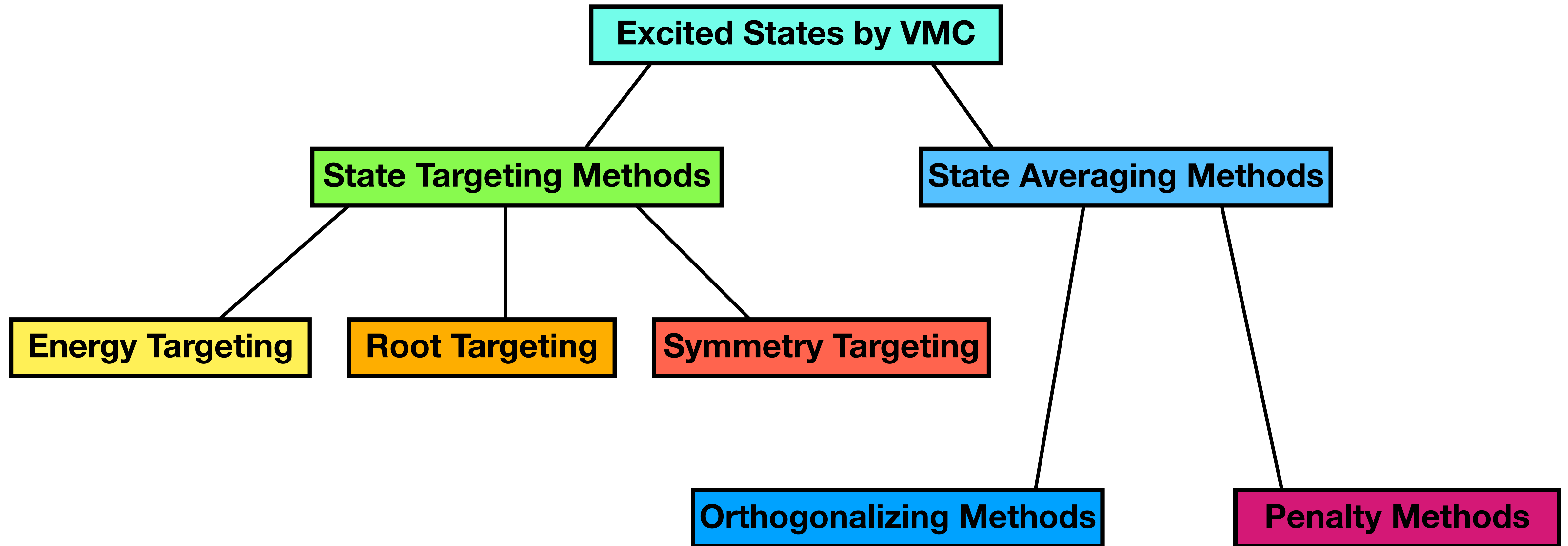
$$\text{Tr} [\mathbf{S}^{-1} \mathbf{H}]$$

$$\mathbf{S} = \begin{pmatrix} \langle \psi_1^2 \rangle & \dots & \langle \psi_1 \psi_K \rangle \\ \vdots & & \vdots \\ \langle \psi_K \psi_1 \rangle & \dots & \langle \psi_K^2 \rangle \end{pmatrix}$$

$$\mathbf{H} = \begin{pmatrix} \langle \psi_1 \hat{H} \psi_1 \rangle & \dots & \langle \psi_1 \hat{H} \psi_K \rangle \\ \vdots & & \vdots \\ \langle \psi_K \hat{H} \psi_1 \rangle & \dots & \langle \psi_K \hat{H} \psi_K \rangle \end{pmatrix}$$

Part II: Existing Excited States Methods

Taxonomy of Excited State Methods



State Targeting Methods

- Compute a *single* excited state
- Generally similar computational cost to computing ground state

State Averaging Methods

- Compute *all* states up to a certain number
- Higher computational cost the more states you compute, but a more “complete” solution

Energy Targeting Methods

- Compute the lowest excited state above a certain energy
- Energy or variance minimization versions
- Requires *a priori* knowledge of desired energy range
- Variance minimization tends to drift away from desired state

$$W(\Psi) = \frac{\langle \Psi | (\omega - \hat{H})^2 | \Psi \rangle}{\langle \Psi | \Psi \rangle} = (\omega - E)^2 + \sigma^2$$

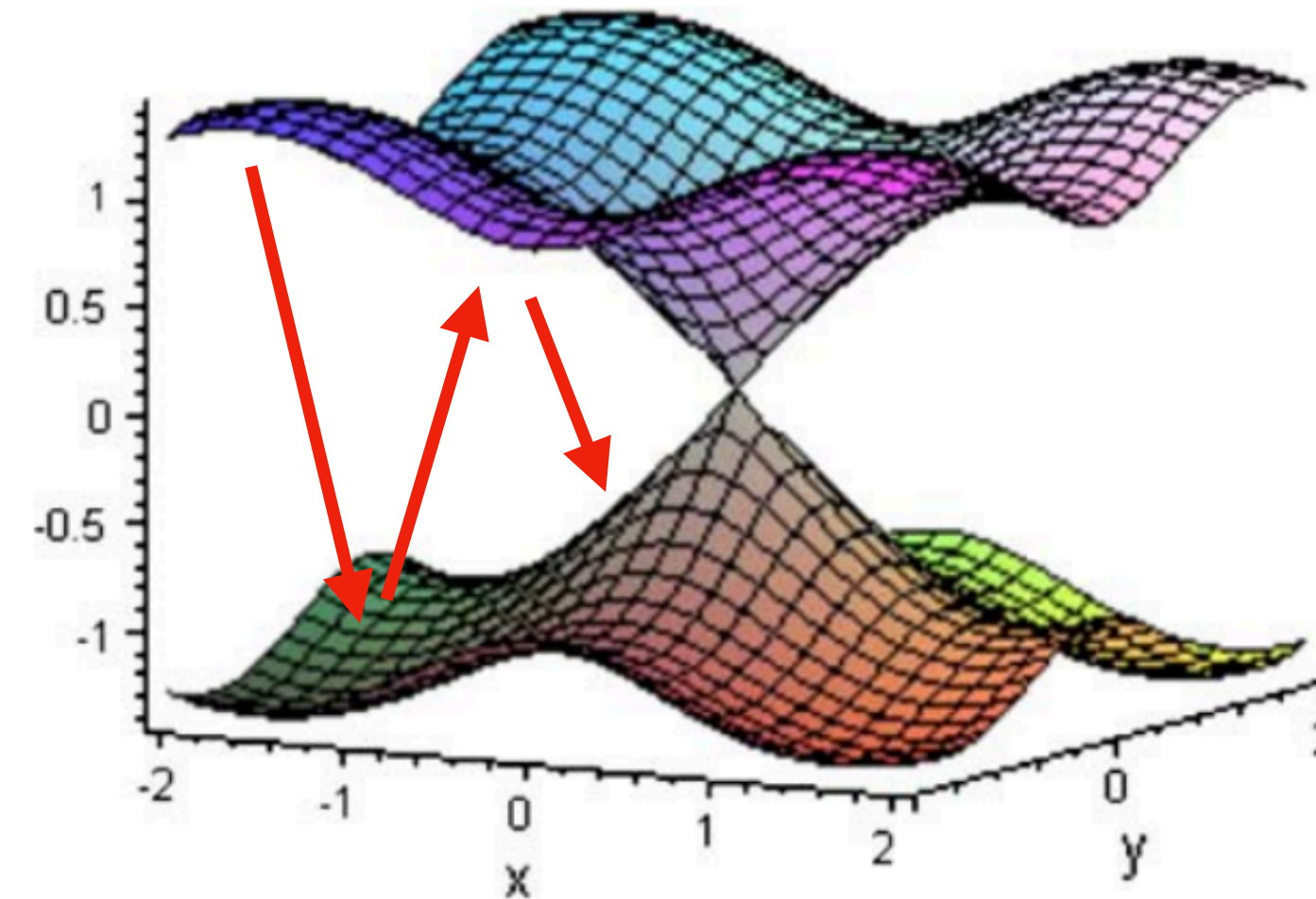
$$\Omega(\Psi) = \frac{\langle \Psi | (\omega - \hat{H}) | \Psi \rangle}{\langle \Psi | (\omega - \hat{H})^2 | \Psi \rangle} = \frac{\omega - E}{(\omega - E)^2 + \sigma^2},$$

Shea and Neuscammann, JCTC (2017)
Cuzzocrea *et al.*, JCTC (2020)

Root Targeting Methods

- Target the K th root of generalized eigenvalue equation
- Still relies on noisy Monte Carlo estimates, so eigenvalues will be biased
- Prone to *root flipping* - order of roots may change during optimization, may not converge

$$\begin{pmatrix} E_0 & \mathbf{g}_R^T/2 \\ \mathbf{g}_L/2 & \bar{\mathbf{H}} \end{pmatrix} \begin{pmatrix} 1 \\ \Delta \mathbf{p} \end{pmatrix} = E \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \bar{\mathbf{S}} \end{pmatrix} \begin{pmatrix} 1 \\ \Delta \mathbf{p} \end{pmatrix}$$



Zimmerman *et al.*, J. Chem. Phys. (2009)

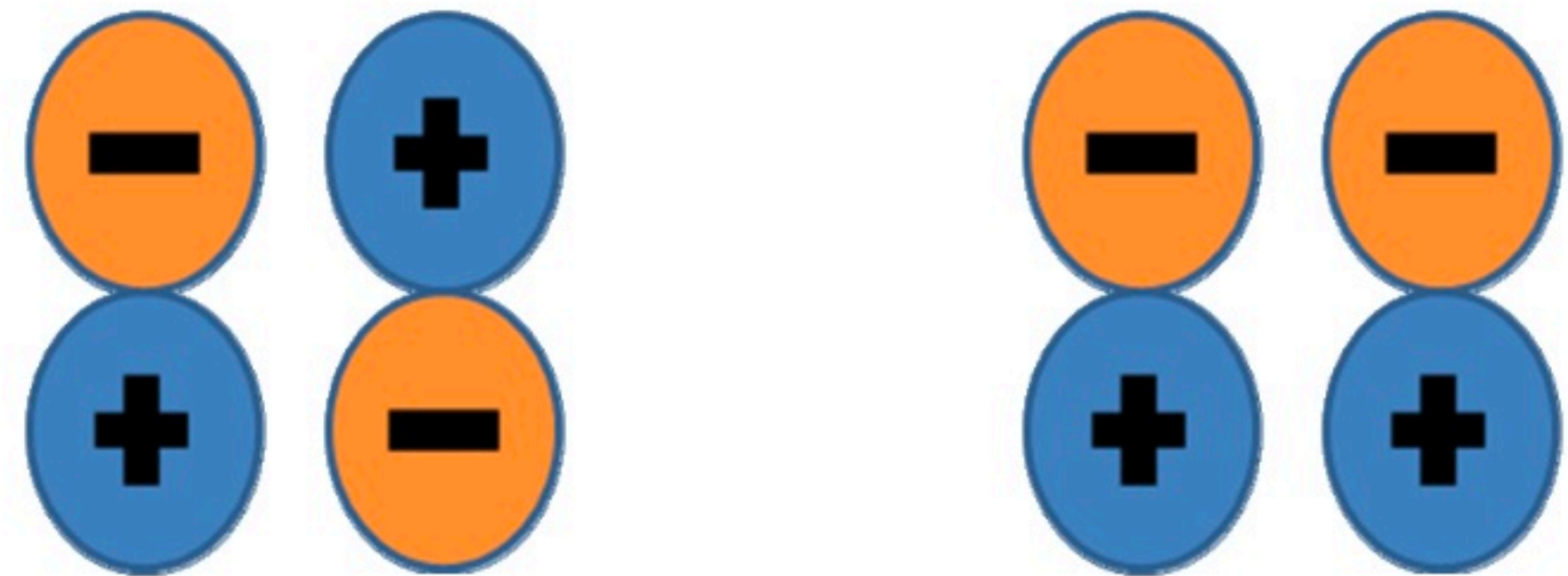
Dorando, Hachmann and Chan, J. Chem. Phys. (2007)

Lewin, J. Math. Chem. (2008)

Symmetry Targeting Methods

- Restrict the Ansatz to obey certain symmetries other than ground state (e.g. spin, parity, reflection, rotation)
- Requires *a priori* knowledge of the desired symmetry
- Only works for lowest state of the desired symmetry

$$\log \Psi(\boldsymbol{\sigma}) = \sum_{i=1}^{\nu} r_{i,\boldsymbol{\sigma}} \log \omega_i + \log [\tilde{\Psi}(\boldsymbol{\sigma}_{\text{canonical}})],$$



LibreText Chemistry

$$\psi(\mathbf{x}) = \tilde{\psi}(\mathbf{x}) \pm \tilde{\psi}(-\mathbf{x})$$

State Averaging Methods

- Minimize the total (weighted) energy
- Need to keep the states orthogonal somehow
- Need to choose a sampling distribution

$$\min_{\psi_1, \dots, \psi_K} \sum_k w_k \frac{\langle \psi_k \hat{H} \psi_k \rangle}{\langle \psi_k^2 \rangle}$$

“Orthogonalizing” Methods

- Usually what people mean when they talk about state averaging approaches
- Either design states to always be orthogonal, or periodically “reorthogonalize” them
- Usually assumes Ansatz is a linear combination of basis functions - not applicable to neural networks
- Still relies on noisy Monte Carlo estimates for nonlinear orthogonalization step

$$\sum_j H_{ij} c_j^I = E_I \sum_j S_{ij} c_j^I$$

$$H_{ij} = \langle \mathcal{J} D_i | \hat{H} | \mathcal{J} D_j \rangle \quad \text{and} \quad S_{ij} = \langle \mathcal{J} D_i | \mathcal{J} D_j \rangle$$

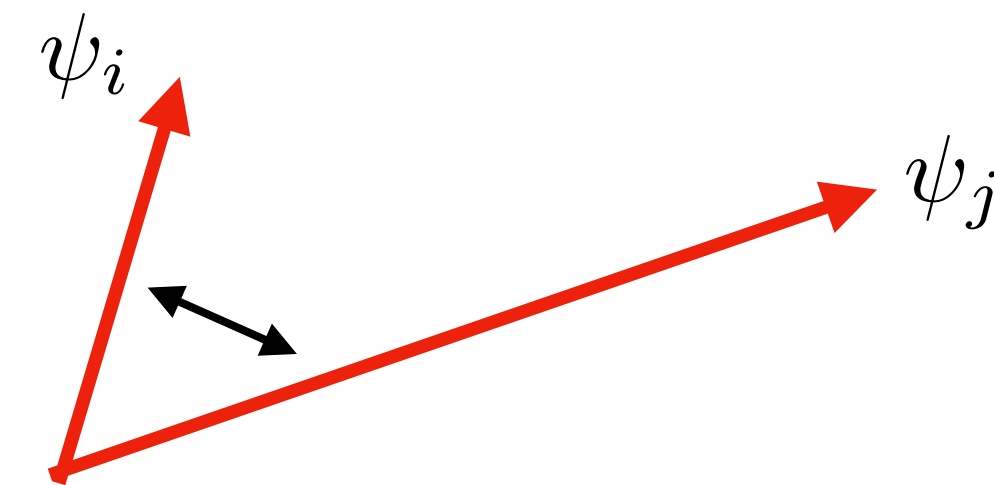
$$\mathbf{x} \sim \frac{1}{K} \sum_k \psi_k^2(\mathbf{x})$$

Penalty Methods

- Add penalty term that pushes states to be orthogonal
- Works with neural networks
- Mixes very differently behaved terms together in the loss, plus the penalty weight must be tuned
- States can collapse onto each other if not careful
- Changing the loss to push states apart more strongly introduces bias

$$\min_{\psi_1, \dots, \psi_K} \sum_k w_k \frac{\langle \psi_k \hat{H} \psi_k \rangle}{\langle \psi_k^2 \rangle} + \lambda \sum_{i < j} f \left(\frac{\langle \psi_i \psi_j \rangle}{\sqrt{\langle \psi_i^2 \rangle} \sqrt{\langle \psi_j^2 \rangle}} \right)$$

$$\mathbf{x}^k \sim \psi_k^2(\mathbf{x}^k)$$



Choo *et al.*, PRL (2018)

Pathak *et al.*, J. Chem. Phys. (2021)

Entwistle *et al.*, Nat. Comm. (2023)

Part III: Natural Excited States

Energy Minimisation - Excited States

Generalized Rayleigh Quotient

Can we generate *unbiased* samples of the total energy and its gradient?

Can we recover individual energies and other observables in addition to the total?

$$\text{Tr} [\mathbf{S}^{-1} \mathbf{H}]$$

$$\mathbf{S} = \begin{pmatrix} \langle \psi_1^2 \rangle & \dots & \langle \psi_1 \psi_K \rangle \\ \vdots & & \vdots \\ \langle \psi_K \psi_1 \rangle & \dots & \langle \psi_K^2 \rangle \end{pmatrix}$$

$$\mathbf{H} = \begin{pmatrix} \langle \psi_1 \hat{H} \psi_1 \rangle & \dots & \langle \psi_1 \hat{H} \psi_K \rangle \\ \vdots & & \vdots \\ \langle \psi_K \hat{H} \psi_1 \rangle & \dots & \langle \psi_K \hat{H} \psi_K \rangle \end{pmatrix}$$

Converting Back to a Ground State Problem

$$\text{Tr} [\mathbf{S}^{-1} \mathbf{H}] = \frac{\langle \Psi \hat{\mathcal{H}} \Psi \rangle}{\langle \Psi^2 \rangle}$$

$$\hat{\mathcal{H}} = \hat{H}_1 \oplus \dots \oplus \hat{H}_K$$

$$\Psi (\mathbf{x}^1, \dots, \mathbf{x}^K) = \det \mathbf{\Psi} (\mathbf{x}^1, \dots, \mathbf{x}^K)$$

$$\mathbf{\Psi} (\mathbf{x}^1, \dots, \mathbf{x}^K) = \begin{pmatrix} \psi_1(\mathbf{x}^1) & \dots & \psi_K(\mathbf{x}^1) \\ \vdots & & \vdots \\ \psi_1(\mathbf{x}^K) & \dots & \psi_K(\mathbf{x}^K) \end{pmatrix}$$

Convert the problem of finding many excited states of one system into the problem of finding the ground state of a “meta”-system

Converting Back to a Ground State Problem

$$\text{Tr} [\mathbf{S}^{-1} \mathbf{H}] = \frac{\langle \Psi \hat{\mathcal{H}} \Psi \rangle}{\langle \Psi^2 \rangle}$$

$$\hat{\mathcal{H}} = \hat{H}_1 \oplus \dots \oplus \hat{H}_K$$

$$\Psi (\mathbf{x}^1, \dots, \mathbf{x}^K) = \det \mathbf{\Psi} (\mathbf{x}^1, \dots, \mathbf{x}^K)$$

$$\mathbf{\Psi} (\mathbf{x}^1, \dots, \mathbf{x}^K) = \begin{pmatrix} \psi_1(\mathbf{x}^1) & \dots & \psi_K(\mathbf{x}^1) \\ \vdots & & \vdots \\ \psi_1(\mathbf{x}^K) & \dots & \psi_K(\mathbf{x}^K) \end{pmatrix}$$

The determinant keeps the states from collapsing onto each other, without an orthogonalization penalty!

Natural Excited States

$$\frac{\langle \Psi \hat{\mathcal{H}} \Psi \rangle}{\langle \Psi^2 \rangle} = \mathbb{E}_{\mathbf{x} \sim \Psi^2} \left[\Psi^{-1}(\mathbf{x}) \hat{\mathcal{H}} \Psi(\mathbf{x}) \right] = \text{Tr} \left[\mathbb{E}_{\mathbf{x} \sim \Psi^2} \left[\Psi^{-1}(\mathbf{x}) \hat{H} \Psi(\mathbf{x}) \right] \right]$$

Local Energy Matrix

$$\hat{H} \Psi(\mathbf{x}) = \begin{pmatrix} \hat{H} \psi_1(\mathbf{x}^1) & \dots & \hat{H} \psi_K(\mathbf{x}^1) \\ \vdots & & \vdots \\ \hat{H} \psi_1(\mathbf{x}^K) & \dots & \hat{H} \psi_K(\mathbf{x}^K) \end{pmatrix}$$

Applying a similar derivation to the Monte Carlo objective yields a simple generalization of the local energy

Natural Excited States

~~Can we generate *unbiased* samples of the total energy and its gradient?~~

Can we recover individual energies and other observables in addition to the total?

Natural Excited States - Energies

$$\psi_i = \sum_j A_{ij} \psi_j^*$$

We only get a *linear combination* of solutions

$$\Psi^{*-1}(\mathbf{x}) \hat{H} \Psi^*(\mathbf{x}) = \mathbf{A}^{-1} \begin{pmatrix} E_1 & & \\ & \ddots & \\ & & E_K \end{pmatrix} \mathbf{A}$$

Eigenvalues of local energy matrix are energies!

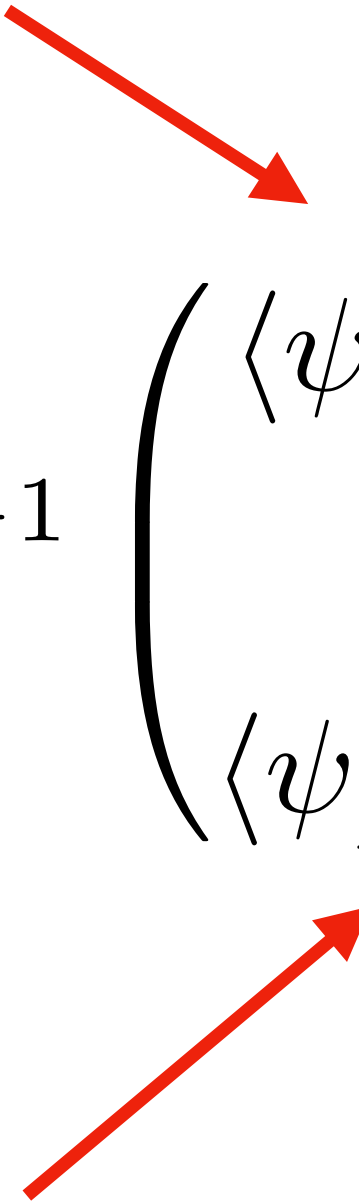
$$\mathbb{E}_{\Psi^2} \left[\Psi^{*-1}(\mathbf{x}) \hat{H} \Psi^*(\mathbf{x}) \right] = \mathbf{U}^{-1} \begin{pmatrix} E_1 & & \\ & \ddots & \\ & & E_K \end{pmatrix} \mathbf{U}$$

Can recover mixing matrix *up to a constant in each column*

$$\mathbf{U} = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_K \end{pmatrix} \mathbf{A}$$

Natural Excited States - Observables

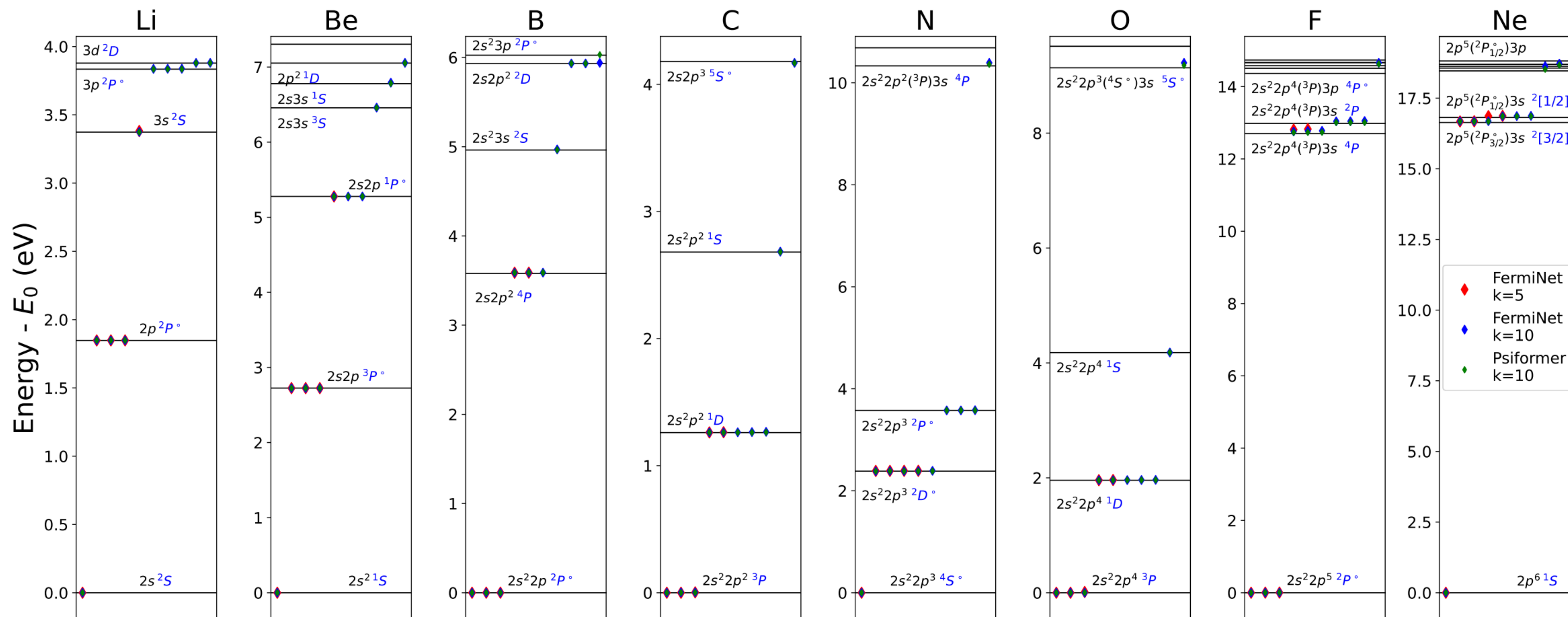
For *on-diagonal* observables, changing the coordinates of the local observable matrix is enough


$$\mathbb{E}_{\Psi_2} \left[\mathbf{\Psi}^{*-1}(\mathbf{x}) \hat{O} \mathbf{\Psi}^*(\mathbf{x}) \right] = \mathbf{U}^{-1} \begin{pmatrix} \langle \psi_1^* \hat{O} \psi_1^* \rangle & \dots & \langle \psi_1^* \hat{O} \psi_K^* \rangle \\ \vdots & & \\ \langle \psi_K^* \hat{O} \psi_1^* \rangle & \dots & \langle \psi_K^* \hat{O} \psi_K^* \rangle \end{pmatrix} \mathbf{U}$$

For *off-diagonal* observables, multiplying by the transpose cancels out the non-identifiable scale factor

Part IV: Results

Atomic Spectra



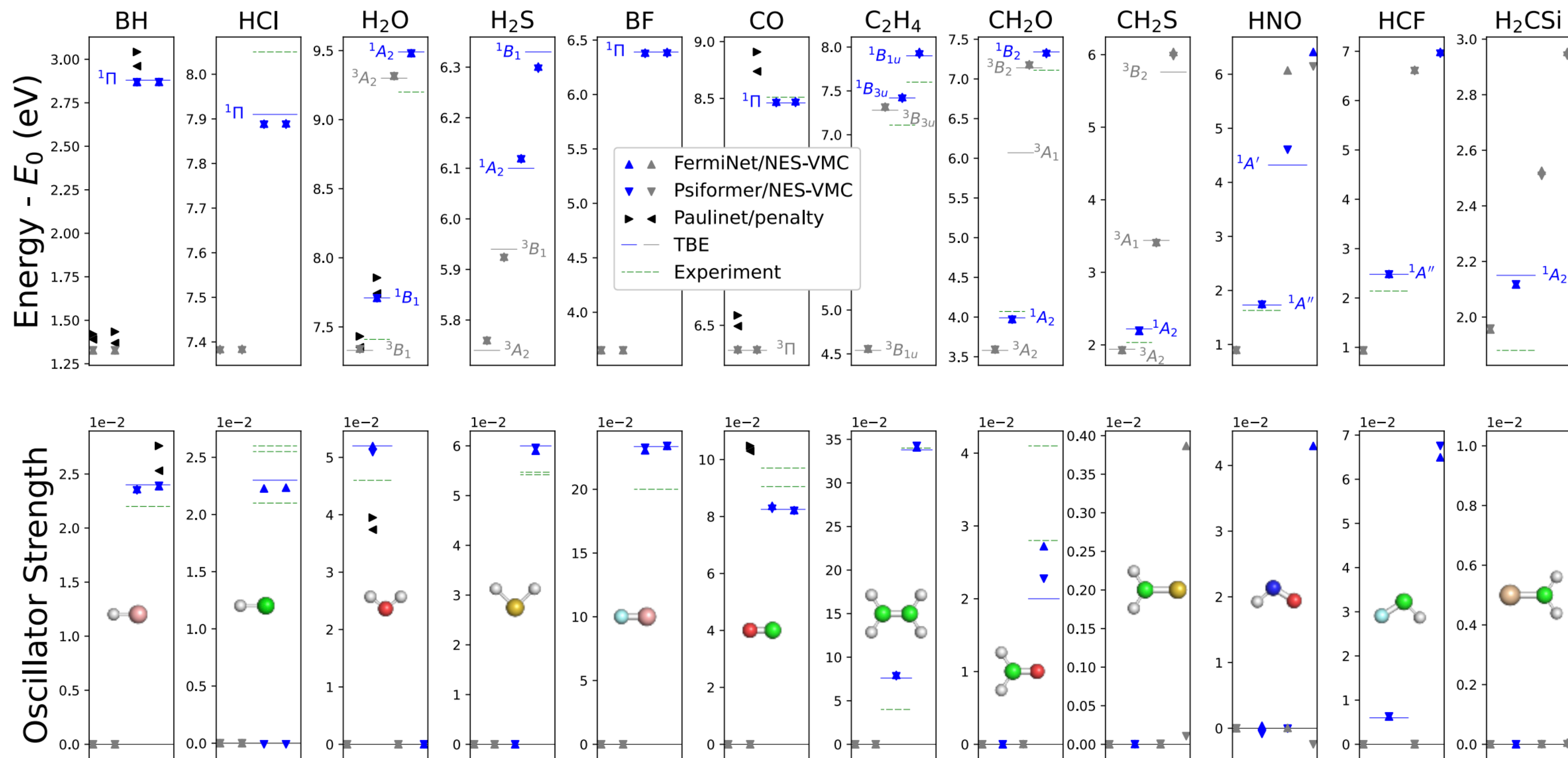
Small Molecules and Oscillator Strengths

- Transition dipole moment - vector that measures the dipole moment *between* states. Not directly observable
- Oscillator strength - dimensionless measure of transition probability between states. Observable from experiment

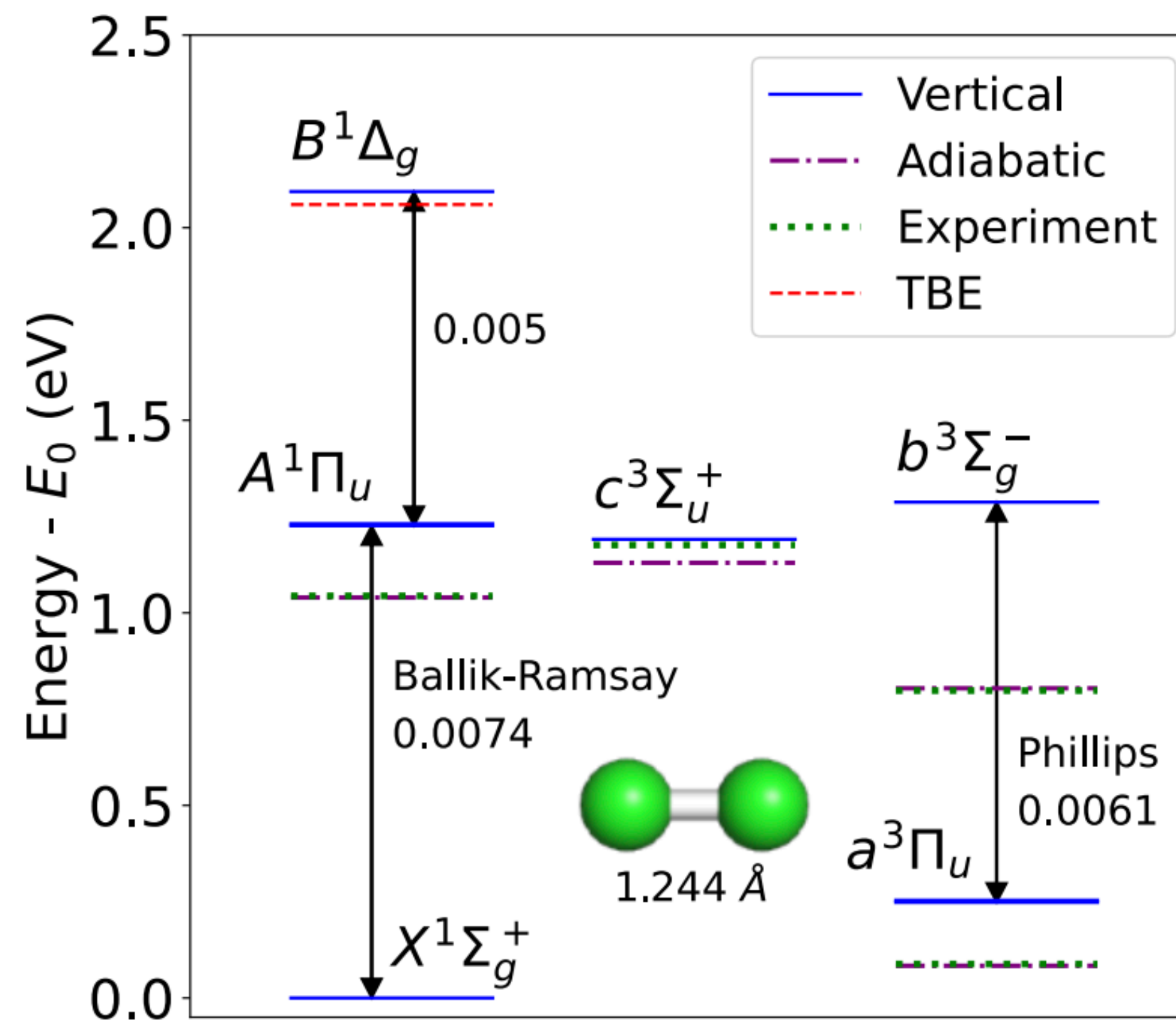
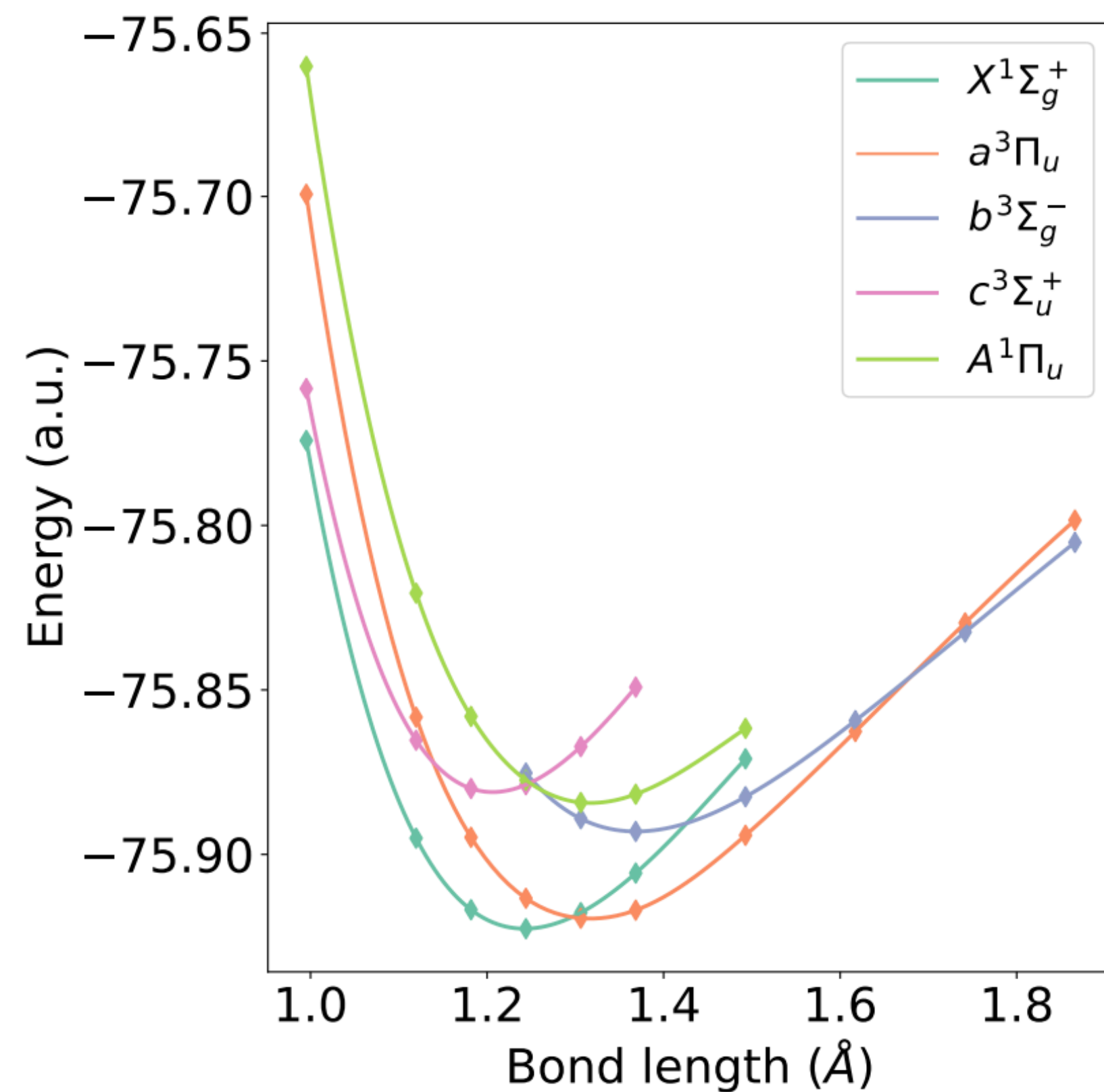
$$\mathbf{d}_{ij} = \left\langle \psi_i \sum_k q_k \mathbf{r}_k \psi_j \right\rangle$$

$$f_{ij} = \frac{2}{3} \frac{m}{\hbar^2} (E_i - E_j) |\mathbf{d}_{ij}|^2$$

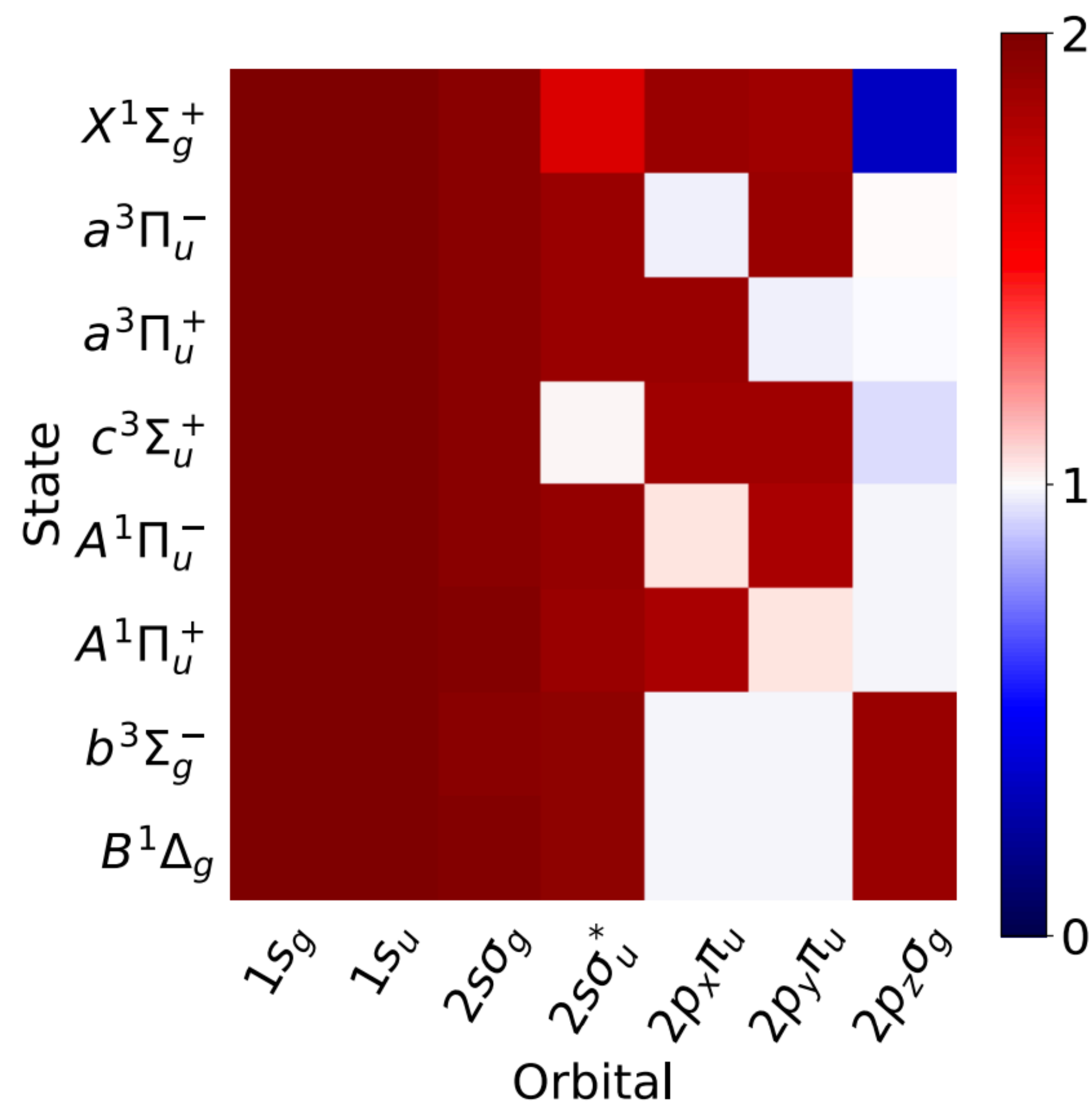
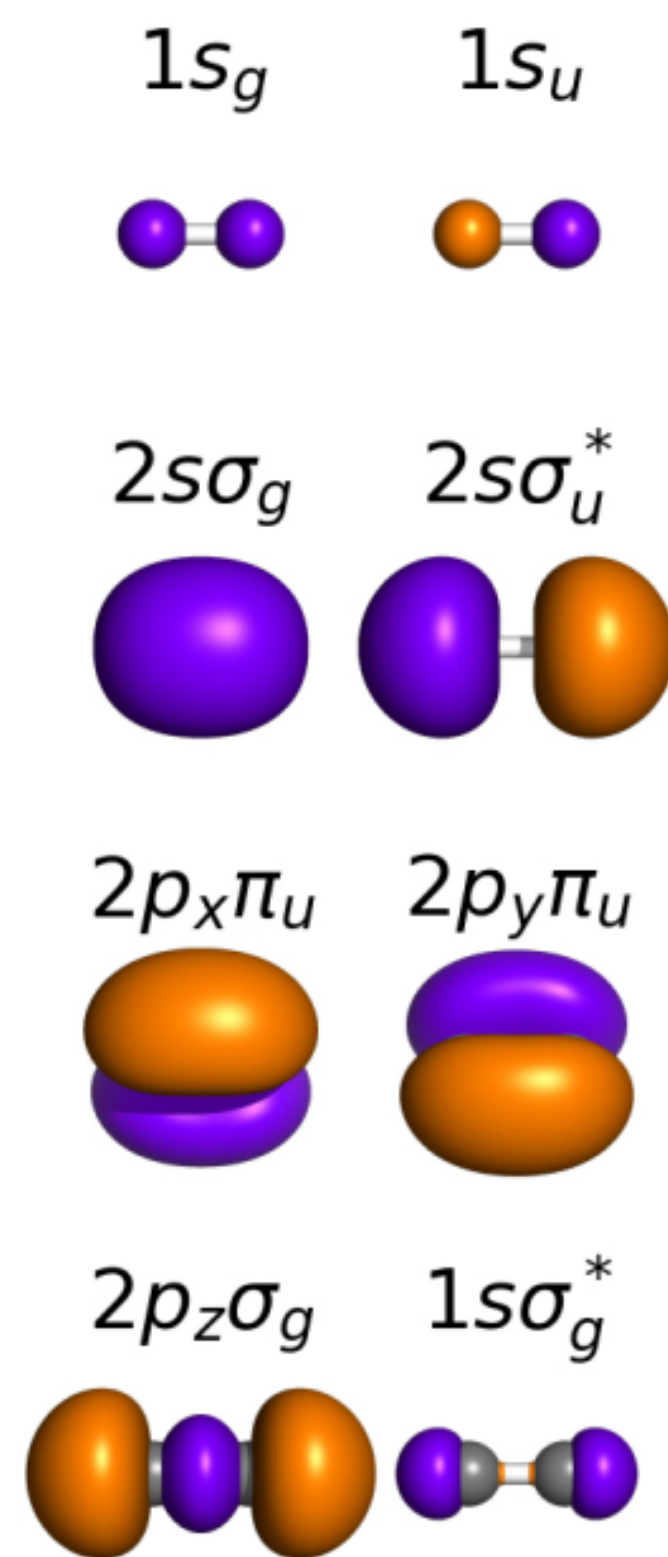
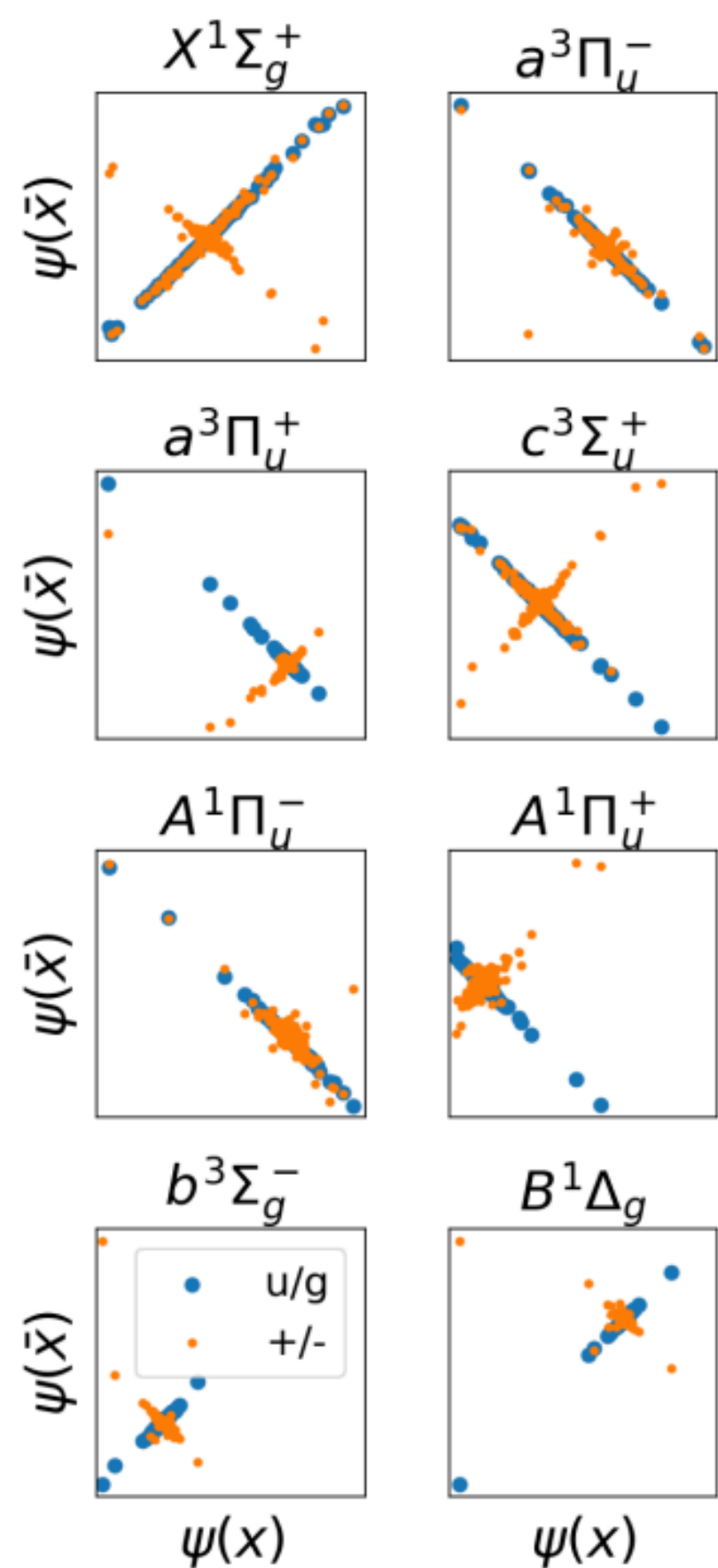
Small Molecules and Oscillator Strengths



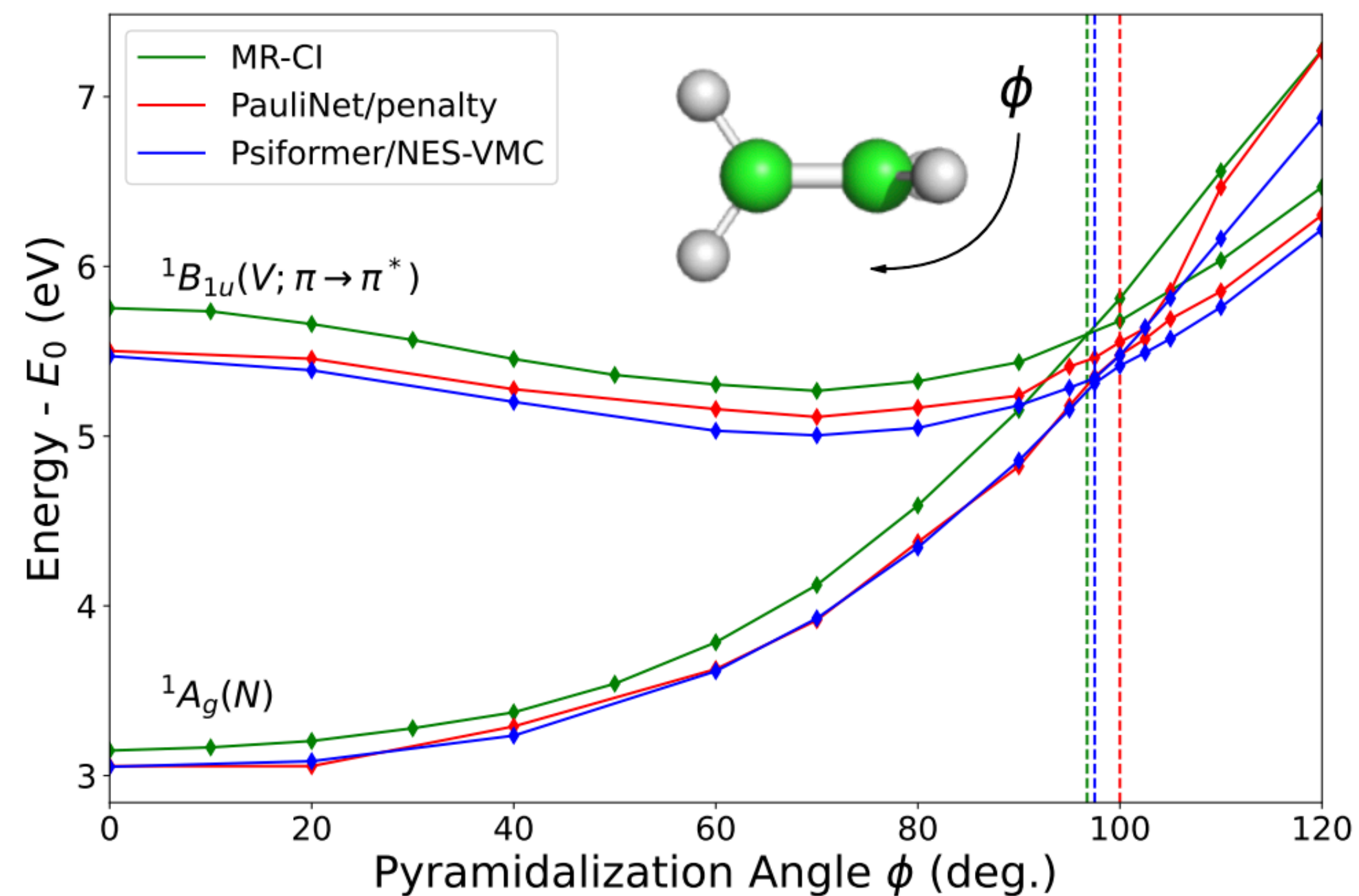
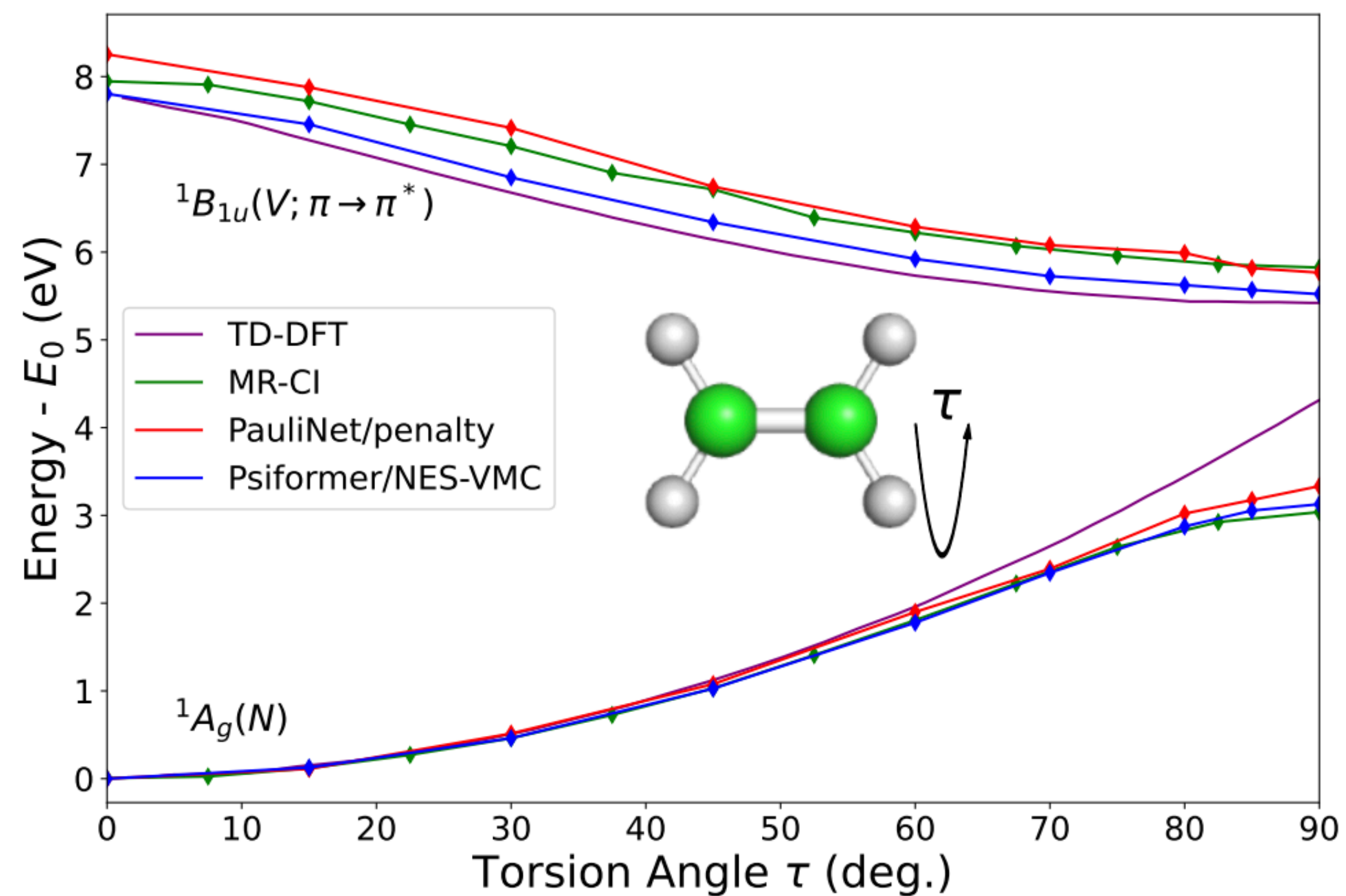
Carbon Dimer



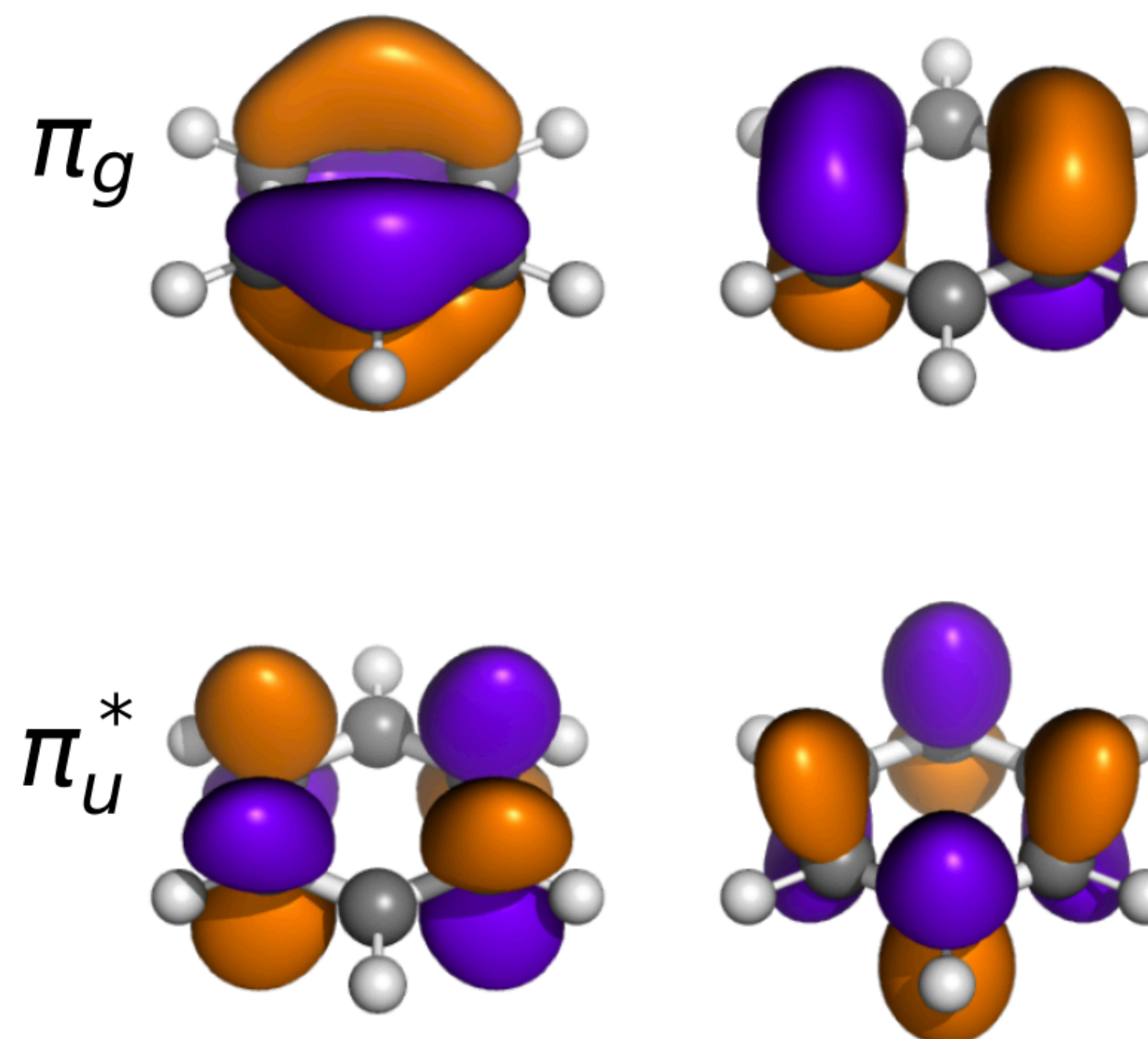
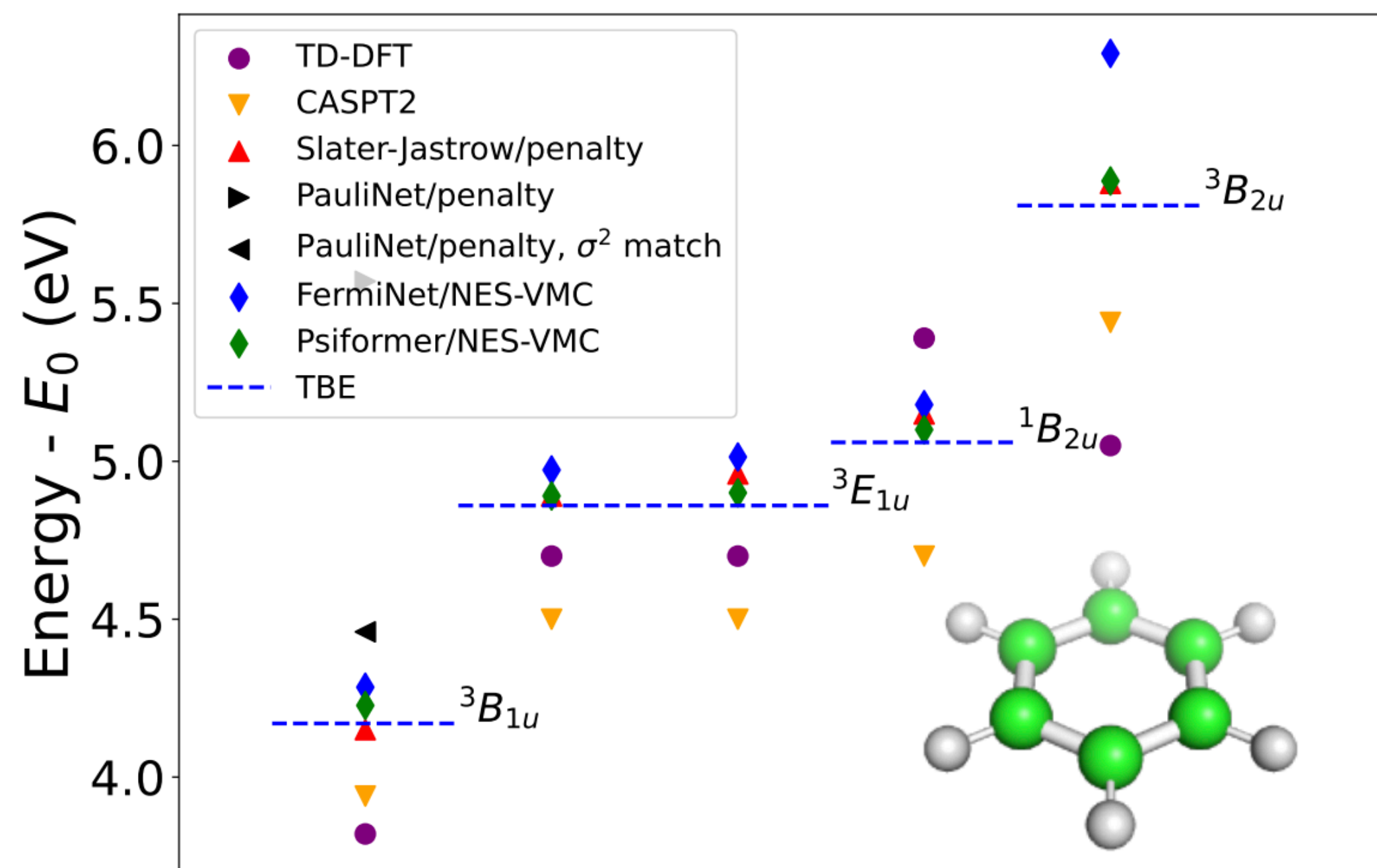
Carbon Dimer



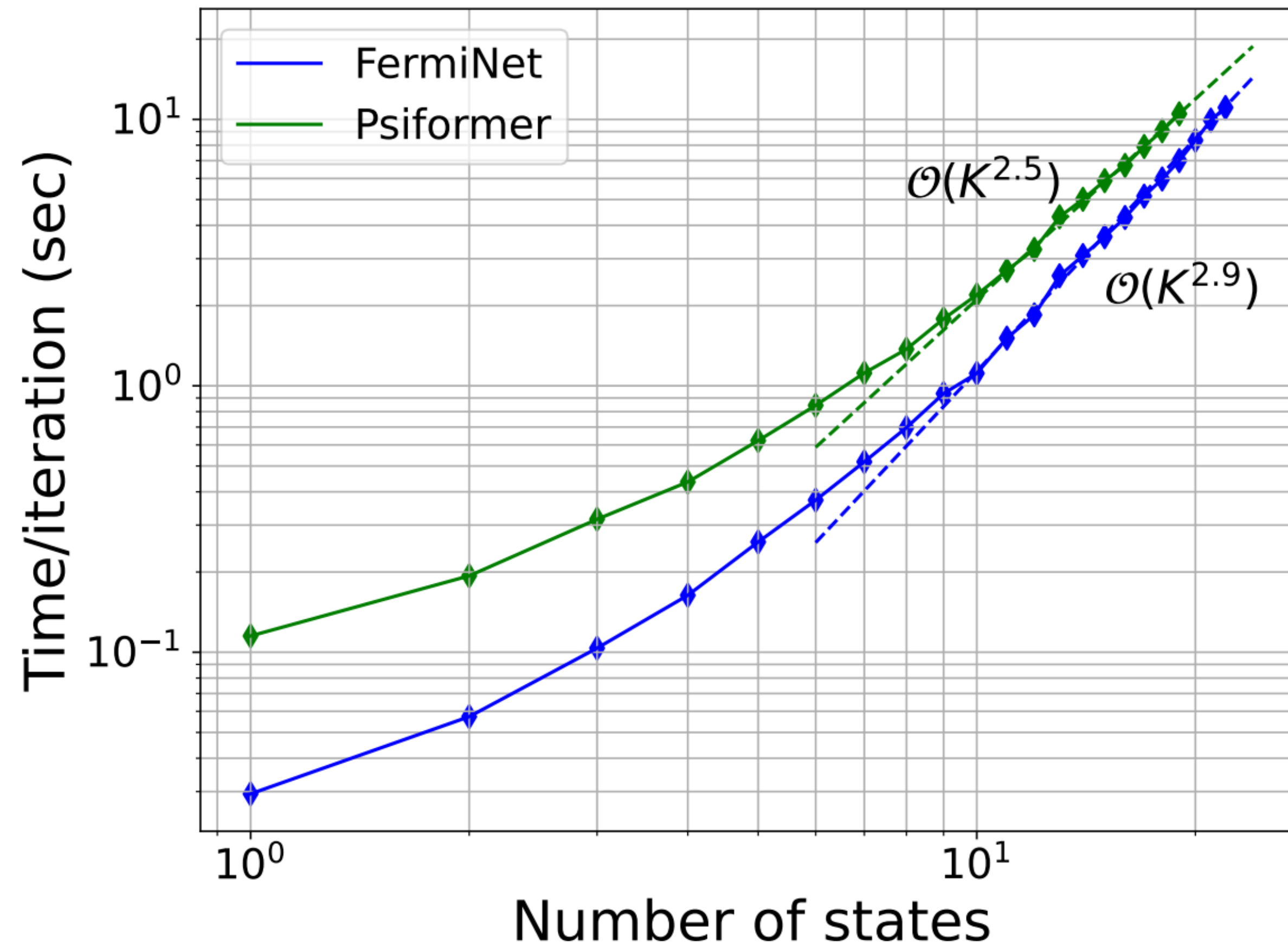
Twisted Ethene



Benzene



Computational Complexity



Summary

- Natural Excited States is a method for VMC computation of excited states without biased gradients, free parameters, orthogonal Ansätze or any restrictions on the functional form of the Ansatz
- NES addresses many of the limitations of previous approaches that prevented the computation of excited states with neural network Ansätze
- Works well for computing energies, oscillator strengths and other properties of molecules with neural networks
- But, NES is *fully general* - could be used with *any* variational Ansatz and *any* quantum system!

Thanks



Simon Axelrod
James Spencer
Ingrid von Glehn
Alex Matthews
Alex Botev
Annette Obika
James Kirkpatrick



Halvard Sutterud
Gino Cassella
Tonny Lou
Andres Perez Fadon
Matthew Foulkes

Paper:

D. Pfau, S. Axelrod, H. Sutterud, I. von Glehn, J. S. Spencer (2023)

Natural Quantum Monte Carlo Computation of Excited States. arXiv:2308.16848

Code:

Will be added to our GitHub repo: [**https://github.com/deepmind/ferminet**](https://github.com/deepmind/ferminet)