

## Deep-Learning Approach for the Atomic Configuration Interaction Problem on Large Basis Sets

Pavlo Bilous<sup>1,2,\*</sup>, Adriana Pálffy<sup>3,2</sup> and Florian Marquardt<sup>1,4</sup>

<sup>1</sup>Max Planck Institute for the Science of Light, Staudtstraße 2, 91058 Erlangen, Germany

<sup>2</sup>Max Planck Institute for Nuclear Physics, Saupfercheckweg 1, 69117 Heidelberg, Germany

<sup>3</sup>Institute of Theoretical Physics and Astrophysics, University of Würzburg, Am Hubland, 97074 Würzburg, Germany

<sup>4</sup>Department of Physics, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany



(Received 13 September 2022; revised 29 June 2023; accepted 9 August 2023; published 27 September 2023)

High-precision atomic structure calculations require accurate modeling of electronic correlations typically addressed via the configuration interaction (CI) problem on a multiconfiguration wave function expansion. The latter can easily become challenging or infeasibly large even for advanced supercomputers. Here, we develop a deep-learning approach which allows us to preselect the most relevant configurations out of large CI basis sets until the targeted energy precision is achieved. The large CI computation is thereby replaced by a series of smaller ones performed on an iteratively expanding basis subset managed by a neural network. While dense architectures as used in quantum chemistry fail, we show that a convolutional neural network naturally accounts for the physical structure of the basis set and allows for robust and accurate CI calculations. The method was benchmarked on basis sets of moderate size allowing for the direct CI calculation, and further demonstrated on prohibitively large sets where the direct computation is not possible.

DOI: 10.1103/PhysRevLett.131.133002

Precise knowledge of atomic structure is indispensable for frequency standards in metrology, spectral analysis in astrophysics, understanding of nuclear phenomena involving atomic electrons, or physics beyond the standard model, e.g., space and time variation of fundamental constants [1]. High performance *ab initio* atomic structure codes provide a wide range of electronic properties of atoms and ions, such as energy levels, radiative transition rates,  $g$  factors, or hyperfine structure constants. The practical difficulty is many-body effects in atoms or ions with high atomic number  $Z$  and many electrons. The electronic correlations are typically tackled by the configuration interaction (CI) method based on the multiconfiguration wave function expansion  $|\Psi\rangle = \sum_{\alpha} c_{\alpha} |\Phi_{\alpha}\rangle$  with unknown coefficients  $c_{\alpha}$  obtained as a solution of the Hamiltonian diagonalization problem  $\hat{H}|\Psi\rangle = E|\Psi\rangle$  [2]. The size of the involved basis set  $\{|\Phi_{\alpha}\rangle\}$  can easily become challenging even for state-of-the-art parallelized codes running on supercomputer systems; see, e.g., recent calculations for Th<sup>35+</sup> [3–5], Ir<sup>17+</sup> [6,7], or Fe<sup>16+</sup> [6,8].

Instead of computations on the full basis, “selected CI” methods were applied to atomic and molecular systems using selection criteria based on perturbation theory [9,10] or the Monte Carlo approach [11,12]. However, perturbation theory still requires computations on the entire basis, whereas the random selection completely disregards the properties of the basis states. Fortunately, the application of machine learning techniques has led in recent years to significant progress in selected CI in quantum chemistry [13–16]. For importing this success in the field of large-scale atomic computations, the neural network (NN) approach would be the first choice due to its established scalability and flexibility [17]. Used in the active learning algorithm as presented in Refs. [13,14], it would allow us to iteratively construct a compact wave function that delivers accurate observables without the computational effort on the full basis. However, we show here that the usual dense architectures as applied in Refs. [13,14] lack sufficient accuracy and often fail. In this work we apply instead a convolutional NN (CNN)—the architecture well known from image recognition applications [17,18]. We demonstrate that this is the natural choice considering the physical structure of the atomic basis states, and leads to robustness of the approach and a strong improvement of the computational results.

In this Letter, we develop an efficient deep-learning approach to iteratively construct a compact approximative wave function for high- $Z$  atoms and ions with many electrons. We address the problem in the coupled basis

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Open access publication funded by the Max Planck Society.

of configuration state functions (CSF) [1,2] characterized by electronic orbital occupations, and the angular momenta couplings within and between the orbitals. Typically, the physical properties of a CSF are determined predominantly by a few orbitals (different for each CSF). The others form a “background” consisting of low-energy fully occupied and high-energy empty orbitals. In analogy to image recognition applications, the applied CNN efficiently suppresses this background and highlights the “useful image” of the physically relevant orbitals in each CSF. We demonstrate that this natural choice accounting for the physical structure of CSFs leads to significant improvements in comparison to NNs of the usual dense type. Our CI solver is based on an iterative scheme employing the CNN together with the general relativistic atomic structure package GRASP2018 [19].

The CSFs  $|\Phi_\alpha\rangle$  based on  $N$  orbitals are uniquely characterized by the set of  $3N$  quantum numbers generically denoted as  $\alpha$ . For orbital  $k$ , they consist of the population  $n_k$ , the total angular momentum of its electrons  $J_k$ , and the angular momentum  $J_k^{\text{cpl}}$  representing the coupling of  $J_k$  and  $J_{k-1}^{\text{cpl}}$  [20,21]. We normalize the populations of the orbitals  $n_k$  to their maximal capacity, and the angular momenta  $J_k$  and  $J_k^{\text{cpl}}$  to the total angular momentum of the considered energy level (see also Supplemental Material [22] for the CSF basis construction). The three classes of input data  $\tilde{n}_k$ ,  $\tilde{J}_k$ ,  $\tilde{J}_k^{\text{cpl}}$  are interpreted as color channels of a 1D convolutional input layer. In Fig. 1(b) we show graphically this color representation for an exemplary CSF. The value of each parameter is encoded by the length of the corresponding vertical bar, whereas the gray horizontal strips indicate the unity bar length. We consider here the natural ordering of the orbitals which is default in GRASP2018.

Our network architecture is shown in Fig. 1(a). The input layer (A) consists of three color channels [see Fig. 1(b)] of size  $N$ . The input is processed with a filter kernel of size 3 (B) resulting in 96 feature maps (C) each of size  $N - 2$ . The latter are mapped to 16 feature maps (D) of size  $N - 2$  by application of a filter kernel of size 1 (thus representing a purely local transformation). The CNN part (D) is monitored for observation of the background suppression effect which we show in Fig. 1(c) and discuss further on. The obtained output of  $16 \times (N - 2)$  values is then flattened and forwarded to a network of four dense layers (E) with 150, 120, 90, and 2 neurons, respectively. The rectified linear unit (ReLU) defined as  $\max(0, x)$  was used as the activation function throughout the NN apart from the two-neuron output layer (F), where the softmax function is applied yielding the probabilities of the CSF to be “important” or “unimportant.” The categorical cross-entropy was chosen as the loss function. The NN is trained on batches using the Adam optimization algorithm with early stopping based on the classification accuracy evaluated on a validation set (20% of data excluded in advance from the training set). We used the PYTHON library KERAS [23,24] with TENSORFLOW [25] in the back end.

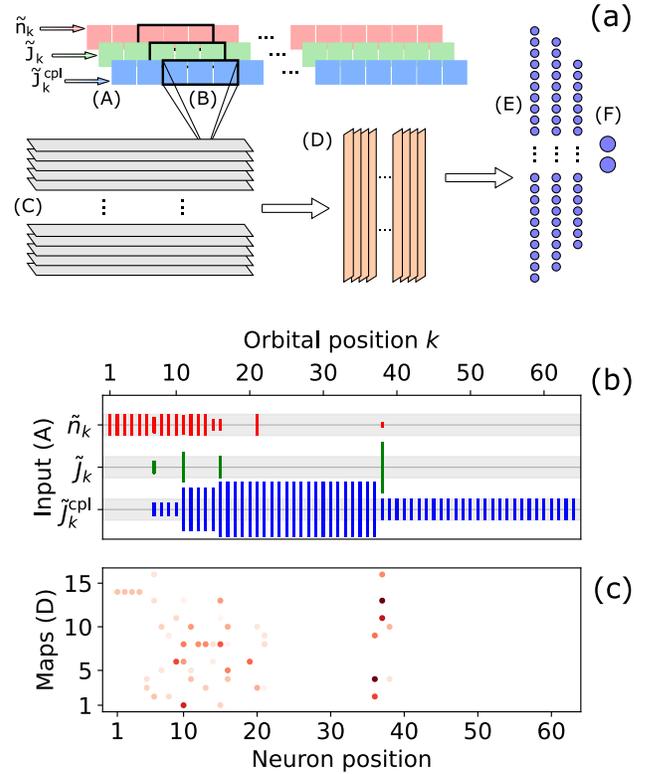


FIG. 1. (a) NN architecture used in the present work. (b) Color representation of the exemplary CSF from the set  $SD^*(3p, 9h)$  of the Re atom as the NN input (A). (c) The background suppression observed in the neurons of the feature maps (D) of the NN. See the text and Supplemental Material [22] for details.

The described NN is employed in an iterative active-training algorithm based on the scheme from Refs. [13,14]. Each CSF is either important or unimportant: its weight in the CI wave function  $w_\alpha = |c_\alpha|^2$  either exceeds or does not exceed a cutoff value  $w^0$  chosen in advance. CSFs are included in the CI expansion iteratively in relatively small portions based on the NN prediction of their importance. The diagonalization using GRASP2018 yields the energy and the coefficients  $c_\alpha$ . The latter are used for a feedback and additional training of the NN, monitoring the energy to stop when the targeted precision is achieved. In each iteration, CSFs which turned out to be unimportant are excluded from the CI expansion, but are considered again in later iterations. Instead of using a fixed cutoff as in Refs. [13,14], we use a running cutoff taking at the  $i$ th iteration a new value  $w^i < w^{i-1} < \dots < w^1 < w^0$ . This approach is crucial for avoiding energy convergence to an unwanted value which does not correspond to the full set, but to a smaller set of CSFs having weights that exceed the fixed cutoff value. This point was also observed in Ref. [15].

The NN needs feedback not only on the selected but also on rejected CSFs. Therefore, we include in the CI expansion the same amount of randomly picked disregarded

CSFs as the selected ones in every iteration (but the last one since no NN training follows). Most of these balancing CSFs are automatically excluded in the next iteration due to their small weights. Before feeding into the NN, the training data are reshuffled for avoiding the ordering bias due to the CSF construction procedure. Some important CSFs for a particular electronic level are known from the start and should always be included in the CI set. They form the primary subset and we do not expose them to the NN at any stage. We have checked that inclusion of the primary subset in the training set does not bring any improvement for the method. At the starting point, the NN is trained on a random selection of CSFs from the considered basis set (excluding the primary subset). At the same time, the distribution of these CSFs over their weights is used to choose the running cutoff values  $w^i$ . Throughout this work we use 1% of CSFs for this starting iteration.

For demonstration, we choose the case of the  $^{187}\text{Re}$  ( $[\text{Xe}]4f^{14}5d^56s^2$ ) and  $^{187}\text{Os}$  ( $[\text{Xe}]4f^{14}5d^66s^2$ ) atoms and calculate their ground state energies. These energies have been recently evaluated with GRASP2018 to extract the  $\beta$ -decay energy of the  $^{187}\text{Re}$  nucleus from experimentally determined masses of  $^{187}\text{Re}^{29+}$  and  $^{187}\text{Os}^{29+}$  [26]. The basis sets in Ref. [26] contain states stemming also from additional configurations obtained by allowing for electronic excitations from the main configuration. Single (S) and double (D) excitations from the filled orbitals down to  $3p$  to the vacant (virtual) orbitals up to  $9h$  were considered in [26] resulting in over  $90 \times 10^6$  CSFs (see Supplemental Material [22] for details on the basis construction). We denote these sets here as  $\text{SD}(3p, 9h)$ . Because of the prohibitively large basis set size, the authors of Ref. [26] had to preselect about  $5 \times 10^6$  most important CSFs by evaluating transition and ionization energies and fitting them to experimental values [27].

For benchmarking, we apply our method to a smaller  $^{187}\text{Re}$  basis set  $\text{SD}^*(3p, 9h)$  of 4 267 362 CSFs in which only part of the double excitations are allowed, with the restriction that each virtual orbital can be either doubly occupied or empty. The moderate size of  $\text{SD}^*(3p, 9h)$  allows for comparison of our approach to direct GRASP2018 computations. The primary subset consists of 37 220 CSFs constructed from SD excitations to the valence orbitals and S excitations to the virtual orbitals. The radial electronic wave functions are obtained with GRASP2018 in advance on the primary CSF set using the layer-by-layer procedure as described in Ref. [28]. Table I shows the results obtained in each iteration: the energy  $E^{\text{part}}$  on the current partial CI set with respect to the exact value  $E^{\text{all}} = -454\,661.1637$  eV (obtained separately in a direct calculation) and the number of CSFs in the GRASP2018 run. We note that for CI on a partial basis the energies always satisfy  $E^{\text{part}} > E^{\text{all}}$  [29]. The iterations are labeled by  $\log_{10} w^i$  where  $w^i$  is the running cutoff value at the  $i$ th iteration. After the very last iteration, CSFs unimportant with respect to the value

TABLE I. Results of approximate energy calculations on the  $\text{SD}^*(3p, 9h)$  basis set for the Re atom ground state using our deep-learning-based approach. Iterations are labeled by  $\log_{10} w^i$  where  $w^i$  is the running cutoff value at the  $i$ th iteration. The row labeled as “Initial” represents the initial iteration on 1% randomly chosen CSFs.

$\log_{10} w^i$	CSFs in GRASP	$E^{\text{part}} - E^{\text{all}}$ (meV)
Initial	79 521	17 223.3
-8.6	178 901	6431.2
-9.2	364 562	802.9
-9.8	515 289	140.3
-10.4	723 540	31.4
-11.0	755 766	6.4

$\log_{10} w^i = -11.6$  as calculated by GRASP2018 are excluded from the CI wave function yielding the final CI set with 729 451 instead of 755 766 CSFs. The latter step is helpful for further calculations on the state, e.g., refining of the radial wave functions or evaluation of QED corrections and isotope shifts. Both the direct and the NN-supported computations could be carried out in a few hours on a few hundred cores. The peak memory and disk space consumption which are the bottleneck in larger GRASP2018 computations (see Supplemental Material [22]) could be decreased in this case from a TB to a few hundred GB.

Figure 2 illustrates the growth of the CI wave function expansion. We plot the number of CSFs from the set  $\text{SD}^*(3p, 9h)$  not (yet) included in the CI expansion as a function of the weight  $\log_{10} w_\alpha$  for each iteration immediately after the unimportant CSFs are excluded. The distribution is normalized with respect to the total size of the  $\text{SD}^*(3p, 9h)$  set and the weights  $w_\alpha$  are taken from the full

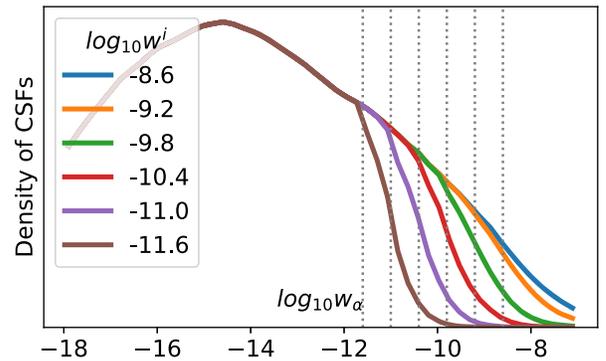


FIG. 2. The distribution of CSF from the set  $\text{SD}^*(3p, 9h)$  not (yet) included in the CI expansion of the wave function as a function of their respective weights  $\log_{10} w_\alpha$  for each iteration. The distributions are normalized with respect to the total size of the  $\text{SD}^*(3p, 9h)$  set. The iterations are labeled by the cutoff values  $\log_{10} w^i$  which are additionally illustrated by the vertical dotted lines.

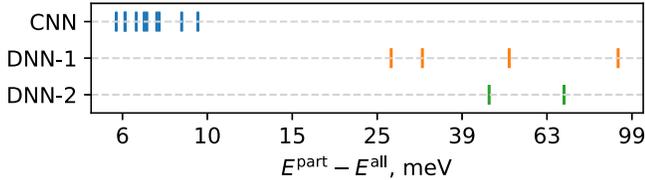


FIG. 3. Deviation of the final energy on the partial set  $E^{\text{part}}$  with respect to the “full” energy  $E^{\text{all}}$  obtained using the CNN (10 runs) and dense networks DNN-1 and DNN-2 (five runs each). Note the logarithmic scale on the horizontal axis. The missing DNN runs have failed.

GRASP2018 calculation. In each iteration, CSFs are included in the CI expansion (and thus removed from the depicted distributions) from the right. The right edge of the distributions is not completely sharp, meaning that not all CSFs important with respect to the current cutoff are included in the CI expansion. The NN selection ensures that the slope becomes stable in the first iterations and moves then from right to left.

We have performed the same computation replacing the CNN by a usual dense NN. We considered two dense NN architectures: DNN-1 is the dense part (E)—(F) of the applied CNN [see Fig. 1(a)]; DNN-2 has three hidden layers with 192, 384, 192 neurons, respectively, possessing in total a similar number of trainable parameters as the considered CNN. In Fig. 3 we show the final energy  $E^{\text{part}} - E^{\text{all}}$  in milli-electron-volts obtained in a few computation repetitions using the CNN and dense networks DNN-1 and DNN-2. Often, the DNNs fail, and this takes place already in the first iteration due to the randomly selected and thus strongly disbalanced training set. In contrast to the DNNs, our CNN is more robust against this disbalance and failed only a few times in hundreds of runs. In case the DNNs do cope with the first iteration, they are still strongly outperformed by the CNN. The processing of the input data using a kernel in the CNN plays a twofold role: (a) application of the same weights along the input neurons (independent of the orbital ordering); and (b) revealing mutual data dependencies for the neighboring input neurons (sensitive to the orbital ordering). A careful analysis on how these mechanisms contribute to the performance of the CNN reveals that (a) plays the most important role (see Supplemental Material [22]).

The CNN tends to treat the fully occupied low-lying orbitals as well as the completely vacant high-energy orbitals as a common background. This background is suppressed and the remaining “useful image” corresponding to the physically relevant partially occupied orbitals is highlighted. We demonstrate this effect in Fig. 1(c) which shows the values of the 62 neurons in each of the 16 feature maps (D) for the exemplary CSF. These neurons are in positional correspondence with the 64 input neurons (A). The values in (D) are plotted at the moment when the

computation is completed and the NN is in its final state. The color intensity of the red dots indicates the (always non-negative) values normalized to the maximal value in all the feature maps (higher intensity corresponds to a larger value). Almost all neurons in the region of the common background to the left and to the right turn out to have zero values after the NN training. Further discussion of this effect and more examples can be found in Supplemental Material [22].

We switch now to calculations on the large basis sets  $\text{SD}(3p, 9h)$  for the Re and Os neutral atoms relevant for the determination of the  $^{187}\text{Re}$   $\beta$ -decay energy in Ref. [26]. These calculations involve basis sets of over  $90 \times 10^6$  CSFs each and cannot be performed directly using GRASP2018. However, it is sufficient to retain for each basis set only the most important CSFs that deliver a 1 eV precision for the calculated energy. Using our deep-learning approach, we could achieve the targeted accuracy in a few days by performing partial GRASP2018 runs on up to about  $5 \times 10^6$  CSFs which required 5 TB of memory and 7 TB of disk space. The primary CSF subsets were constructed as in the previous example resulting in 37 220 and 32 660 CSFs for the Re and Os atom, respectively. The radial electronic wave functions were also here obtained on the primary CSF sets. Table II shows the Re and Os energies and basis set sizes at the diagonalization stage in each iteration. The energy values obtained in the last iteration satisfy our precision target. We carried out additional verifications by running the computation on other cutoffs and made sure that they lead to the same energy value within the required accuracy. The obtained binding energies of the Re and Os neutral atoms are  $E_{\text{Re}}^{\text{atom}} = -454\,703.55$  eV and  $E_{\text{Os}}^{\text{atom}} = -470\,036.60$  eV, respectively.

In order to compare our results with Ref. [26] where the electronic binding energy differences  $\delta E = E^{\text{atom}} - E^{\text{ion}}$  between a neutral atom and a 29+ ion for Re and Os were

TABLE II. Results of approximate energy calculations on the  $\text{SD}(3p, 9h)$  basis set for the ground state of the Re and Os atoms using our deep-learning-based approach. Iterations are labeled by  $\log_{10} w^i$  where  $w^i$  is the running cutoff value at the  $i$ th iteration. The row labeled as “Initial” represents the initial iteration on 1% randomly chosen CSFs.

	Re		Os	
	CSFs	$-E^{\text{part}}$ , eV	CSFs	$-E^{\text{part}}$ , eV
$\log_{10} w^i$		454 000+		470 000+
Initial	971 011	644.55	985 571	-23.64
-8.0	578 018	668.16	628 961	9.07
-8.5	1 609 943	684.00	972 374	25.60
-9.0	2 055 985	697.82	1 345 026	33.04
-9.5	2 550 922	701.65	2 046 765	35.71
-10.0	3 607 689	702.97	2 397 010	36.36
-10.5	4 028 106	703.55	3 185 458	36.60

provided, we evaluate the energies of the  $\text{Re}^{29+}$  and  $\text{Os}^{29+}$  ions on the basis sets  $\text{SD}(3p, 9h)$ . Since these consist of only 53 885 and 2 455 449 CSFs, respectively, we carry out the GRASP diagonalization on the full sets directly. The radial wave functions for the ions were obtained using the layer-by-layer procedure [28] on the full set  $\text{SD}(3p, 9h)$  for the  $\text{Re}^{29+}$  ion and on a partial set constructed as a union  $\text{SD}(3p, 5g) \cup \text{SD}^*(3p, 6h) \cup \text{S}(3p, 9h)$  for the  $\text{Os}^{29+}$  ion. The obtained ion energies are  $E_{\text{Re}}^{\text{ion}} = -443\,804.16$  eV and  $E_{\text{Os}}^{\text{ion}} = -459\,083.43$  eV. Based on the calculated atom and ion energies, we find  $\delta E_{\text{Re}} = -10\,899.39$  eV and  $\delta E_{\text{Os}} = -10\,953.17$  eV, which agree with the values  $\delta \tilde{E}_{\text{Re}} = -10\,894.5 \pm 25.9$  eV and  $\delta \tilde{E}_{\text{Os}} = -10\,947.9 \pm 24.6$  eV from Ref. [26]. For the difference  $\Delta E = \delta E_{\text{Re}} - \delta E_{\text{Os}}$  relevant for the computation of the  $^{187}\text{Re}$   $\beta$ -decay energy, we obtain  $\Delta E = 53.78$  eV, whereas in Ref. [26] the value  $\Delta \tilde{E} = 53.4 \pm 1.0$  eV was reported. Our approach allowed us to achieve the same results as in Ref. [26] without relying on additional experimental information not always available. We note that the individual atom and ion energies change upon inclusion of the QED corrections and further improvement of the radial wave functions. At the same time, the obtained value  $\Delta E$  does not change significantly due to the cancellation effects originating from similarity of the Re and Os electronic shells—a fact observed also in Ref. [26].

In conclusion, we have applied a CNN combined with the “color” representation of the CSFs for atomic CI calculations with large basis sets—a natural choice leading to robustness of the method and a strong improvement in precision. Analogously to image recognition, our CNN recognizes and suppresses the background corresponding to the fully occupied low-lying and the vacant high-energy orbitals, whereas the “useful image” related to the physically relevant partially filled orbitals is highlighted. We believe that this approach can be useful also in other areas where the CI method is applied. The code for NN-supported GRASP2018 computations is available in Ref. [30] and can be adapted for the upcoming new GRASP version [31].

We are thankful to Chunhai Lyu, Marianna Safronova, Sergey Porsev, and Charles Cheung for useful discussions. A. P. gratefully acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG) in the framework of the Heisenberg Program.

---

\* pavlo.bilous@mpl.mpg.de

- [1] C. Froese Fischer, M. Godefroid, T. Brage, P. Jönsson, and G. Gaigalas, *J. Phys. B* **49**, 182004 (2016).  
 [2] I. P. Grant, *Relativistic Calculation of Molecular Properties* (Springer, New York, NY, 2007).

- [3] P. V. Bilous, H. Bekker, J. C. Berengut, B. Seiferle, L. von der Wense, P. G. Thirolf, T. Pfeifer, J. R. C. López-Urrutia, and A. Pálffy, *Phys. Rev. Lett.* **124**, 192502 (2020).  
 [4] S. G. Porsev, C. Cheung, and M. S. Safronova, *Quantum Sci. Technol.* **6**, 034014 (2021).  
 [5] E. Peik, T. Schumm, M. S. Safronova, A. Pálffy, J. Weitenberg, and P. G. Thirolf, *Quantum Sci. Technol.* **6**, 034002 (2021).  
 [6] C. Cheung, M. Safronova, and S. Porsev, *Symmetry* **13**, 4 (2021).  
 [7] M. G. Kozlov, M. S. Safronova, J. R. Crespo López-Urrutia, and P. O. Schmidt, *Rev. Mod. Phys.* **90**, 045005 (2018).  
 [8] S. Kühn *et al.*, *Phys. Rev. Lett.* **124**, 225001 (2020).  
 [9] R. J. Harrison, *J. Chem. Phys.* **94**, 5021 (1991).  
 [10] V. A. Dzuba, V. V. Flambaum, and M. G. Kozlov, *Phys. Rev. A* **54**, 3948 (1996).  
 [11] J. C. Greer, *J. Chem. Phys.* **103**, 7996 (1995).  
 [12] J. Greer, *J. Comput. Phys.* **146**, 181 (1998).  
 [13] J. P. Coe, *J. Chem. Theory Comput.* **14**, 5739 (2018).  
 [14] W. Jeong, C. A. Gaggioli, and L. Gagliardi, *J. Chem. Theory Comput.* **17**, 7518 (2021).  
 [15] S. D. Pineda Flores, *J. Chem. Theory Comput.* **17**, 4028 (2021).  
 [16] J. J. Goings, H. Hu, C. Yang, and X. Li, *J. Chem. Theory Comput.* **17**, 5482 (2021).  
 [17] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, MA, 2016).  
 [18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, in *Advances in Neural Information Processing Systems*, edited by F. Pereira, C. Burges, L. Bottou, and K. Weinberger (Curran Associates, Inc., New York, 2012), Vol. 25.  
 [19] C. Froese Fischer, G. Gaigalas, P. Jönsson, and J. Bieroń, *Comput. Phys. Commun.* **237**, 184 (2019).  
 [20] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University Press, Princeton, NJ, 1996).  
 [21] G. E. Drake, *Springer Handbook of Atomic, Molecular, and Optical Physics* (Springer, New York, 2006).  
 [22] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.131.133002> for details on solution of the atomic CI problem in the CSF basis using GRASP2018, and processing of the input data by our CNN with background suppression.  
 [23] F. Chollet *et al.*, Keras, <https://keras.io> (2015).  
 [24] A. Gulli and S. Pal, *Deep Learning with Keras* (Packt Publishing Ltd., Birmingham, 2017).  
 [25] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, in *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, Savannah, GA (USENIX Association, Savannah, 2016), pp. 265–283.  
 [26] P. Filianin, C. Lyu, M. Door, K. Blaum, W. J. Huang, M. Haverkort, P. Indelicato, C. H. Keitel, K. Kromer, D. Lange, Y. N. Novikov, A. Rischka, R. X. Schüssler, C. Schweiger, S. Sturm, S. Ulmer, Z. Harman, and S. Eliseev, *Phys. Rev. Lett.* **127**, 072502 (2021).  
 [27] C. Lyu (private communication).  
 [28] J. Bieroń, C. F. Fischer, G. Gaigalas, I. Grant, and P. Jönsson, A practical guide to grasp2018—A collection of

- Fortran 95 programs with parallel computing using MPI, <https://github.com/compas/grasp> (2018).
- [29] A. Szabo and N. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Dover Books on Chemistry (Dover Publications, New York, 2012).
- [30] P. Bilous, Code for neural-network-supported GRASP computations, [https://github.com/pavlobilous/neural\\_grasp](https://github.com/pavlobilous/neural_grasp) (2023).
- [31] Y.T. Li, K. Wang, R. Si, M. Godefroid, G. Gaigalas, C. Y. Chen, and P. Jönsson, *Comput. Phys. Commun.* **283**, 108562 (2023).