Real-space machine learning of correlation density functionals

Elias Polak, Heng Zhao, and Stefan Vuckovic*)
Department of Chemistry, University of Fribourg, Chemin de Musée 9, CH-1700 Fribourg, Switzerland.

(Dated: 9 July 2025)

Machine learning (ML) plays a pivotal role in extending the reach of quantum chemistry methods for both molecules and materials. However, leveraging ML to improve upon human-designed density functional approximations (DFAs), the primary workhorse for quantum simulations, remains challenging due to severely limited transferability to unseen chemical systems. Here we address this challenge through real-space ML, where energies are learned point by point via correlation energy densities per particle obtained from regularized perturbation theory. We pursue two crucial strategies that enable the construction of highly transferable DFAs, grounded in the Møller–Plesset adiabatic connection framework, for correlation energies defined with respect to the Hartree–Fock reference. First, we introduce the Local Energy Loss, whose data efficiency (expanding each system's single energy into thousands of datapoints) dramatically enhances transferability when combined with a physically informed ML construction. Second, we construct a real-space, machine-learned, and regularized extension of Spin-Component-Scaled second-order Møller–Plesset perturbation theory, opening new avenues for developing transferable DFAs particularly suited for overcoming self-interaction errors common to traditional DFAs.

INTRODUCTION

Machine learning (ML) is driving a paradigm shift across scientific disciplines, including quantum chemistry (QC), where it reshapes the landscape of used methods[1–6]. The recent surge in ML has further increased the importance of density functional approximations (DFAs), already a cornerstone of quantum simulations in materials science and chemistry (see the recent review by von Lilienfeld and co-workers[7]). On one hand, DFAs generate vast amounts of data to train ML models[7–15], greatly extending their reach in time and length scales[16–25]. On the other hand, ML techniques provide improved DFAs[26–38] (DFAs), with the DM21 functional by DeepMind[39] being a prominent example.

A remaining critical problem with ML in QC is their limited transferability—the ability to generalize to unseen data[40–43]. These limitations hinder the applicability of ML models in QC, making users cautious and often leading them to stick with well-established *old-school* methods[44] over *new-school* ML counterparts. This situation has lead to a *no man's land* between old-school and new-school DFAs[45]: the promised revolution of ML-based DFAs is hampered by the far broader applicability of old-school DFAs[45], such as B3LYP[46–49] or PBE[50]. For example, while DM21's training on systems with fractional charges and spin addresses some limitations of old-school functionals[39], its catastrophic transferability failures have recently become evident[45]. Namely, it has been shown that DM21 does not converge for certain transition metal (TM) atoms—a convergence task easily handled by reputable old-school functionals[45]. Thus, it is no wonder that organic chemists still prefer old-school DFAs over DM21 or other new-school models for TM-catalyzed reaction mechanisms[51], despite major TM shortcomings of the former[52,53].

To move from this no man's land and leverage the power of ML for DFAs design, we need to solve the underlying transferability problem. Helping strategies are the use of physical constraints[54–57] (a mix of old- and new-school methods), more diverse data in the training set[41], or the engineering of new features[58]. Yet, an angle in ML of DFAs that requires more attention is the training data efficiency, particularly since feeding more data to ML models can become a never-ending game due to the data hunger in ML models and the vastness of chemical space[2,3,5]. The purpose of this work is to critically examine data efficiency in training DFAs and maximize it in order to embed transferability in ML of DFAs.

The primary goal when training DFAs is to learn how a given electronic density translates into energy. However, ML practices in QC currently treat energies as not very informative, following a 1 *system* = 1 *energy data point* approach (see Refs. [30–32, 34, 37, 39, 55, 59, and 60]). For example, when learning force fields, forces (energy gradients) are much more informative than energies[19,61]. Similarly, recent works have shown that in current ML DFA practices, electronic densities are also far more informative than energies[31,34,55,59,60] (each grid point is a density datapoint). While using electronic densities effectively enhances the transferability of ML-based DFAs, applying a point-by-point learning strategy to energies (typically the primary target of simulations) offers a distinct direction for improving ML-based DFAs. Identifying transferability as the key issue in ML DFAs, here we establish a framework for making energy training more data-efficient and address the challenges that must be overcome to leverage this data efficiency to embed transferability into ML DFAs.

In view of our objective to enhance data efficiency for energies in ML-based DFAs, we introduce real-space energy learning and apply it to machine-learn a DFA for

*)email:stefan.vuckovic@unifr.ch

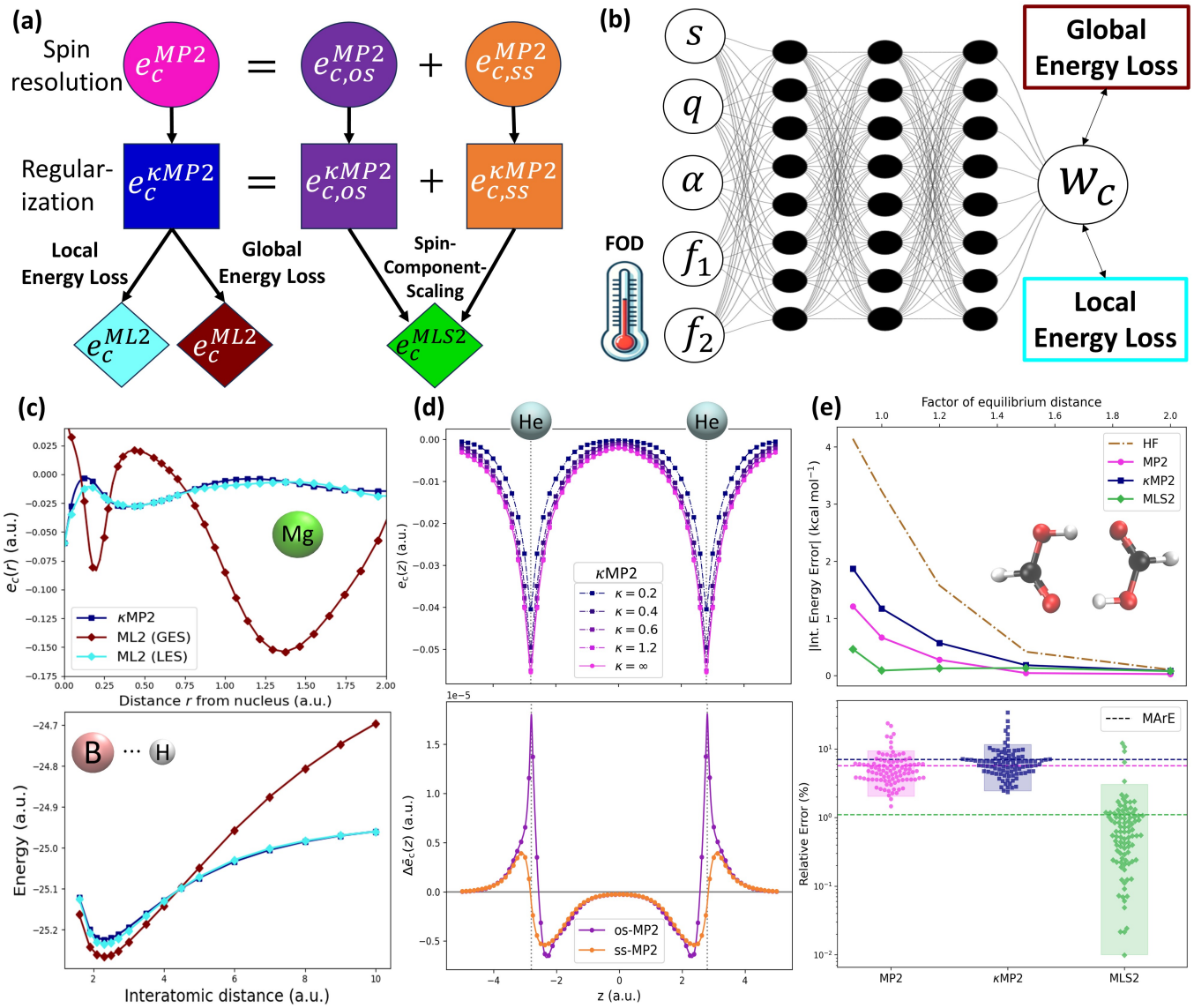

Fig. 1: Real-space machine learning correlation energy densities (a): Overview of methods used here and their connections. **(b):** Neural network illustrations with real-space features defined in Sec. S6 (reduced density gradient s, reduced density Laplacian q, regularized energy kinetic variable α and temperature-dependent fractional occupation number weighted densities[62,63] f_1 and f_2), and LES/GES strategies used to create DFAs in this work. **(c,top):** Correlation energy density per particle of the Mg atom for GES-learned and LES-learned ML2, and κMP2 as a proxy reference. **(c,bottom):** Corresponding dissociation curve of the BH diatomic system. **(d,top):** Correlation energy density per particle of the Helium dimer at interatomic distance of 5.6 Bohr plotted along the principal axis of the system for different κ-regularization values. **(d,bottom):** Spin-resolved interaction correlation energy density of the Helium dimer (dimer energy density minus that of atoms). **(e,top):** Errors in interaction energies (kcal mol^{-1}) along the dissociation of the formic acid dimer (geometries and reference values are taken from the S22x5 database[64]) for HF, MP2, κMP2 and MLS2. **(e,bottom):** Relative absolute correlation energy errors in log-scale of available systems from the W4-11 test dataset[65] (see Sec. S12 in the SI for a detailed list of test data points) for MP2, κMP2 and MLS2.

correlation energy (a crucial target for DFAs). This approach expands each system's single energy data point into thousands of energy data points for the training. During the learning of our DFAs, each point in space contributes to the loss function, which we call Local Energy Loss (LES). Crucially, LES penalizes error cancellations between energy contributions from different regions in space, thereby enhancing transferability. With LES, every system in the training set becomes an entire energy dataset, and we show here that a careful, physically-informed application of LES is essential to fully realize its data-efficiency potential and embed transferability in ML of DFAs.

Applying LES for ML of correlation DFAs requires two crucial ingredients: (i) a well-defined correlation energy contributions at each point in space $e_c(\mathbf{r})$ (i.e., correlation energy density per particle, see Eq. 1 below) and (ii) a robust strategy for generating accurate $e_c(\mathbf{r})$ training

https://doi.org/10.26434/chemrxiv-2024-zk6hp-v3 **ORCID:** https://orcid.org/0000-0002-0768-9176

data. To meet these requirements simultaneously, we develop here LES-based ML DFAs for correlation energies defined with respect to the Hartree–Fock (HF) reference. While historically DFA development has been tied to Kohn-Sham density functional theory (KS DFT)[66], recent theoretical advances based on the Møller–Plesset adiabatic connection (MPAC) formally ground the developments of correlation DFAs evaluated on HF densities (see Ref. 67). Constructing DFAs on fixed (HF) densities within the MPAC framework enables us to isolate and focus specifically on real-space energy learning strategies, complementing (see Discussion) existing real-space density learning approaches[31,34,55,59,60]. As this work employs local energy quantities, we note that these quantities provide valuable chemical insights when well-defined (see, e.g., Refs. 68–70). Since $e_c(\mathbf{r})$ is not uniquely defined, here we adopt a physically transparent definition arising from the MPAC framework[67] and demonstrate its advantages for LES.

An overview of the key methods presented in this paper is given in Fig. 1. To demonstrate the power of the LES-based approach, and more generally real-space ML for DFAs, we first construct a robust proxy reference for $e_c(\mathbf{r})$ that preserves the original MPAC-based definition and efficiently implement it to enable direct training of our ML models for $e_c(\mathbf{r})$. This proxy reference $e_c(\mathbf{r})$ is built by combining second-order perturbation theory[67] (PT2) [magenta circle in Fig. 1**(a)**] and the specific PT2 regularization[71,72] [blue square in Fig. 1**(a)**], which is crucial for making our proxy reference sufficiently accurate. The effect of regularization on the PT2's $e_c(\mathbf{r})$ for the helium dimer (going from the magenta circle to the blue square in Fig. 1**(a)**) is displayed in Fig. 1**(d,top)**. We implement a numerical data generator of this proxy reference $e_c(\mathbf{r})$ by leveraging modern Python libraries (e.g., JAX[73]). New input features tailored to the problem, such as Grimme's real-space electronic correlation measures[62,63], enable us to construct a robust neural network (NN) for $e_c(\mathbf{r})$ [Fig. 1**(b)**]. We then contrast the transferability of our LES strategy with the common global energy loss (GES) that adopts standard 1 system = 1 energy data point approach [cyan and maroon diamonds in Fig. 1**(a)**]. Keeping other factors in the DFA training the same allows us to isolate how the transferability is affected when we move from GES to LES [Fig. 1**(c)** shows how well the two models trained on small atoms transfer to the dissociation curve of BH within spin-restricted calculations]. Similar transferability tests reveal subtle yet crucial requirements for successful and robust LES applications: it should be defined in terms of $e_c(\mathbf{r})$ rather than alternative quantities (e.g., its density-weighted counterpart), and coupled with a physically-informed ML model trained on a physically-informed $e_c(\mathbf{r})$ definition. We also derive the contributions at each point in space for different spin channel pairs of our proxy $e_c(\mathbf{r})$ [purple and orange circles or squares in Fig. 1**(a)**] and we show these spin-resolved interaction components for the helium dimer in Fig. 1**(d,bottom)**. We then use spin-resolved energy densities per particle to build a real-space, machine-learned and regularized extension of spin-component-scaled[74,75] (SCS) PT2 correlation energy (the performance of this ML strategy for the formic acid dimer is shown in Fig. 1**(e,top)**). The resulting model [green diamond in Fig. 1**(a)**] opens up avenues for DFAs construction and enables us to bridge the gap between our proxy reference correlation energies (regularized PT2) and their exact counterpart.

RESULTS

Local and global energy loss (LES vs GES)

Distinguishing between LES and GES is a crucial point of this work when training ML DFAs. To define LES and GES generally, consider the reference (i.e. exact) energy

$$E^{\mathrm{ref}} = \int e^{\mathrm{ref}}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} \tag{1}$$

with a corresponding reference energy density[58] per particle, $e^{\mathrm{ref}}(\mathbf{r})$, and electronic density $\rho(\mathbf{r})$. An ML energy quantity defined in the same way is indicated using the ML superscript. Then, GES reads

$$\mathcal{L}_{\mathrm{GES}} \sim \left|E^{\mathrm{ref}} - E^{\mathrm{ML}}\right|. \tag{2}$$

In contrast, with LES, we consider the pointwise difference of the reference and ML energy densities per particle weighted by the density:

$$\mathcal{L}_{\mathrm{LES}} \sim \int \left|e^{\mathrm{ref}}(\mathbf{r}) - e^{\mathrm{ML}}(\mathbf{r})\right| \rho(\mathbf{r})\, d\mathbf{r}. \tag{3}$$

Minimizing LES, strictly defined in terms of the energy density per particle, instead of GES turns each point in space into an energy data point, and as we shall see, moving from GES to LES dramatically enhances the transferability of the underlying ML DFA even with very small training sets.

Improving and deriving PT2 correlation energy densities per particle

To demonstrate the difference between LES and GES, we will target correlation energy approximations. While correlation DFAs are typically developed within the KS DFT framework, recent work has shown that the Møller–Plesset adiabatic connection (MPAC) formally grounds the construction of DFAs mapping Hartree–Fock (HF) densities directly to correlation energies (defined here w.r.t. the HF energies). As detailed in Ref. 67, this ground distinguishes the MPAC-based correlation DFAs from density-corrected DFT, where HF densities are introduced heuristically to improve DFAs developed within KS DFT[76–78]. Leveraging this MPAC formalism

and its recently introduced correlation energy densities[67], we construct DFAs as NN-based functionals of HF densities, enabling practical use of HF orbitals to compute all energy terms and input features for our NNs.

To briefly introduce our energy density per particle targets for LES, we define correlation energy as,

$$E_c = E^{\text{ref}} - \langle\Phi|\hat{H}|\Phi\rangle = \int e_c(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} = \int \bar{e}_c(\mathbf{r})d\mathbf{r}, \quad (4)$$

where E^{ref} is the exact ground-state energy, $\hat{H}$ is the corresponding exact Hamiltonian, and Φ is the HF wavefunction (a single Slater determinant minimizer of $\hat{H}$, that yields $\rho^{\text{HF}}(\mathbf{r}) = \rho(\mathbf{r})$). In Eq. 4, we distinguish the correlation energy density, $\bar{e}_c(\mathbf{r}) = e_c(\mathbf{r})\rho(\mathbf{r})$, from the correlation energy per particle, $e_c(\mathbf{r})$, as this distinction is crucial for our subsequent LES analysis. As $e_c(\mathbf{r})$ is not uniquely defined, we adopt here a specific definition (i.e., *gauge*) for $e_c(\mathbf{r})$, derived in Ref. 67 from the MPAC theory. This gauge is designed as an MPAC-based analogue[67] of the conventional DFT gauge for correlation energies, i.e., the *electrostatic potential of the correlation hole*[79–82], known for transparent physical interpretation and advantages in DFA construction[82,83]. Since computing the exact $e_c(\mathbf{r})$ is costly[67], we approximate it using its weakly interacting limit determined by MP2, preserving the original MPAC gauge,

$$e_c^{\text{MP2}}(\mathbf{r}) = \frac{1}{4\rho(\mathbf{r})}\int \frac{P_2^{\text{MP2}}(\mathbf{r},\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}', \quad (5)$$

where $P_2^{\text{MP2}}(\mathbf{r},\mathbf{r}')$ is the first-order MPAC correction to the pair density that yields the MP2 correlation energy (for its formal definition and derivation of Eq. 5, see Sec. S1 in the SI). Crucially, $P_2^{\text{MP2}}(\mathbf{r},\mathbf{r}')$ isolates the correlation contribution to the pair density and is analogous to the DFT correlation hole[67].

Expressing Eq. 5 in terms of HF orbitals yields,

$$e_c^{\text{MP2}}(\mathbf{r}) = \frac{1}{\rho(\mathbf{r})}\left[\sum_{ijab} V_{ijab}(\mathbf{r})\left(\frac{1}{2}T_{ijba} - T_{ijab}\right) + \sum_{ijab} V_{ijba}(\mathbf{r})\left(\frac{1}{2}T_{ijab} - T_{ijba}\right)\right], \quad (6)$$

where i,j are occupied, and a,b are virtual KS orbital ($\phi(\mathbf{r})$) indices. T_{ijab} are the *partial* MP2 doubles amplitude,

$$T_{ijab} = \frac{\langle ij|ab\rangle}{\varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j}, \quad (7)$$

where ε are orbital energies, and $V_{ijab}(\mathbf{r})$ is the orbital potential,

$$V_{ijab}(\mathbf{r}) = \phi_i(\mathbf{r})\phi_a(\mathbf{r})\int \frac{\phi_j(\mathbf{r}')\phi_b(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}'. \quad (8)$$

As $e_c^{\text{MP2}}(\mathbf{r})$ by Eq. 4 integrates to the MP2 correlation energy, it can be easily argued that it is not a sufficiently good proxy reference for LES-based applications given general MP2 limitations[71,72,84], particularly for small orbital-gap systems[85] (see the MP2 dissociation curve relevant to this work in Fig. **S1** in the SI). To address this, we apply Head-Gordon's κ-regulizaration[71,72] ($\kappa \geq 0$) to $e_c^{\text{MP2}}(\mathbf{r})$ of Eq. 5 by regularizing its partial MP2 doubles amplitudes,

$$T_{ijab}^{\kappa} = T_{ijab}\left(1 - e^{-\kappa(\varepsilon_a+\varepsilon_b-\varepsilon_i-\varepsilon_j)}\right)^2. \quad (9)$$

The regularized $e_c^{\kappa\text{MP2}}(\mathbf{r})$ is obtained from Eq. 6 by the replacement $T_{ijab} \to T_{ijab}^{\kappa}$, which also regularizes the underlying pair density (see Sec. S1 in the SI):

$$e_c^{\kappa\text{MP2}}(\mathbf{r}) = \frac{1}{4\rho(\mathbf{r})}\int \frac{P_2^{\kappa\text{MP2}}(\mathbf{r},\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}'. \quad (10)$$

When $\kappa = 0$, $e_c^{\kappa\text{MP2}}(\mathbf{r}) = 0$, and when $\kappa \to \infty$, $e_c^{\kappa\text{MP2}}(\mathbf{r}) = e_c^{\text{MP2}}(\mathbf{r})$. In Fig. **1(d,top)**, we observe how $e_c^{\kappa\text{MP2}}(\mathbf{r})$ evolves to $e_c^{\text{MP2}}(\mathbf{r})$ for the helium dimer as κ increases. Using the regularized $e_c^{\kappa\text{MP2}}(\mathbf{r})$ [blue square in Fig. **1(a)**] instead of $e_c^{\text{MP2}}(\mathbf{r})$ [magenta circle in Fig. **1(a)**] as the proxy reference for LES-based learning is crucial, as the former significantly improves dissociation curves of diatomic systems, which are central to our LES vs. GES comparison. In these curves, κ-regularization removes the MP2 divergence at large bond lengths arising from small orbital energy gaps. Here we set $\kappa = 2.0$ to generate $e_c^{\kappa\text{MP2}}(\mathbf{r})$ proxy reference, as it better balances improvements over MP2 for stretched bonds while maintaining accuracy near equilibrium, compared to the originally proposed $\kappa = 1.4$[71] (see Fig. **S1** in the SI for a clarifying example of N_2 dissociation).

After regularization, we perform the spin-resolution of $e_c^{\kappa\text{MP2}}(\mathbf{r})$ into same-spin (ss) and opposite-spin (os) components: $e_c^{\kappa\text{MP2}}(\mathbf{r}) = e_{c,\text{ss}}^{\kappa\text{MP2}}(\mathbf{r}) + e_{c,\text{os}}^{\kappa\text{MP2}}(\mathbf{r})$, enabling a spin-resolved real-space analysis of electron correlation [purple and orange in Fig. **1(a)**]. The more compact of the two, $e_{c,\text{os}}^{\kappa\text{MP2}}(\mathbf{r})$, we derive from Eq. 6 by considering only os electronic pairs:

$$e_{c,\text{os}}^{\kappa\text{MP2}}(\mathbf{r}) = -\frac{1}{2\rho(\mathbf{r})}\sum_{ijab}\left[T_{ijab}^{\kappa}V_{ijab}(\mathbf{r}) + T_{ijab}^{\kappa}V_{ijba}(\mathbf{r})\right].$$

Later in Sec., we will use ML for real-space scaling of $e_{c,\text{os}}^{\kappa\text{MP2}}(\mathbf{r})$ and $e_{c,\text{ss}}^{\kappa\text{MP2}}(\mathbf{r})$ separately to build a correction from κMP2 to the true E_c.

Central to our data generator for real-space ML of $e_c(\mathbf{r})$ are the κ-regularization and spin resolution of MP2 correlation energy densities, both efficiently implemented by combining density fitting[86,87] with the power of JAX[73] and modern tensor libraries[88] for optimizing tensor contractions required to obtain $e_c^{\kappa\text{MP2}}(\mathbf{r})$. Further implementation details are provided in Methods and Sec. S3 of the SI.

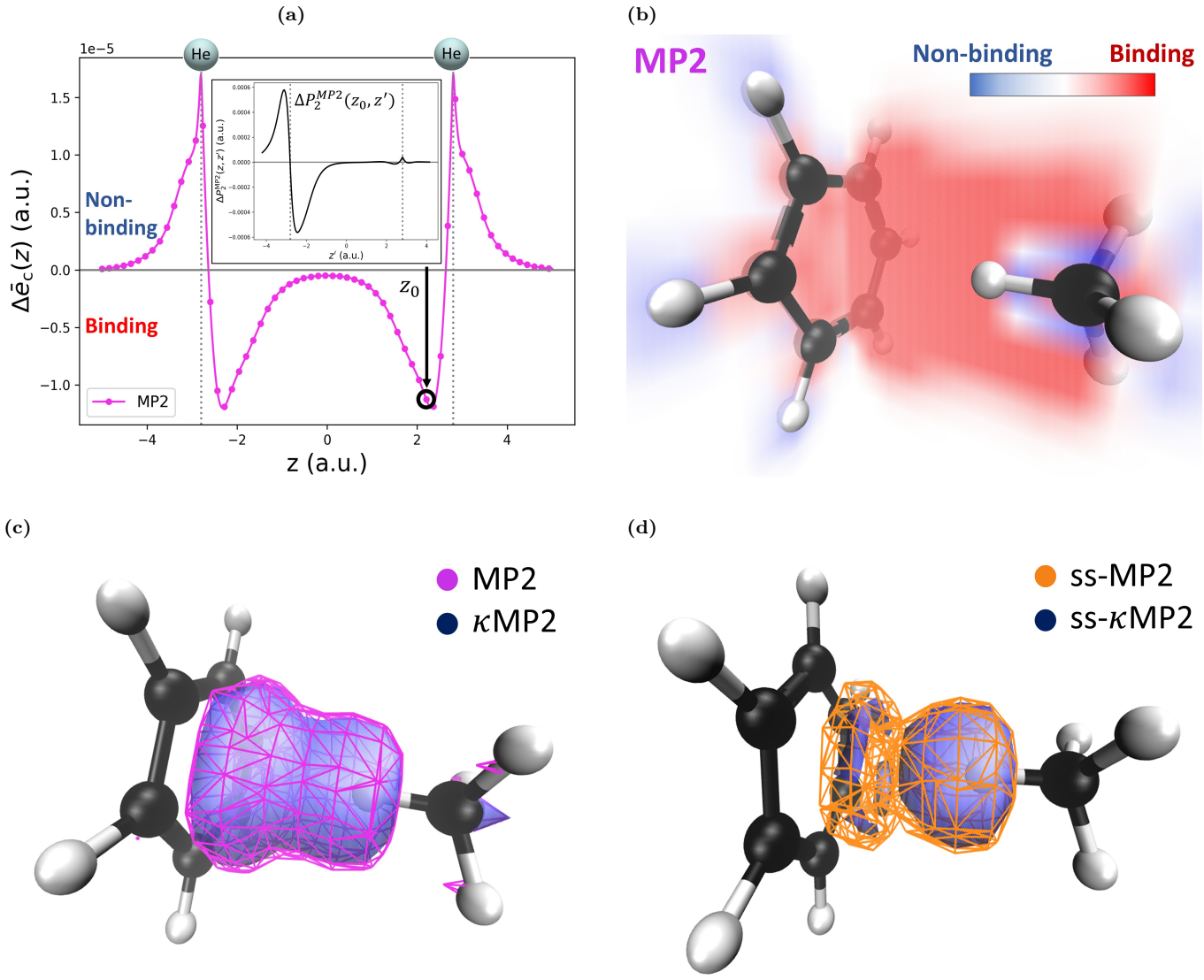

Fig. 2: Visualization of $\Delta\bar{e}_c(\mathbf{r})$ (total correlation energy density minus the one from the individual subsystems). (a): MP2 plot along the inter-nuclear axis of the Helium dimer (position of the nuclei are denoted by the spheres) at an interatomic distance of 5.6 Bohr. Inset shows the corresponding $\Delta P_2^{\mathrm{MP2}}(\mathbf{r},\mathbf{r}')$ at $\mathbf{r}$ =z_0 on the same axis. **(b)**: MP2 volume slice plots for the benzene–CH_4 complex[85] along planes perpendicular and parallel to the benzene ring, highlighting binding and non-binding regions. **(c)**: Isosurface visualization of MP2 and κMP2 ($\kappa = 1.4$) for the same complex at binding isovalue (−2.5e-5 (a.u.)). **(d)**: same as (c), but for the ss part.

Regularized PT2 correlation energy densities for interaction energies

Now we move to a real-space analysis of the interaction correlation energy density, $\Delta\bar{e}_c(\mathbf{r})$, defined as the total system's $\bar{e}_c(\mathbf{r})$ minus the sum of the $\bar{e}_c(\mathbf{r})$ of the isolated subsystems (e.g., dimer minus monomers). A subtle point here is that while for ML purposes $e_c(\mathbf{r})$ is vastly superior to $\bar{e}_c(\mathbf{r})$ (see below), spatial visualization of interactions requires $\Delta\bar{e}_c(\mathbf{r})$ instead of $\Delta e_c(\mathbf{r})$, as the former directly integrates to ΔE_c, while the latter lacks a clear density factor to do so.

From Eq. 10, $\Delta\bar{e}_c^{\mathrm{MP2}}(\mathbf{r})$ is given by the electrostatic potential of $\Delta P_2^{\kappa\mathrm{MP2}}(\mathbf{r},\mathbf{r}')$,

$$\Delta\bar{e}_c^{\kappa\mathrm{MP2}}(\mathbf{r}) = \frac{1}{4}\int \frac{\Delta P_2^{\kappa\mathrm{MP2}}(\mathbf{r},\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}', \qquad (11)$$

where $\Delta P_2^{\kappa\mathrm{MP2}}(\mathbf{r},\mathbf{r}')$ is the interaction component of $P_2^{\kappa\mathrm{MP2}}(\mathbf{r},\mathbf{r}')$. As $P_2^{\kappa\mathrm{MP2}}(\mathbf{r},\mathbf{r}')$ isolates the correlation part of the underlying pair density, $\Delta P_2^{\kappa\mathrm{MP2}}(\mathbf{r},\mathbf{r}')$ further isolates how this quantity is deformed by the interaction between fragments, making this quantity crucial for describing the physics of weak interactions, particularly dispersion effects (see, e.g., Refs. 89–91).

While Eq. 11 represents just one possible gauge for $\Delta\bar{e}_c^{\kappa\mathrm{MP2}}(\mathbf{r})$, we emphasize that it encodes the

physics of weak interactions through its direct link to $\Delta P_2^{\kappa\mathrm{MP2}}(\mathbf{r},\mathbf{r}')$, a fundamental quantity for describing the physics of weak interactions, such as dispersion effects[89–91]. To illustrate this for dispersion, we first consider a simple example: the helium dimer, for which spin-resolved interaction correlation energy densities $\Delta\bar{e}_\mathrm{c}^{\mathrm{MP2}}(\mathbf{r})$ are shown in Fig. **1(d,bottom)** along the internuclear axis. We now show their sum in Fig. **2(a)**, highlighting distinct binding (negative) and non-binding contributions for stretched He_2. To illustrate the interaction physics encoded in $\Delta\bar{e}_\mathrm{c}^{\mathrm{MP2}}(\mathbf{r})$, we fix the reference electron at one of the nuclei (z_0) in He_2, and in the inset of Fig. **2(a)** we show $\Delta P_2^{\mathrm{MP2}}(z_0,z')$ along the internuclear axis as a function of the second electron position z'. The plot reveals spatially nonlocal polarization in $\Delta P_2^{\mathrm{MP2}}$, characteristic of dispersion: when z' is near the second helium nucleus, $\Delta P_2^{\mathrm{MP2}}(z_0,z')$ exhibits a negative accumulation closer to z_0 and a corresponding positive region further away. As the negative part of $\Delta P_2^{\mathrm{MP2}}(z_0,z')$ lies closer to z_0 than the positive part, the resulting electrostatic potential, i.e., $\Delta\bar{e}_\mathrm{c}(z_0)$, is negative. If the polarization pattern is reversed, i.e., the positive part of $\Delta P_2^{\mathrm{MP2}}(z_0,z')$ lies closer to z_0 than the negative part, as when z_0 is in the outer region, then $\Delta\bar{e}_\mathrm{c}(z_0)$ becomes positive (see Fig. S3 for additional plots in the SI). Still, the negative (i.e., binding) regions dominate, and MP2 correlation yields net binding in He_2 (Table **S1** in the SI). In this way, Eq. 11 condenses information from $\Delta P_2(\mathbf{r},\mathbf{r}')$, a key two-body quantity encoding interaction physics, into $\Delta\bar{e}_\mathrm{c}(\mathbf{r})$, a one-body correlation quantity for interaction energies. Thus, even though it is not unique, the specific gauge defined by Eq. 11 yields a physically interpretable local correlation energy quantity for describing weak interactions between fragments.

Having established the direct connection between $\Delta\bar{e}_\mathrm{c}^{\mathrm{MP2}}(\mathbf{r})$ and dispersion physics, we now analyze $\Delta\bar{e}_\mathrm{c}^{\mathrm{MP2}}(\mathbf{r})$ for the benzene–methane complex in the remaining panels of Fig. **2**. In Fig. **2b**, volume slices along planes perpendicular and parallel to the benzene ring distinguish regions between the fragments (typically binding regions) from those outside (typically non-binding). MP2 overbinds this complex, whereas κMP2 reduces this overbinding (Table S2 in the SI). This reduction is visually reflected in Fig. 2c, which compares MP2 and κMP2 $\Delta\bar{e}_\mathrm{c}(\mathbf{r})$ isosurfaces for the binding region, with the κMP2 isosurface confined within the MP2 counterpart. The spatial confinement is even more pronounced when focusing on just the ss component of $\Delta\bar{e}_\mathrm{c}^{\kappa\mathrm{MP2}}(\mathbf{r})$ [the electrostatic potential of the ss component of the underlying $\Delta P_2^{\kappa\mathrm{MP2}}(\mathbf{r},\mathbf{r}')$], as shown in Fig. 2d.

ML2 model via machine-learning of regularized $e_\mathrm{c}^{\kappa\mathrm{MP2}}(\mathbf{r})$

We will now present the ML2 model, a machine-learned correlation energy density based on the regularized MP2 proxy reference, which we obtain by mapping a set of pointwise features using neural networks (NNs) [going from the blue square to the cyan or maroon diamonds in Fig. **1(a)**]. This mapping is illustrated in Fig. **1(b)**. In addition to the established features used in ML of DFAs[31,57]—the reduced density gradient $s(\mathbf{r})$, the reduced density Laplacian $q(\mathbf{r})$, and the regularized kinetic energy variable $\alpha(\mathbf{r})$ from the r^2SCAN DFA[92]—we introduce Grimme's electronic temperature-dependent fractional occupation number weighted density[62,63] (FOD) as the crucial feature (a detailed list of features is given in Sec. S6 in the SI). FOD differentiates strongly correlated and weakly correlated regions in molecules. While both $e_\mathrm{c}^{\kappa\mathrm{MP2}}(\mathbf{r})$ and FOD provide insights into electronic correlation through the interaction between occupied and unoccupied orbitals, FOD is computationally much cheaper, making it an excellent feature for ML of the correlation energy density.

Even though $\rho(\mathbf{r})$-based features (e.g., the Wigner–Seitz radius) might seem like a natural choice, we intentionally omit them from the ML2 feautes. This is because we found that $e_\mathrm{c}^{\mathrm{MP2}}(\mathbf{r})$ and its κ-counterpart are scaling invariant for a uniformly scaled[93] density $\rho_\lambda(\mathbf{r})=\lambda^3\rho(\lambda\mathbf{r})$,

$$e_\mathrm{c}^{\kappa\mathrm{MP2}}[\rho_\lambda](\mathbf{r}) = e_\mathrm{c}^{\kappa\mathrm{MP2}}[\rho](\mathbf{r}). \tag{12}$$

A detailed derivation is given in Sec. S7 in the SI. Following this scaling invariance, we construct $e_\mathrm{c}^{\mathrm{ML2}}(\mathbf{r})$, the ML2 analog of $e_\mathrm{c}^{\kappa\mathrm{MP2}}(\mathbf{r})$, as

$$e_\mathrm{c}^{\mathrm{ML2}}(\mathbf{r}) = \mathrm{w_c}(\mathbf{r})\ e_\mathrm{x}(\mathbf{r})\ \rho^{-1/3}(\mathbf{r}), \tag{13}$$

where $e_\mathrm{x}(\mathbf{r})$ is the exchange energy (we use the same Python code to implement both $e_\mathrm{c}^{\mathrm{MP2}}(\mathbf{r})$ and $e_\mathrm{x}(\mathbf{r})$ on the same footing), and $\mathrm{w_c}(\mathbf{r})$ are the ML2 weights obtained from the NN (see Fig. **1(b)** for the illustration and Methods for NN architecture details). With the use of HF-based ingredients, the computational cost of ML2 is comparable to that of an HF calculation and cannot be lower. As we shall see later, embedding the physics into LES-based ML2 through Eq. 13 is crucial for the robustness of the model.

Using the same features, functional form of Eq. 13, and NN architecture for mapping the features at a given $\mathbf{r}$ to $\mathrm{w_c}(\mathbf{r})$, we can now isolate the difference between using LES and GES for NN training of a DFA [cyan vs. maroon diamonds in Fig. **1(a)**]. For simplicity and to create a challenging transferability test, we train our ML2-based NN only on eight small closed-shell atoms/ions (H^-, He, Be, Ne, Mg, Ar, Ca, and Kr). The total loss is calculated as the mean over these eight datapoints, as detailed in Sec. S8 in the SI. We validate our training on comparable small closed-shell atoms/ions (see Fig. **S4**) in the SI.

In Fig. **1(c)**, we explore GES-based vs. LES-based ML2 results (cyan vs. maroon diamonds). Both energy densities are plotted against the $e_\mathrm{c}^{\kappa\mathrm{MP2}}(\mathbf{r})$ reference in Fig. **1(c,top)** for the Mg atom as one of the training datapoints. We can see that $e_\mathrm{c}^{\mathrm{ML2}}(\mathbf{r})$ based on LES closely follows the $e_\mathrm{c}^{\kappa\mathrm{MP2}}(\mathbf{r})$ (proxy) reference, whereas the GES-based $e_\mathrm{c}^{\mathrm{ML2}}(\mathbf{r})$ completely misses the shape of

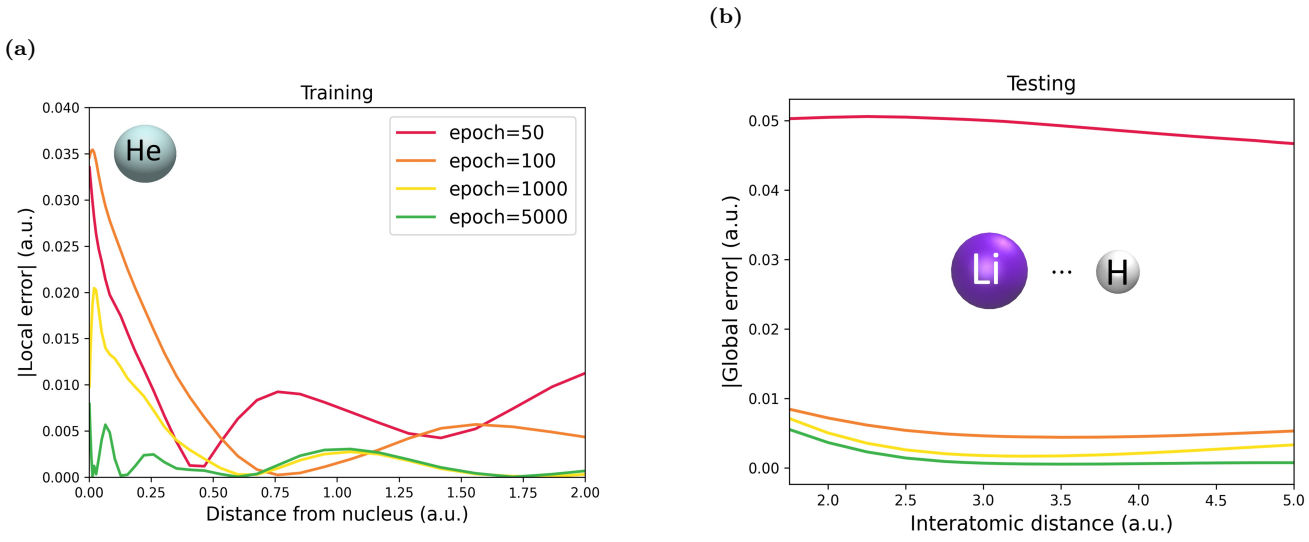

Fig. 3: Errors of LES-based ML2 with respect to κMP2 at different learning epochs. (a): Absolute (local) error in correlation energy densities for the Helium atom. **(b):** Absolute global energy error of ML2 (LES) for LiH along its dissociation curve.

the reference. But, can we, solely based on this observation, conclude that LES is better than GES? We need to be careful here, as these correlation energy densities are not observables even within our physically sound gauge (see previous section). Instead, they are used here to enhance energy training data efficiency in ML of DFAs, expanding from one to thousands of energy datapoints per system. Thus, what ultimately matters for judging the quality of GES vs. LES training are the integrated correlation energies (Eq. 1) once we go outside of the training set.

For the Mg case in Fig. 1**(c,top)**, all energy densities integrate to nearly the same correlation energy by Eq. 1. However, the correlation energies from our GES-based and LES-based ML2 models differ dramatically when tested for transferability. The first such test is shown in Fig. 1**(c,bottom)**, where we see major differences in accuracy when applied to stretching the BH diatomic system along its dissociation curve. The LES-based model closely follows the $e_c^{\kappa MP2}(\mathbf{r})$ (proxy) reference result, while the GES-based model is not sufficiently accurate even at equilibrium and breaks down completely as the bond is stretched.

The advantages of LES-based learning process

After seeing in Fig. 1**(c,bottom)** that LES-based ML2 trained only on atomic systems successfully transfers to the BH diatomic along its dissociation curve, we now investigate this atom to diatomics transferability in more detail. Namely, a closer look at the learning process of the LES-based ML2 is given in Fig. **3a**, where panel (a) shows the absolute error in $e_c^{ML2}(\mathbf{r})$ for the He atom (one of ML2's training points) at different epochs. On the other hand, panel (b) focuses on the test and shows the absolute error in the ML2 correlation energy along the LiH dissociation curve for the same set of epochs. Overall, Fig. **3a** shows that the pointwise improvement in $e_c^{ML2}(\mathbf{r})$ for the He atom during training translates epoch by epoch into improved correlation energies for the unseen LiH system, as indicated by the decreasing errors with increasing epochs in panel(b). In contrast to that, the GES-based model yields no improvements of the transferability from atoms as the learning process progresses (see Fig. **S5** in the SI exposing poor transferability to diatomics, especially at large distances even at large epochs). Furthermore, we observe another crucial difference between GES-based and LES-based learning: when LES is used as the loss function, the learning process is much smoother and it has a much faster convergence with respect to learning steps (*epochs*) compared to GES (see Fig. **S6** in the SI).

Finishing off, by making energies more information-rich through LES in the training, we equip our models with a high level of transferability, ensuring that any information learned from atoms lead to better performance on molecules. LES reveals information completely washed away with GES, showing that combinations of features and corresponding energy densities per particle, even for atoms, are highly relevant for molecules. Thus, LES provides a powerful strategy for ML of transferable DFAs, and it what follows we will explore more subtle details linked to the LES-based training.

Uniqueness and Robustness of our LES-based ML2 model

Building on the successful LES-based ML2 transferability from atoms to challenging BH and LiH dissoci-

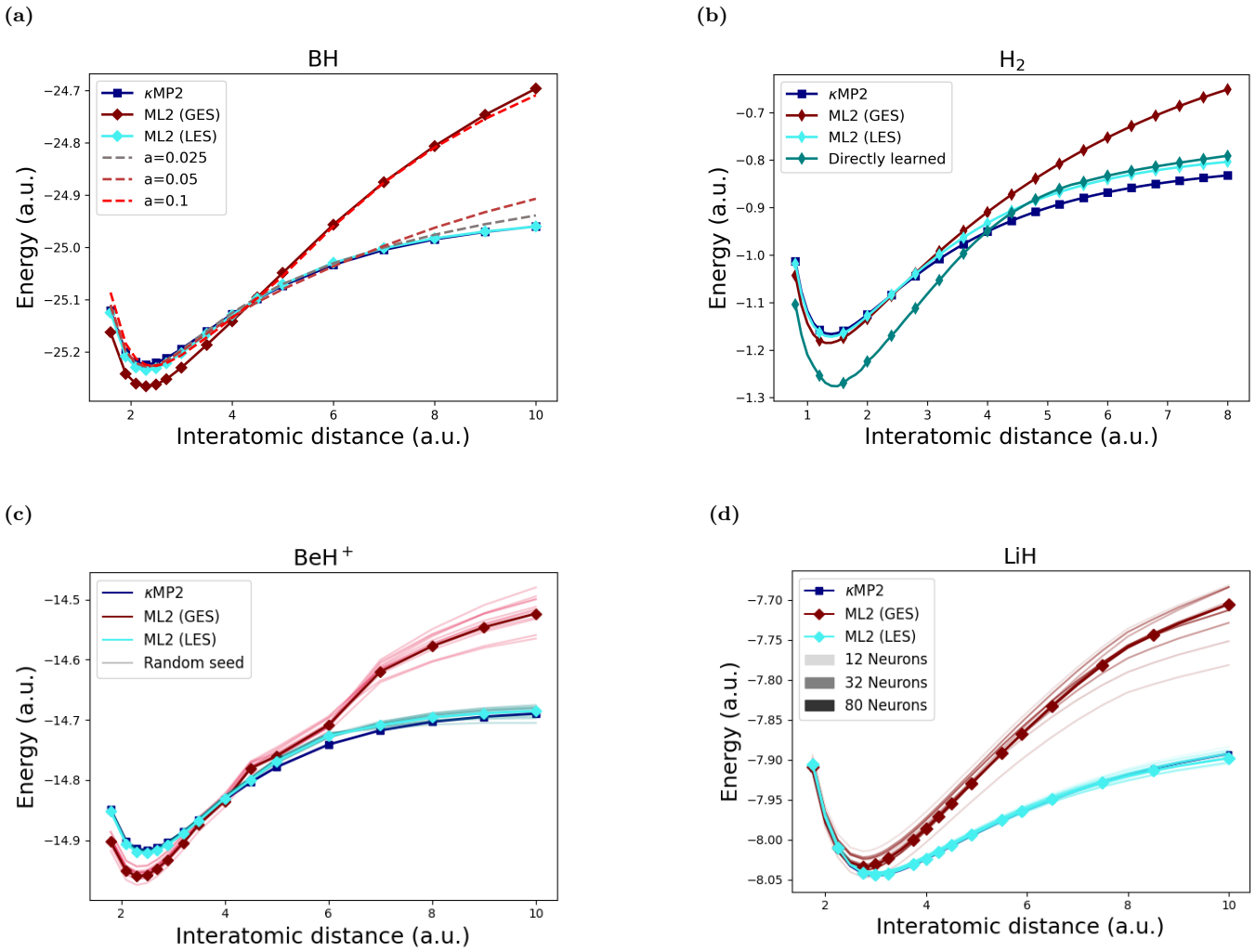

Fig. 4: Dissociation curves as in Fig. 1(c,bottom), but for four different systems with additional data for comparison. **(a):** BH curves including ML2 results that employ LES with proxy correlation energy densities coming from different $a-$parameter dependent gauges (see Eq. 14). **(b):** H_2 curves including 'directly learned' model, that learns $\bar{e}_c^{\kappa MP2}(\mathbf{r})$ of κMP2, based on 'direct loss' (see text). **(c):** BeH^+ curves including GES-based and LES-based ML2 results coming from different random seed initializations. **(d):** LiH curves including additional ML2 results coming from NNs with different total number of neurons (see Fig. S11 in the SI for the distribution of neurons per hidden layer).

ations, Fig. 4 shows dissociation curves for four additional diatomics, further confirming the LES's advantage over GES, which can yield unphysical curves. The dissociation curves for closed-shell diatomics are obtained using a spin-restricted formalism to avoid artificial energy lowering from spin-symmetry breaking (see Refs. 94–96 for discussions on the challenge of describing bond breaking without spin-symmetry breaking). We now present four analyses, each applied to a different curve from Fig. 4, to highlight distinct aspects of LES-based ML2 transferability. Importantly, the conclusions from each analysis are robust, consistently holding when cross-checked against other diatomics, as demonstrated in the SI (Figs. S7–S11).

We first test whether the superior transferability of LES over GES from atoms to diatomics arises purely from LES's higher data efficiency. Specifically, panels (a) and (b) of Fig. 4 show that this transferability is lost if the LES-based ML2 model is constructed without the physics encoded in Eq. 3 (LES defined via energy densities per particle), Eq. 10 (specific $e_c(\mathbf{r})$ gauge), and Eq. 13 (physically constrained ML2 functional form).

To examine the sensitivity of ML2 training to the gauge choice of $e_c^{\kappa MP2}(\mathbf{r})$ (Eq. 5), we introduce the following gauge transformation:

$$e_c^a(\mathbf{r}) = e_c^{\kappa MP2}(\mathbf{r}) + a\ q(\mathbf{r})\rho^{2/3}(\mathbf{r}), \tag{14}$$

with $q(\mathbf{r}) = [\nabla^2\rho(\mathbf{r})]/[4(3\pi^2)^{2/3}\rho(\mathbf{r})^{5/3}]$, and where the real parameter a does not affect the integrated correlation energy (Eq. 4) for exponentially decaying densities.

Instead of our original target, $e_c^{\kappa MP2}(\mathbf{r})$, we now repeat the LES-based ML2 training using $e_c^a(\mathbf{r})$ as the target at various a values. In Fig. 4a, we test the transferability of the underlying a-dependent model by using the BH dissociation curve previously shown in Fig. 1(c,bottom). These curves become worse as we move away from $a = 0$ (original LES-based ML2) and even become unphysical at larger a (the results for other diatomic dissociation curves follow similar trends as shown in Fig. S7 in the SI). Observing $e_c^a(\mathbf{r})$ for atoms (train) and BH (test) in Fig. S8 in the SI, we see that at $a = 0$, the range and shape of $e_c^a(\mathbf{r})$ varies much less between train and test systems compared to larger a values, explaining why the original LES-based ML2 ($a = 0$) exhibits the best transferability. While Eq. 14 does not cover all possible gauges, our tests show that the transferability from atoms to diatomics achieved by our gauge is highly nontrivial, and that superior performance of LES over GES in our ML2 model is not solely due to higher data efficiency but critically depends on the gauge choice (Eq. 10) for the $e_c^{ML}(\mathbf{r})$ training target (observe how κMP2 is more amenable for ML in Fig. S8 in the SI).

To further demonstrate that data efficiency alone is insufficient for ML2's success, we compare dissociation curves for H_2 from various models in Fig. 4b. The figure highlights the importance of defining LES in terms of energy densities per particle (Eq. 3) and employing the physically constrained LES-based ML2 form (Eq. 13). If, instead of LES, the loss is defined directly via energy densities (not per particle)[97,99], $\mathcal{L}_{\text{direct}} \sim \int \left| \bar{e}_c^{\text{ref}}(\mathbf{r}) - \bar{e}_c^{\text{ML}}(\mathbf{r}) \right| d\mathbf{r}$, then the resulting model ("directly learned") performs as bad as the GES-based model (see Fig. 4b). Additional examples of even more drastic failures of $\mathcal{L}_{\text{direct}}$-based learning, illustrating the subtle yet crucial importance of learning $e_c(\mathbf{r})$ rather than $\bar{e}_c(\mathbf{r})$, are shown in Fig. S9 in the SI. The poor model's transferability when direct loss is used (learning $\bar{e}_c^{ML}(\mathbf{r})$) in place of LES (learning $e_c^{ML}(\mathbf{r})$) is unsurprising, given that both $e_c(\mathbf{r})$ and the ML2 weights, $w_c(\mathbf{r}) = e_c(\mathbf{r})/(e_x(\mathbf{r})\rho^{-1/3}(\mathbf{r}))$ (Eq. 13), are far less sensitive to variations in system size than $\bar{e}_c(\mathbf{r}) = e_c(\mathbf{r})\rho(\mathbf{r})$. This comparison further emphasizes that LES-based ML2's transferability success does not rely solely on higher data efficiency than GES, but critically depends on the synergy between this efficiency and the physics embedded in Eqs. 3, 10, and 13.

Finally, we demonstrate the robustness of LES-based ML2 with respect to NN training conditions: unlike GES, LES-based ML2 remains stable under variations in random initialization seeds (Fig. 4c for BeH^+; additional examples in Fig. S10 in the SI) and NN architecture, including the number of neurons (Fig. 4d for LiH; additional examples in Fig. S11 in the SI). We also show in Fig. S12 in the SI that the use of mean square-based LES (i.e. LES^2 instead of the original absolute differences-based LES of Eq. 3 has little effect on the ML2 results. Additionally, we show that when progressively increasing the training dataset size, LES remains more robust than GES for diatomic dissociation curves (Fig. S13 in the SI). Overall, the robustness and uniqueness of the LES-based ML2 model demonstrated in this section further emphasize the advantages of LES-based ML2.

SPIN-RESOLVED AND REGULARIZED MODELLING OF THE CORRELATION ENERGY DENSITY

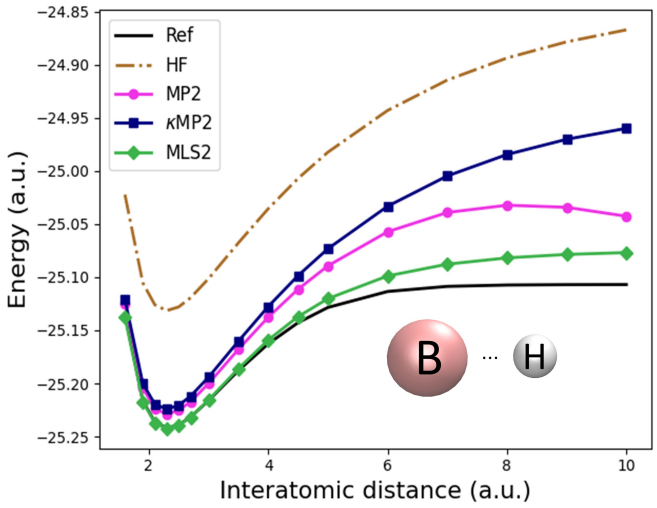

Fig. 5: Test result of MLS2. Dissociation energy curve of the BH diatomic system, as in Fig. 1(c,bottom), but with additional models (reference values (Ref) taken from Ref. 100).

In the previous section we have shown that LES enhances the transferability of ML DFAs. Yet, κMP2 correlation has been the proxy reference in place of its exact counterpart. In this section, we use our regularized PT2-based generator for a NN-based combination of the spin-resolved κMP2 energy densities to bridge the gap between κMP2 and true correlation energies. For this purpose, our ML model for correlation energy densities per particle is defined as

$$e_c^{MLS2}(\mathbf{r}) = w_{os}(\mathbf{r}) e_{c,os}^{\kappa MP2}(\mathbf{r}) + w_{ss}(\mathbf{r}) e_{c,ss}^{\kappa MP2}(\mathbf{r}), \quad (15)$$

where $w_{os}(\mathbf{r})$ and $w_{ss}(\mathbf{r})$ are machine-learned weights at every point in space. We call it MLS2, which represents a real-space, machine-learned and regularized extension of SCS MP2[74,75]. Its construction is represented by the step from orange and purple squares to the green diamond in Fig. 1(a). By leveraging our implementation of spin-resolved $w_{os}(\mathbf{r})e_{c,os}^{\kappa MP2}(\mathbf{r})$ and $w_{ss}(\mathbf{r})e_{c,ss}^{\kappa MP2}(\mathbf{r})$, Eq. 15 yields a regularized and real-space extension of SCS MP2, thus opening up avenues for DFAs creation.

To obtain $w_{os}(\mathbf{r})$ and $w_{ss}(\mathbf{r})$ in MLS2 using a NN, we employ a similar architecture as for the ML2 model (see Fig. 1(b)) with some crucial differences. First, unlike the MP2 correlation energy, the true correlation energy is generally not scale invariant[93]. This allows us contrary to ML2 (see Eq. 12 and the preceding paragraph)

to incorporate also density-based features, specifically, the Wigner-Seitz radius. Second, a sigmoid activation function is applied to the NN's output layer (more details on the activation functions between layers are given in Methods), constraining the MLS2 weights between 0 and 1. We scale the resulting NN weights $w_{os}(\mathbf{r})$ and $w_{ss}(\mathbf{r})$ by a constant factor of 10 before applying them in Eq. 15. This scaling enables the MLS2 model to be accurate for the cases where the true correlation energy (in absolute terms) is much larger than what κMP2 predicts (see Sec. S11 in the SI for more details). Finally, we use $e_c^{\kappa MP2}(\mathbf{r})$ and $e_{c,os}^{\kappa MP2}(\mathbf{r})$ quantities normalized by $\rho^{-1/3}(\mathbf{r})e_x(\mathbf{r})$ as extra MLS2 dimensionless features (see Sec. S6 in the SI for a detailed list of features).

Ideally, with access to a robust data generator for the exact $e_c(\mathbf{r})$, we could use LES to train $e_c^{MLS2}(\mathbf{r})$. However, due to the severe computational limitations of such a generator[81,101] to very small systems and basis sets, we settle with a GES-based training of MLS2 (the resulting loss function is detailed in Sec. S8 in the SI). Using GES here requires more data for training. Thus, we employ the eight atoms/ions training dataset from ML2 combined with 13 small closed-shell molecules, mainly dimers, and correlation energies of H_2, N_2 and Li_2 at five geometries of large interatomic distances. In addition to these total energies, our MLS2 NN is also trained on interaction energies from the RG18 dataset[102], including dispersion-bound complexes with noble gases. We elaborate in Sec. S8 of the SI on how we combine the total energy-based GES and the interaction energy-based GES when training MLS2. Furthermore, a detailed MLS2's list of all training data points is given in Sec. S12 of the SI.

To test MLS2, we go back to Fig. 1**(d,bottom)**, which includes results for 96 total correlation energies from the W4-11 database[65] not present in the training set (see Sec. S12 in the SI for a full list and how we obtain the underlying reference total correlation energies). Specifically, Fig. 1**(d,bottom)** shows the relative absolute correlation energy errors for κMP2, MP2 and MLS2. Note the log-scale in the y-axis and the dashed lines representing mean absolute relative errors (MArEs). Going from MP2 to κMP2 (from magenta circles to blue squares), we can see that the introduction of the κ= 2.0 regularization slightly increases the MP2 errors. On the other hand, our MLS2 model (green diamonds) yields here far more accurate correlation energies than MP2, with MArE reduced by from ∼10% to 1%.

In Fig. 5, we revisit the dissociation curve of BH to test MLS2 as the bond stretches and include additional methods beyond those shown in Fig. 1**(c,bottom)**. Unsurprisingly, MP2 (magenta curve) fails to capture the correct physics of BH bond stretching due to the divergence of its correlation energies when the HOMO-LUMO orbital gap closes (see Eq. 7). κMP2 (blue curve; Eq. 9) eliminates this divergence, but its energies are much higher than the exact ones when the bond stretches. In contrast, MLS2 is more accurate than κMP2 not only when the unseen BH bond stretches but also at equilibrium (see also Fig. **S15** in the SI for training results of N_2 and H_2 bond stretches). This demonstrates the power of MLS2 to successfully dissociate covalent bonds without breaking spin symmetries, which is, as said, a crucial challenge for quantum chemistry methods[94–96,103]. In MLS2, this is achieved by first employing κMP2 to eliminate the divergence present in MP2, followed by real-space NN-based enhancements of its spin-resolved energy densities (see Eq. 15).

Finally, revisiting Fig. 1**(e, top)** with the formic acid dimer interaction energy curve, we test the performance of MLS2 for hydrogen-bonded systems. The formic acid dimer is selected because of its two hydrogen bonds with a strong electrostatic component, making it a system that starkly contrasts the dispersion-bonded RG18 systems on which MLS2 was trained. Overall, MP2 overbinds the formic acid dimer, with κMP2 overbinding even more, while MLS2 outperforms both in predicting the dimer's interaction energies.

We observe similar improvements of MLS2 over κMP2 for other dissociation curves of noncovalent systems from the S22x5 dataset[64] (see Fig. **S16** in the SI). These results further confirm MLS2's transferability, as it successfully extrapolates from dispersion-bound systems (RG18) to distinctly different noncovalent interactions, such as hydrogen bonds.

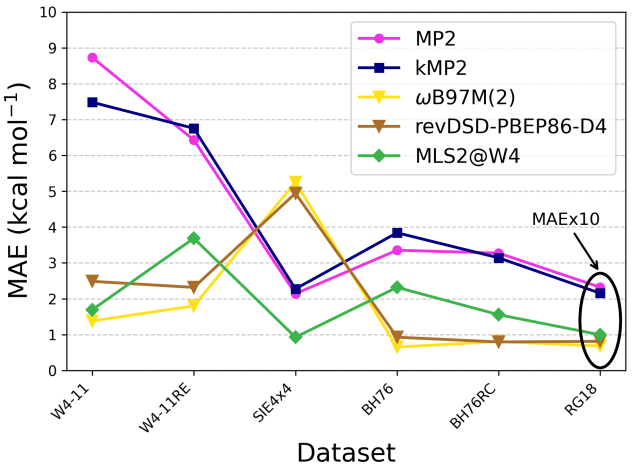

Fig. 6: Energy MAEs (kcal mol^{-1}) of various models for subsets of the GMTKN55 database[104] and the additional W4-11RE set[105]. MLS2 is trained on W4-11 atomization energies[65] (MLS2@W4). The results for RG18 dataset[102] are scaled by a factor of 10 for better visibility.

To go beyond the closed-shell systems considered so far and demonstrate MLS2's broader applicability, we retrain it on the W4-11 atomization energy dataset[65] and test its performance on representative subsets of GMTKN55 (a large main-group database)[104]. We denote this variant as MLS2@W4. Since the MLS2@W4 training and test sets include open-shell systems, we supplement the existing MLS2 features (Sec. S6 in the

SI) with the spin polarization function $\zeta(\mathbf{r})$ (Eq. S12 in the SI). In Fig. 6, we summarize MLS2@W4 performance with the MAEs testing its transferability from W4-atomizations to unseen energy types (see Sec. **S13** in the SI for further details): reaction energies (W4-11RE) with ∼11k reaction data points derived from the W4-11 total energies[105], dissociation energies of small open-shell cationic dimers (SIE4x4)[106], barrier heights (BH76) and reaction energies (BH76RC)[104], and non-covalent interaction energies[104] (RG18). Results for MP2, κMP2, and state-of-the-art double-hybrid DFAs (ωB97M(2)[107] and revDSD-PBEP86-D4[108]) are included for comparison. First, Fig. 6 shows that MLS2@W4 clearly outperforms MP2 and κMP2 with a very good transferability from atomization to other energies. Second, although the two double hybrids perform better for barrier heights and reaction energies (energy types included in their training and unseen by MLS2@W4), MLS2@W4 still achieves reasonable accuracy for these sets. However, for the SIE4x4 dataset, which is very difficult for the standard DFAs due to *self-interaction errors*[104], with MAE slightly below 1 kcal/mol, MLS2@W4 impressively outperforms both double hybrids by a factor of ~ 5. Overall, MLS2@W4 stands out as the most robust method considered here, being the only one that achieves MAE below 4 kcal mol^{-1} across all tested datasets. This further confirms that the general MLS2 framework combining ML with Eq. 15 is highly promising for developing future DFAs.

In view of the good MLS2 performance, it is important to note the regularizing role of Eq. 15. In addition to using a full amount of exact exchange (correlation here modeled relative to the HF reference), Eq. 15 ensures that MLS2 is exact for one-electron systems ($E_c^{\text{MLS2}} = 0$). This is because, $e_c^{\text{MLS2}}(\mathbf{r}) = e_{c,ss}^{\kappa\text{MP2}}(\mathbf{r}) = e_{c,os}^{\kappa\text{MP2}}(\mathbf{r}) = 0$ for $N = 1$, irrespective of the weights produced by the NN. This good property of MLS2 and likely its good transferability would be easily lost if other terms (e.g., exchange energy density or semilocal quantities multiplied by corresponding ML weights) are added to Eq. 15. Crucially, through the specific use of Eq. 15 and NNs, MLS2 provides a way of employing semilocal features [e.g., $s(\mathbf{r})$] while still satisfying the one-electron constraint, in contrast to double hybrids, which violate this constraint due to their way of employing semilocal features. Consequently, MLS2@W4 is not only exact for one-electron systems such as H_2^+ (by contrast, the two double hybrids yield substantial errors upon stretching H_2^+ as shown in Fig. S17 in the SI), but also performs well for self-interaction cases involving more electrons given its excellent performance on SIE4x4.

DISCUSSION

To address the urgent need for transferable ML DFAs, we introduce and analyze several key strategies based on real-space energy learning. By leveraging our regularized, spin-resolved PT2-based correlation energy density generator, we pursue two directions, each demonstrating a distinct aspect of the power of real-space ML for DFAs.

ML2 demonstrates the power of LES. The first direction, ML2, leverages $e_c^{\kappa\text{MP2}}(\mathbf{r})$ (Eq. 5) as a proxy reference for LES-based learning. While LES intrinsically expands a single energy datapoint of GES into thousands per molecule (each grid point becoming a distinct training datapoint), this data efficiency advantage is fully realized only when specific physical considerations are accounted for. These include the use of energy density per particle as the learning target (Eq. 3), adopting the physically meaningful gauge (Eq. 5), and employing a physically-informed ML model (Eq. 13). When these conditions are met, LES provides significantly greater transferability compared to commonly used GES. In particular, our ML2 model trained solely on a small set of atoms effectively generalizes to diatomic dissociation curves. Moreover, under these physically-informed conditions, LES not only enhances transferability but also leads to smoother and faster learning convergence, as well as robustness with respect to variations in ML training conditions compared to GES.

MLS2 leverages our local quantities to open up ML DFAs avenues. In the second direction, MLS2, we leverage our $e_c^{\kappa\text{MP2}}(\mathbf{r})$ and its decomposition into same-spin and opposite-spin channels to open new avenues for ML DFAs. Specifically, MLS2 generalizes spin-component-scaled MP2 by scaling each spin channel locally with NN weights (Eq. 15), combined with Head-Gordon's κ-regularization. MLS2 improves over κMP2 across diverse systems, achieves competitive accuracy compared to modern double hybrids, and outperforms them for challenging systems affected by self-interaction errors.

Going back to the demonstrated power of LES over GES in the proof-of-principle ML2 model based on the κMP2 reference calls for developing robust energy density generators at higher levels of theory. This will be a crucial objective in our future work, with the first step already taken in Ref. 67, which enables obtaining $e_c(\mathbf{r})$ from full configuration interaction (FCI) wavefunctions, with the procedure being easily adaptable to other variational wavefunctions. To learn higher-level $e_c(\mathbf{r})$, one can adapt the LES-based ML2 model, currently designed for $e_c^{\kappa\text{MP2}}(\mathbf{r})$, by adding r_s in the features list and modifying Eq. 13. Nevertheless, we believe that a more controlled approach for ML of higher-level $e_c(\mathbf{r})$ can be achieved by integrating ML2 and MLS2 as follows. Using higher-level energy densities per particle, we can implement LES-based training for MLS2, while simultaneously replacing $e_{c,os}^{\kappa\text{MP2}}(\mathbf{r})$ and $e_{c,ss}^{\kappa\text{MP2}}(\mathbf{r})$ of Eq. 15 with their ML2-based surrogates. This replacement avoids these two more expensive quantities in post-training calculations while leveraging existing ML2 physics (Eq. 13).

Within the MPAC framework that we use here to demonstrate the advantages of LES[67], the input density is fixed to HF, and the learning target is exclusively the energy, allowing us to clearly focus on the effects of LES as a real-space learning strategy for DFAs. At the

same time, this does **not** imply that density learning, as a complementary real-space learning strategy, should be abandoned within frameworks where both densities and energies are learning targets. E.g., in DFA development within KS DFT, where both energies and densities are learning targets, loss functions can incorporate the LES term alongside density term (see, e.g., Refs. 31 and 34), or even further sophisticate such loss functions by using the specific link between energy densities and *correlation potential*[109]. Such combined strategies would leverage strengths of both real-space approaches, which we plan to explore in future work.

While we demonstrate here several advantages of LES over GES, the strength of GES is that it can readily leverage existing global energy data from extensive chemical databases (e.g., GMTKN55). Computing high-level energy densities per particle for every system in such large databases would be impractical and likely unnecessary, particularly given the high transferability potential of LES demonstrated here using data from just eight atoms. Thus, a practical approach would be to design a loss function incorporating both GES and LES terms: employing GES for existing global energy data while strategically complementing it with LES-based training on carefully selected subsets (e.g., a dozen atoms and representative molecules from "Slim" subsets[110] of GMTKN55) to leverage its power for embedding transferability into ML DFAs.

METHODS

Computational details

All electronic structure calculations were performed using the PySCF 2.3.0 program package[111,112] within the Python coding environment[113] v3.11.4.

For the evaluations in Fig. **1(d)**, Fig. **2**, Fig. **6** and Fig. **S3**, Table S1, Table **S2**, Fig. **S17**, Fig. **S18**, Fig. **S19** and Fig. **S20** in the SI, we have used *def2-QZVPPD* basis set[114,115], while for the rest, unless specified otherwise, we have used *def2-QZVP* basis set[114]. Our implementation of the MP2 correlation energy density generator uses the Python package for optimizing tensor contractions[88], together with JAX[73] to enable parallelization and high-performance platform agnostic evaluation of the energy densities. For the energy density generation, we also employ the density fitting[86,87] (DF) approximation for MP2 with the *def2-QZVP(PD)-RI* auxiliary basis set. We adapt the same DF code and combine it with the *def2-universal-jkfit* auxiliary basis set to calculate the exchange energy density. In Sec. S3 in the SI, we discuss the effect on the accuracy from the numerical integration with respect to the DFT grid[116] and from the use of DF.

For the regularized correlation energy density data generation, we set $\kappa = 2$ throughout this work unless otherwise specified.

Reference correlation energies are taken from specified references or obtained from CCSD(T) calculations in PySCF.

Neural Network training

The neural network training was performed with the Pytorch 2.1.2 deep learning library[117]. The input features (Sec. S6 in the SI) were obtained in Python from the HF PySCF output of the given chemical system and pointwise evaluated on the DFT grid[116]. In particular, two FOD features at electronic temperatures $T_1 = 10000K$ and $T_2 = 25000K$ were employed for every system. Their implementation is based on the formula from Ref. 62, which we divide by the density to obtain a dimensionless quantity. We employed the Adam optimizer[118] in Pytorch for NN training with custom learning rates incorporating a warm-up period and a fixed number of training steps (epochs).

ML2 was trained on eight small closed-shell atoms/ions (H^-, He, Be, Mg, Ne, Ar, Ca, and Kr). The architecture of the neural network for ML2 has three hidden layers, each consisting of 16 neurons. In ML2, we apply hyperbolic tangent (tanh) activation functions to the NN output layers, bounding the ML2 weights between -1 and 1 (for a concrete example illustrating the range of LES-based ML2 $w_c(\mathbf{r})$ weights that justifies this choice, see Fig. S20 in the SI). For simplicity, the same tanh activation is also used for all hidden layers. All models in this work are optimized by using either the LES-based or the GES-based loss function (see Sec. S8 in the SI for specific details). In ML2, the learning rate follows an exponential decay that continues over 5000 epochs including the warm-up phase.

A detailed overview of the training set for the first MLS2 example is given in Sec. S12 in the SI. For the architecture of the MLS2's NN, we used four hidden layers with 64 neurons each. Following the ML2 architecture, in MLS2 we also use tanh activation function for the hidden layers, while the sigmoid function is applied to the output layer to constrain the MLS2 weights between 0 and 1 before scaling them by a factor of 10 (see Sec. **S11** in the SI for further details on MLS2 construction). The total loss function contains total and interaction correlation energy errors, as detailed in Sec. S8 in the SI. Here, the learning rate follows also an exponential decay over 1000 epochs.

The original MLS2 architecture (4×64) is also used for MLS2@W4, and in Fig. S18 in the SI, we compare alternative NN architectures and show that 4×64 offers modest improvements in validation loss over smaller models, motivating its use. Additionally, we performed ten training runs of MLS2@W4 with different random seed initializations (see Fig. **S19** in the SI for results of each run). From these ten runs, the final MLS2@W4 NN shown in Fig. **6** was selected based on the lowest MAE on the entire W4-11 set used for training and validation. Since MLS2@W4 also involves open-shell systems,

we have used unrestricted Hartree–Fock (UHF) orbitals. Following Ref. 119 of Sim and co-workers, constrained-UHF (CUHF) was employed for reactions where UHF orbitals exhibited spin contamination following the criterion from Ref. 120. For further MLS2@W4 training and testing details, see Sec. **S13** in the SI..

DATA AVAILABILITY

The data that support the key findings of this study are available from the manuscript and its supporting information. Further data will be made available on Zenodo repository.

CODE AVAILABILITY

The Python code for the energy density generators is available at `https://github.com/vuckovic-lab/energy_densities.git`.

REFERENCES

[1] M. R. Hush, Science **355**, 580 (2017).
[2] R. Ramakrishnan and O. A. von Lilienfeld, Machine learning, quantum chemistry, and chemical space, in *Reviews in Computational Chemistry* (John Wiley & Sons, Ltd, 2017) Chap. 5, pp. 225–256, https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781119356059.ch5.
[3] O. A. von Lilienfeld, Angewandte Chemie International Edition **57**, 4164 (2018).
[4] K. T. Schütt, M. Gastegger, A. Tkatchenko, K.-R. Müller, and R. J. Maurer, Nature Communications **10**, 5024 (2019).
[5] A. Tkatchenko, Nature Communications **11**, 4125 (2020).
[6] P. Zheng, R. Zubatyuk, W. Wu, O. Isayev, and P. O. Dral, Nature Communications **12**, 7022 (2021).
[7] B. Huang, G. F. von Rudorff, and O. A. von Lilienfeld, Science **381**, 170 (2023).
[8] R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, Scientific Data **1**, 140022 (2014).
[9] R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, Journal of Chemical Theory and Computation **11**, 2087 (2015).
[10] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, The Journal of Chemical Physics **148**, 241722 (2018).
[11] C. Chen, W. Ye, Y. Zuo, C. Zheng, and S. P. Ong, Chemistry of Materials **31**, 3564 (2019).
[12] J. S. Smith, R. Zubatyuk, B. Nebgen, N. Lubbers, K. Barros, A. E. Roitberg, O. Isayev, and S. Tretiak, Scientific Data **7**, 134 (2020).
[13] I. Batatia, D. P. Kovacs, G. Simm, C. Ortner, and G. Csanyi, Mace: Higher order equivariant message passing neural networks for fast and accurate force fields, in *Advances in Neural Information Processing Systems*, Vol. 35, edited by S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Curran Associates, Inc., 2022) pp. 11423–11436.
[14] A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon, and E. D. Cubuk, Nature **624**, 80 (2023).
[15] S. Zhang, M. Z. Makoś, R. B. Jadrich, E. Kraka, K. Barros, B. T. Nebgen, S. Tretiak, O. Isayev, N. Lubbers, R. A. Messerly, and J. S. Smith, Nature Chemistry **16**, 727 (2024).
[16] A. Chandrasekaran, D. Kamal, R. Batra, C. Kim, L. Chen, and R. Ramprasad, npj Computational Materials **5**, 22 (2019).
[17] K. Mills, K. Ryczko, I. Luchak, A. Domurad, C. Beeler, and I. Tamblyn, Chemical Science **10**, 4129 (2019).
[18] O. A. von Lilienfeld and K. Burke, Nature Communications **11**, 4895 (2020).
[19] O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, and K.-R. Müller, Chemical Reviews **121**, 10142 (2021).
[20] A. Nandy, C. Duan, M. G. Taylor, F. Liu, A. H. Steeves, and H. J. Kulik, Chemical Reviews **121**, 9927 (2021).
[21] R. Pederson, B. Kalita, and K. Burke, Nature Reviews Physics **4**, 357 (2022).
[22] H. J. Kulik, Israel Journal of Chemistry **62**, e202100016 (2022).
[23] L. Fiedler, K. Shah, M. Bussmann, and A. Cangi, Physical Review Materials **6**, 040301 (2022).
[24] J. A. Rackers, L. Tecot, M. Geiger, and T. E. Smidt, Machine Learning: Science and Technology **4**, 015027 (2023).
[25] L. Fiedler, N. A. Modine, S. Schmerler, D. J. Vogel, G. A. Popoola, A. P. Thompson, S. Rajamanickam, and A. Cangi, npj Computational Materials **9**, 115 (2023).
[26] D. J. Tozer, V. E. Ingamells, and N. C. Handy, The Journal of Chemical Physics **105**, 9200 (1996).
[27] J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller, and K. Burke, Physical Review Letters **108**, 253002 (2012).
[28] L. Li, T. E. Baker, S. R. White, and K. Burke, Physical Review B **94**, 245129 (2016).
[29] J. Schmidt, C. L. Benavides-Riveros, and M. A. L. Marques, The Journal of Physical Chemistry Letters **10**, 6425 (2019).
[30] S. Dick and M. Fernandez-Serra, Nature Communications **11**, 3509 (2020).
[31] R. Nagai, R. Akashi, and O. Sugino, npj Computational Materials **6**, 43 (2020).
[32] M. Bogojeski, L. Vogt-Maranto, M. E. Tuckerman, K.-R. Müller, and K. Burke, Nature Communications **11**, 5223 (2020).
[33] B. Kalita, L. Li, R. J. McCarty, and K. Burke, Accounts of Chemical Research **54**, 818 (2021).
[34] L. Li, S. Hoyer, R. Pederson, R. Sun, E. D. Cubuk, P. Riley, and K. Burke, Physical Review Letters **126**, 036401 (2021).
[35] J. T. Margraf and K. Reuter, Nature Communications **12**, 344 (2021).
[36] K. Bystrom and B. Kozinsky, Journal of Chemical Theory and Computation **18**, 2180 (2022).
[37] M. Ruth, D. Gerbig, and P. R. Schreiner, Journal of Chemical Theory and Computation **19**, 4912 (2023).
[38] S. Riemelmoser, C. Verdi, M. Kaltak, and G. Kresse, Journal of Chemical Theory and Computation **19**, 7287 (2023).
[39] J. Kirkpatrick, B. McMorrow, D. H. P. Turban, A. L. Gaunt, J. S. Spencer, A. G. D. G. Matthews, A. Obika, L. Thiry, M. Fortunato, D. Pfau, L. R. Castellanos, S. Petersen, A. W. R. Nelson, P. Kohli, P. Mori-Sánchez, D. Hassabis, and A. J. Cohen, Science **374**, 1385 (2021).
[40] P. Zheng, R. Zubatyuk, W. Wu, O. Isayev, and P. O. Dral, Nature Communications **12**, 7022 (2021).
[41] T. Gould, B. Chan, S. G. Dale, and S. Vuckovic, Chemical Science , 11122 (2024).
[42] H. Shimakawa, A. Kumada, and M. Sato, npj Computational Materials **10**, 11 (2024).
[43] K. Li, A. N. Rubungo, X. Lei, D. Persaud, K. Choudhary, B. DeCost, A. B. Dieng, and J. Hattrick-Simpers, Probing out-of-distribution generalization in machine learning for materials (2024).
[44] C. D. Sherrill, D. E. Manolopoulos, T. J. Martinez, and A. Michaelides, The Journal of Chemical Physics **153**, 070401 (2020).
[45] H. Zhao, T. Gould, and S. Vuckovic, Physical Chemistry Chemical Physics **26**, 12289 (2024).
[46] A. D. Becke, Physical Review A **38**, 3098 (1988).
[47] C. Lee, W. Yang, and R. G. Parr, Physical Review B **37**, 785 (1988).

[48] A. D. Becke, The Journal of Chemical Physics **98**, 5648 (1993).

[49] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, The Journal of Physical Chemistry **98**, 11623 (1994).

[50] J. P. Perdew, K. Burke, and M. Ernzerhof, Physical Review Letters **77**, 3865 (1996), the 1996 exchange functional and gradient-corrected correlation functional of Perdew, Burke and Ernzerhof (PBE) as the generalized gradient approximation of the exchange-correlation functional.

[51] V. Butera, Physical Chemistry Chemical Physics **26**, 7950 (2024).

[52] C. J. Cramer and D. G. Truhlar, Physical Chemistry Chemical Physics **11**, 10757 (2009).

[53] F. Liu, C. Duan, and H. J. Kulik, The Journal of Physical Chemistry Letters **11**, 8067 (2020).

[54] J. Hollingsworth, L. Li, T. E. Baker, and K. Burke, The Journal of Chemical Physics **148**, 241743 (2018).

[55] S. Dick and M. Fernandez-Serra, Physical Review B **104**, L161109 (2021).

[56] R. Nagai, R. Akashi, and O. Sugino, Physical Review Research **4**, 013106 (2022).

[57] K. Pokharel, J. W. Furness, Y. Yao, V. Blum, T. J. P. Irons, A. M. Teale, and J. Sun, The Journal of Chemical Physics **157**, 174106 (2022).

[58] S. Vuckovic and H. Bahmann, Journal of Chemical Theory and Computation **19**, 6172 (2023).

[59] M. F. Kasim and S. M. Vinko, Physical Review Letters **127**, 126403 (2021).

[60] P. A. M. Casares, J. S. Baker, M. Medvidović, R. dos Reis, and J. M. Arrazola, The Journal of Chemical Physics **160**, 062501 (2024).

[61] S. Mohanty, S. Yoo, K. Kang, and W. Cai, Computer Physics Communications **288**, 108723 (2023).

[62] S. Grimme and A. Hansen, Angewandte Chemie International Edition **54**, 12308 (2015).

[63] C. A. Bauer, A. Hansen, and S. Grimme, Chemistry - A European Journal **23**, 6150 (2017).

[64] L. Gráfová, M. Pitoňák, J. Rezáč, and P. Hobza, Journal of Chemical Theory and Computation **6**, 2365 (2010).

[65] A. Karton, S. Daon, and J. M. Martin, Chemical Physics Letters **510**, 165 (2011).

[66] W. Kohn and L. J. Sham, Physical Review **140**, A1133 (1965).

[67] K. J. Daas, H. Zhao, E. Polak, and S. Vuckovic, Journal of Chemical Theory and Computation **21**, 5501 (2025).

[68] P. Szarek and A. Tachibana, Journal of molecular modeling **13**, 651 (2007).

[69] P. Szarek, Y. Sueda, and A. Tachibana, The Journal of chemical physics **129** (2008).

[70] A. Tachibana, Journal of Molecular Structure: THEOCHEM **943**, 138 (2010).

[71] J. Lee and M. Head-Gordon, Journal of Chemical Theory and Computation **14**, 5203 (2018).

[72] J. Shee, M. Loipersberger, A. Rettig, J. Lee, and M. Head-Gordon, The Journal of Physical Chemistry Letters **12**, 12084 (2021).

[73] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, Jax: composable transformations of Python+NumPy programs (2018).

[74] S. Grimme, The Journal of Chemical Physics **118**, 9095 (2003).

[75] S. Grimme, L. Goerigk, and R. F. Fink, WIREs Computational Molecular Science **2**, 886 (2012).

[76] M.-C. Kim, E. Sim, and K. Burke, Physical Review Letters **111**, 073003 (2013).

[77] S. Vuckovic, S. Song, J. Kozlowski, E. Sim, and K. Burke, Journal of Chemical Theory and Computation **15**, 6636 (2019).

[78] S. Song, S. Vuckovic, Y. Kim, H. Yu, E. Sim, and K. Burke, Nature Communications **14**, 799 (2023).

[79] F. G. Cruz, K.-C. Lam, and K. Burke, The Journal of Physical Chemistry A **102**, 4911 (1998).

[80] K. Burke, F. G. Cruz, and K.-C. Lam, The Journal of Chemical Physics **109**, 8161 (1998).

[81] S. Vuckovic, T. J. P. Irons, A. Savin, A. M. Teale, and P. Gori-Giorgi, Journal of Chemical Theory and Computation **12**, 2598 (2016).

[82] S. Vuckovic, T. J. P. Irons, L. O. Wagner, A. M. Teale, and P. Gori-Giorgi, Physical Chemistry Chemical Physics **19**, 6169 (2017).

[83] S. Vuckovic, M. Levy, and P. Gori-Giorgi, The Journal of Chemical Physics **147**, 214107 (2017).

[84] S. Vuckovic, E. Fabiano, P. Gori-Giorgi, and K. Burke, Journal of Chemical Theory and Computation **16**, 4141 (2020).

[85] P. Jurečka, J. Sponer, J. Cerný, and P. Hobza, Phys. Chem. Chem. Phys. **8**, 1985 (2006).

[86] B. I. Dunlap, Physical Chemistry Chemical Physics **2**, 2113 (2000).

[87] H.-J. Werner, F. R. Manby, and P. J. Knowles, The Journal of Chemical Physics **118**, 8149 (2003).

[88] D. G. A. Smith and J. Gray, Journal of Open Source Software **3**, 753 (2018).

[89] M. Piris, X. Lopez, and J. M. Ugalde, The Journal of Chemical Physics **128**, 10.1063/1.2883959 (2008).

[90] O. Gritsenko and E. J. Baerends, The Journal of Chemical Physics **124**, 10.1063/1.2165183 (2006).

[91] M. Via-Nadal, M. Rodríguez-Mayorga, and E. Matito, Physical Review A **96**, 050501 (2017).

[92] J. W. Furness, A. D. Kaplan, J. Ning, J. P. Perdew, and J. Sun, The Journal of Physical Chemistry Letters **11**, 8208 (2020).

[93] M. Levy and J. P. Perdew, Physical Review A **32**, 2010 (1985).

[94] Y. Shi, Y. Shi, and A. Wasserman, The Journal of Physical Chemistry Letters **15**, 826 (2024).

[95] N. Q. Su, C. Li, and W. Yang, Proceedings of the National Academy of Sciences **115**, 9678 (2018).

[96] S. Vuckovic, L. O. Wagner, A. Mirtschink, and P. Gori-Giorgi, Journal of Chemical Theory and Computation **11**, 3153 (2015).

[97] T. Nudejima, Y. Ikabata, J. Seino, T. Yoshikawa, and H. Nakai, The Journal of Chemical Physics **151**, 10.1063/1.5100165 (2019).

[98] Y. Ikabata, R. Fujisawa, J. Seino, T. Yoshikawa, and H. Nakai, The Journal of Chemical Physics **153**, 10.1063/5.0021281 (2020).

[99] R. Han, M. Rodríguez-Mayorga, and S. Luber, Journal of Chemical Theory and Computation **17**, 777 (2021).

[100] Ł. M. Mentel, R. van Meer, O. V. Gritsenko, and E. J. Baerends, The Journal of Chemical Physics **140**, 214105 (2014).

[101] T. J. P. Irons and A. M. Teale, Molecular Physics , 1 (2015).

[102] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, The Journal of Chemical Physics **132**, 154104 (2010).

[103] S. Vuckovic, Journal of Chemical Theory and Computation **15**, 3580 (2019).

[104] L. Görigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, and S. Grimme, Physical Chemistry Chemical Physics **19**, 32184 (2017).

[105] J. T. Margraf, D. S. Ranasinghe, and R. J. Bartlett, Physical Chemistry Chemical Physics **19**, 9798 (2017).

[106] L. Goerigk and S. Grimme, Journal of Chemical Theory and Computation **6**, 107 (2010).

[107] N. Mardirossian and M. Head-Gordon, The Journal of Chemical Physics **148**, 10.1063/1.5025226 (2018).

[108] G. Santra, M. Cho, and J. M. L. Martin, The Journal of Physical Chemistry A **125**, 4614 (2021).

[109] S. Giarrusso, S. Vuckovic, and P. Gori-Giorgi, Journal of Chemical Theory and Computation **14**, 4151–4167 (2018).

[110] T. Gould and S. Vuckovic, Journal of Chemical Theory and Computation 10.1021/acs.jctc.5c00512 (2025).

[111] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K. Chan, WIREs Computational Molecular Science **8**, e1340 (2018).

[112] Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui, J. J. Eriksen, Y. Gao, S. Guo, J. Hermann, M. R. Hermes, K. Koh, P. Koval, S. Lehtola, Z. Li, J. Liu, N. Mardirossian, J. D. McClain, M. Motta, B. Mussard, H. Q. Pham, A. Pulkin, W. Purwanto, P. J. Robinson, E. Ronca, E. R. Sayfutyarova, M. Scheurer, H. F. Schurkus, J. E. T. Smith, C. Sun, S.-N. Sun, S. Upadhyay, L. K. Wagner, X. Wang, A. White, J. D. Whitfield, M. J. Williamson, S. Wouters, J. Yang, J. M. Yu, T. Zhu, T. C. Berkelbach, S. Sharma, A. Y. Sokolov, and G. K.-L. Chan, The Journal of Chemical Physics **153**, 024109 (2020).

[113] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual* (CreateSpace, Scotts Valley, CA, 2009).

[114] F. Weigend, F. Furche, and R. Ahlrichs, The Journal of Chemical Physics **119**, 12753 (2003).

[115] D. Rappoport and F. Furche, The Journal of Chemical Physics **133**, 10.1063/1.3484283 (2010).

[116] V. I. Lebedev and D. N. Laikov, Doklady Mathematics **59**, 477 (1999).

[117] J. Ansel, E. Yang, H. He, N. Gimelshein, A. Jain, M. Voznesensky, B. Bao, P. Bell, D. Berard, E. Burovski, G. Chauhan, A. Chourdia, W. Constable, A. Desmaison, Z. DeVito, E. Ellison, W. Feng, J. Gong, M. Gschwind, B. Hirsh, S. Huang, K. Kalambarkar, L. Kirsch, M. Lazos, M. Lezcano, Y. Liang, J. Liang, Y. Lu, C. Luk, B. Maher, Y. Pan, C. Puhrsch, M. Reso, M. Saroufim, M. Y. Siraichi, H. Suk, M. Suo, P. Tillet, E. Wang, X. Wang, W. Wen, S. Zhang, X. Zhao, K. Zhou, R. Zou, A. Mathews, G. Chanan, P. Wu, and S. Chintala, in *29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2 (ASPLOS '24)* (ACM, 2024).

[118] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization (2014).

[119] H. Yu, S. Song, S. Nam, K. Burke, and E. Sim, The Journal of Physical Chemistry Letters **14**, 9230–9237 (2023).

[120] S. Song, S. Vuckovic, E. Sim, and K. Burke, Journal of Chemical Theory and Computation **18**, 817 (2022).

ACKNOWLEDGMENTS

We thank for fruitful discussions and helpful comments from Kimberly Daas and Tim Gould. We also thank Prof. Eunji Sim for giving us access to their code[119], which we have used to calculate CUHF orbitals in PySCF. The authors acknowledge funding from the SNSF Starting Grant project (TMSGI2_211246).

AUTHOR CONTRIBUTIONS

E.P. ran all ML simulations. The electronic structure dataset was generated by H.Z. S.V. and E.P. did the data analysis and co-wrote the paper. S.V. supervised the project and conceptualized the idea together with E.P.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at .

Supplementary Information for: Real-space machine learning of correlation density functionals

Elias Polak[1], Heng Zhao[1], and Stefan Vuckovic[1,a]

[1]*Department of Chemistry, University of Fribourg, Fribourg, Switzerland.*

[a)]Corresponding author: stefan.vuckovic@unifr.ch

CONTENTS

S1. MP2-BASED CORRELATION ENERGY DENSITY

In this section, we give a brief summary of how how our recently introduced Møller–Plesset adiabatic connection (MPAC) correlation energy density framework[1] yields the MP2 correlation energy density per particle (Eq. 5). Consider the following coupling constant λ-dependent Hamiltonian[1,2] for N-electron system:

$$\hat{H}_\lambda = \hat{T} + \hat{V}_{\text{ext}} + \lambda \hat{V}_{ee} + (1-\lambda)(\hat{J} + \hat{K}), \tag{S1}$$

where $\hat{T}$ is the kinetic energy operator, $\hat{V}_{\text{ext}}$ the external potential, $\hat{V}_{ee}$ the electron-electron interaction, and $\hat{J}$ and $\hat{K}$ are the standard HF Coulomb and exchange operators, defined in terms of the HF density and orbitals, respectively. Let Ψ_λ be the ground-state wavefunction of $\hat{H}_\lambda$. Then, the MPAC integral expression for correlation energy (true - HF energy) reads,

$$E_{\text{c}} = \int_0^1 E_{\text{c}}^\lambda d\lambda, \tag{S2}$$

where E_{c}^λ is defined as

$$E_{\text{c}}^\lambda = \left\langle \Psi_\lambda \middle| \hat{V}_{ee} - \hat{J} - \hat{K} \middle| \Psi_\lambda \right\rangle - \left\langle \Psi_0 \middle| \hat{V}_{ee} - \hat{J} - \hat{K} \middle| \Psi_0 \right\rangle. \tag{S3}$$

In Ref. 1, the expression for $e_{\text{c}}^\lambda(\mathbf{r})$ was derived using a gauge analogous to the correlation hole in DFT, which, upon λ-integration [Eq. S2], yields the corresponding $e_{\text{c}}(\mathbf{r})$,

$$E_{\text{c}} = \int e_{\text{c}}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} = \int \left[\int_0^1 e_{\text{c}}^\lambda(\mathbf{r}) d\lambda \right] \rho(\mathbf{r}) d\mathbf{r}. \tag{S4}$$

Expanding $e_{\text{c}}^\lambda(\mathbf{r})$ in the small-λ limit (see Ref. 1 for further details), yields,

$$e_{\text{c}}^\lambda(\mathbf{r}) \approx e_{\text{c}}'(\mathbf{r})\lambda = \frac{1}{2\rho(\mathbf{r})} \int \frac{P_2'(\mathbf{r},\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' \lambda \qquad \text{for } \lambda \to 0, \tag{S5}$$

where,

$$e_{\text{c}}'(\mathbf{r}) = \frac{d}{d\lambda} e_{\text{c}}^\lambda(\mathbf{r}) \Big|_{\lambda=0} \tag{S6a}$$

$$P_2^{\text{MP2}}(\mathbf{r},\mathbf{r}') = P_2'(\mathbf{r},\mathbf{r}') = \frac{d}{d\lambda} P_2^\lambda(\mathbf{r},\mathbf{r}') \Big|_{\lambda=0} \tag{S6b}$$

$$P_2^\lambda(\mathbf{r},\mathbf{r}') = N(N-1) \int |\Psi_\lambda(\mathbf{r},\mathbf{r}',\mathbf{r}_3,\ldots,\mathbf{r}_N)|^2 \, d\mathbf{r}_3 \cdots d\mathbf{r}_N \tag{S6c}$$

with $P_2^\lambda(\mathbf{r},\mathbf{r}')$ being the λ- dependent pair density, where real Ψ_λ is assumed in Eq. S6c. By virtue of Eq. S6b, $P_2^{\text{MP2}}(\mathbf{r},\mathbf{r}') = P_2'(\mathbf{r},\mathbf{r}')$ corresponds to the first-order derivative of the pair density $P_2^\lambda(\mathbf{r},\mathbf{r}')$ with respect to λ, evaluated at $\lambda = 0$. Therefore, it contains only the correlation component, and resolving it in terms of HF orbitals yields[1],

$$P_2^{\text{MP2}}(\mathbf{r},\mathbf{r}') = -2 \sum_{ijab} t_{ij}^{ab} \big(\phi_i(\mathbf{r})\phi_j(\mathbf{r}')\phi_a(\mathbf{r}')\phi_b(\mathbf{r})\delta_{ia}\delta_{jb} - \phi_i(\mathbf{r})\phi_j(\mathbf{r})\phi_a(\mathbf{r}')\phi_b(\mathbf{r}')\delta_{ib}\delta_{ja} \big), \tag{S7}$$

where δ is the Kronecker δ over two spin indices and t_{ij}^{ab} are the MP2 double amplitudes, $t_{ij}^{ab} = T_{ijab}\delta_{ia}\delta_{jb} - T_{ijba}\delta_{ib}\delta_{ja}$, with $T_{ijab} = \frac{\langle ij|ab\rangle}{\varepsilon_a+\varepsilon_b-\varepsilon_i-\varepsilon_j}$ defined as the partial MP2 double amplitudes in Eq. (7) using orbital energies, ε, and two-electron integrals, $\langle ij|ab\rangle$. Finally, taking the λ-integral from Eq. S4 of Eq. S5 yields,

$$e_{\mathrm{c}}(\mathbf{r}) \approx \frac{1}{2}e'_{\mathrm{c}}(\mathbf{r}) = e_{\mathrm{c}}^{\mathrm{MP2}}(\mathbf{r}) = \frac{1}{4\rho(\mathbf{r})}\int \frac{P_2^{\mathrm{MP2}}(\mathbf{r},\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}'. \tag{S8}$$

Applying the κ-regularization[3,4] to the partial MP2 double amplitudes (see Eq. (9)),

$$T_{ijab}^{\kappa} = T_{ijab}\left(1 - e^{-\kappa(\varepsilon_a+\varepsilon_b-\varepsilon_i-\varepsilon_j)}\right)^2,$$

results in $P_2^{\kappa\mathrm{MP2}}(\mathbf{r},\mathbf{r}')$, which is the regularized expression of Eq. S7, and thus, we have $e_{\mathrm{c}}^{\kappa\mathrm{MP2}}(\mathbf{r})$ from Eq. 10.

S2. EFFECT OF REGULARIZATION ON MP2 RESULTS

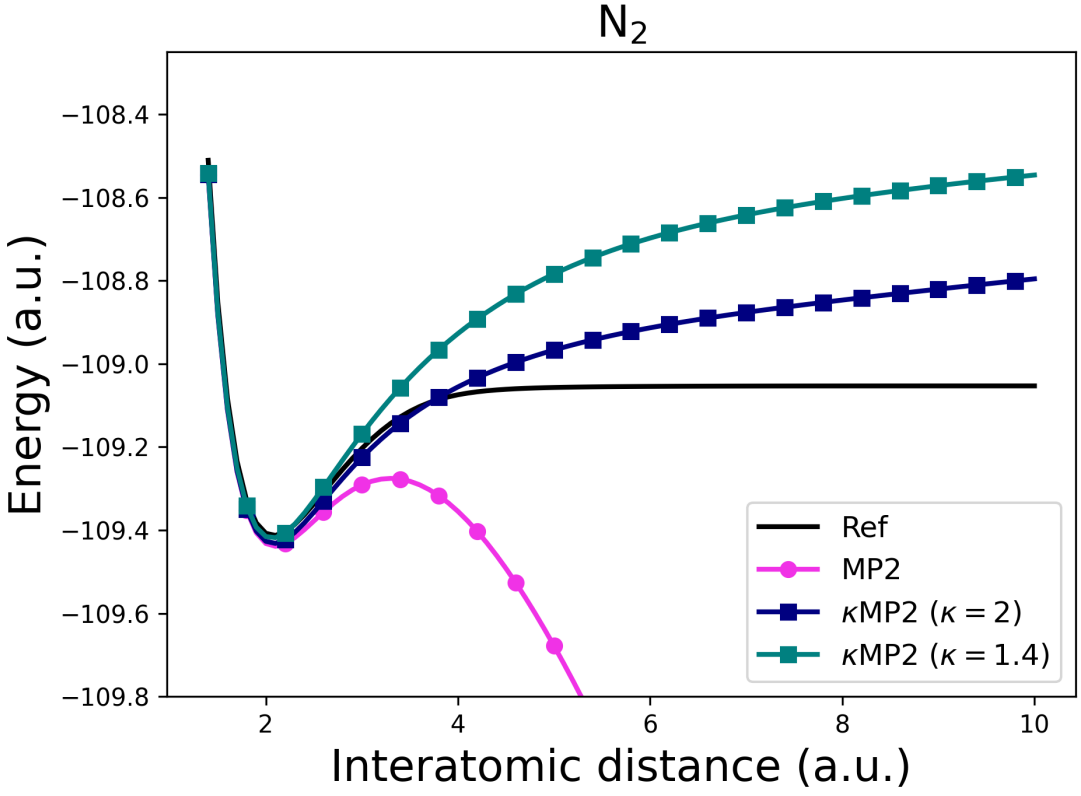

Fig. S1: κMP2-based dissociation energy curve of N_2 with different κ values. MP2 corresponds to $\kappa \to \infty$. Reference (Ref) taken from Ref. 5.

S3. IMPLEMENTATION TECHNIQUES OF THE MP2 CORRELATION ENERGY DENSITY GENERATOR

This section is devoted to the numerical settings of our MP2 correlation energy density implementation in the Python coding language[6]. The resulting generator reads the relevant atomic orbital functions from a HF treatment using the PySCF library[7,8] and performs the necessary MP2-based integrations and tensor contractions. This procedure scales similar to other second-order perturbation theory derived models, and thus, we can employ some approximation techniques to enhance the applicability and efficiency of our code. The resulting accuracy can be assessed using the correlation energy result from the already implemented MP2 module in PySCF (refMP2).

For the following discussion, we introduce some notation for the orbitals:

- the molecular orbital functions (MOs) ϕ_i;
- the atomic orbital functions (AOs) χ_n;
- the auxiliary basis functions ψ_t.

Expanding the MOs inside the integral of $V_{ijab}(\mathbf{r})$ in AO basis, $\phi_i(\mathbf{r}) = \sum_n C_{ni}\chi_n(\mathbf{r})$, yields

$$\begin{aligned} V_{ijab}(\mathbf{r}) &= \phi_i(\mathbf{r})\phi_a(\mathbf{r}) \sum_{m,n} C_{mj}C_{nb} \int \frac{\chi_m(\mathbf{r}')\chi_n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' \\ &= \phi_i(\mathbf{r})\phi_a(\mathbf{r}) \sum_{m,n} C_{mj}C_{nb}A_{mn}(\mathbf{r}), \end{aligned} \tag{S9}$$

where we define the tensor integral

$$A_{mn}(\mathbf{r}) = \int \frac{\chi_m(\mathbf{r}')\chi_n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}'.$$

Density Fitting[9,10] (DF) is a common strategy for MP2 to improve the scaling of the two-electron integral evaluations. It uses an auxiliary basis to expand the direct product of AOs, $\chi_m(\mathbf{r})\chi_n(\mathbf{r}) = \sum_t Q_{mnt}\psi_t(\mathbf{r})$. This allows a one order of magnitude speed increase of our code since the resulting integral evaluation only scales linearly with the basis set (or number of electrons). Specifically, the orbital potential $V_{ijab}(\mathbf{r})$ from Eq. S9 then becomes

$$V_{ijab}(\mathbf{r}) = \phi_i(\mathbf{r})\phi_a(\mathbf{r}) \sum_{m,n,t} C_{mj}C_{nb}Q_{mnt} \int \frac{\psi_t(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}'. \tag{S10}$$

Employing a tensor notation for the integral

$$I_t(\mathbf{r}) = \int \frac{\psi_t(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}'$$

results in the following density fitted orbital potential formulation:

$$V_{ijab}(\mathbf{r}) = \phi_i(\mathbf{r})\phi_a(\mathbf{r}) \sum_{m,n,t} C_{mj}C_{nb}Q_{mnt}I_t(\mathbf{r}). \tag{S11}$$

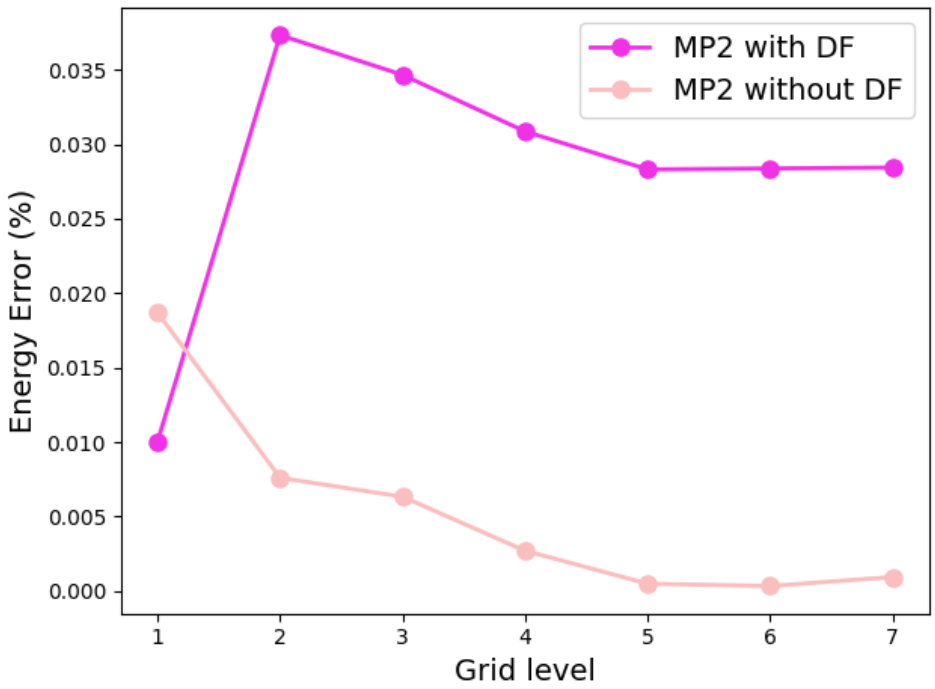

Fig. S2: Relative interaction energy error. The shown data is evaluated with respect to the refMP2 value (-0.0029 kcal mol^{-1}) at different DFT discretization grid-levels[11] for Ne_2 at equilibrium (geometry taken from Ref. 12). Plotted are the results with and without the DF approximation.

We assess correlation energy error due to DF using interaction energies. For example, the relative absolute interaction energy error with respect to refMP2 for Ne_2 is plotted in Fig. **S2** at different DFT discretization grid levels[11].

As expected, the error in correlation energy with respect to refMP2 vanishes as the grid size increases. It shows furthermore how the default grid-level (3) is already very accurate with an error below 0.01% in the interaction energy. Using the DF approximation results in a constant shift with higher integration accuracy. Nevertheless, this difference is still negligible relative to the refMP2 interaction energy. Throughout the rest of this work, we always employ the DF approximation, expecting negligible errors arising from it.

S4. ADDITIONAL $\Delta P_2^{\mathbf{MP2}}(\mathbf{r}, \mathbf{r}')$ PLOTS FOR STRETCHED HELIUM DIMER

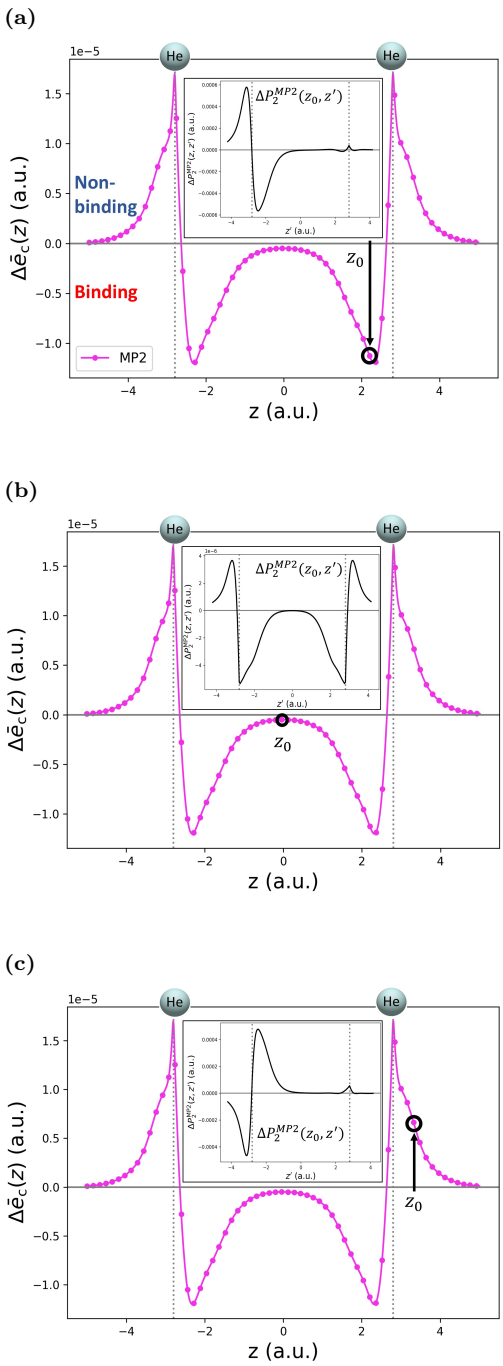

Fig. S3: Visualization of $\Delta\bar{e}_c^{\mathrm{MP2}}(z)$ plotted as in Fig. 2a. Panels **S3a**–**S3c** differ in their $\Delta P_2^{\mathrm{MP2}}(z_0, z')$ inset plots, which are given at three different z_0 positions.

S5. REGULARIZED MP2 INTERACTION ENERGIES

κ	0.2	0.4	0.6	0.8	1.0	1.2	1.4	2.0	∞
$E^{\kappa\text{MP2}}$	0.0089	0.0011	-0.0024	-0.0044	-0.0057	-0.0065	-0.0070	-0.0073	-0.0074

Table S1: Interaction energies (kcal mol^{-1}) of the Helium dimer for different κ-regularizations. Data corresponds to the energy density plots in Figs. 1**(d,top)** and 2a. $\kappa \to \infty$ represents the MP2 result without regularization. Additional interaction energies for comparison: $E^{\text{CCSD(T)}} = -0.013$ and $E^{\text{HF}} = 0.018$.

Method	Ref	HF	MP2	κMP2
E	-1.5	1.13	-1.87	-1.58

Table S2: Interaction energies (kcal mol^{-1}) of the benzene-methane complex. Data corresponds to the interaction energy density visualizations in Figs. 2b– 2d. Reference value at CCSD(T)/CBS level of theory taken from the S22 database[13].

S6. ML FEATURES FOR ML2 AND MLS2

We list in this section formulas of the DFT-based ML features that were used for the training of the ML2 and the MLS2 models. They are all implemented in Python as dimensionless objects and then evaluated at every point of the DFT grid for the input data.

ML features for ML2:

- Reduced density gradient,

$$s(\mathbf{r}) = \frac{|\nabla\rho(\mathbf{r})|}{2\left(3\pi^2\right)^{1/3}\rho^{4/3}(\mathbf{r})}.$$

- Reduced density Laplacian,

$$q(\mathbf{r}) = \frac{\nabla^2\rho(\mathbf{r})}{4\left(3\pi^2\right)^{2/3}\rho^{5/3}(\mathbf{r})}.$$

- Regularized kinetic energy variable from the r^2SCAN DFA[14],

$$\alpha(\mathbf{r}) = \frac{\tau(\mathbf{r}) - \tau_{\mathrm{w}}(\mathbf{r})}{3\left(3\pi^2\right)^{2/3}\rho^{5/3}(\mathbf{r})/10 + \eta\tau_{\mathrm{w}}(\mathbf{r})},$$

  where $\eta = 10^{-3}$, and

$$\tau(\mathbf{r}) = \frac{1}{2}\sum_i |\nabla\phi_i(\mathbf{r})|^2 \text{ and } \tau_{\mathrm{w}}(\mathbf{r}) = \frac{|\nabla\rho(\mathbf{r})|^2}{8\rho(\mathbf{r})}$$

  are the kinetic energy density and the von Weizsäcker kinetic energy density.

- Fractional occupation number weighted density[15,16] (FOD) normalized by the density,

$$\rho^{\mathrm{FOD}}(\mathbf{r}) = \frac{1}{\rho(\mathbf{r})}\sum_i(\delta_1 - \delta_2 f_i)|\phi_i(\mathbf{r})|^2,$$

  where δ_1 and δ_2 are chosen such that only fractionally occopied ϕ_i contribute to the sum. The weights f_i are given by the Fermi-Dirac distribution,

$$f_i = \frac{1}{e^{(\varepsilon_i - E_F)/kT} + 1}.$$

  Here, ε_i are orbital energies, E_F is the Fermi energy and $k = 3.166811563 \times 10^{-6}$ is the Boltzmann constant in Hartree per Kelvin. The electronic temperature T is set to $T_1 = 10000K$ or $T_2 = 25000K$, yielding two different FOD features for the neural network, ρ_1^{FOD} and ρ_2^{FOD}.

Additional features for MLS2 (on top of the ones used for ML2):

- Wigner-Seitz radius,

$$r_s(\mathbf{r}) = \left(\frac{3}{4\pi\rho(\mathbf{r})}\right)^{1/3}.$$

- Normalized κMP2 correlation energy density,

$$\tilde{e}_{\mathrm{c}}^{\kappa\mathrm{MP2}}(\mathbf{r}) = \frac{e_{\mathrm{c}}^{\kappa\mathrm{MP2}}(\mathbf{r})}{\rho^{-1/3}(\mathbf{r})e_{\mathrm{x}}(\mathbf{r})},$$

where $e_{\mathrm{x}}(\mathbf{r})$ is the exchange energy density coming from the same gauge as $e_{\mathrm{c}}^{\mathrm{MP2}}(\mathbf{r})$.

- Normalized os-κMP2 correlation energy density,

$$\tilde{e}_{\mathrm{c,os}}^{\kappa\mathrm{MP2}}(\mathbf{r}) = \frac{e_{\mathrm{c,os}}^{\kappa\mathrm{MP2}}(\mathbf{r})}{\rho^{-1/3}(\mathbf{r})e_{\mathrm{x}}(\mathbf{r})}.$$

- Spin polarization (for open-shell systems),

$$\zeta(\mathbf{r}) = \frac{\rho_\alpha(\mathbf{r}) - \rho_\beta(\mathbf{r})}{\rho(\mathbf{r})}, \tag{S12}$$

defined in terms of the the spin densities $\rho_\alpha(\mathbf{r})$ and $\rho_\beta(\mathbf{r})$.

S7. SCALING INVARIANCE OF THE MP2 CORRELATION ENERGY DENSITY

In this section, we derive the scaling invariance of the MP2 correlation energy density defined in Eq. 6 once evaluated on either KS or HF orbitals.

Uniform density scaling[17] scales the density with respect to a scalar $\gamma > 0$ as follows: $\rho_\gamma(\mathbf{r}) = \gamma^3 \rho(\gamma \mathbf{r})$. A direct consequence is the scaling of the corresponding orbitals:

$$\phi[\rho_\gamma](\mathbf{r}) = \gamma^{3/2} \phi[\rho](\gamma \mathbf{r}). \tag{S13}$$

For the scaling invariance of $e_\mathrm{c}^\mathrm{MP2}(\mathbf{r})$, we need to show that the only real-valued exponent p for which

$$e_\mathrm{c}^\mathrm{MP2}[\rho_\gamma](\mathbf{r}) = \gamma^p e_\mathrm{c}^\mathrm{MP2}[\rho](\gamma \mathbf{r})$$

holds, is $p = 0$ (see Eq. 12).

First, we derive the scaling of T_{ijab} from Eq. 7. Its nominator has the same linear scaling ($p = 1$) as the electron-electron interaction density functional[17], while the denominator depends on the scaling of orbital energies (see Ref. 18):

$$\varepsilon_i[\rho_\gamma] = \gamma^2 \varepsilon_i[\rho] \tag{S14}$$

Combining now the scaling of the nominator and of the denominator of Eq. 7 gives

$$\begin{aligned} T_{ijab}[\rho_\gamma] &= \frac{\gamma \langle ij|ab\rangle}{\varepsilon_i[\rho_\gamma] + \varepsilon_j[\rho_\gamma] - \varepsilon_a[\rho_\gamma] - \varepsilon_b[\rho_\gamma]} \\ &= \frac{\gamma \langle ij|ab\rangle}{\gamma^2 \varepsilon_i[\rho] + \gamma^2 \varepsilon_j[\rho] - \gamma^2 \varepsilon_a[\rho] - \gamma^2 \varepsilon_b[\rho]} \\ &= \gamma^{-1} \frac{\langle ij|ab\rangle}{\varepsilon_i[\rho] + \varepsilon_j[\rho] - \varepsilon_a[\rho] - \varepsilon_b[\rho]} \\ &= \gamma^{-1} T_{ijab}[\rho], \end{aligned} \tag{S15}$$

which shows the $p = -1$ scaling property of T_{ijab}.

Next, we derive the scaling of $V_{ijab}(\mathbf{r})$ from Eq. 8 using again the scaling property of orbitals and the transformation of the integral variable:

$$\begin{aligned} V_{ijab}[\rho_\gamma](\mathbf{r}) &= \phi_i[\rho_\gamma](\mathbf{r}) \phi_j[\rho_\gamma](\mathbf{r}) \int \frac{\phi_a[\rho_\gamma](\mathbf{r}') \phi_b[\rho_\gamma](\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \\ &= \gamma^{3/2} \phi_i[\rho](\gamma \mathbf{r}) \gamma^{3/2} \phi_j[\rho](\gamma \mathbf{r}) \int \frac{\gamma^{3/2} \phi_a[\rho](\gamma \mathbf{r}') \gamma^{3/2} \phi_b[\rho](\gamma \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \\ &= \gamma^3 \phi_i[\rho](\gamma \mathbf{r}) \phi_j[\rho](\gamma \mathbf{r}) \int \frac{\phi_a[\rho](\gamma \mathbf{r}') \phi_b[\rho](\gamma \mathbf{r}')}{|\gamma \mathbf{r} - \gamma \mathbf{r}'|} \gamma^4 d\mathbf{r}' \\ &= \gamma^3 \phi_i[\rho](\gamma \mathbf{r}) \phi_j[\rho](\gamma \mathbf{r}) \gamma \int \frac{\phi_a[\rho](\mathbf{s}) \phi_b[\rho](\mathbf{s})}{|\gamma \mathbf{r} - \mathbf{s}|} d\mathbf{s} \\ &= \gamma^4 V_{ijab}[\rho](\gamma \mathbf{r}). \end{aligned} \tag{S16}$$

Finally, combining the results for T_{ijab} ($p = -1$) from Eq. S15 and for $V_{ijab}(\mathbf{r})$ ($p = 4$) from Eq. S16 with the uniform density scaling ($p = 3$), concludes the scaling invariance of the

MP2 correlation energy density:

$$
\begin{aligned}
e_{\mathrm{c}}^{\mathrm{MP2}}[\rho_\gamma](\mathbf{r}) &= -\frac{1}{4\rho_\gamma(\mathbf{r})}\sum_{ijab}\left[(T_{ijab}[\rho_\gamma]\delta_{ia}\delta_{jb} - T_{ijba}[\rho_\gamma]\delta_{ib}\delta_{ja})(V_{ijab}[\rho_\gamma](\mathbf{r})\delta_{ia}\delta_{jb} - V_{ijba}[\rho_\gamma](\mathbf{r})\delta_{ib}\delta_{ja}\right] \\
&= -\frac{1}{4\gamma^3\rho(\gamma\mathbf{r})}\sum_{ijab}\left[(\gamma^{-1}T_{ijab}[\rho]\delta_{ia}\delta_{jb} - \gamma^{-1}T_{ijba}[\rho]\delta_{ib}\delta_{ja})(\gamma^4 V_{ijab}[\rho](\gamma\mathbf{r})\delta_{ia}\delta_{jb}\right. \\
&\quad \left.-\gamma^4 V_{ijba}[\rho](\gamma\mathbf{r})\delta_{ib}\delta_{ja}\right] \\
&= -\frac{1}{4\rho(\gamma\mathbf{r})}\sum_{ijab}\left[(T_{ijab}[\rho]\delta_{ia}\delta_{jb} - T_{ijba}[\rho]\delta_{ib}\delta_{ja})(V_{ijab}[\rho](\gamma\mathbf{r})\delta_{ia}\delta_{jb} - V_{ijba}[\rho](\gamma\mathbf{r})\delta_{ib}\delta_{ja}\right] \\
&= e_{\mathrm{c}}^{\mathrm{MP2}}[\rho](\gamma\mathbf{r}).
\end{aligned}
\tag{S17}
$$

The same scaling invariance also applies to $e_{\mathrm{c}}^{\kappa\mathrm{MP2}}$ and to its spin-resolved components.

S8. LOSS FUNCTIONS FOR ML2 AND MLS2

For the LES-based training of the ML2 neural network, we employ the mean of $\mathcal{L}_{\text{LES}}$ from Eq. 3 scaled by the number of electrons:

$$\text{Mean(LES)} = \frac{1}{M}\sum_{k=1}^{M}\frac{\mathcal{L}_{\text{LES}}^{k}}{N_k}, \tag{S18}$$

where M is the number of systems ($M = 8$ in our case: H^-, He, Be, Mg, Ne, Ar, Ca, and Kr) and N_k indicates the number of electrons of the k-th system with $\mathcal{L}_{\text{LES}}^{k}$ being the corresponding k-th LES.

The GES-based training of the ML2 neural network uses a similar mean of $\mathcal{L}_{\text{GES}}$ from Eq. 2:

$$\text{Mean(GES)} = \frac{1}{M}\sum_{k=1}^{M}\mathcal{L}_{\text{GES}}^{k}, \tag{S19}$$

where $\mathcal{L}_{\text{GES}}^{k}$ is the GES of the k-th system.

In the case of MLS2, we employ a GES combined with the interaction correlation energy error. First, the mean absolute relative error (MArE) is calculated via

$$\text{MArE} = \frac{1}{M}\sum_{k=1}^{M}\frac{\mathcal{L}_{\text{GES}}^{k}}{|E_{\text{c}}^{\text{ref}}[\rho_k]|}, \tag{S20}$$

with the density ρ_k indicating the reference correlation energy value of the k-th system. Second, we evaluate interaction GES ($\mathcal{L}_{\text{int-GES}}$) using reference interaction energies ($E_{\text{c}}^{\text{int-Ref}}$) and employ the corresponding mean absolute relative interaction error (int-MArE) as

$$\text{int-MArE} = \frac{1}{M}\sum_{k=1}^{M}\frac{\mathcal{L}_{\text{int-GES}}^{k}}{|E_{\text{c}}^{\text{int-ref}}[\rho_k]|}. \tag{S21}$$

The final total loss function is the average between MArE from Eq. S20 and int-MArE from Eq. S21.

S9. ADDITIONAL RESULTS FOR ML2

In this section, additional data on validation, convergence, uniqueness and robustness are presented for the GES-based and LES-based ML2 models.

As noted in the main paper, we use eight small closed-shell atoms/ions (H^-, He, Be, Mg, Ne, Ar, Ca, and Kr) to train ML2. For ML2's validation, we select ions of similar size to the training set (Li^+, Ca^{2+}, and Ne^{2+}). Fig. S4 shows that the validation loss closely follows the training loss for LES-based ML2.

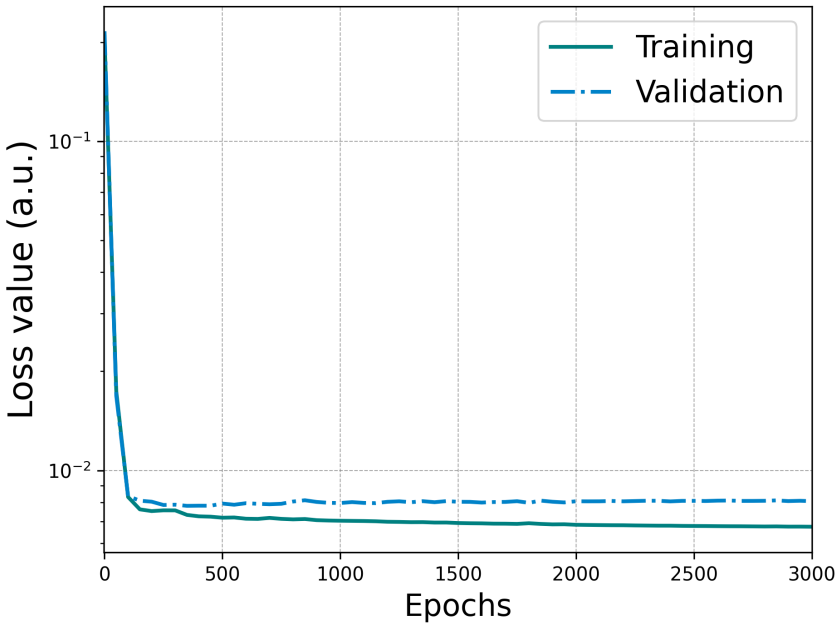

Fig. S4: Training loss (a.u.) of the LES-based ML2 as a function of training steps (epochs). The training dataset is H^-, He, Be, Mg, Ne, Ar, Ca, and Kr. The validation dataset is Li^+, Ca^{2+} and Ne^{2+}.

In Fig. S5, we show the absolute global energy error result for GES-based ML2 as a counterpart to Fig. 3b. Due to the sparse training dataset for the GES, our NN prediction of correlation energies gets worse at larger interatomic distances with increasing training epochs. Furthermore, the energy error is one order of magnitude higher in comparison to the results from the LES-based predictions in Fig. 3b, underscoring again the advantage of LES over GES.

To study the learning process of the ML2 model, we define the following Relative Loss:

$$\text{Relative Loss} \sim \left| \frac{\text{Loss(epochs)}}{\text{Loss(zeroth epoch)}} \right| \times 100, \tag{S22}$$

In Fig. S6, we plot the Relative Loss with respect to training steps (epochs) of the GES-based and LES-based ML2 training. LES shows much faster and smoother convergence compared to the GES counterpart. This is another crucial advantage of LES over GES, in addition to data efficiency.

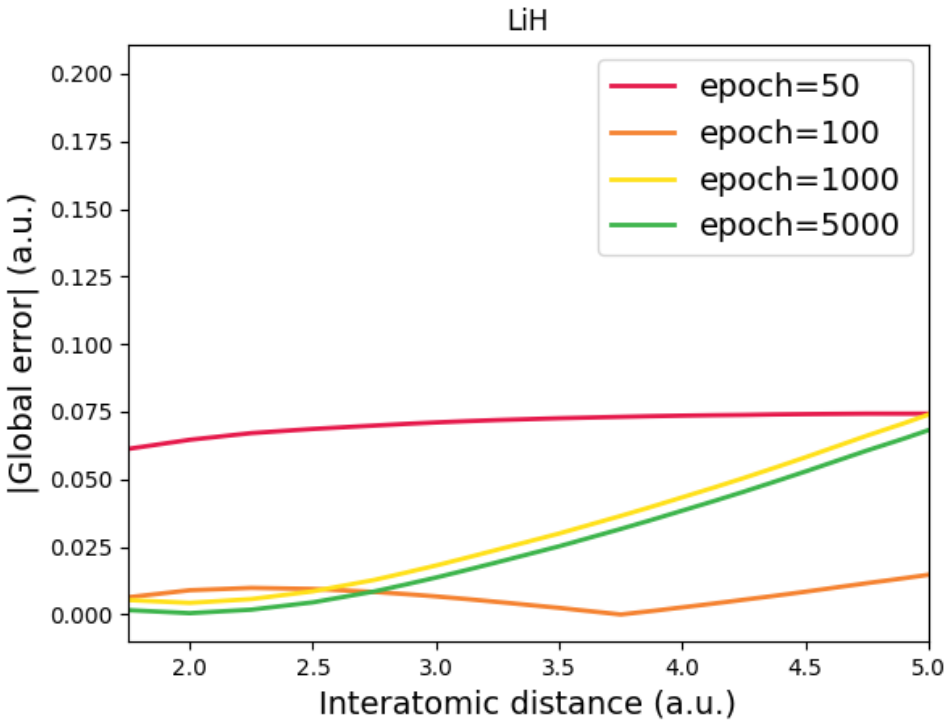

Fig. S5: GES-based ML2 global error as in Fig. 3b

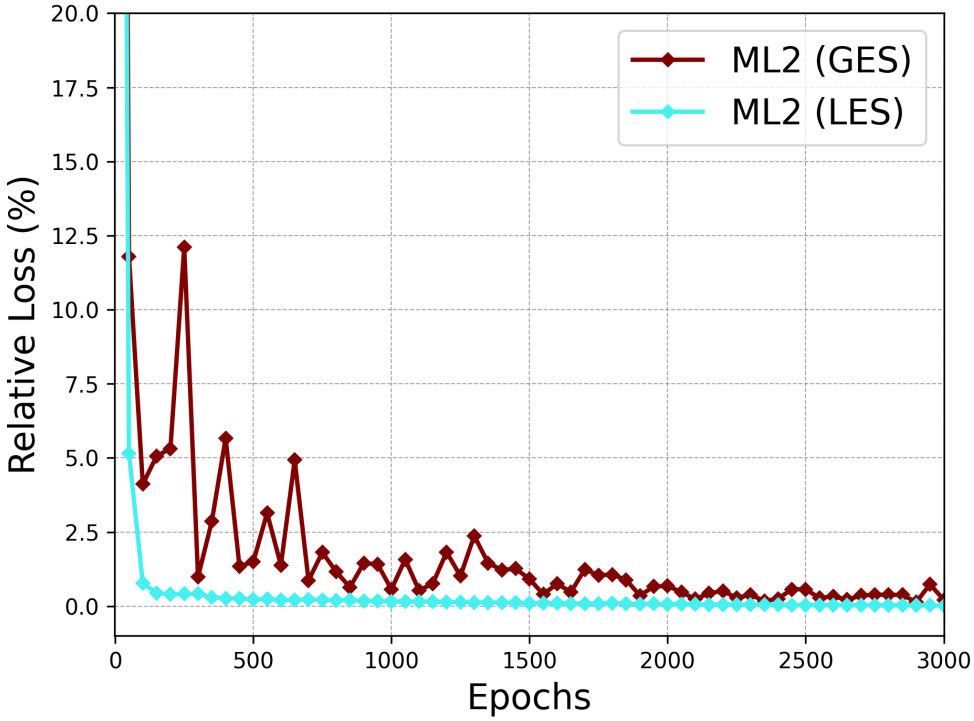

Fig. S6: Relative Loss (see Eq. S22) of ML2 with respect to the initial loss value.

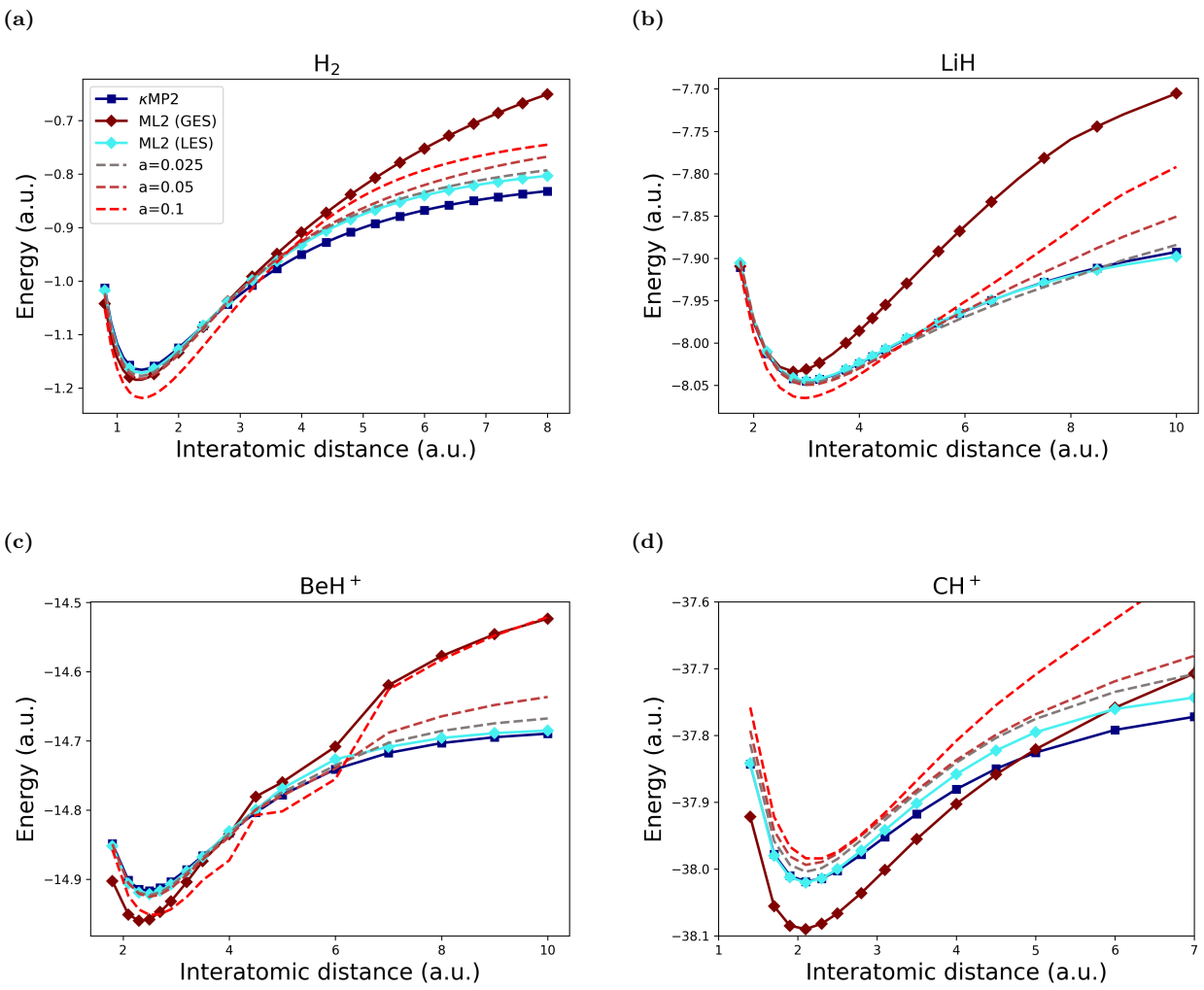

Fig. S7: Dissociation curves as in Fig. 4a for additional diatomic systems. Since the $w_c(\mathbf{r})$ weights from Eq. 13 for $|a| > 0$ are not necessarily bounded between -1 and 1, we learn them as $x(\mathbf{r}) = \tanh\left(\frac{e_c^a(\mathbf{r})}{e_x(\mathbf{r})\rho^{-1/3}(\mathbf{r})}\right)$, retrieving the ML correlation energies by applying $\tanh^{-1}$ to $(x(\mathbf{r}))$, which numerically reduces to ML2 (LES) when $a = 0$.

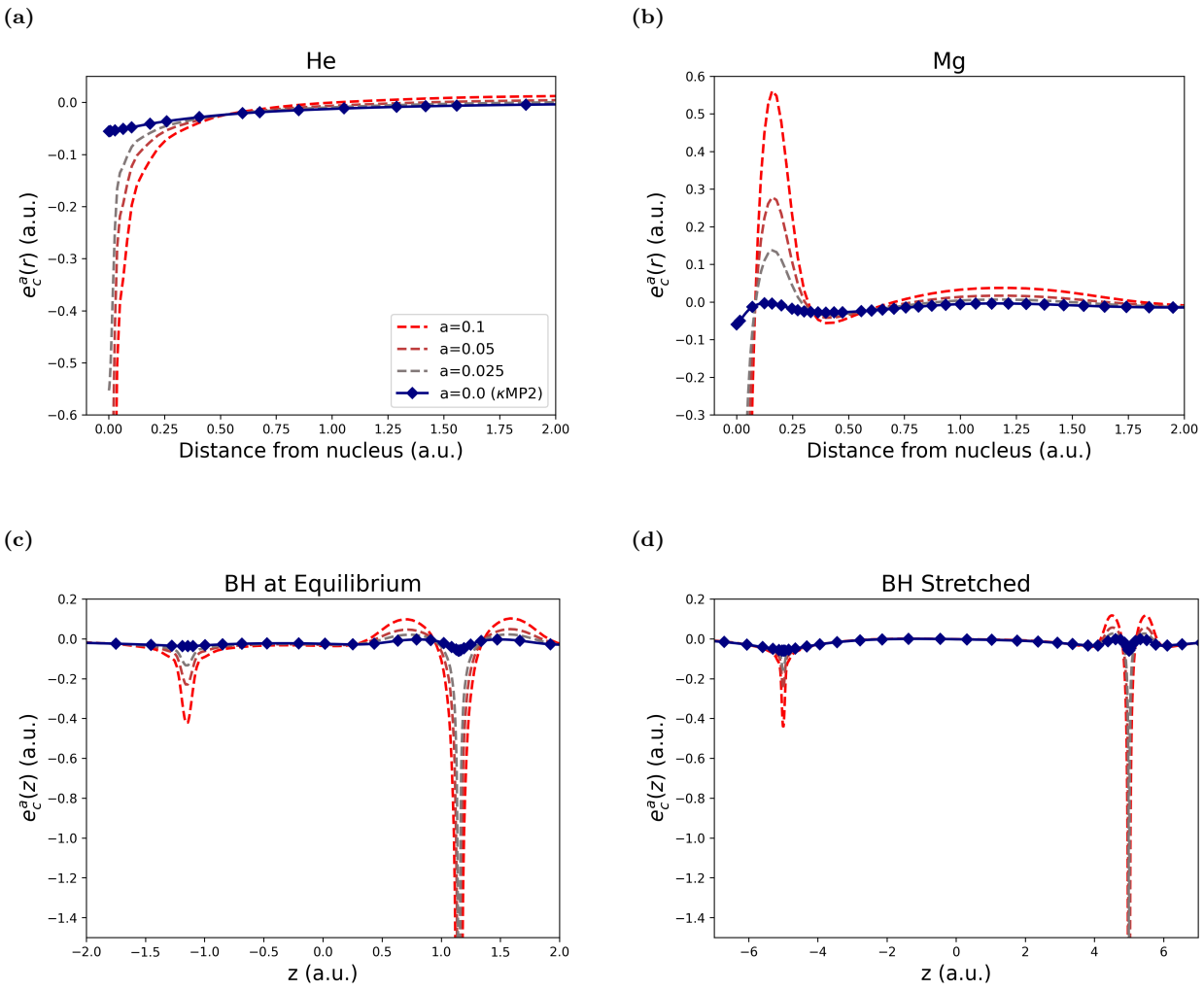

Fig. S8: Gauge transformed correlation energy densites per particle ($e_c^a(r)$ from Eq. (14)). **(a)** and **(b): Atoms used in Training.** $e_c^a(r)$ as in Figs. 1**(c,top)** of the Helium atom and the Magnesium atom for different a values. **(c)** and **(d): Diatomics used in Testing.** $e_c^a(z)$ as in 1**(d,top)** of BH at equilibrium interatomic distance (2.3 a.u.) and stretched (10 a.u.) for different a values.

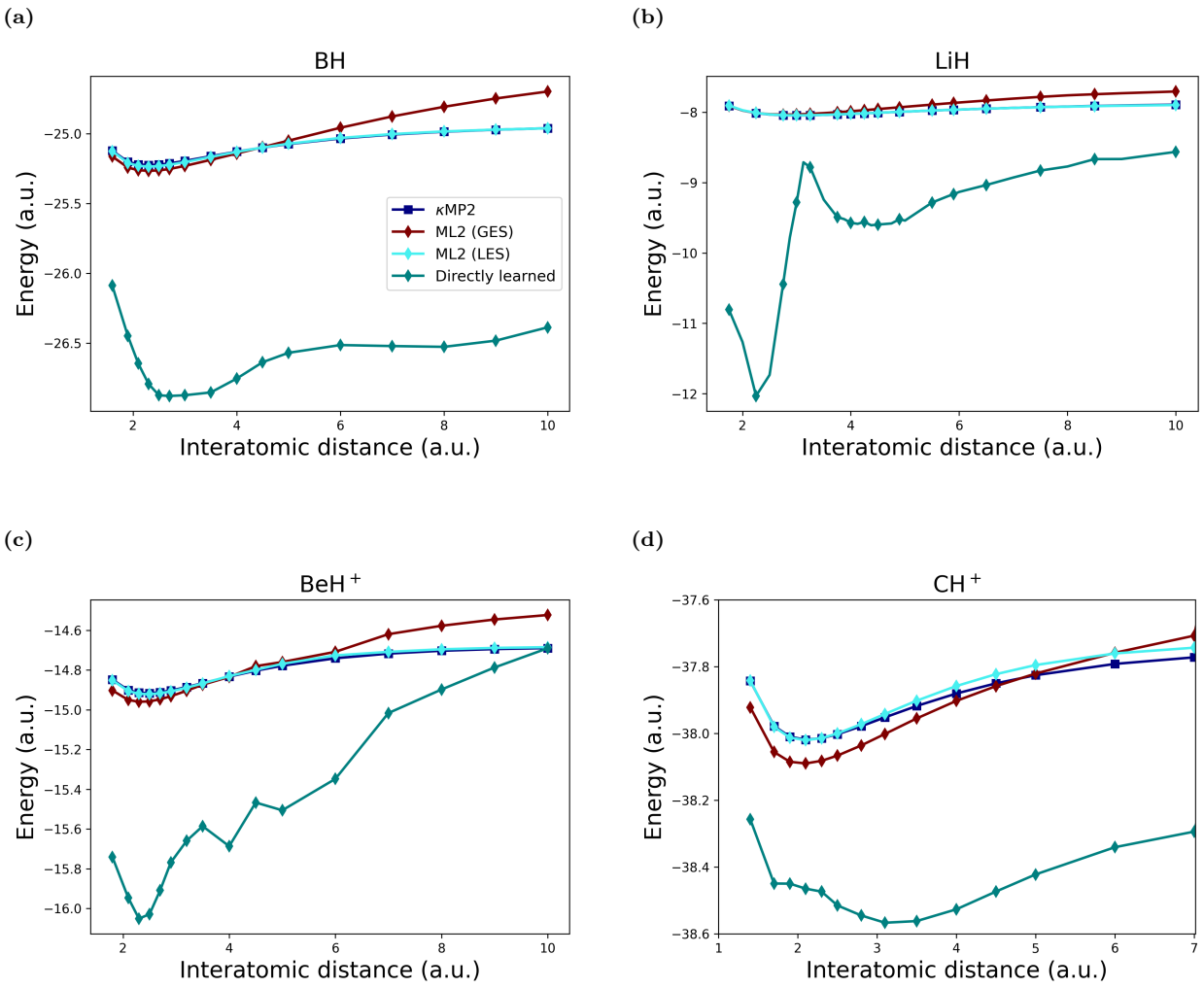

Fig. S9: Dissociation curves as in Fig. 4b for additional diatomic systems. The ML2 model results are based on the GES-learned and LES-learned ML weights from Eq. 13, while the directly learned predictions adopt a 'direct loss' (see the main text) and learn the NN weights via tanh preprocessing (similarly as in Fig. **S7**)

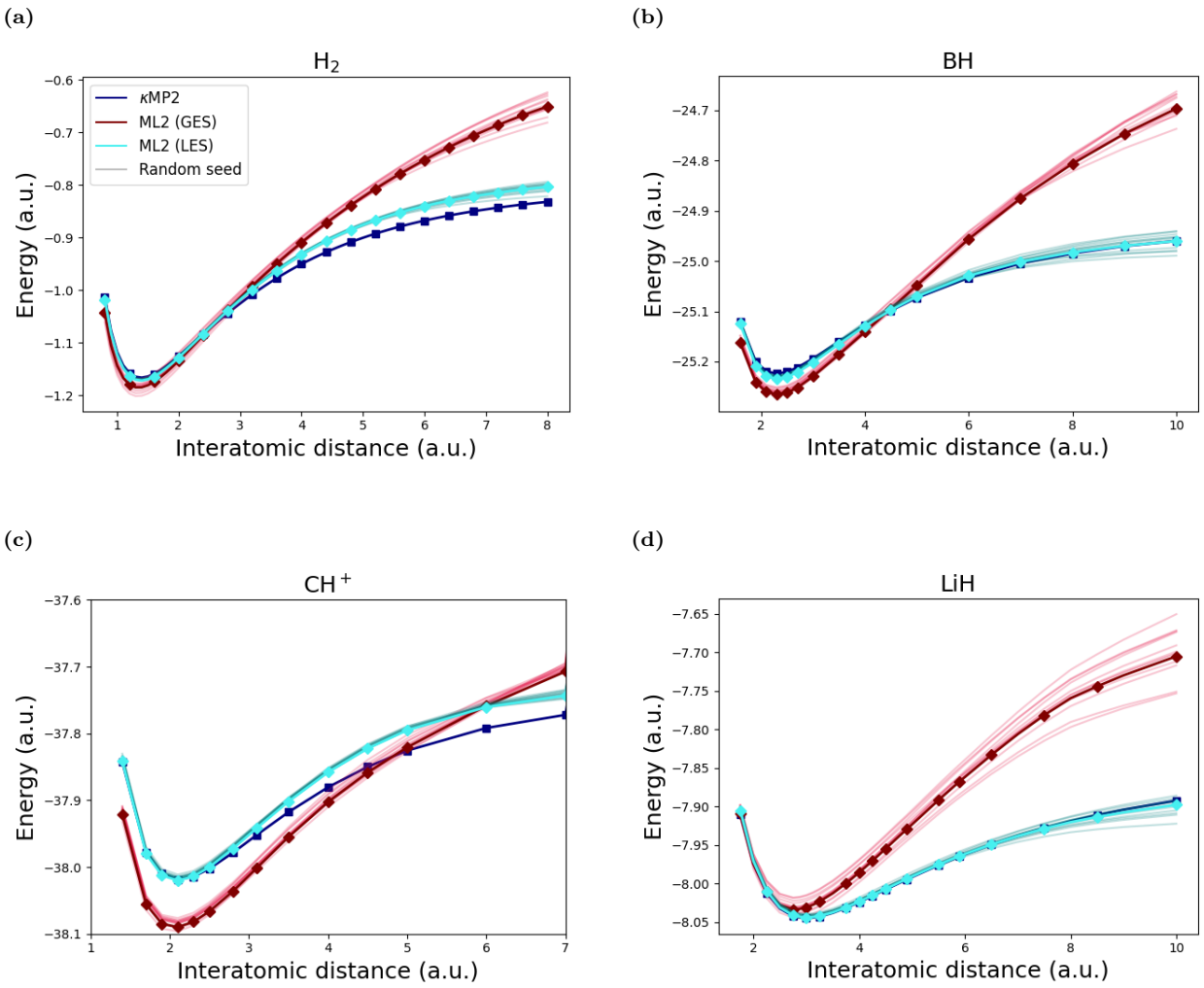

Fig. S10: Dissociation curves as in Fig. 4c for additional diatomic systems.

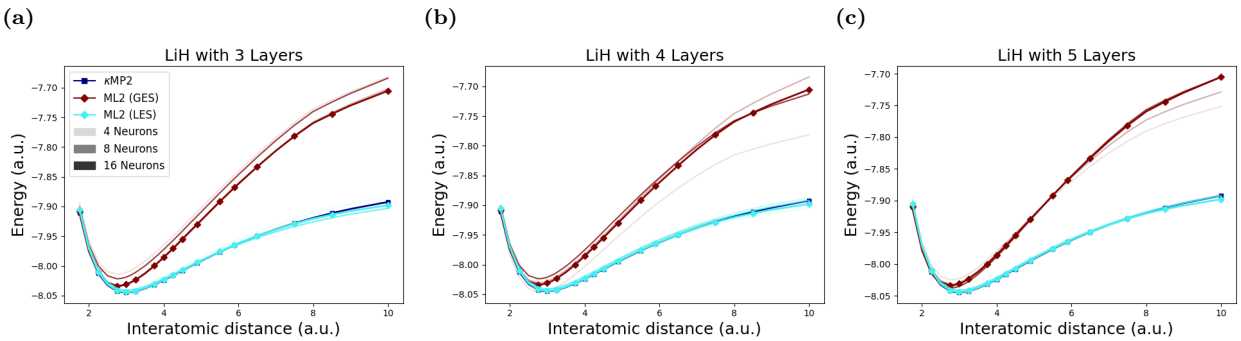

Fig. S11: Dissociation curves as in Fig. 4d with different NN architectures (the number of neurons are given per layer).

In Fig. **S12**, we show dissociation curves for six diatomic systems similar to Fig. **1(c,bottom)**. Here, the GES-based and LES-based ML2 predictions are supplemented by ML2 results coming from a squared LES (see Eq. 3),

$$\mathcal{L}^2_{\text{LES}} \sim \int |e^{\text{ref}}(\mathbf{r}) - e^{\text{ML}}(\mathbf{r})|^2 \rho(\mathbf{r})\, d\mathbf{r}. \tag{S23}$$

We can observe in all plots in Fig. **S12** that the transferability of the LES-based ML2 is not affected by the choice of the loss function.

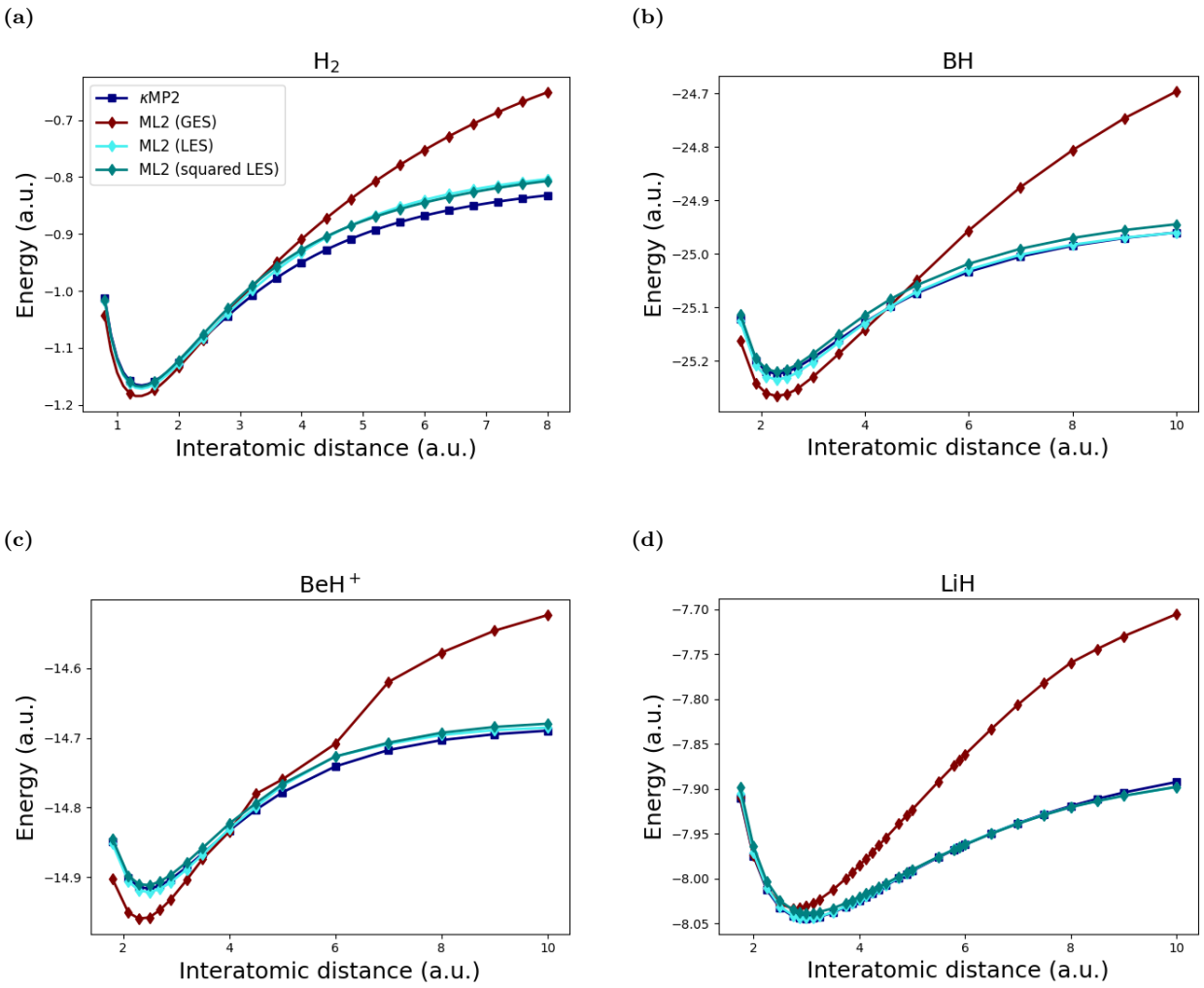

Fig. S12: Dissociation curves as in Fig. 1(c,bottom) for additional diatomic systems. ML2 energy predictions in teal diamonds come from a squared LES training strategy (see Eq. S23).

S10. TRAINING SET SIZE DEPENDENCE ON ML2 TRANSFERABILITY TO DIATOMICS

The dissociation curve of BeH^+ in Fig. **S13a** compares the proxy reference, κMP2 and the GES-based and LES-based ML2 models once more data is progressively added to the training. Each "ML2 (GES)+" curve corresponds to a training set containing of the included eight small closed-shell atoms/ions and seven extra energy data points coming from geometries at and around the equilibrium structure of the labeled diatomic systems. These are (in order) H_2, LiH, Li_2, BH, CH^+, N_2 and BeH^+. This means that at each step (each new addition), the NNs are retrained and applied to dissociation curves. While the final training set and additions at each step may not be large in absolute terms, they are major relative to the test (e.g., for the full BeH^+ dissociation curve test, the final model includes in the training its own BeH^+ points near-equilibrium) and the original atomic training set.

LES-based ML2 trained on atoms already gives highly accurate dissociation curves and shows virtually no changes as more training data is added to LES-based ML2. In contrast, GES-based ML2, even after these extra points are added to the training set, is still outperformed by LES. While GES-based curves are improving after more data are added to the training, one can still observe unphysical discontinuities and "bumps"[19] in some of the GES-based curves. In the final batch, we construct the most data-rich model, which even includes equilibrium BeH^+ datapoints in its own training set. Yet, despite this expanded training, the GES-based ML2 remains insufficiently accurate for stretched BeH^+ geometries, unlike the LES-based approach. Using these same models with increasing training data and applying them to the dissociation curves of the other three dimers shown in Fig. **S13a**, similar trends are observed.

Finally, it should be noted that although the GES-based ML2 employs the global energy loss (Eq. 2) for training, it still uses the ML2 energy density per particle ansatz given by Eq. 13. This ansatz has been constructed based on the scaling constraint (Eq. 12) satisfied by $e_c^{\kappa MP2}(\mathbf{r})$ within our chosen gauge. Consequently, GES-based ML2 is still partially informed by our energy densities per particle gauge, which in turn improves its transferability. Crucially, if instead of this GES that we use here, one employs "raw" GES—i.e., using Eq. 2 to learn energy densities without the ansatz of Eq. 13, the resulting GES transferability would be even worse than shown here.

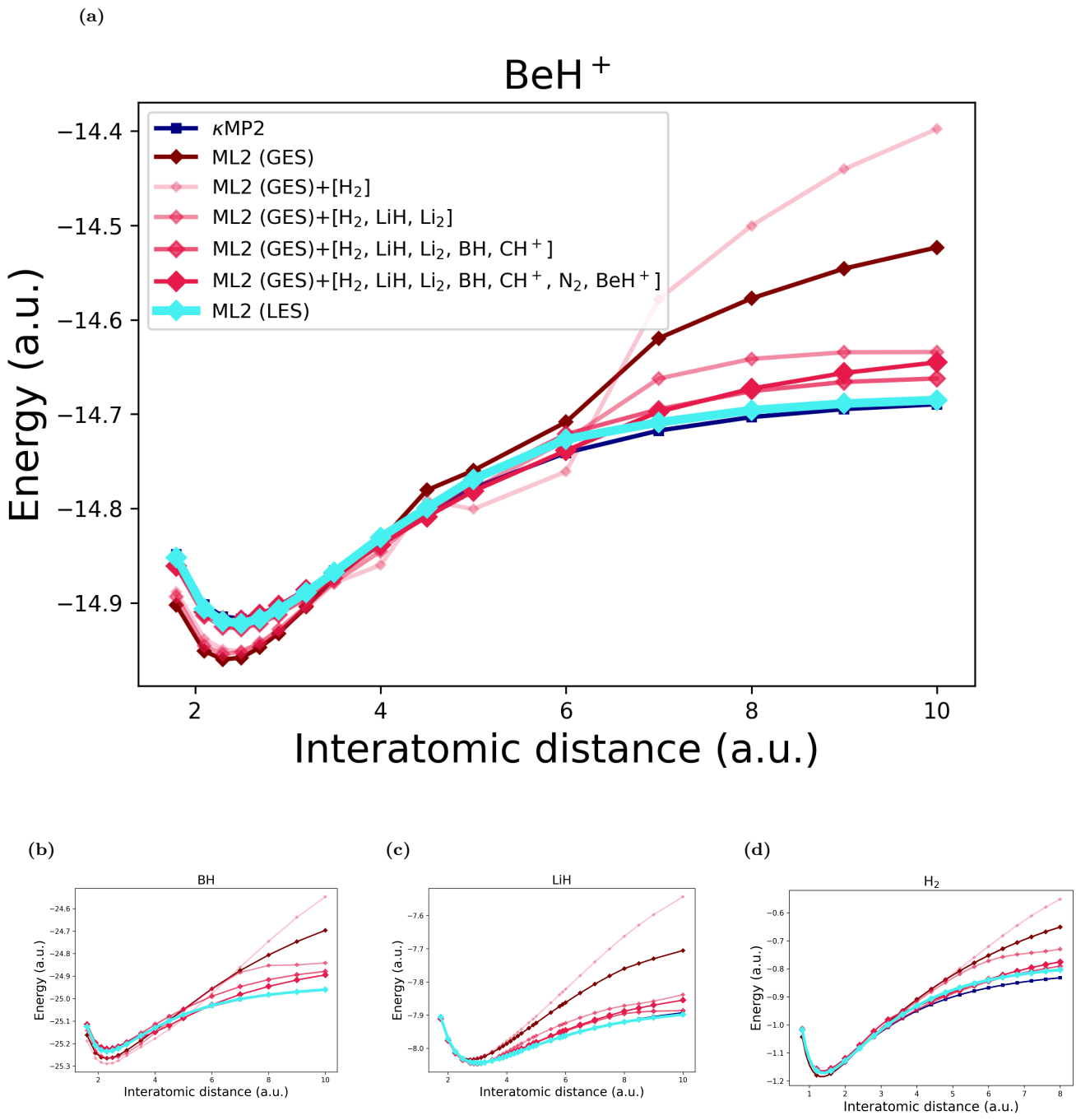

Fig. S13: Dissociation curve of BeH^+ as in Fig. 1(c,bottom). The GES-based ML2 energies are obtained from NN models obtained by progressively increasing training set size. For each labeled diatomic molecule, seven energy training datapoints near its equilibrium geometry are added to the training set.

S11. SCALING OF MLS2 WEIGHTS

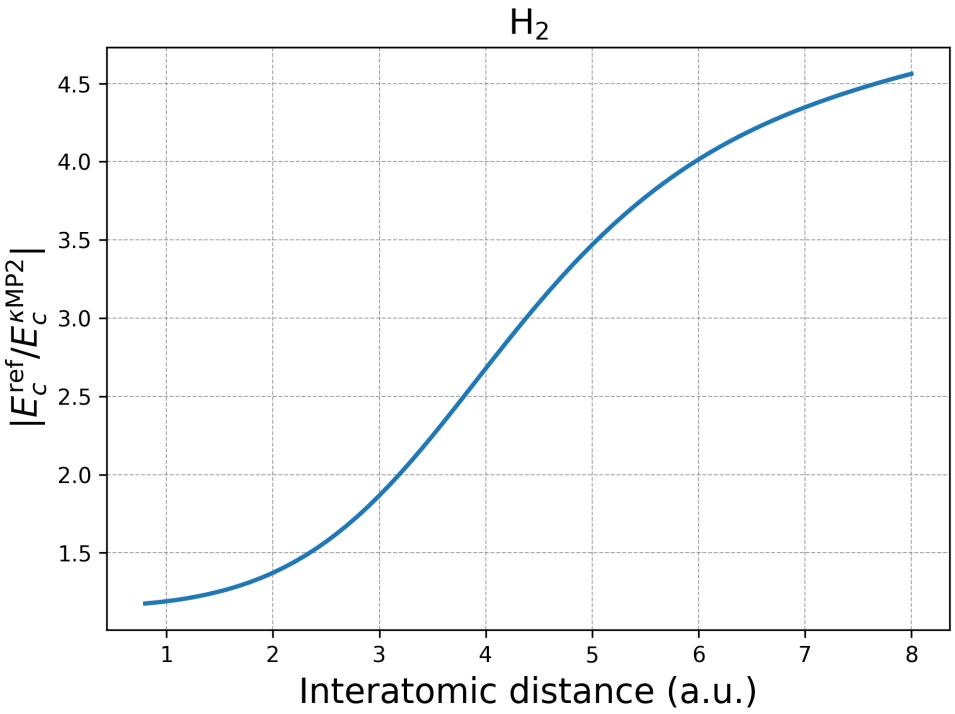

Fig. S14: Absolute ratio between correlation energies. Plotted is the energy ratio for H_2 as a function of interatomic distance.

The scaling of the machine-learned weights in Eq. 15 is important for the ability to predict correlation energies when the true value is much greater than those of κMP2. This situation is particularly pronounced when stretching bonds, see Fig. **S14**, where the ratio between the true and κMP2 correlation energies easily exceeds a factor of 4 once the H_2 bond is stretched. Without this adjustment, the MLS2's correlation energies could never be larger than those of κMP2, as $w_{os}(\mathbf{r})$ and $w_{ss}(\mathbf{r})$ derived from the sigmoid activation function are bound between 0 and 1.

S12. TRAINING AND TEST DATASETS OF MLS2

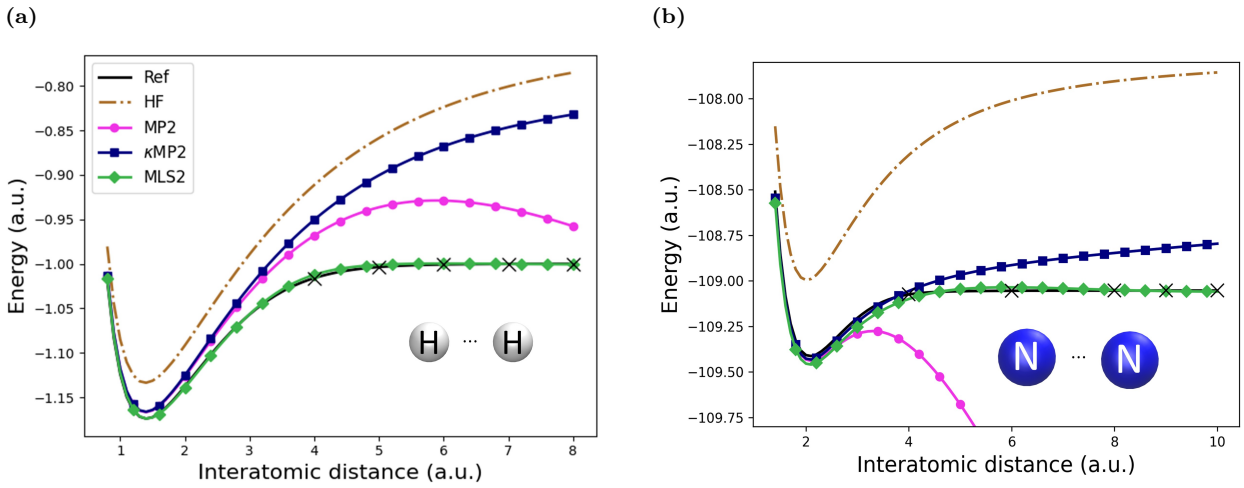

Fig. S15: Dissociation energy curves as in Fig. S1 with additional HF and MLS2 results for comparison. MLS2's energy training data indicated as the cross-marked points on the Ref curve. **(a):** Dissociation energy curve of H_2. **(b):** Dissociation energy curve of N_2.

For correlation energy data, the same eight small closed-shell atoms/ions (H^-, He, Be, Mg, Ne, Ar, Ca, and Kr) as for ML2 are employed. Additionally, the following collection of 13 small closed-shell complexes from the MB16-43 database[20] is incorporated: Cl_2, BeH_2, NaH, H_2, BH_3, AlH_3, MgH_2, N_2, SiH_4, LiH, F_2, CH_4 and P_2. Here, we calculate the reference correlation energies with CCSD(T)/*def2-QZVP* in PySCF[7,8] for structures taken from the original dataset. We further train the MLS2 model on correlation energy data points from H_2 (calculated at full configuration interaction level of theory), N_2 (taken from Ref. 5) and Li_2 (taken from Ref. 21) at five different large interatomic distances, namely (in a.u.) $4, 5, 6, 7$ and 8 for H_2 (equilibrium at 1.4), $4, 6, 8, 9$ and 10 for N_2 (equilibrium at 2.1) and $5, 6, 7, 8.5$ and 10 for Li_2 (equilibrium at 5.05). They are visualized, for instance, as cross-marked points in Fig. **S15**. We also plot the HF, MP2, and κMP2-based results for comparison in Figs. **S15a** and **S15b** respectively. Adding the training data points at large distances helps MLS2 to efficiently bridge the gap between regularized PT2 and true correlation energies.

Finally, the reference data for the interaction energy training is obtained from all 18 dispersion-bonded complexes in the RG18 dataset[22]. We employ the interaction correlation energies for the loss in Eq. S21 coming from the difference of HF interaction energies and the database reference values, which are obtained at the CCSD(T) level of theory using the complete basis set limit[23,24] (CBS).

From the W4-11 database[25], we select for correlation energy testing the following subset of 96 closed-shell systems that are not included in our training set from above: AlF_3, S_4, cis-N_2H_2, HN_3, HONC, trans-N_2H_2, HF, HCL, HCN, ethanol, CF_4, BeF_2, AlH, CS_2, Cl_2O, P_4, HNC, S_2O, CH_2F_2, CH_2NH, BH, CCl_2, N_2H_4, SO_3, F_2O, FCCF, BHF_2, acetic acid, BF, NCCN, PH_3, Be_2, O_3, B_2H_6, SiH_3F, HCOF, AlF, BN, C_2H_6, CO, Si_2H_6, OCS, H_2O, HOCN, CO_2, allene, SO_2, $BeCl_2$, CF_2, propene, acetaldehyde, NH_3, CS, BF_3, CH_3F, NH_2Cl, trans-HCOH, propane, C_2H_5F, SiF_4, C_2H_3F, H_2CO, oxirane, cis-HONO, formic

acid, methanol, $AlCl_3$, S_3, ClCN, CH_2, ClF, HCCF, H_2S, cis-HCOH, HOF, C_2, SiO, C_2H_4, ketene, CH_2C, HCNO, HNO, propyne, oxirene, dioxirane, C_2H_2, CH_3NH_2, trans-HONO, FOOF, N_2O, F_2CO, HNCO, AlCl, HOOH, HOCl and glyoxal. The corresponding reference correlation energies are also obtained with CCSD(T)/*def2-QZVP* in PySCF[7,8].

The BH test data is taken from Ref. [26] and employed similar to the training reference correlation data points of the stretched diatomic systems. Lastly, the interaction energies of the formic acid dimer, and other systems of Fig. **S16** from the S22x5 database[27], are calculated analogously to RG18 at the CCSD(T)/CBS level of theory.

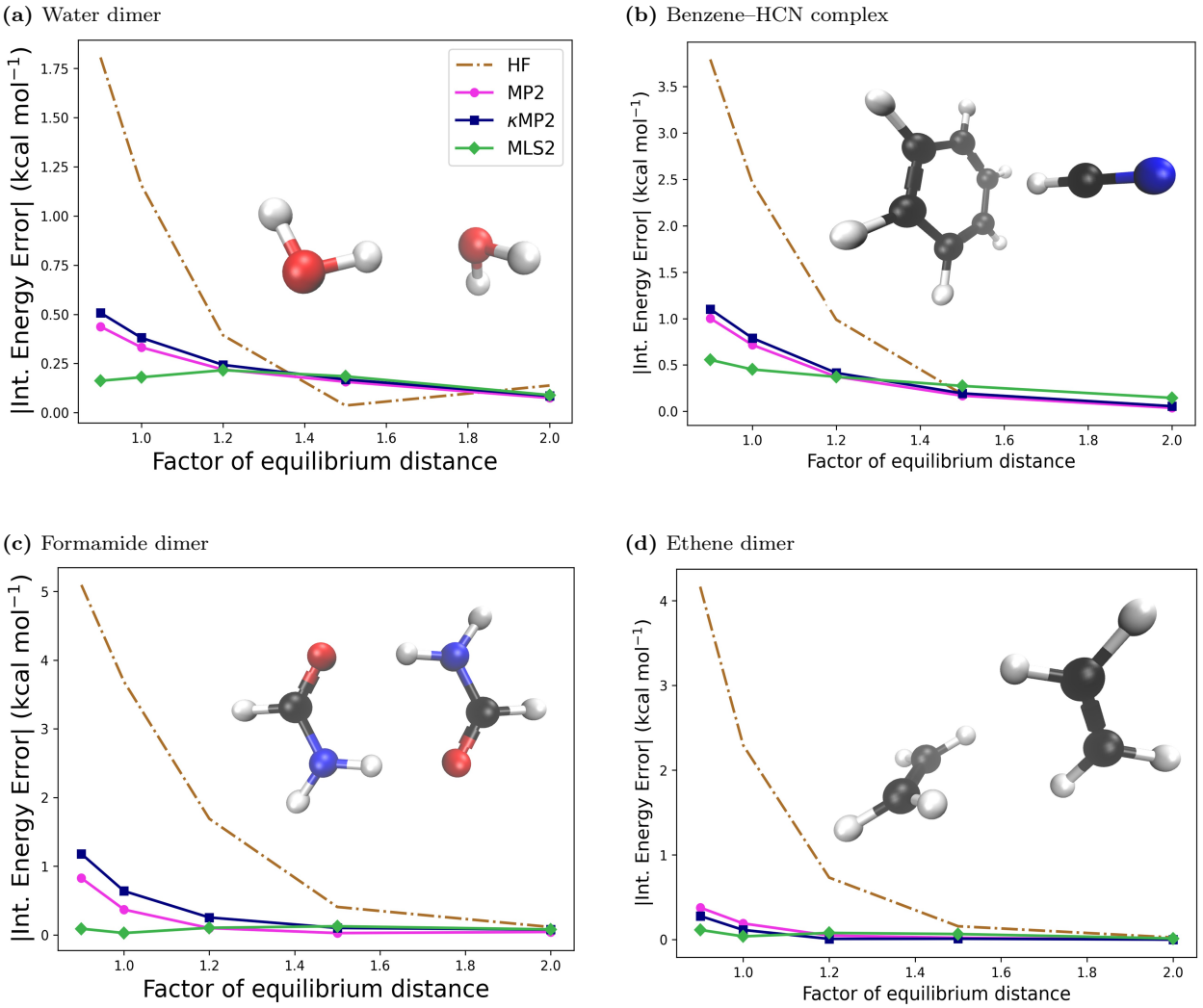

Fig. S16: Interaction (Int.) energy error curves similar to Fig. 1(e,top). Reference and geometries taken from the S22x5 database[27].

S13. TRAINING AND TEST DATASETS OF MLS2@W4

For the second MLS2 extrapolation test [MLS2@W4], we employed atomization energies from the W4-11 database[25]. For strictly positive features $r_s(\mathbf{r})$ and $s(\mathbf{r})$ (see Sec. S6), we use a $\log(\tanh(\cdot))$ preprocessing. Since the training data consist of atomization energies (interaction), we used the MAE of an interaction-based GES (see Eq. S19) for the training. With 140 training datapoints from W4-11, a randomly chosen subset of 10% (14 atomizations) was used for validation. The MAE calculation for all methods in Fig. 6 considered the following respective reaction energy datapoints: W4-11RE has 11247 total reactions[28], from which 4134 were determined to be SC, SIE4x4 has 16 reactions[29], from which 2 were determined to be SC, BH76 has 76 reactions[30], from which 13 were determined to be SC, BH76RC has 30 reactions[30], from which 1 was determined to be SC and RG18 has 18 reactions[22], from which none was determined to be SC. For the DHs (ωB97M(2)[31] and revDSD-PBEP86-D4[32]), we used total energies from Ref. 32 to compute the MAEs for the individual datasets.

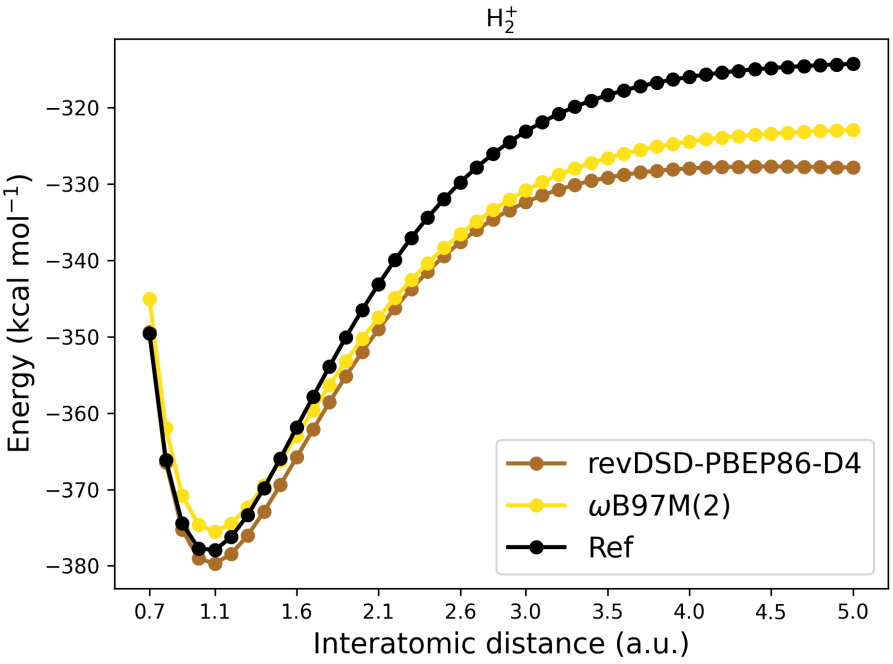

Fig. S17: Dissociation energy curve of H_2^+. Shown are the energy results for ωB97M(2)[31], revDSD-PBEP86-D4[32] and the exact reference (HF = MLS2 = κMP2 = MP2) – all evaluated with QChem[33] 6.2.2.

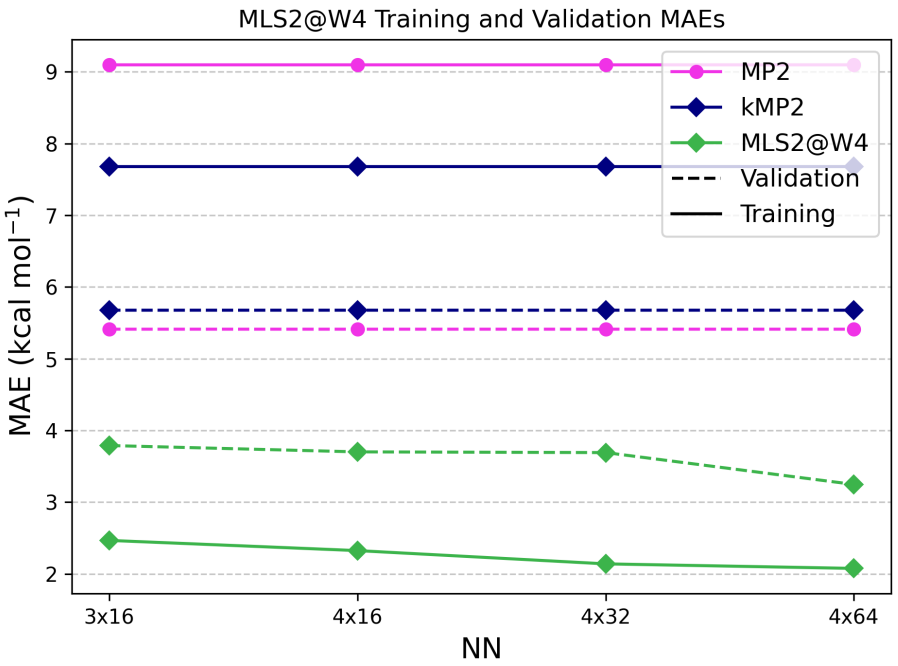

Fig. S18: Energy MAEs (kcal mol^{-1}) as in Fig. 6 for validation and training at different NN architectures of the MLS2@W4 model. Each NN employs the same training and validation subsets from W4-11, as detailed in Sec. **S13**. The MP2 and κMP2 results are shown for comparison.

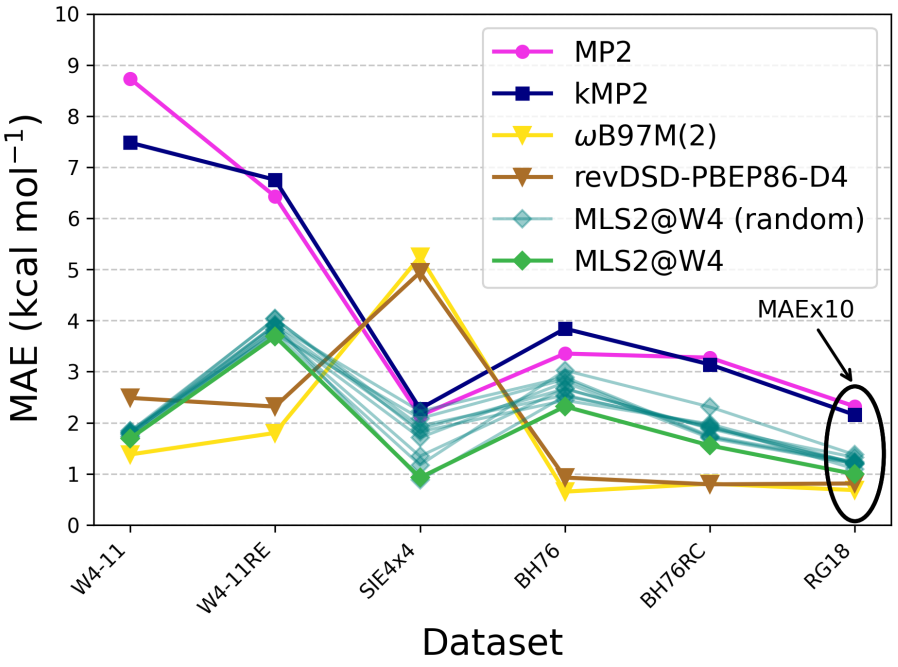

Fig. S19: Energy MAEs (kcal mol^{-1}) as in Fig. 6 with results from multiple ML runs. MAEs of MLS2@W4 from 10 ML runs with different seed initializations (random). MLS2@W4 corresponds to the ML run with the lowest MAE for W4-11.

S14. ADDITIONAL PLOTS FOR ML(S)2 TRAINING DETAILS

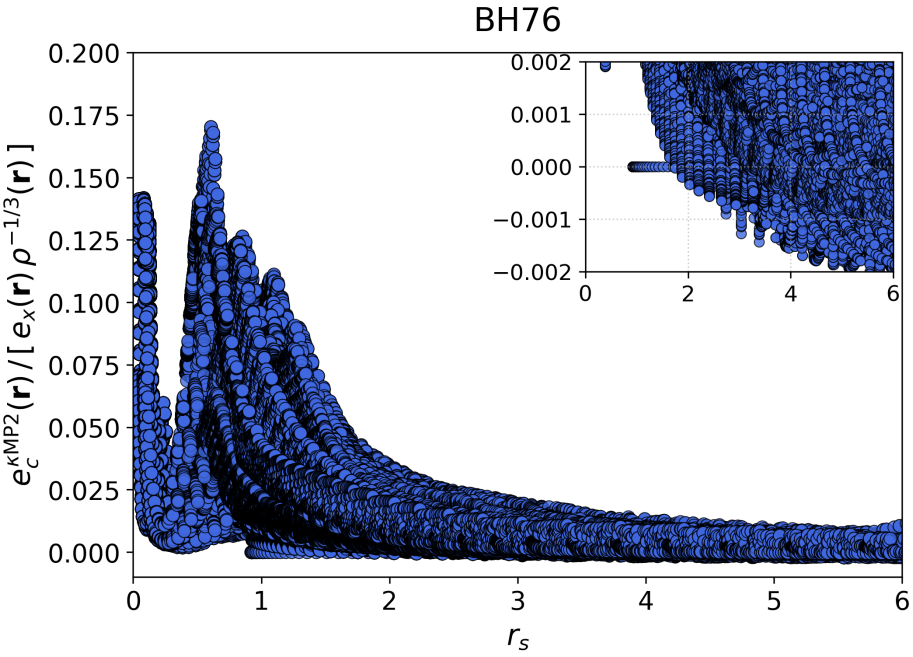

Fig. S20: Scatter plot of "exact" LES-based ML2 $w_c(\mathbf{r})$ weights, obtained by inverting Eq. 13, illustrating the range of relevant w_c values by employing the BH76 set. The weights are shown as a function of r_s to highlight energetically important regions ($0 \leq r_s \leq 6$), even though r_s is not used as an input in ML2, for reasons discussed in the main text. The observed range of $w_c(\mathbf{r})$ values, mostly between 0 and 0.2, with some slightly negative values close to zero shown in the inset, motivates the choice of a bounded (tanh) activation function, whose application to the final output layer, with the resulting $(-1, 1)$ output range, readily contains the observed $w_c(\mathbf{r})$ range.

REFERENCES

[1] K. J. Daas, H. Zhao, E. Polak, and S. Vuckovic, Journal of Chemical Theory and Computation **21**, 5501 (2025).

[2] K. J. Daas, J. Grossi, S. Vuckovic, Z. H. Musslimani, D. P. Kooi, M. Seidl, K. J. H. Giesbertz, and P. Gori-Giorgi, The Journal of Chemical Physics **153**, 10.1063/5.0029084 (2020).

[3] J. Lee and M. Head-Gordon, Journal of Chemical Theory and Computation **14**, 5203 (2018).

[4] J. Shee, M. Loipersberger, A. Rettig, J. Lee, and M. Head-Gordon, The Journal of Physical Chemistry Letters **12**, 12084 (2021).

[5] D. Kats and D. P. Tew, Journal of Chemical Theory and Computation **15**, 13 (2019).

[6] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual* (CreateSpace, Scotts Valley, CA, 2009).

[7]Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K. Chan, WIREs Computational Molecular Science **8**, e1340 (2018).

[8]Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui, J. J. Eriksen, Y. Gao, S. Guo, J. Hermann, M. R. Hermes, K. Koh, P. Koval, S. Lehtola, Z. Li, J. Liu, N. Mardirossian, J. D. McClain, M. Motta, B. Mussard, H. Q. Pham, A. Pulkin, W. Purwanto, P. J. Robinson, E. Ronca, E. R. Sayfutyarova, M. Scheurer, H. F. Schurkus, J. E. T. Smith, C. Sun, S.-N. Sun, S. Upadhyay, L. K. Wagner, X. Wang, A. White, J. D. Whitfield, M. J. Williamson, S. Wouters, J. Yang, J. M. Yu, T. Zhu, T. C. Berkelbach, S. Sharma, A. Y. Sokolov, and G. K.-L. Chan, The Journal of Chemical Physics **153**, 024109 (2020).

[9]B. I. Dunlap, Physical Chemistry Chemical Physics **2**, 2113 (2000).

[10]H.-J. Werner, F. R. Manby, and P. J. Knowles, The Journal of Chemical Physics **118**, 8149 (2003).

[11]V. I. Lebedev and D. N. Laikov, Doklady Mathematics **59**, 477 (1999).

[12]R. Hellmann, E. Bich, and E. Vogel, Molecular Physics **106**, 133 (2008), equilibrium distance for Ne_2 at 5.585 Bohr.

[13]P. Jurečka, J. Šponer, J. Černý, and P. Hobza, Phys. Chem. Chem. Phys. **8**, 1985 (2006).

[14]J. W. Furness, A. D. Kaplan, J. Ning, J. P. Perdew, and J. Sun, The Journal of Physical Chemistry Letters **11**, 8208 (2020).

[15]S. Grimme and A. Hansen, Angewandte Chemie International Edition **54**, 12308 (2015).

[16]C. A. Bauer, A. Hansen, and S. Grimme, Chemistry - A European Journal **23**, 6150 (2017).

[17]M. Levy and J. P. Perdew, Physical Review A **32**, 2010 (1985).

[18]R. V. Leeuwen, Kohn-sham potentials in density functional theory (1994).

[19]S. Vuckovic, T. J. Irons, A. Savin, A. M. Teale, and P. Gori-Giorgi, Journal of chemical theory and computation **12**, 2598 (2016).

[20]M. Korth and S. Grimme, Journal of Chemical Theory and Computation **5**, 993 (2009), mB16-43 benchmark.

[21]G. C. Lie and E. Clementi, The Journal of Chemical Physics **60**, 1288 (1974).

[22]S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, The Journal of Chemical Physics **132**, 154104 (2010).

[23]A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, and A. K. Wilson, Chemical Physics Letters **286**, 243 (1998).

[24]A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, and J. Olsen, Chemical Physics Letters **302**, 437 (1999).

[25]A. Karton, S. Daon, and J. M. Martin, Chemical Physics Letters **510**, 165 (2011).

[26]Ł. M. Mentel, R. van Meer, O. V. Gritsenko, and E. J. Baerends, The Journal of Chemical Physics **140**, 214105 (2014).

[27]L. Gráfová, M. Pitoňák, J. Řezáč, and P. Hobza, Journal of Chemical Theory and Computation **6**, 2365 (2010).

[28]J. T. Margraf, D. S. Ranasinghe, and R. J. Bartlett, Physical Chemistry Chemical Physics **19**, 9798 (2017).

[29]L. Goerigk and S. Grimme, Journal of Chemical Theory and Computation **6**, 107 (2010).

[30]L. Görigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, and S. Grimme, Physical Chemistry Chemical Physics **19**, 32184 (2017).

[31]N. Mardirossian and M. Head-Gordon, The Journal of Chemical Physics **148**, 10.1063/1.5025226 (2018).

[32]G. Santra, M. Cho, and J. M. L. Martin, The Journal of Physical Chemistry A **125**, 4614 (2021).
[33]Y. Shao, Z. Gan, E. Epifanovsky, A. T. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kuś, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. Woodcock, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. O. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. DiStasio, H. Do, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. Hanson-Heine, P. H. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyaev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. D. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S.-P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, E. Neuscamman, C. M. Oana, R. Olivares-Amaya, D. P. O'Neill, J. A. Parkhill, T. M. Perrine, R. Peverati, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, S. M. Sharada, S. Sharma, D. W. Small, A. Sodt, T. Stein, D. Stück, Y.-C. Su, A. J. Thom, T. Tsuchimochi, V. Vanovschi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, J. Yang, S. Yeganeh, S. R. Yost, Z.-Q. You, I. Y. Zhang, X. Zhang, Y. Zhao, B. R. Brooks, G. K. Chan, D. M. Chipman, C. J. Cramer, W. A. Goddard, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. Schaefer, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xu, A. Aspuru-Guzik, R. Baer, A. T. Bell, N. A. Besley, J.-D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. V. Voorhis, J. M. Herbert, A. I. Krylov, P. M. Gill, and M. Head-Gordon, Molecular Physics **113**, 184 (2015).