



Heriot-Watt University  
Research Gateway

# Machine learning configuration interaction for ab initio potential energy curves

## Citation for published version:

Coe, JP 2019, 'Machine learning configuration interaction for ab initio potential energy curves', *Journal of Chemical Theory and Computation*, vol. 15, no. 11, pp. 6179-6189. <https://doi.org/10.1021/acs.jctc.9b00828>

## Digital Object Identifier (DOI):

[10.1021/acs.jctc.9b00828](https://doi.org/10.1021/acs.jctc.9b00828)

## Link:

[Link to publication record in Heriot-Watt Research Portal](#)

## Document Version:

Peer reviewed version

## Published In:

Journal of Chemical Theory and Computation

## Publisher Rights Statement:

This document is the Accepted Manuscript version of a Published Work that appeared in final form in The Journal of Chemical Theory and Computation, copyright © American Chemical Society after peer review and technical editing by the publisher. To access the final edited and published work see: <https://doi.org/10.1021/acs.jctc.9b008287xICeRwJFp8nKTCDecNs>

## General rights

Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

## Take down policy

Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [open.access@hw.ac.uk](mailto:open.access@hw.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

# Machine learning configuration interaction for *ab initio* potential energy curves

J. P. Coe\*

*Institute of Chemical Sciences, School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, United Kingdom*

E-mail: J.Coe@hw.ac.uk

## Abstract

The concept of machine learning configuration interaction (MLCI) [*J. Chem. Theory Comput.* **2018**, *14*, 5739], where an artificial neural network (ANN) learns on the fly to select important configurations, is further developed so that accurate *ab initio* potential energy curves can be efficiently calculated. This development includes employing the artificial neural network also as a hash function for the efficient deletion of duplicates on the fly so that the singles and doubles space does not need to be stored and this barrier to scalability is removed. In addition configuration state functions are introduced into the approach so that pure spin states are guaranteed, and the transferability of data between geometries is exploited. This improved approach is demonstrated on potential energy curves for the nitrogen molecule, water, and carbon monoxide. The results are compared with full configuration interaction values, when available, and different transfer protocols are investigated. It is shown that, for all of the considered systems, accurate potential energy curves can now be efficiently computed with MLCI. For the potential curves of N<sub>2</sub> and CO, MLCI can achieve lower errors than stochastically selecting configurations while also using substantially less processor hours.

# 1 Introduction

The use and success of machine learning is rapidly increasing in many fields, and quantum chemistry is no exception to this. In this context its impressive applications have often involved interpolating and extrapolating potential energy surfaces for dynamics calculations. This includes recent work using Gaussian process regression<sup>1-3</sup> or artificial neural networks (ANNs).<sup>4</sup>

Alternatively one may employ machine learning, often trained on density-functional theory (DFT) results, to predict properties from electronic structure data. For example, using DFT data, an efficient deep learning approach has been developed and demonstrated on molecules up to the size of salicylic acid where it accurately predicts energies.<sup>5</sup> ANNs have also been trained to predict spin-state splittings of transition metal complexes.<sup>6</sup> Various machine learning methods were investigated for the prediction of atomization energies from molecular geometries in Ref. 7. Then a new representation for the data was introduced<sup>8</sup> to enable machine learning based on kernel ridge regression to achieve chemical accuracy for organic compounds. A novel machine learning algorithm similar to an ANN was put forward in Ref. 9 together with an active learning approach that chooses new additions to the training set to reduce errors, which enabled chemical accuracy in prediction to be reached faster.

Machine learning, trained to predict properties, has also been used to create functionals for DFT. For example, a convolutional neural network was developed to predict the kinetic energy for hydrocarbons when using the electron density as an input.<sup>10</sup> Machine learning was harnessed in Ref. 11 to predict the electron density from the external potential and also the energy from the electron density. A kinetic energy density functional has also been created using a multilayer ANN where the density and its gradients are used as the input to give predictions that are more accurate than existing functionals.<sup>12</sup> While an approach has been developed<sup>13</sup> that trains an ANN to accurately predict the energies of homogeneous Hubbard models then uses this as a functional for lattice-DFT calculations.

The elegant wavefunction approaches of second-order Møller-Plesset perturbation theory (MP2)<sup>14</sup> and coupled-cluster singles and doubles (CCSD)<sup>15</sup> have also been used to provide training data for machine learning. However, although efficient and accurate for problems that are well-described by small corrections to a single determinant, MP2 and CCSD can exhibit pathological behavior<sup>16</sup> when confronted with multireference situations such as stretched bonds. The  $\Delta$ -Machine learning approach, trained on large numbers of organic molecules, has been demonstrated to accurately predict MP2 and coupled-cluster correlation energies.<sup>17</sup> Ref. 18 used an ANN trained on CCSD(T) data to improve MP2 results by predicting corrections to the energies. The CCSD or MP2 amplitudes have also been used as inputs for machine learning to predict the CCSD correlation energy.<sup>19</sup> Ref. 20 developed a machine learning approach using Gaussian process regression to predict correlation energies for MP2 and CCSD when using properties of the Hartree-Fock molecular orbitals as input data. There the transferability of the approach was demonstrated across geometries and when varying the structure or elements of molecules.

When using wavefunction methods, rather than DFT, generating sufficient data for training can be challenging, particularly if more computationally expensive methods that can cope with multireference problems are used. A third way is then to embed a machine learning approach within an electronic structure method with the aim of accelerating the *ab initio* calculation and improving its accuracy. This was done in the approach of machine learning configuration interaction (MLCI)<sup>21</sup> where an ANN learns on the fly to choose important configurations in an iterative selected configuration interaction (CI) scheme. This was demonstrated to significantly reduce the number of iterations to convergence and, for higher-accuracy single-point calculations, required less time than other ways of selecting configurations. A machine learning approach that successfully accelerates CCSD calculations has also recently been developed.<sup>22</sup> There the k-nearest neighbors algorithm is used to predict the initial amplitudes from electronic structure properties when using training data from different geometries. In addition, the concept of using machine learning to predict adaptive

basis sets has been put forward<sup>23</sup> and demonstrated to substantially accelerate DFT calculations. In this work we build upon the approach of MLCI<sup>21</sup> with the aim of accelerating convergence whilst maintaining accuracy across a range of geometries when calculating *ab initio* potential energy curves. Consistently describing the energy across geometries that range from single reference to strongly multireference problems is a much greater challenge than making a single-point energy calculation more efficient. The new developments to overcome this include using the ANN also as a hash function when removing duplicates to increase the scalability of MLCI, employing configuration state functions to ensure pure spin states, and exploiting the transferability of data between geometries. These improvements aim to allow this selected CI approach to efficiently and accurately calculate *ab initio* potential energy curves.

Selected CI approaches seek to iteratively construct a compact wavefunction that sufficiently captures the properties of the full configuration interaction (FCI) wavefunction. Although the FCI wavefunction is the most accurate for a given basis, it is computationally intractable for all but the smallest systems and basis sets. Yet often the majority of configurations have negligible coefficients in the FCI wavefunction. It is this sparsity that selected CI seeks to exploit so that, in principle, it can describe both single reference and multireference problems sufficiently well using only a very small fraction of the FCI space without recourse to choosing an active space. One early example of selected CI was the method of Configuration Interaction using a Perturbative Selection made Iteratively (CIPSI).<sup>24</sup> Recently the development and application of selected CI approaches has become increasingly popular. For example, rather than the exact first-order perturbation calculation of the wavefunction, Heat bath CI<sup>25</sup> uses an approximation to this to choose configurations and has been successfully applied to the potential energy curve for C<sub>2</sub> and a geometry of the chromium dimer. The adaptive sampling CI approach<sup>26</sup> solely considers configurations with large enough coefficients when creating the singles and doubles space from which configurations are then selected using perturbation theory. In the MCI3 method,<sup>27</sup> the first-order correction to the

wavefunction is sampled using diffusion Monte Carlo in configuration space<sup>28,29</sup> to select determinants. MCI3 has been successfully applied to excited potential curves of the carbon dimer.<sup>27</sup> This approach has been further developed<sup>30</sup> to enable various pure spin states to be calculated simultaneously then demonstrated on multiple states of C<sub>2</sub> and an iron-sulfur cluster. In the Quantum Package 2.0 program<sup>31</sup> an efficient CIPSI is presented where the second-order perturbation of the energy is used to select configurations. This approach was employed for the calculation of benchmark vertical transition energies for a range of small compounds.<sup>32</sup> It has also been used to construct the nodal surfaces for diffusion Monte Carlo to enable very accurate calculations that, for example, gave the correct ground spin state of FeS in Ref. 33 and have provided excitation energies for formaldehyde and water.<sup>34</sup> This implementation of CIPSI has been further developed to efficiently create pure spin states<sup>35</sup> and to provide a dressed perturbation approach.<sup>36</sup> Iterative-Configuration Expansion Configuration Interaction (ICE-CI) builds upon CIPSI<sup>24</sup> to use configuration state functions so that the wavefunctions are more compact and has been implemented in the ORCA package.<sup>37</sup>

Alternative methods to using perturbation theory to select configurations have also been created. For example, the  $\Lambda$ -CI approach<sup>38</sup> uses the expectation value of the configuration's energy. The adaptive configuration interaction approach<sup>39</sup> then considers both a perturbative estimate of the energy and the configuration's coefficient in the resultant wavefunction. Further development of this method has enabled the efficient computation of excitation energies.<sup>40</sup> In Monte Carlo configuration interaction (MCCI)<sup>41-43</sup> configurations are chosen randomly and those whose absolute coefficient in the resultant wavefunction is less than a cutoff are eventually removed from it. This approach has been actively further developed to successfully model, for example, vertical excitation energies,<sup>44</sup> hyperpolarizabilities,<sup>45</sup> X-ray emission and absorption,<sup>46</sup> molecular tunnel junctions,<sup>47</sup> dissociation energies via perturbation,<sup>48</sup> and spin-orbit splitting.<sup>49</sup> Instead of using standard diagonalization, CDFCI<sup>50</sup> reformulates the FCI equation and uses coordinate descent where new determinants are only included if their update in the scheme is larger than a threshold. CDFCI is shown to com-

pute highly accurate results for systems including water using  $1.8 \times 10^6$  determinants in the cc-pVDZ basis, and all-electron  $\text{Cr}_2$ . Related to the concept of selected CI is the @CC method<sup>51</sup> which uses truncated CI calculations to estimate the importance of orbitals and decide which to include in the coupled cluster calculations. While in the *a priori* identification of configurational deadwood approach<sup>52</sup> the form of the wavefunction is assumed to be similar to truncated coupled cluster then used to estimate the configurations to include based on truncated CI calculations. There are also the recently introduced approaches of incremental FCI<sup>53</sup> and many-body expanded FCI<sup>54-56</sup> where the latter method allows the FCI energy be approached in a controlled and intrinsically parallel way by implementing multiple complete active space configuration interaction computations using a screening procedure as the numbers of orbitals are increased. For selected CI, if we knew which of the interacting configurations (single and double substitutions) were actually important then we could quickly build a compact and accurate wavefunction. This is exactly what the ANN in MLCI<sup>21</sup> attempts to learn to do as it orchestrates the iterative selected CI calculation.

In this Article, we briefly recap the approach of MLCI then present the use of the ANN as a hash function to delete duplicates on the fly and make the method more scalable by removing the requirement to store all of the single and double substitutions. The adaptation of MLCI to use configuration state functions (CSFs) is then discussed followed by the three transfer protocols investigated in this Article. We first use Slater determinants and consider a single geometry of  $\text{N}_2$  in the cc-pVDZ basis. Here MLCI when using the ANN as a hash function to remove duplicates is compared with sorting all of the single and double substitutions, and also with stochastic selection (MCCI). We then use this more scalable MLCI with CSFs to calculate the potential energy curve of  $\text{N}_2$  in the cc-pVDZ basis which we compare with FCI. Transferability is investigated for each of the three protocols and the accuracy is quantified. This and the time required is compared with previous MCCI results. The multireference character at selected geometries is also calculated and we then investigate the performance of MLCI when the configuration space is significantly enlarged

by using the cc-pVTZ basis. Next we turn to the potential energy curve for the double hydrogen dissociation of water in the cc-pVDZ basis and investigate the transfer protocols for MLCI when contrasted with FCI results. Finally we consider the potential energy curve for carbon monoxide in a cc-pVDZ basis and compare the FCI data, when available, with MLCI calculations using each of the transfer protocols.

## 2 Methods

### 2.1 Machine learning configuration interaction

The MLCI method<sup>21</sup> trains an ANN on the fly to predict important configurations in an iterated selected CI scheme. The inputs to the ANN are the occupations of the spin orbitals in the configuration of interest which are either zero or one, together with one constant input also known as a bias node which is fixed at a value of one. Hence for a basis set of  $M$  orbitals there are  $2M + 1$  inputs. We use a single hidden layer of  $n_h$  nodes plus one bias node, and one output which gives the predicted transformed coefficient of the configuration. The value of the hidden nodes and output are given by logistic functions which produce values on  $(0, 1)$  and their weights are found when training on the results of each diagonalization by using backpropagation with stochastic gradient descent. This attempts to minimize the error  $\frac{1}{2}(\text{output} - o_t)^2$  where  $o_t$  is the target transformed coefficient value for each configuration in the training set, which is shuffled after each complete pass through. Previously<sup>21</sup> the training in MLCI used 2000 passes through the data then kept the weights that gave the lowest error on the verification set. As a MLCI calculation progresses the best weights often occur towards the start of these 2000 passes so now, for efficiency, the algorithm checks every 10 passes to see if the lowest error on the verification set has reduced, if it has not then the training ends for this iteration. When updating the weights we use the same protocol as Ref. 21 by employing an initial learning rate of 0.1 that drops to 0.01 after the second iteration of MLCI or if we are using weights from a previous calculation.



MLCI begins by adding all single and double substitutions to the Hartree-Fock single determinant wavefunction while the weights for the ANN are initialized to random values on  $[-0.1, 0.1]$  if there are no other data provided. We briefly describe the steps in the MLCI algorithm used in this work below:

- The Hamiltonian matrix is constructed in the current configuration space then diagonalized using the Davidson algorithm<sup>57</sup> to give the wavefunction.
- Configurations with  $|c_i| < c_{\min}$  that are newly added are removed from the configuration space but retained in a reject set. All of the current configuration space is considered for removal every ten iterations.
- The ANN is then trained to predict the transformed coefficients  $|\tilde{c}_i|$  by combining the reject set with the wavefunction and dividing these data randomly and equally between training and verification sets.
- All symmetry-allowed single and double substitutions are generated from the current configuration space with their ANN prediction calculated on the fly and duplicates efficiently removed.
- Those  $L$  configurations, where  $L$  is size of the current wavefunction, with the largest predicted  $|\tilde{c}_i|$ , are used to enlarge the current configuration space.
- The procedure is repeated until the energy satisfies the convergence criterion of Ref. 44 which is tested on every iteration after the first six.

The transformed coefficients are used to increase the contrast between important and unimportant configurations when training, by ensuring that coefficients greater than the cutoff for removing configurations ( $c_{\min}$ ) become greater than 0.6 as the logistic functions used in the ANN give a value of 0.5 if all weights are zero. In contrast those coefficients less than  $c_{\min}$  are set to zero. The coefficients are normalized so that  $\sum_i |c_i|^2 = 1$  which means

that  $|c_i| \leq 1$ . The transformation for those greater than or equal to  $c_{\min}$  is given by<sup>21</sup>

$$|\tilde{c}_i| = \frac{0.4|c_i| + 0.6 - c_{\min}}{1 - c_{\min}}. \quad (1)$$

so that coefficients on  $[c_{\min}, 1]$  are mapped to  $[0.6, 1]$ . We use Molpro<sup>58</sup> for the calculation of Hartree-Fock molecular orbitals together with the one-electron and two-electron integrals.

## 2.2 Artificial neural network as a hash function

The original MLCI algorithm<sup>21</sup> used the procedure of Ref. 59 to remove duplicates when generating all single and double substitutions from the current configuration space. This entailed storing all of them then sorting the list using the quicksort algorithm. This was followed by one pass through the sorted list to delete replicas. If the size of the singles and doubles space is  $N_{sd}$  then this tends to scale as  $O(N_{sd} \log N_{sd})$ . Hence it is efficient but by having to store all of the generated singles and doubles then the scalability of the method is restricted to a maximum size for the singles and doubles space as this has to be stored in memory.

In this Article we remove this impediment by storing only the  $L$  configurations that are currently predicted to be the most important. This is achieved by generating single and double substitutions on the fly and applying the ANN. If the predicted value is less than the lowest predicted value of the current  $L$  most important then the next configuration is considered. Otherwise the configuration is checked to see if it is a duplicate. As we have already calculated it, we use the output of the ANN for this configuration as a hash function to check for duplicates in an efficient way. A hash function maps data to a smaller set of integers (the hash values) corresponding to locations in an array so that the data can be stored and efficiently accessed using these hash values. In this Article the hash value is given as  $\lfloor \text{Output}2L \rfloor + 1$  where  $2L$  was chosen to keep the size of the array similar to the size of the wavefunction while reducing collisions, i.e., different configurations with

the same hash value. For example, if the ANN predicts 0.8643 as the importance of the configuration and there are 1000 configurations in the wavefunction then the hash value is  $\lfloor 0.8643 \times 2 \times 1000 \rfloor + 1 = 1729$ . Hence we only have to check that the new configuration is not a duplicate of the existing configurations, if any, stored at 1729 in the hash table for the  $L$  most important configurations and for the current wavefunction. If it is not a duplicate then the least important of the  $L$  stored configurations is deleted, the new configuration is added to the ordered list and the hash table is updated. The next single and double substitution is then generated and the process is continued until the space has been exhausted.

While using the ANN also as a hash function to delete duplicates on the fly removes the memory bottleneck of storing all of the singles and doubles, it could be slower than the previous approach as now the ANN is applied to all of the duplicates as well as the unique singles and doubles space. However the hypothesis is that because applying the ANN is fast then this should not impose too much of a burden and we shall see in the Results section that using the ANN as a hash function can actually enable MLCI to run faster than when using the quicksort approach to delete duplicates.

## 2.3 Configuration state functions

The MLCI procedure is also built upon to work with configuration state functions (CSFs) so that the computed wavefunctions will be pure states for a given total spin. This uses the framework of the MCCI program<sup>41-43</sup> to calculate the now more involved matrix elements for the Hamiltonian matrix  $\mathbf{H}$  and, as the CSFs created using the procedure in MCCI of projection followed by a random walk through the spin path diagram are not necessarily orthonormal,<sup>42</sup> their overlap matrix  $\mathbf{S}$ . Then the matrix equation  $\mathbf{H}\vec{c} = \mathbf{E}\mathbf{S}\vec{c}$  is solved to give the energy and coefficients.

When using projection to create CSFs, then for a chosen total spin, the  $N$  spin orbitals can define an  $N$  electron CSF, but the same  $N$  spatial orbitals with spins swapped around may correspond to another CSF that is linearly dependent on the first. To prevent this

the procedure of MCCI<sup>43</sup> then uses a random walk through the spin path diagram<sup>42</sup> to put the list of  $N$  orbitals into a common form by randomly swapping spins. This means that the corresponding CSF is either a linearly independent spin eigenfunction or has the same spin orbitals as another CSF so can later be recognized as a duplicate. In the resulting wavefunction, the CSFs are made to be approximately orthonormal using the method of Ref. 44 which scales the coefficients of the wavefunction using  $c'_i = c_i \sqrt{S_{ii}} / \sqrt{\sum_j c_j S_{jj} c_j}$  and means that  $\sum_i |c'_i|^2 = 1$ .

As the occupied spin orbitals now uniquely define a linearly independent CSF then it should be possible for the ANN to use these occupations as an input to learn to predict the importance of CSFs and the Results section demonstrates that this can indeed be successfully accomplished.

## 2.4 Transferability

The idea of transferability is of interest for the efficient application of machine learning, i.e., how well does the trained algorithm transfer to a different but related task. This concept, for example, has been investigated<sup>20</sup> in electronic structure when predicting CCSD and MP2 correlation energies for transfers that range from between geometries to among molecules and elements. In this work we consider transferability between geometries and take a broader view, as in MLCI the ANN is embedded inside a selected CI scheme, that encompasses other data not just the weights of the ANN. We investigate whether transferring data from the previous geometry as we stretch a bond can offer any benefits such as accelerating an MLCI calculation or improving its accuracy. Three transfer protocols are considered which are compared with standard MLCI. In the first we transfer the weights of the ANN, the reject set and the wavefunction as a starting point for the next MLCI calculation. The second protocol only transfers the weights of the ANN while the third only transfers the wavefunction ( $\Psi$ ). We emphasize that the ANN keeps learning in all of these protocols, but its initial weights are the trained weights of a previous geometry when the ANN weights are

transferred and the initial training rate is lowered to 0.01 in this case.

### 3 Results

We first compare MLCI using the ANN as a hash function with the previous approach that sorted all of the singles and doubles to remove duplicates. For this initial comparison we use Slater determinants, 30 hidden nodes and fix the seed for the random number generator to allow a fair comparison between the ways to remove duplicates. We consider the ground  $A_g$  state when using  $D_{2h}$  of  $N_2$  in the cc-pVDZ basis set with two frozen orbitals and use a cutoff of  $c_{\min} = 5 \times 10^{-4}$ .

For a stretched bond length of 2.2225 Å we see in Fig. 1 that the more scalable new approach to duplicates in MLCI gives the same results as before, but in less time. The final energies were identical yet using the ANN prediction as a hash function required 1816 seconds compared with 2409 seconds when using quicksort to remove duplicates. Both versions are also significantly faster than stochastically selecting configurations using Monte Carlo configuration interaction (MCCI)<sup>41–43</sup> with Slater determinants when running on a single processor: MCCI used 9504 seconds to give an energy that is slightly higher by 1.1 kcal/mol and required 13502 determinants compared with 12456 in MLCI.

#### 3.1 Nitrogen molecule potential curve

We now continue with the new version of MLCI that employs the ANN as a hash function, and use CSFs, rather than Slater determinants, to guarantee that the wavefunction is a pure spin state. In addition as a single CSF will represent numerous Slater determinants then we would expect MLCI calculations using CSFs at a given cutoff to be more accurate than when using Slater determinants. We apply this approach to potential energy curves beginning with the  $A_g$  ground-state singlet of  $N_2$  in the cc-pVDZ basis set with two frozen orbitals when using the  $D_{2h}$  point group. The FCI results used for comparison were calculated for this

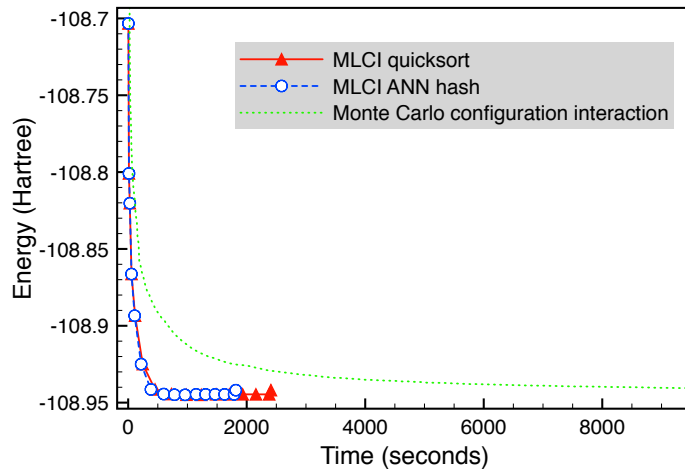


Figure 1: Energy (Hartree) against time (seconds) for serial calculations on  $N_2$  at a bond length of  $2.2225 \text{ \AA}$  in the cc-pVDZ basis with two frozen orbital using MLCI with quicksort to remove duplicate configurations, MLCI with the ANN as a hash function to remove duplicate configurations, or MCCI. Here all methods employ Slater determinants and a cutoff of  $c_{\min} = 5 \times 10^{-4}$ .

system in Refs. 60–62. The potential curve of  $N_2$  is known to be a challenging problem as, for example, coupled cluster methods that use a single reference can perform poorly when the bond is stretched.<sup>16</sup>

The number of hidden nodes is first considered and we run calculations with a cutoff of  $c_{\min} = 5 \times 10^{-4}$ . In Fig. 2 we plot the energy error for the MLCI result when compared with the FCI data for the 15 bond lengths. We see that increasing the number of hidden nodes from 30 to 40 noticeably reduces the error at the most stretched geometries, but slightly increases the error at some intermediate bond lengths. Both MLCI potential curves used around 4900 CSFs on average while the FCI wavefunctions required  $\sim 5 \times 10^8$  determinants. Hence it is not expected that the compact wavefunctions from MLCI will give energies that are extremely close to FCI, but rather that errors will be sufficiently balanced and so relative energies will be much more accurate. As a potential curve can be shifted by a constant energy, we can aim for MLCI potential energy curves that are essentially FCI quality.

If we shift the potential curves so that they all have their minimum at zero, we see in Fig. 3 that with  $c_{\min} = 5 \times 10^{-4}$  and 40 hidden nodes the MLCI potential curve is practically

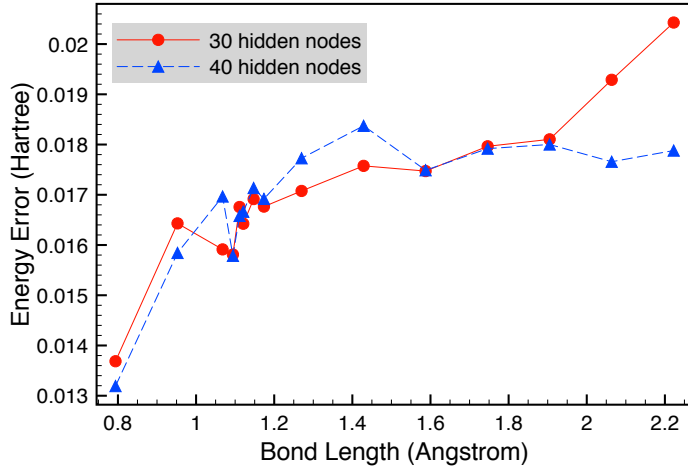


Figure 2: Difference between the FCI and MLCI energy (Hartree) for  $N_2$  in the cc-pVDZ basis set with two frozen orbitals where for MLCI the number of hidden nodes is varied, CSFs are used and  $c_{\min} = 5 \times 10^{-4}$ .

indistinguishable from the FCI result on the scale of the graph. To indicate the multireference character of the MLCI wavefunction for the given orbitals we use

$$MR = \sum_i |c_i|^2 - |c_i|^4 \quad (2)$$

from Refs. 63 and 64.  $MR$  is zero for a single determinant and approaches one as the number and importance of the configurations increases. Here the  $c_i$  are the coefficients of the configuration interaction wavefunction normalized so that  $\sum_i |c_i|^2 = 1$ . This approach is approximate when using non-orthonormal CSFs. Although the CSFs are made approximately orthonormal in MLCI using the procedure of Ref. 44, the time cost is not too important for individual  $MR$  calculations so we make the CSFs exactly orthonormal for use with this indicator. To do this we use the approach of Ref. 65, which means that the new coefficients are  $\vec{c}_{\text{new}} = \sqrt{\mathbf{S}}\vec{c}$  where  $\mathbf{S}$  is the overlap matrix for the original non-orthonormal CSFs. We find that for the shortest bond length considered then  $MR = 0.121$  for the orthonormal CSFs suggesting that single-reference approaches are likely to work sufficiently well while this increases to 0.908 at the longest bond length indicating that the problem, with Hartree-Fock

orbitals, is now very strongly multireference.

To enable a quantitative comparison of the accuracy of approximations to the FCI curve and take into account that the potential curves can be shifted by a constant we use the non-parallelity error<sup>66</sup> (NPE) and the standard deviation of the difference in energies<sup>59</sup>  $\sigma_{\Delta E}$ .

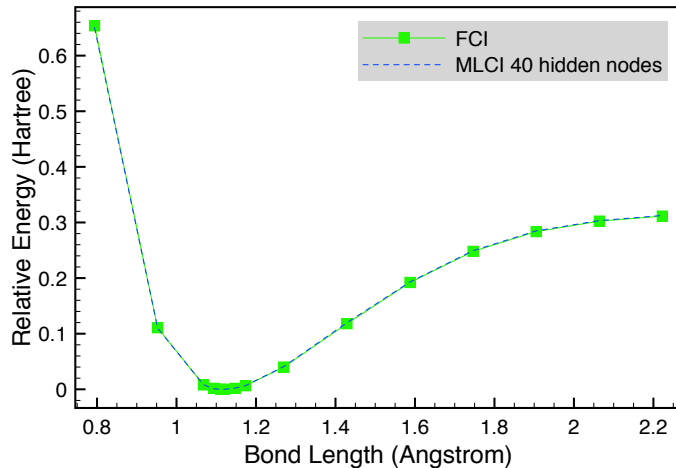


Figure 3: FCI and MLCI energies relative to their minimum values against bond length for  $N_2$  in the cc-pVDZ basis set with two frozen orbitals when using MLCI with 40 hidden nodes, CSFs and  $c_{\min} = 5 \times 10^{-4}$ .

If we write  $\Delta E_i = E_{i,approx} - E_{i,FCI}$  then

$$NPE = \max_i |\Delta E_i| - \min_i |\Delta E_i| \quad (3)$$

where  $i$  ranges over all points in the curve. Hence if the approximate curve when shifted by a constant is equal to the FCI curve then  $NPE = 0$ . The standard deviation of the difference in energies is also equal to zero when there is a constant difference between the curves and furthermore depends on all points in the curve not just the maximum and minimum  $\Delta E_i$ . It is defined as

$$\sigma_{\Delta E} = \sqrt{\frac{1}{d} \sum_{i=1}^d (\Delta E_i - \overline{\Delta E})^2} \quad (4)$$

where  $d$  is the number of points in the curve and  $\overline{\Delta E}$  is the mean of  $\Delta E$ .



In Table 1 we show the effect of varying the number of hidden nodes then of transferring all of the ANN, reject list and the wavefunction as the starting point for the next calculation at increasing bond length. We also look at transferring just the ANN or only the wavefunction. We see that the errors for the curves when quantified by  $\sigma_{\Delta E}$  are generally less than 1 kcal/mol. The time required clearly increases, but not drastically, with the number of hidden nodes which fits in with the training steps taking longer. However the accuracy does not necessarily follow this trend although it is always sufficiently high. The results using 40 hidden nodes are the most accurate rather than those with 50, which could be due to the training becoming less successful when using a single hidden layer as the number of hidden nodes becomes sufficiently large. On increasing the number of hidden nodes from 30 to 40 the NPE drops by around 1 kcal/mol with only a modest increase in calculation time. The accuracy of the curve is further improved by transferring all of the wavefunction ( $\Psi$ ), reject list and ANN. However this increases the calculation time by around 1 hour. Interestingly if only the ANN is transferred then, although the calculation time is essentially the same as without transferring, the error in the curve is slightly higher than that from the other approaches at a still reasonable 1.12 kcal/mol for  $\sigma_{\Delta E}$ . The highest accuracy and longest calculation time is from transferring only the wavefunction and this give  $\sigma_{\Delta E} = 0.56$  kcal/mol but took almost 1.5 hours longer than the standard  $n_h = 40$  calculation. This increase in accuracy and time appears due to a greater number of important configurations being found as on average there are around 1000 more CSFs than in the other approaches. Another contribution to the increase in time when transferring the wavefunction is that there are more configurations when the calculation is initialized. This is due to using the MLCI wavefunction from the previous geometry as the starting point rather than the CISD wavefunction.

We note that in previous work<sup>59</sup> the stochastic approach of MCCI required 7.55 hours with  $c_{\min} = 5 \times 10^{-4}$  on the same hardware when using the Hartree-Fock molecular orbitals to reach an accuracy of 1.06 kcal/mol for  $\sigma_{\Delta E}$ . This was on 12 processors so equates to

Table 1: Results for MLCI with  $c_{\min} = 5 \times 10^{-4}$  when varying the number of hidden nodes ( $n_h$ ) and the transfer protocol when applied to the potential curve of  $N_2$  in the cc-pVDZ basis set with two frozen orbitals.

	NPE (kcal/mol)	$\sigma_{\Delta E}$ (kcal/mol)	Mean CSFs	Time (Hours)
$n_h = 20$	3.63	0.81	4840	2.72
$n_h = 30$	4.23	0.94	4904	2.90
$n_h = 40$	3.25	0.78	4856	3.12
$n_h = 50$	4.20	0.96	4874	3.63
$n_h = 40$ Transfer all	3.09	0.71	5094	4.15
$n_h = 40$ Transfer ANN	4.86	1.12	4748	3.20
$n_h = 40$ Transfer $\Psi$	2.00	0.56	5972	4.57

90.6 processor hours. The results show that for the nitrogen dimer, MLCI can produce a more accurate curve in less time compared with randomly selecting configurations using MCCI. It appears that the ANN does not transfer particularly well as a starting point for the geometries of  $N_2$  considered here. This suggests that the most accurate approach is to use the wavefunction from the previous geometry as a starting point and train a completely fresh ANN each time.

Although currently beyond a FCI calculation, we also consider  $N_2$  using the cc-pVTZ basis with two frozen core orbitals to investigate the performance of MLCI when confronted with a larger configuration space. Now the number of determinants in the FCI space would be around  $3 \times 10^{12}$  yet we see in Fig. 4 that MLCI with  $c_{\min} = 5 \times 10^{-4}$  and when transferring the wavefunction gives a smooth curve that behaves similarly to the cc-pVDZ result but is noticeably lower in energy at all points. This used 9724 CSFs on average compared with 5972 for cc-pVDZ while 47.2 hours were required which is around ten times that needed for cc-pVDZ. However when one considers that the FCI space is around 6000 times larger the time scaling for this implementation of MLCI appears promising. The size of the reject set was also seen to increase but again by much less than the scaling of the FCI space: the reject set finally contained 21627 configurations for the longest bond length considered in cc-pVDZ which rose to 64645 in cc-pVTZ.

We continue with 40 hidden nodes and next investigate the performance of MLCI together

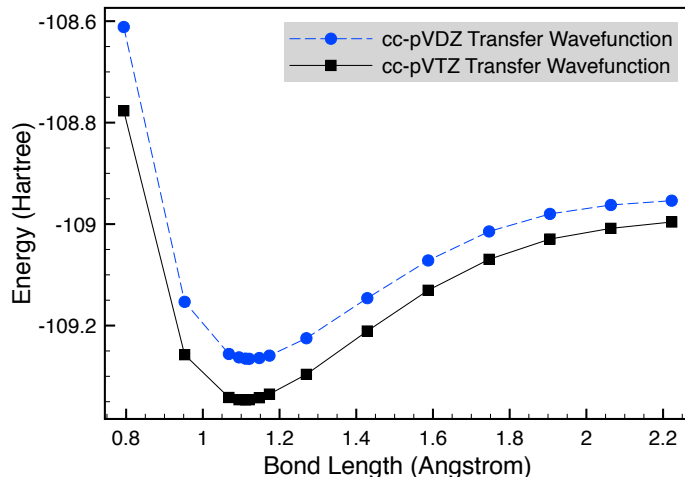


Figure 4: Comparison of MLCI  $N_2$  results for cc-pVDZ with cc-pVTZ both with two frozen orbitals when using 40 hidden nodes, CSFs and  $c_{\min} = 5 \times 10^{-4}$ .

with transfer protocols on the potential curve for the double hydrogen dissociation of the water molecule.

### 3.2 Water potential curve

For the  $H_2O$  potential energy curves of the ground-state singlet the bond angle is fixed at 104.5 degrees as we simultaneously vary both bond lengths when using the cc-pVDZ basis with one frozen orbital. We therefore look at the lowest  $A_1$  singlet state in  $C_{2v}$ . For comparison we use the FCI results from Ref. 59. When initially using  $c_{\min} = 10^{-3}$  and transferring  $\Psi$  we see in Fig. 5 that the agreement with FCI is very good on the scale of the graph when both curves are shifted so their minimum is at zero. However there is a slight difference that can be noticed at intermediate bond lengths.

The lower accuracy at intermediate bond lengths is confirmed in the plot of differences with the FCI result (Fig. 6) where we see that the largest difference when transferring the wavefunction is at 2.8 Bohr. Fig. 6 also reveals that if we do not transfer the wavefunction then the errors generally continue rising as the bonds lengthen.

We first consider the multireference character using  $MR$  when we do not transfer the

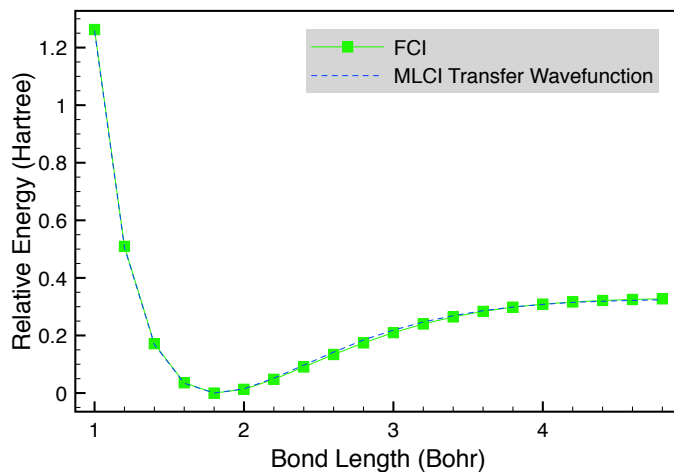


Figure 5: FCI and MLCI energies relative to their minimum values for  $\text{H}_2\text{O}$  as both bonds are varied in the cc-pVDZ basis set with one frozen orbital when using MLCI with transfer of the wavefunction, 40 hidden nodes, CSFs and  $c_{\min} = 10^{-3}$ .

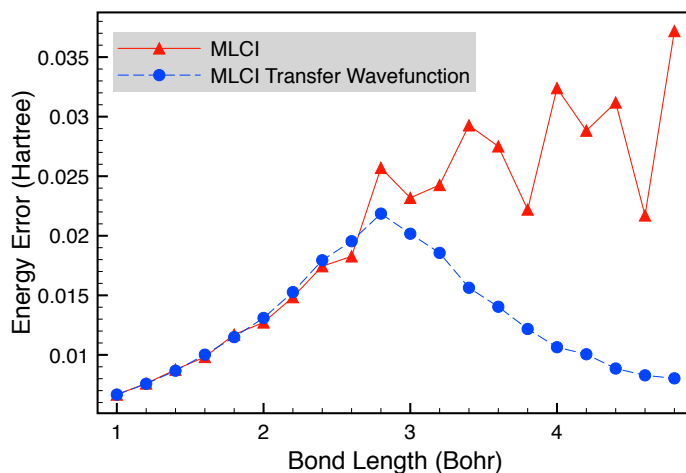


Figure 6: Difference between the FCI and MLCI energies with and without transference of the wavefunction for  $\text{H}_2\text{O}$  when both bonds are varied in the cc-pVDZ basis set with one frozen orbital when using MLCI with 40 hidden nodes, CSFs and  $c_{\min} = 10^{-3}$ .

MLCI wavefunction. At 1 Bohr this is 0.065, by 2.8 Bohr it has risen to 0.210 then at the longest bond length considered (4.8 Bohr) it is 0.784. We compare this to the results when the wavefunction is transferred, at 1 Bohr there is nothing to transfer so the value remains the same, at 2.8 Bohr it is 0.228 then at 4.8 Bohr it is 0.808. Hence at the shortest bond length then the problem is certainly single-reference, the  $MR$  value then suggests that by

2.8 Bohr there may now be a small amount of multireference character, while the system with Hartree-Fock orbitals has become strongly multireference by 4.8 Bohr. Although the energy differences are apparent at long bond lengths between the two approaches, both give approximately the same multireference character. That MLCI when transferring the wavefunction performs better after 2.6 Bohr could be due to the increase in multireference character. Here it appears the algorithm has a better chance to find a sufficient number of the important configurations if it is given the previous set of configurations as a starting point to build and learn from rather than beginning anew from a single determinant.

In table 2 we compare the accuracy of the curves from different transfer protocols when using  $c_{\min} = 10^{-3}$  and see that, as for  $N_2$ , transferring the wavefunction substantially improves the results while the least accurate curve is due to only transferring the ANN. The improvement is more apparent as the NPE values are higher than for  $N_2$  which we attribute to the larger cutoff used. The most accurate result used the most configurations on average and consequently required the longest time. We note that previous work<sup>59</sup> with MCCI produced a curve with an error of  $\sigma_{\Delta E} = 0.92$  kcal/mol in 0.52 hours which equates to 6.2 processor hours. Hence for the water molecule at this cutoff then MLCI is less accurate which may be due to the multireference character being smaller than  $N_2$  with regards to the shortest and longest bonds considered, and the FCI space ( $\sim 2 \times 10^7$  determinants) being around an order of magnitude lower. Hence fewer configurations are required for the MLCI wavefunction thus providing less opportunity for learning. To investigate this latter conjecture we next consider lowering the cutoff.

Table 2: Results for MLCI with a cutoff of  $c_{\min} = 10^{-3}$  and 40 hidden nodes when the transfer protocol is varied when applied to the potential curve for the double hydrogen dissociation of  $H_2O$  in the cc-pVDZ basis set with one frozen orbital.

	NPE (kcal/mol)	$\sigma_{\Delta E}$ (kcal/mol)	Mean CSFs	Time (Hours)
MLCI	19.15	5.57	1032	0.61
MLCI Transfer all	24.26	7.74	937	0.63
MLCI Transfer ANN	26.55	8.39	930	0.57
MLCI Transfer $\Psi$	9.53	2.87	1212	0.74

We see in table 3 that at the lower cutoff of  $5 \times 10^{-4}$  then the errors are much lower with  $\sigma_{\Delta E}$  less than 2 kcal/mol for all the MLCI approaches and the mean number of CSFs is larger but still less than  $N_2$ . Now transferring all the data, not just the wavefunction, gives the most accurate curve but the differences between the approaches are small. Transferring the wavefunction required the most CSFs on average and although every individual energy is lower than just transferring the ANN the  $\sigma_{\Delta E}$  error is higher. This serves to remind us that it is the consistency in describing multiple points on the potential curve that is important not just getting all points to be lower in energy. Although the results for this system are now sufficiently accurate for all of the MLCI approaches, the maximum  $\sigma_{\Delta E}$  of 1.75 kcal/mol is still higher than the  $c_{\min} = 10^{-3}$  MCCI value<sup>59</sup> of 0.92 kcal/mol. We note that we can achieve a more accurate potential curve if we further lower the cutoff to  $c_{\min} = 2 \times 10^{-4}$  where now MLCI when only transferring the wavefunction gives an error of  $\sigma_{\Delta E} = 0.70$  kcal/mol but takes longer (13.3 hours) than the  $c_{\min} = 10^{-3}$  MCCI result. In this case more CSFs were used on average (9744) than for the  $N_2$  results, which fits in with the idea that there needs to a sufficient number of configurations in the wavefunction for the training of the ANN of to work well. Although the MLCI  $c_{\min} = 5 \times 10^{-4}$  results were not quite as accurate as the MCCI  $c_{\min} = 10^{-3}$  values, they required less processor hours but the performance gain was not as substantial as for the nitrogen dimer. We attribute this to the lower multireference character of this system and fewer configurations being needed for a given cut-off so the MCCI calculations did not need much time and there is less scope for improvement.

Table 3: Results for MLCI with a cutoff of  $c_{\min} = 5 \times 10^{-4}$  and 40 hidden nodes when the transfer protocol is varied when applied to the potential energy curve for the double hydrogen dissociation of  $H_2O$  in the cc-pVDZ basis set with one frozen orbital.

	NPE (kcal/mol)	$\sigma_{\Delta E}$ (kcal/mol)	Mean CSFs	Time (Hours)
MLCI	6.68	1.54	2256	1.59
MLCI Transfer all	5.74	1.39	2430	2.15
MLCI Transfer ANN	7.27	1.67	2226	1.60
MLCI Transfer $\Psi$	6.05	1.75	2892	2.28

### 3.3 Carbon monoxide potential curve

Finally we consider the potential curve for the  $A_1$  ground-state singlet of carbon monoxide when using  $C_{2v}$  and the cc-pVDZ basis with two frozen orbitals. We see in Fig. 7 that, when using  $c_{\min} = 5 \times 10^{-4}$  and transferring the wavefunction, the MLCI potential curve is essentially FCI quality for the points where FCI values are available. Due to the lower symmetry of CO compared with  $N_2$ , we note that there are now  $\sim 10^9$  determinants in the FCI space which makes it the largest considered in this Article. At a geometry of 6 Bohr we could not run a Molpro<sup>58</sup> FCI calculation<sup>67</sup> to completion as it required more than 1TB of disk space which we partly attribute to the strong multireference character. This would suggest that the starting guess vectors for the Davidson diagonalization algorithm<sup>57</sup> are poor in this situation so more iterations and vectors are required for convergence resulting in larger subspace matrices that have to be diagonalized. Hence we use the FCI values from Ref. 59 which are available to a maximum bond length of 3.4 Bohr.

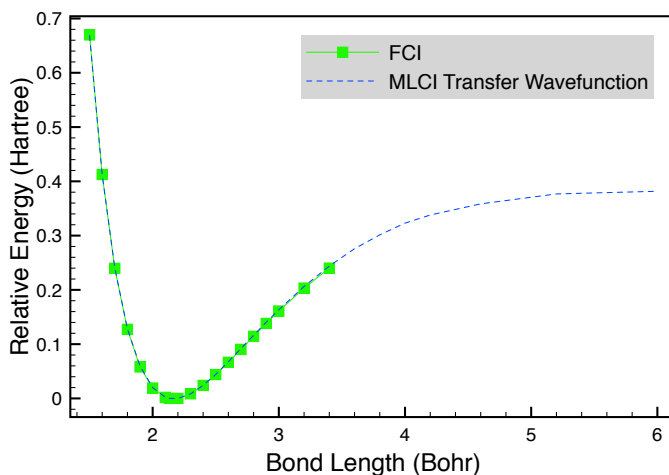


Figure 7: FCI and MLCI energies relative to their minimum values against bond length for CO in the cc-pVDZ basis set with two frozen orbitals when using MLCI with transfer of the wavefunction, 40 hidden nodes, CSFs and  $c_{\min} = 5 \times 10^{-4}$ .

In Fig. 8 we see that transferring the wavefunction in MLCI gives lower and more consistent errors in comparison to standard MLCI. This improvement in performance appears

better than that for the potential curve of water when the geometries are not too stretched (Fig. 6).

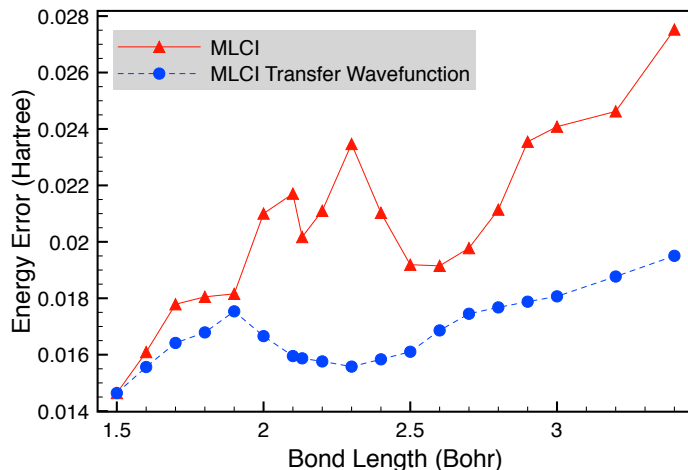


Figure 8: Difference between the FCI and MLCI energies against bond length with and without transference of the wavefunction for CO in the cc-pVDZ basis set with two frozen orbitals when using MLCI with 40 hidden nodes, CSFs and  $c_{\min} = 5 \times 10^{-4}$ .

Table 4 shows that, as with  $N_2$ , transferring just the wavefunction gives the most accurate curve with a  $\sigma_{\Delta E}$  error that is less than half the value for the other approaches. There is a noticeable time cost in calculations involving transferring the wavefunction which require around 19 processor hours. Now transferring just the ANN is the second most accurate approach, but the errors for transferring all or just the ANN are both reasonably similar to standard MLCI. We note that errors when transferring all data from each point to the next are slightly higher than standard MLCI despite using almost 1000 more CSFs on average. This is due to the error calculation only being possible for the first 17 points where we have FCI values. For example the MLCI results when transferring all data are actually 0.02 Hartree lower at  $R = 6$  than the value from standard MLCI. The effect of transferring the ANN from the  $N_2$  40 node results is also considered. Because the geometries used are not the same, we partition the potential curve into four regimes: repulsive wall (1.5 to 1.9 Bohr), equilibrium (2 to 2.5 Bohr), intermediate (2.6 to 3.8 Bohr) and stretched (4 to 6 Bohr). These use the  $N_2$  ANN from a geometry in the corresponding regime as a starting point: 0.9525



Å for the repulsive wall, 1.1208 Å for equilibrium, 1.4288 Å for intermediate and 2.0638 Å for stretched. We see in Table 4 that this scheme is the fastest, which can be attributed to the mean number of CSFs being noticeably lower than the other approaches. Although this method is not as accurate as transferring the wavefunction it appears similar in accuracy to the remaining approaches. However this does not reveal the accuracy at stretched bonds where the FCI data are not available. Here transferring the ANN from N<sub>2</sub> gives an energy that at times is more than 0.1 Hartree higher than the other approaches. Hence it appears here that transferring the ANN from an isoelectronic molecule is not appropriate in the strongly multireference region.

We also compare these data with previous results<sup>59</sup> using MCCI. There the  $\sigma_{\Delta E}$  error was 0.89 kcal/mol when using Hartree-Fock orbitals, and the time for the entire curve was 20.72 hours which equates to 248.6 processor hours. Hence by transferring only the wavefunction in MLCI we can achieve slightly higher accuracy with slightly less wall time and with a very significant reduction in the number of processor hours.

Table 4: Results for MLCI with a cutoff of  $c_{\min}=5 \times 10^{-4}$  and 40 hidden nodes when the transfer protocol is varied when applied to the potential curve of CO in the cc-pVDZ basis set with two frozen orbitals. Errors are with regards to the available 17 FCI values with bond lengths less than or equal to 3.4 Bohr, while CSFs and timings are for the entire curve.

	NPE (kcal/mol)	$\sigma_{\Delta E}$ (kcal/mol)	Mean CSFs	Time (Hours)
MLCI	8.08	1.91	5586	12.17
MLCI Transfer all	8.54	2.19	6438	18.55
MLCI Transfer ANN	7.25	1.72	5880	12.31
MLCI Transfer N <sub>2</sub> ANN	7.17	2.01	4766	10.25
MLCI Transfer $\Psi$	3.05	0.76	7496	19.84

As for the previous two systems we again make the CSFs orthonormal to calculate  $MR$  values to indicate the multireference character for MLCI when transferring the wavefunction. We find  $MR = 0.116$  at 1.5 Bohr and 0.192 at 2.1316 Bohr suggesting that it would not yet be classified as multireference at these geometries but there may be a small amount of multireference character at the equilibrium geometry. However by 3.4 Bohr this has risen

to 0.736 indicating that it is now a multireference problem and  $MR$  reaches the highest value in this Article of 0.931 at 6 Bohr demonstrating that the problem is very strongly multireference by this point.

## 4 Summary

In this Article, machine learning configuration interaction (MLCI) was further developed so that the singles and doubles space did not need to be stored when efficiently removing duplicates. This removed a restriction on the scalability of the method and was achieved by using the artificial neural network (ANN) prediction as a hash function to allow duplicates to be efficiently found when configurations are generated on the fly. This enhancement of MLCI to make it more scalable was demonstrated to give the same result as the original MLCI in less time for a single-point energy of the nitrogen molecule when using Slater determinants. It was also substantially faster than stochastically selecting configurations using Monte Carlo configuration interaction (MCCI) when run in serial.

MLCI was then applied to the problem of calculating *ab initio* potential curves. Here configuration state functions were used in the procedure to ensure that the wavefunction was a pure spin state. The transferability of data from the previous geometry as a starting point to improve the next calculation was also investigated.

For the nitrogen molecule in a cc-pVDZ basis we found that the MLCI potential curve was practically indistinguishable to the full configuration interaction (FCI) result when energies were plotted relative to their minimum value. This accuracy was quantified and supported by using the non-parallelity error and  $\sigma_{\Delta E}$  which were as low as 2.00 and 0.56 kcal/mol respectively when transferring the wavefunction. This was more accurate than previous results using MCCI to randomly select configurations,<sup>59</sup> used less wall time despite MLCI being run in serial while MCCI was run on 12 processors, and substantially lowered the number of processor hours. Transferring the wavefunction was more accurate than transferring all of the

data from the previous geometry, while only transferring the ANN gave the largest error although not by much and this was still a respectable 1.12 kcal/mol for  $\sigma_{\Delta E}$ . We also investigated the performance of MLCI when the configuration space was enlarged by using the cc-pVTZ basis. Here a smooth curve was produced that behaved similarly to the cc-pVDZ result but with a lower energy at all points. Although more time was required the scaling was substantially less than that of the FCI space.

The water molecule was then considered and the potential curve from MLCI when using a larger cutoff of  $10^{-3}$  could not be distinguished from FCI on the scale of the graph except for a region around intermediate bond lengths. Transferring the wavefunction gave a noticeable reduction in errors but they were higher than using MCCI. The errors were lowered on decreasing the cutoff to  $5 \times 10^{-4}$  yet the curves remained slightly less accurate than when using MCCI with a cutoff of  $10^{-3}$ , although less processor hours were needed for MLCI. The cutoff had to be lowered to  $2 \times 10^{-4}$  to surpass the accuracy of the MCCI  $10^{-3}$  result and now MLCI required more processor hours. It was suggested that the smaller configuration space and lower multireference character of this system meant that there was less opportunity for the ANN to learn and not so much room for improvement on the already relatively fast MCCI calculations.

Finally we looked at carbon monoxide when using a cc-pVDZ basis and found that when transferring the wavefunction then MLCI gave essentially a FCI quality potential curve for the points where FCI data were available. This MLCI curve was of slightly higher accuracy than previous results using random selection of configurations with MCCI and used slightly less wall time. Furthermore the number of processor hours used was significantly lowered by MLCI from 248.6 to 19.8. Again transferring solely the wavefunction gave the best results while now only transferring the ANN was the second most accurate protocol, but there was not much difference in accuracy between this and either transferring all data or nothing.

We have seen that this improved, more scalable machine learning configuration interaction can cope with geometries ranging from single reference to very strongly multireference

problems to produce potential curves that are practically FCI quality. MLCI generally performed best when transferring just the wavefunction. This protocol, when running in serial, used less wall time and processor hours but gave a more accurate curve than when stochastically selecting configurations in parallel for  $N_2$  and CO. This was not the case however for the water molecule, which was attributed to its configuration space and multireference character being lower. It was generally seen that transferring the ANN from a near geometry as a starting point did not confer any advantage in accuracy but nor did it significantly change the calculation time, while transferring just the wavefunction could noticeably improve accuracy but required more time. Hence it appeared that the best approach was to have the ANN start learning afresh but begin with the configurations of a near geometry. As only a single hidden layer with 40 nodes was used here then training of the ANN does not represent a significant cost, it is possible that transferring the ANN could offer an improvement in time and accuracy when using deep nets trained on graphical processing units which will be investigated in future work when applying the approach to larger molecules. A parallel MLCI will also be considered where, for example, the configurations that have been predicted to be important are not added all at once for a single diagonalization, but shared between smaller concurrent diagonalizations to provide approximate training data on their importance in the wavefunction.

## Acknowledgement

JPC thanks the EPSRC for support via the platform grant EP/P001459/1.

## References

- (1) Qu, C.; Yu, Q.; Van Hoozen Jr., B. L.; Bowman, J. M.; Vargas-Hernández, R. A. Assessing Gaussian Process Regression and Permutationally Invariant Polynomial Ap-

- proaches To Represent High-Dimensional Potential Energy Surfaces. *J. Chem. Theory Comput.* **2018**, *14*, 7.
- (2) Schmitz, G.; Christiansen, O. Gaussian process regression to accelerate geometry optimizations relying on numerical differentiation. *J. Chem. Phys.* **2018**, *148*, 241704.
  - (3) Wiens, A. E.; Copan, A. V.; Schaefer, H. F. Multi-fidelity Gaussian process modeling for chemical energy surfaces. *Chem. Phys. Lett.: X* **2019**, *3*, 100022.
  - (4) Gastegger, M.; Marquetand, P. High-Dimensional Neural Network Potentials for Organic Reactions and an Improved Training Algorithm. *J. Chem. Theory Comput.* **2015**, *11*, 2187.
  - (5) Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A. Quantum-chemical insights from deep tensor neural networks. *Nat. Comm.* **2017**, *8*, 13890.
  - (6) Janet, J. P.; Kulik, H. J. Predicting electronic structure properties of transition metal complexes with neural networks. *Chem. Sci.* **2017**, *8*, 5137.
  - (7) Hansen, K.; Montavon, G.; Biegler, F.; Fazli, S.; Rupp, M.; Scheffler, M.; von Lilienfeld, O. A.; Tkatchenko, A.; Müller, K. R. Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies. *J. Chem. Theory Comput.* **2013**, *9*, 3404.
  - (8) Faber, F. A.; Christensen, A. S.; Huang, B.; von Lilienfeld, O. A. Alchemical and structural distribution based representation for universal quantum machine learning. *J. Chem. Phys.* **2018**, *148*, 241717.
  - (9) Gubaev, K.; Podryabinkin, E. V.; Shapeev, A. V. Machine learning of molecular properties: Locality and active learning. *J. Chem. Phys.* **2018**, *148*, 241727.
  - (10) Yao, K.; Parkhill, J. Kinetic Energy of Hydrocarbons as a Function of Electron Density and Convolutional Neural Networks. *J. Chem. Theory Comput.* **2016**, *12*, 1139.

- (11) Brockherde, F.; Vogt, L.; Li, L.; Tuckerman, M. E.; Burke, K.; Müller, K.-R. Bypassing the Kohn-Sham equations with machine learning. *Nat. Comm.* **2017**, *8*, 872.
- (12) Seino, J.; Kageyama, R.; Fujinami, M.; Ikabata, Y.; Nakai, H. Semi-local machine-learned kinetic energy density functional with third-order gradients of electron density. *J. Chem. Phys.* **2018**, *148*, 241705.
- (13) Custódio, C. A.; Filletti, E. R.; França, V. V. Artificial neural networks for density-functional optimizations in fermionic systems. *Sci. Rep.* **2019**, *9*, 1886.
- (14) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618.
- (15) Purvis, G. D.; Bartlett, R. J. A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *J. Chem. Phys.* **1982**, *76*, 1910–1918.
- (16) Bartlett, R. J.; Musiał, M. Coupled-cluster theory in quantum chemistry. *Rev. Mod. Phys.* **2007**, *79*, 291.
- (17) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big Data Meets Quantum Chemistry Approximations: The  $\Delta$ -Machine Learning Approach. *J. Chem. Theory Comput.* **2015**, *11*, 2087–2096.
- (18) McGibbon, R. T.; Taube, A. G.; Donchev, A. G.; Siva, K.; Hernández, F.; Hargus, C.; Law, K.-H.; Klepeis, J. L.; Shaw, D. E. Improving the accuracy of Møller-Plesset perturbation theory with neural networks. *J. Chem. Phys.* **2017**, *147*, 161725.
- (19) Margraf, J. T.; Reuter, K. Making the Coupled Cluster Correlation Energy Machine-Learnable. *J. Phys. Chem. A* **2018**, *122*, 6343.
- (20) Welborn, M.; Cheng, L.; Miller, III, T. F. Transferability in Machine Learning for Electronic Structure via the Molecular Orbital Basis. *J. Chem. Theory Comput.* **2018**, *14*, 4772.

- (21) Coe, J. P. Machine learning configuration interaction. *J. Chem. Theory Comput.* **2018**, *14*, 5739.
- (22) Townsend, J.; Vogiatzis, K. D. Data-Driven Acceleration of the Coupled-Cluster Singles and Doubles Iterative Solver. *J. Phys. Chem. Lett.* **2019**, *10*, 4129.
- (23) Schütt, O.; VandeVondele, J. Machine Learning Adaptive Basis Sets for Efficient Large Scale Density Functional Theory Simulation. *J. Chem. Theory Comput.* **2018**, *14*, 4168.
- (24) Huron, B.; Malrieu, J. P.; Rancurel, P. Iterative perturbation calculations of ground and excited state energies from multiconfigurational zeroth-order wavefunctions. *J. Chem. Phys.* **1973**, *58*, 5745.
- (25) Holmes, A. A.; Tubman, N. M.; Umrigar, C. J. Heat-Bath Configuration Interaction: An Efficient Selected Configuration Interaction Algorithm Inspired by Heat-Bath Sampling. *J. Chem. Theory Comput.* **2016**, *12*, 3674.
- (26) Tubman, N. M.; Lee, J.; Takeshita, T. Y.; Head-Gordon, M.; Birgitta Whaley, K. A deterministic alternative to the full configuration interaction quantum Monte Carlo method. *J. Chem. Phys.* **2016**, *145*, 044112.
- (27) Ohtsuka, Y.; Hasegawa, J. Selected configuration interaction using sampled first-order corrections to wave functions. *J. Chem. Phys.* **2017**, *147*, 034102.
- (28) Ohtsuka, Y.; Nagase, S. Projector Monte Carlo method based on configuration state functions. Test applications to the H<sub>4</sub> system and dissociation of LiH. *Chem. Phys. Lett.* **2008**, *463*, 431.
- (29) Booth, G. H.; Thom, A. J. W.; Alavi, A. Fermion Monte Carlo without fixed nodes: A game of life, death, and annihilation in Slater determinant space. *J. Chem. Phys.* **2009**, *131*, 054106.

- (30) Ohtsuka, Y. Spin-symmetry adaptation to the Monte Carlo correction configuration interaction wave functions. *J. Chem. Phys.* **2019**, *151*, 034115.
- (31) Garniron, Y.; Applencourt, T.; Gasperich, K.; Benali, A.; Ferté, A.; Paquier, J.; Pradines, B.; Assaraf, R.; Reinhardt, P.; Toulouse, J.; Barbaresco, P.; Renon, N.; David, G.; Malrieu, J.-P.; Vénil, M.; Caffarel, M.; Loos, P.-F.; Giner, E.; Scemama, A. Quantum Package 2.0: An Open-Source Determinant-Driven Suite of Programs. *J. Chem. Theory Comput.* **2019**, *15*, 3591.
- (32) Loos, P.-F.; Scemama, A.; Blondel, A.; Garniron, Y.; Caffarel, M.; Jacquemin, D. A Mountaineering Strategy to Excited States: Highly Accurate Reference Energies and Benchmarks. *J. Chem. Theory Comput.* **2018**, *14*, 4360.
- (33) Scemama, A.; Garniron, Y.; Caffarel, M.; Loos, P.-F. Deterministic Construction of Nodal Surfaces within Quantum Monte Carlo: The Case of FeS. *J. Chem. Theory Comput.* **2018**, *14*, 1395.
- (34) Scemama, A.; Benali, A.; Jacquemin, D.; Caffarel, M.; Loos, P.-F. Excitation energies from diffusion Monte Carlo using selected configuration interaction nodes. *J. Chem. Phys.* **2018**, *149*, 034108.
- (35) Applencourt, T.; Gasperich, K.; Scemama, A. Spin adaptation with determinant-based selected configuration interaction. *arXiv:1812.06902* **2018**,
- (36) Garniron, Y.; Scemama, A.; Giner, E.; Caffarel, M.; Loos, P.-F. Selected configuration interaction dressed by perturbation. *J. Chem. Phys.* **2018**, *149*, 064103.
- (37) Neese, F. Software update: the ORCA program system, version 4.0. *WIREs Comput Mol Sci* **2018**, *8*, e1327.
- (38) Evangelista, F. A. Adaptive multiconfigurational wave functions. *J. Chem. Phys.* **2014**, *140*, 124114.



- (39) Schriber, J. B.; Evangelista, F. A. Communication: An adaptive configuration interaction approach for strongly correlated electrons with tunable accuracy. *J. Chem. Phys.* **2016**, *144*, 161106.
- (40) Schriber, J. B.; Evangelista, F. A. Adaptive Configuration Interaction for Computing Challenging Electronic Excited States with Tunable Accuracy. *J. Chem. Theory Comput.* **2017**, *13*, 5354.
- (41) Greer, J. C. Estimating full configuration interaction limits from a Monte Carlo selection of the expansion space. *J. Chem. Phys.* **1995**, *103*, 1821.
- (42) Greer, J. C. Monte Carlo Configuration Interaction. *J. Comp. Phys.* **1998**, *146*, 181.
- (43) Tong, L.; Nolan, M.; Cheng, T.; Greer, J. C. A Monte Carlo configuration generation computer program for the calculation of electronic states of atoms, molecules, and quantum dots. *Comp. Phys. Comm.* **2000**, *131*, 142, see <https://github.com/MCCI/mcci>.
- (44) Győrffy, W.; Bartlett, R. J.; Greer, J. C. Monte Carlo configuration interaction predictions for the electronic spectra of Ne, CH<sub>2</sub>, C<sub>2</sub>, N<sub>2</sub>, and H<sub>2</sub>O compared to full configuration interaction calculations. *J. Chem. Phys.* **2008**, *129*, 064103.
- (45) Coe, J. P.; Paterson, M. J. Approaching exact hyperpolarizabilities via sum-over-states Monte Carlo configuration interaction. *J. Chem. Phys.* **2014**, *141*, 124118.
- (46) Coe, J. P.; Paterson, M. J. Multireference X-ray emission and absorption spectroscopy calculations from Monte Carlo configuration interaction. *Theor. Chem. Acc.* **2015**, *134*, 58.
- (47) Szepieniec, M.; Yeriskin, I.; Greer, J. C. Quasiparticle energies and lifetimes in a metallic chain model of a tunnel junction. *J. Chem. Phys.* **2013**, *138*, 144105.

- (48) Kelly, T. P.; Perera, A.; Bartlett, R. J.; Greer, J. C. Monte Carlo configuration interaction with perturbation corrections for dissociation energies of first row diatomic molecules: C<sub>2</sub>, N<sub>2</sub>, O<sub>2</sub>, CO, and NO. *J. Chem. Phys.* **2014**, *140*, 084114.
- (49) Murphy, P.; Coe, J. P.; Paterson, M. J. Development of spin-orbit coupling for stochastic configuration interaction techniques. *J. Comp. Chem.* **2018**, *39*, 319.
- (50) Wang, Z.; Li, Y.; Lu, J. Coordinate Descent Full Configuration Interaction. *J. Chem. Theory Comput.* **2019**, *15*, 3558.
- (51) Lyakh, D. I.; Bartlett, R. J. An adaptive coupled-cluster theory: @CC approach. *J. Chem. Phys.* **2010**, *133*, 244112.
- (52) Bytautas, L.; Ruedenberg, K. A priori identification of configurational deadwood. *Chem. Phys.* **2009**, *356*, 64.
- (53) Zimmerman, P. M. Incremental full configuration interaction. *J. Chem. Phys.* **2017**, *146*, 104102.
- (54) Eriksen, J. J.; Lipparini, F.; Gauss, J. Virtual Orbital Many-Body Expansions: A Possible Route towards the Full Configuration Interaction Limit. *J. Phys. Chem. Lett.* **2017**, *8*, 4633.
- (55) Eriksen, J. J.; Gauss, J. Many-Body Expanded Full Configuration Interaction. I. Weakly Correlated Regime. *J. Chem. Theory Comput.* **2018**, *14*, 5180.
- (56) Eriksen, J. J.; Gauss, J. Many-Body Expanded Full Configuration Interaction. II. Strongly Correlated Regime. *J. Chem. Theory Comput.* **2019**, *15*, 4873.
- (57) Davidson, E. R. The Iterative Calculation of a Few of the Lowest Eigenvalues and Corresponding Eigenvectors of Large Real-Symmetric Matrices. *J. Comp. Phys.* **1975**, *17*, 87.

- (58) H.-J. Werner,; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M.; Celani, P.; Györffy, W.; Kats, D.; Korona, T.; Lindh, R.; Mitrushenkov, A.; Rauhut, G.; Shamasundar, K. R.; Adler, T. B.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Goll, E.; Hampel, C.; Hesselmann, A.; Hetzer, G.; Hrenar, T.; Jansen, G.; Köppl, C.; Liu, Y.; Lloyd, A. W.; Mata, R. A.; May, A. J.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; O’Neill, D. P.; Palmieri, P.; Peng, D.; Pflüger, K.; Pitzer, R.; Reiher, M.; Shiozaki, T.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Wang, M. MOLPRO, version 2015.1, a package of ab initio programs. 2015; see.
- (59) Coe, J. P.; Paterson, M. J. Development of Monte Carlo configuration interaction: Natural orbitals and second-order perturbation theory. *J. Chem. Phys.* **2012**, *137*, 204108.
- (60) Larsen, H.; Olsen, J.; Jørgensen, P.; Christiansen, O. Full configuration interaction benchmarking of coupled-cluster models for the lowest singlet energy surfaces of N<sub>2</sub>. *J. Chem. Phys.* **2000**, *113*, 6677.
- (61) Chan, G. K.-L.; Kállay, M.; Gauss, J. State-of-the-art density matrix renormalization group and coupled cluster theory studies of the nitrogen binding curve. *J. Chem. Phys.* **2004**, *121*, 6110–6116.
- (62) Gwaltney, S. R.; Byrd, E. F. C.; Voorhis, T. V.; Head-Gordon, M. A perturbative correction to the quadratic coupled-cluster doubles method for higher excitations. *Chem. Phys. Lett.* **2002**, *353*, 359.
- (63) Coe, J. P.; Murphy, P.; Paterson, M. J. Applying Monte Carlo configuration interaction to transition metal dimers: Exploring the balance between static and dynamic correlation. *Chem. Phys. Lett.* **2014**, *604*, 46.

- (64) Coe, J. P.; Paterson, M. J. Investigating Multireference Character and Correlation in Quantum Chemistry. *J. Chem. Theory Comput.* **2015**, *11*, 4189.
- (65) Coe, J. P.; Paterson, M. J. Open-shell systems investigated with Monte Carlo configuration interaction. *Int. J. Quantum Chem.* **2016**, *116*, 1772.
- (66) Li, X.; Paldus, J. Comparison of the open-shell state-universal and state-selective coupled-cluster theories:  $H_4$  and  $H_8$  models. *J. Chem. Phys.* **1995**, *103*, 1024.
- (67) Knowles, P. J.; Handy, N. C. A Determinant Based Full Configuration Interaction Program. *Comp. Phys. Comm.* **1989**, *54*, 75.