Supporting Information for "A symmetry-preserving and transferable representation for learning the Kohn-Sham density matrix"

Liwei Zhang[1], Patrizia Mazzeo[2], Michele Nottoli[3], Edoardo Cignoni[2], Lorenzo Cupellini[2], and Benjamin Stamm[3]

[1]Institut für Geometrie und Praktische Mathematik, RWTH Aachen University, Templergraben 55, 52062 Aachen, Germany
[2]Dipartimento di Chimica e Chimica Industriale, Università di Pisa, Via G. Moruzzi 13, 56124 Pisa, Italy
[3]Universität Stuttgart, Institute of Applied Analysis and Numerical Simulation, Pfaffenwaldring 57, 70569 Stuttgart, Germany

S1 Equivariance of the density matrix

In this appendix, we discuss the equivariance of the density matrix and the rationality of decomposing it in the way that is illustrated in Figure 1(b,c,d).

Given a configuration $\mathbf{R}$, and a set of atomic orbitals $\{\phi_{I\alpha}\}_{I,\alpha}$ with which (2) is discretized, where I ranges from 1 to N_{at} and $\alpha \in \mathcal{I}_{Z_I}$, an index set depending only on Z_I. It is straightforward to see that $N_g = \sum_{I=1}^{N_{\text{at}}} \#\mathcal{I}_{Z_I}$ with $\#A$ denoting the cardinality of the set A. After the discretization, the solution of (2) can be used to approximate the KS orbitals, as

$$\varphi_k(\boldsymbol{r};\mathbf{R}) = \sum_{(I,\alpha)} c_{k,(I,\alpha)}[\mathbf{R}]\phi_{I\alpha}(\boldsymbol{r}), \quad k = 1, 2, \ldots, N.$$

where the collection of all $\{c_{k,(I,\alpha)}[\mathbf{R}]\}$ form exactly the eigenvectors $C_{\mathbf{R}}$ in (2). Define the density matrix operator by

$$D(\boldsymbol{r}, \boldsymbol{r}';\mathbf{R}) = \sum_{k=1}^{N} \varphi_k(\boldsymbol{r};\mathbf{R})^* \varphi_k(\boldsymbol{r}';\mathbf{R}),$$

and denote its diagonal by

$$D(\boldsymbol{r};\mathbf{R}) = D(\boldsymbol{r}, \boldsymbol{r};\mathbf{R}).$$

Then, each sub-block of the (discretized) density matrix is given by

$$\begin{aligned}
\big(D_{\mathbf{R}}\big)_{IJ\alpha\beta} &= \int_{\mathbb{R}^3} \phi_{I\alpha}(\boldsymbol{r})^* D(\boldsymbol{r};\mathbf{R})\phi_{J\beta}(\boldsymbol{r})d\boldsymbol{r} \\
&= \int_{\mathbb{R}^3} \phi_{\alpha}(\boldsymbol{r}-\boldsymbol{r}_I)^* D(\boldsymbol{r};\mathbf{R})\phi_{\beta}(\boldsymbol{r}-\boldsymbol{r}_J)d\boldsymbol{r} \\
&= \int_{\mathbb{R}^3} \phi_{\alpha}(\boldsymbol{r}-\boldsymbol{r}_I)^* \tilde{D}(\boldsymbol{r};\mathbf{R}_{IJ})\phi_{\beta}(\boldsymbol{r}-\boldsymbol{r}_J)d\boldsymbol{r}, \quad 1 \le I, J \le N_{\text{at}},\ \alpha \in \mathcal{I}_{Z_I},\ \beta \in \mathcal{I}_{Z_J}.
\end{aligned} \tag{S1}$$

In the last line, the whole configuration $\mathbf{R}$ is shifted to be centred at a proper position, which will be elaborated in more detail in Section S2 and $\tilde{D}$ is simply the translated density matrix operator. Note that for each orbital index α, there are in fact $2l_\alpha + 1$ orbitals assigned, where l_α is the index indicating the type of the orbital (*e.g.* for the $\alpha = s, p, d, \ldots$, $l_\alpha = 0, 1, 2, \ldots$), hence makes $\big(D_{\mathbf{R}}\big)_{IJ\alpha\beta}$ a $(2l_\alpha + 1) \times (2l_\beta + 1)$ matrix. Additionally, (S1) indicates that the sub-blocks of the density matrix $D_{\mathbf{R}}$ share the same function form as long as they correspond to the same interaction of atoms and the same orbital shells. We follow an extremely similar discussion as in Ref. [1] to obtain the equivariance of the density matrix. More specifically, there holds

$$D_{Q\mathbf{R}} = \boldsymbol{\mathcal{D}}(Q) D_{\mathbf{R}} \boldsymbol{\mathcal{D}}(Q)^*, \quad \forall Q \in \mathrm{O}(3), \tag{S2}$$

where $\boldsymbol{\mathcal{D}}(Q) = \mathrm{Diag}(\{\mathcal{D}^{l_\alpha}(Q)\}_{\alpha \in \mathcal{I}_{Z_I}, I \in \{1,2,\ldots,N_{\text{at}}\}})$. Or equivalently, we may write

$$\big(D_{Q\mathbf{R}}\big)_{IJ\alpha\beta} = \mathcal{D}^{l_\alpha}(Q)\big(D_{\mathbf{R}}\big)_{IJ\alpha\beta}\mathcal{D}^{l_\beta}(Q)^*, \quad \forall (\alpha,\beta) \in \mathcal{I}_{Z_I} \times \mathcal{I}_{Z_J},\ Q \in \mathrm{O}(3), \tag{S3}$$

which is the smallest unit of describing the isometric symmetry of the density matrix.

S2 The construction of the ACE basis

Despite the existence of some other possible ways of construction, all the equivariant ACE bases can be obtained through the following procedure:

$$\text{1-particle basis} \to \text{density projection} \to \nu\text{-body correlation} \to \text{symmetrization}.$$

The main differences that distinguish the ACE bases are the construction of the 1-particle basis and the symmetrization, while the latter remains the same throughout this paper (*cf.* (S2) or (S3)). The focus is thus on the definition of the 1-particle basis. The 1-particle basis is, as its name suggests, a function applied to a single particle σ, which we will specify in detail for both the *onsite* and *offsite* cases, respectively, for completeness. Suppose the configuration is given by $\mathbf{R} = \{\sigma_I\}_{I=1,2,\dots,N_{\mathrm{at}}}$, with $\sigma_I = (Z_I, \boldsymbol{r}_I)$.

***Onsite* basis:** the onsite environment $\mathbf{R}_I$ centering at the I-th atom is defined as

$$\mathbf{R}_I = \{(Z_K, Z_I, \boldsymbol{r}_{KI})\} =: \{\sigma_K\}_{K=1,2,\dots,N_{\mathrm{at}},K\neq I},$$

where $\boldsymbol{r}_{KI} = \boldsymbol{r}_K - \boldsymbol{r}_I$. The second variable in σ_K indicates the center of the system. Given a particle $\sigma_K = (Z_K, Z_I, \boldsymbol{r}) \in \mathbf{R}_I$, we define the 1-particle basis for the *onsite* Z_I model as

$$\phi_v^{\mathrm{on}}(\sigma_K) := \phi_{Znlm}^{\mathrm{on}}(\sigma_K) := \delta_{ZZ_K} P_{nl}(r) Y_{lm}(\hat{\boldsymbol{r}}) f_{\mathrm{cut}}(r)$$

where δ_{ZZ_K} is the kronecker delta, $\boldsymbol{r} = r \cdot \hat{\boldsymbol{r}}$ with $r = |\boldsymbol{r}|$, and we have identified the composite index $v \equiv (Znlm)$ with Z standing for some atomic number. The radial cutoff function $f_{\mathrm{cut}}(r)$ ensures that only the nearby atoms are taken into account (*cf.* Figure 1(e)), which may take different forms. In this work, we choose

$$f_{\mathrm{cut}}(r; r_{\mathrm{cut,Z_I}}) = \begin{cases} (r^2/r_{\mathrm{cut,Z_I}}^2 - 1)^2, & r \le r_{\mathrm{cut,Z_I}}, \\ 0, & r > r_{\mathrm{cut,Z_I}}. \end{cases} \tag{S4}$$

Here, the subscript Z_I indicates that the cutoff radius can be made element-dependent, and will sometimes be neglected for simplicity.

***Offsite* basis:** For the *offsite* interactions, we define the offsite local environment as

$$\mathbf{R}_{IJ} = \{\sigma_{IJ}\} \cup \{\sigma_K\}_{K=1,2,\dots,N_{\mathrm{at}},K\neq I,J},$$

where $\sigma_{IJ} = \{(Z_I, Z_J), \boldsymbol{r}_{IJ}\}$, $\sigma_K = \{Z_K, (Z_I, Z_J), \boldsymbol{r}_{IJ,K}\}$ and

$$\boldsymbol{r}_{IJ,K} = \boldsymbol{r}_K - \boldsymbol{r}_{IJ,\theta}, \text{ with } \boldsymbol{r}_{IJ,\theta} = \boldsymbol{r}_J + \theta(\boldsymbol{r}_I - \boldsymbol{r}_J),\ \theta \in [0,1].$$

The 1-particle basis is then defined as

$$\begin{aligned} \phi_{nlm}^{\mathrm{b}}(\sigma_{IJ}) &= P_{nl}(r_{IJ}) Y_{lm}(\hat{\boldsymbol{r}}_{IJ}) f_{\mathrm{cut}}^{\mathrm{b}}(r_{IJ}), \\ \phi_{Znlm}^{\mathrm{e}}(\sigma_K) &= \delta_{ZZ_K} P_{nl}(r_{IJ,K}) Y_{lm}(\hat{\boldsymbol{r}}_{IJ,K}) f_{\mathrm{cut}}^{\mathrm{e}}(\boldsymbol{r}_{IJ,K}; \boldsymbol{r}_{IJ}). \end{aligned}$$

Here the cutoff function for the bond can be given analogous to those for the *onsite* basis, as

$$f_{\mathrm{cut}}^{\mathrm{b}}(r; r_{\mathrm{bond,Z_IZ_J}}) = f_{\mathrm{cut}}(r; r_{\mathrm{bond,Z_IZ_J}}),$$

with f_{cut} being defined in (S4). On the other hand, the cutoff function $f_{\mathrm{cut}}^{\mathrm{e}}$ for the environmental atom is defined as

$$f_{\mathrm{cut}}^{\mathrm{e}}(\boldsymbol{r}; r_{\mathrm{cut,Z_I}}, r_{\mathrm{cut,Z_J}}) = f_{\mathrm{cut}}\big(\big|\boldsymbol{r} + (1-\theta)(\boldsymbol{r}_J - \boldsymbol{r}_I)\big|; r_{\mathrm{cut,Z_I}}\big) + f_{\mathrm{cut}}\big(\big|\boldsymbol{r} + \theta(\boldsymbol{r}_I - \boldsymbol{r}_J)\big|; r_{\mathrm{cut,Z_J}}\big),$$

where f_{cut} is defined in (S4), $\boldsymbol{r}$ denotes the $\boldsymbol{r}_{IJ,K}$ element in σ_K and $r_{\mathrm{cut,Z_I/Z_J}}$, the cutoff radii of the

two spheres around both the I-th and the J-th atom. Throughout this work, we choose $\theta = 0.5$.

Given the one-particle basis, we can form the density projection and the ν-correlations for the onsite case as

$$\begin{aligned} A_v(\mathbf{R}_I) &:= \sum_{\sigma \in \mathbf{R}_I} \phi_v(\sigma), \\ \boldsymbol{A}_{\boldsymbol{v}}(\mathbf{R}_I) &:= \prod_{t=1}^{\nu} A_{v_t}(\mathbf{R}_I) \qquad \text{for } \boldsymbol{v} = (v_1, \ldots, v_\nu),\ \nu = 1, 2, \ldots, \end{aligned}$$

and for the offsite,

$$\begin{aligned} A_v(\mathbf{R}_{IJ}) &:= \sum_{K \neq I,J} \phi^{\mathrm{e}}_v(\sigma_K), \\ \boldsymbol{A}_{\boldsymbol{v}}(\mathbf{R}_{IJ}) &:= \phi^{\mathrm{b}}_{v^0}(\sigma_{IJ}) \cdot \prod_{t=1}^{\nu} A_{v^t}(\mathbf{R}_{IJ}) \qquad \text{for } \boldsymbol{v} = (v_0, \ldots, v_\nu),\ \nu = 1, 2, \ldots. \end{aligned}$$

Finally, we perform the symmetrization over O(3), by leveraging an averaged integral

$$\mathcal{B}_{\boldsymbol{v},a}(\mathbf{R}_\bullet) = \fint_{\mathrm{O}(3)} D(Q)(\boldsymbol{A}_{\boldsymbol{v}}(Q\mathbf{R}_\bullet)E_a)D(Q)^* dQ,$$

where $\{E_a\}_a$ forms a canonical basis of the matrix space where the density matrices D, or a sub-block thereof, lies in.

By design, the $\mathcal{B}_{\boldsymbol{v},a}$ bases have exactly the same equivariance as (the subblocks) of the density matrix D. To simplify the notation, we absorbed the index a into $\boldsymbol{v}$ in the main text.

S3 Properties of the retraction

In this section, we discuss the retraction operator $\mathcal{P}$ and some of its important properties. Denoting

$$S_{N_g} = \{D \in \mathbb{R}^{N_g \times N_g} : D^T = D, \lambda_{D,N} > \lambda_{D,N+1}\},$$

where $\{\lambda_{D,N}\}_{N=1}^{N_g}$ represents the eigenvalues of D, sorted descendingly, then we have the following proposition.

Proposition 1. *Let $\mathcal{P} : S_{N_g} \to \mathcal{G}^N_{N_g}$ be the retraction operator defined in (8), then:*

(1.a) for $D_{\mathbf{R}}$ satisfying (S2), there holds

$$\mathcal{P}(D_{Q\mathbf{R}}) = \mathcal{D}(Q)\mathcal{P}(D_{\mathbf{R}})\mathcal{D}(Q)^*,\ \forall Q \in \mathrm{O}(3);$$

(1.b) for any $D \in S_{N_g}$,

$$\mathcal{P}(D) \in \operatorname*{arg\,min}_{\tilde{D} \in \mathcal{G}^N_{N_g}} \|\tilde{D} - D\|_F.$$

Proof. We first justify that $\mathcal{P}$ is wx§ell-defined. Let $D \in S_{N_g}$, then its eigenvalue decomposition can be written as $D = U\Sigma U^T$, where the unitary matrix $U = [u_1, u_2, \ldots, u_{N_g}]$ consists of N_g orthonormal eigenvectors of D, Σ is a diagonal matrix containing the eigenvalues of D, sorted decreasingly. By a straight calculation, we have

$$UE^N_{N_g}U^T = \sum_{i=1}^{N} u_i u_i^T.$$

Although U may be non-unique, the result in the RHS of the above equality will not be influenced by the order and signs of the first N orthonormal eigenvectors of D, since $\lambda_{D,N} > \lambda_{D,N+1}$. This shows the well-definedness of $\mathcal{P}$, *i.e.* the image of D via $\mathcal{P}$ is unique regardless of how the eigenvalue decomposition is performed.

Now we can move on to consider (1.a). Assume $D_{\mathbf{R}} = U_{D_{\mathbf{R}}}\Sigma_{D_{\mathbf{R}}}U_{D_{\mathbf{R}}}^T$, then $\mathcal{P}(D_{\mathbf{R}}) = U_{D_{\mathbf{R}}}E_{N_g}^N U_{D_{\mathbf{R}}}^T$. By (S2), we have that for all $Q \in \mathrm{O}(3)$

$$\begin{aligned} D_{Q\mathbf{R}} &= \mathcal{D}(Q)D_{\mathbf{R}}\mathcal{D}(Q)^*, \\ &= \mathcal{D}(Q)U_{D_{\mathbf{R}}}\Sigma_{D_{\mathbf{R}}}U_{D_{\mathbf{R}}}^T\mathcal{D}(Q)^*, \end{aligned}$$

which means that $D_{Q\mathbf{R}}$ has an eigenvalue decomposition as above. As a result,

$$\begin{aligned} \mathcal{P}(D_{Q\mathbf{R}}) &= \mathcal{D}(Q)U_{D_{\mathbf{R}}}E_{N_g}^N U_{D_{\mathbf{R}}}^T\mathcal{D}(Q)^*, \\ &= \mathcal{D}(Q)\mathcal{P}(D_{\mathbf{R}})\mathcal{D}(Q)^*, \quad \forall Q \in \mathrm{O}(3). \end{aligned}$$

This proves (1.a).

As of (1.b), we again suppose $D = U\Sigma U^T \in S_{N_g}$, with U defined as above. We claim that

$$\mathcal{P}(D) = UE_{N_g}^N U^T = \underset{G\in\mathcal{G}_{N_g}^N}{\arg\min}\|G - D\|_F = \underset{\{PE_{N_g}^N P^T:\ P\in\mathcal{O}(N_g)\}}{\arg\min}\|PE_{N_g}^N P^T - D\|_F.$$

Note that the last equality above is based on

$$\mathcal{G}_{N_g}^N = \{PE_{N_g}^N P^T :\ P \in \mathcal{O}(N_g)\}.$$

We now estimate

$$\|PE_{N_g}^N P^T - D\|_F^2 = \|PE_{N_g}^N P^T - U\Sigma U^T\|_F^2 = \mathrm{tr}(PE_{N_g}^N P^T + U\Sigma^2 U^T - 2PE_{N_g}^N P^T U\Sigma U^T). \tag{S5}$$

To minimize (S5), we just need to maximize $\mathrm{tr}(PE_{N_g}^N P^T U\Sigma U^T)$, since the trace of the first two terms in the right hand side is a constant, thus

$$\begin{aligned} \max_{P\in\mathcal{O}(N_g)} \mathrm{tr}(PE_{N_g}^N P^T U\Sigma U^T) &= \max_{P\in\mathcal{O}(N_g)} \mathrm{tr}(U^T PE_{N_g}^N P^T U\Sigma) \\ &= \max_{P\in\mathcal{O}(N_g)} \mathrm{tr}(PE_{N_g}^N P^T\Sigma) \\ &= \max_{G\in\mathcal{G}_{N_g}^N} \mathrm{tr}(G\Sigma) = \max_{G\in\mathcal{G}_{N_g}^N} \sum_{i=1}^{N_g}\sigma_i G_{ii}. \end{aligned}$$

For any $G_0 \in \mathcal{G}_{N_g}^N$, there exists $P_0 \in \mathcal{O}(N_g)$ such that $G_0 = P_0 E_{N_g}^N P_0^T$. Consequently,

$$0 \le G_{0,ii} = \sum_{t=1}^{N} P_{0,it}^2 \le \sum_{t=1}^{N_g} P_{0,it}^2 = 1.$$

In addition, we have $\mathrm{tr}(G_0) = N$. Hence $\frac{1}{N}\sum_{i=1}^{N_g}\sigma_i G_{0,ii} = \sum_{i=1}^{N_g}\sigma_i\frac{G_{0,ii}}{N}$ becomes a convex combinition of $\{\sigma_i\}_{i=1}^{N_g}$, which achieves its maximum at $\sum_{i=1}^{N}\sigma_i$ when G_0 is chosen to be $E_{N_g}^N$. This completes the proof. □

S4 Comparison of fitting the density matrix and the KS matrix

The KS matrix $F_{\mathbf{R}}$ has the same structure and symmetry as the density matrix $D_{\mathbf{R}}$ and the learning of it has been more broadly studied compared to that of the density matrix. In this appendix, we compare the fitting of the two objects using the same method. To make our comparison meaningful and fair, we first fix an ACE basis $\boldsymbol{\mathcal{B}}_{\nu,d}$, where ν and d stand for the correlation order and polynomial degree that define the basis (in addition, the cutoffs are also fixed but is not explicitly written here for the sake of simplicity). That said, the only thing different for the two targeting models (for the KS matrix and the density matrix) is their coefficients. We then pick the same data points, which have the form $\{(\mathbf{R}^{(k)}, F_{\mathbf{R}^{(k)}}, D_{\mathbf{R}^{(k)}})\}_{k=1}^{N_{\text{data}}}$. By solving the least squares problems (7) with the Kohn-Sham matrices and density matrices, respectively, we obtain $\boldsymbol{c}_{\nu,d,H}$ and $\boldsymbol{c}_{\nu,d,D}$ for the two objects. Then we have two routines to get the predicted density matrix of a given configuration $\mathbf{R}$, which lies in the desired manifold (3).

First, with $\boldsymbol{\mathcal{B}}_{\nu,d}$ and $\boldsymbol{c}_{\nu,d,D}$, we obtain directly a feasible approximation of the density matrix

$$\tilde{D}_{\mathbf{R}} = \mathcal{P}\big(\boldsymbol{c}_{\nu,d,D} \cdot \boldsymbol{\mathcal{B}}_{\nu,d}(\mathbf{R})\big),$$

where $\mathcal{P}$ is the retraction operator defined in (8). Alternatively, we can construct

$$\tilde{F}_{\mathbf{R}} = \boldsymbol{c}_{\nu,d,F} \cdot \boldsymbol{\mathcal{B}}_{\nu,d}(\mathbf{R}),$$

following by solving

$$\tilde{F}_{\mathbf{R}}\tilde{C}_{\mathbf{R}} = \tilde{C}_{\mathbf{R}}E_{\mathbf{R}}$$

for its N eigenvectors $\tilde{C}_{\mathbf{R}}$, and finally obtain

$$\bar{D}_{\mathbf{R}} = \tilde{C}_{\mathbf{R}}\tilde{C}_{\mathbf{R}}^{T}.$$

Molecule	Density Matrix		KS Matrix	
	Specific Model	Unified Model	Specific Model	Unified Model
Acetaldehyde	$4.416{\cdot}10^{-5}$	$3.278{\cdot}10^{-4}$	$2.369{\cdot}10^{-5}$	$2.137{\cdot}10^{-4}$
Acrolein	$2.514{\cdot}10^{-4}$	$5.028{\cdot}10^{-4}$	$1.688{\cdot}10^{-4}$	$4.052{\cdot}10^{-4}$
Aniline	$4.300{\cdot}10^{-4}$	$4.868{\cdot}10^{-4}$	$5.400{\cdot}10^{-4}$	$1.316{\cdot}10^{-3}$
o-Toluidine	$5.430{\cdot}10^{-4}$	$5.962{\cdot}10^{-4}$	$4.760{\cdot}10^{-4}$	$3.806{\cdot}10^{-3}$
m-Toluidine	$5.384{\cdot}10^{-4}$	$5.824{\cdot}10^{-4}$	$5.401{\cdot}10^{-4}$	$4.115{\cdot}10^{-3}$
Benzene*	-	$4.058{\cdot}10^{-4}$	-	$4.333{\cdot}10^{-4}$
Toluene*	-	$6.369{\cdot}10^{-4}$	-	$4.893{\cdot}10^{-4}$
Phenol**	-	$4.809{\cdot}10^{-3}$	-	$3.204{\cdot}10^{-2}$
Benzaldehyde**	-	$4.129{\cdot}10^{-3}$	-	$3.101{\cdot}10^{-2}$
p-Toluidine**	-	$2.840{\cdot}10^{-3}$	-	$1.922{\cdot}10^{-2}$
1-Propanol	$3.049{\cdot}10^{-4}$	$4.427{\cdot}10^{-4}$	$4.044{\cdot}10^{-4}$	$3.468{\cdot}10^{-4}$
1-Butanol	$4.510{\cdot}10^{-4}$	$4.921{\cdot}10^{-4}$	$2.801{\cdot}10^{-4}$	$3.579{\cdot}10^{-4}$
2-Butanol	$9.173{\cdot}10^{-4}$	$1.494{\cdot}10^{-3}$	$5.748{\cdot}10^{-4}$	$1.709{\cdot}10^{-3}$
1-Hexanol	$1.031{\cdot}10^{-3}$	$5.324{\cdot}10^{-4}$	$3.264{\cdot}10^{-4}$	$4.193{\cdot}10^{-4}$
Ethanol*	-	$9.644{\cdot}10^{-4}$	-	$1.133{\cdot}10^{-3}$
2-Propanol*	-	$7.701{\cdot}10^{-4}$	-	$2.948{\cdot}10^{-3}$
2-Hexanol*	-	$7.384{\cdot}10^{-4}$	-	$1.219{\cdot}10^{-3}$
1-Heptanol*	-	$8.896{\cdot}10^{-4}$	-	$2.234{\cdot}10^{-3}$

Table S1: The average test set RMSEs on the density matrices obtained by the (3,8) specific models and unified model trained with the density matrices and the KS matrices ($r_{\text{cut}} = 4.0$ Å).

In table S1, we compare the element-wise test set RMSE in the predicted density matrices obtained by the two approaches above, including both the cases of specific models and the unified model. To be fair, we use exactly the same data points as those mentioned in Section 3 for training. It can be observed that fitting the KS matrix gives comparable, or even smaller test set RMSEs on the predicted density matrices with respect to the specific models, whereas it performs poorly on the unified model. This implies that the approach of fitting the KS matrix only by minimizing the element-wise error is either less transferable or requires more careful weighting for different systems. In any case, it is less robust than the same approach for the density matrix.

In addition, to see the performance of the specific models more clearly, we compared the properties derived from the specific models tailored to both the density matrices and the KS matrices. The errors on the properties are illustrated in Figure S1, from which we can see that although the approach of predicting the KS matrices provides smaller element-wise errors, it cannot guarantee that the derived properties are equally effective.

Overall, the proposed approach of fitting the matrix element through a Least-Squares approach is more suitable for the density matrix than for the KS matrix, at least for the molecular systems mentioned in our experiments, especially in terms of the model transferability.

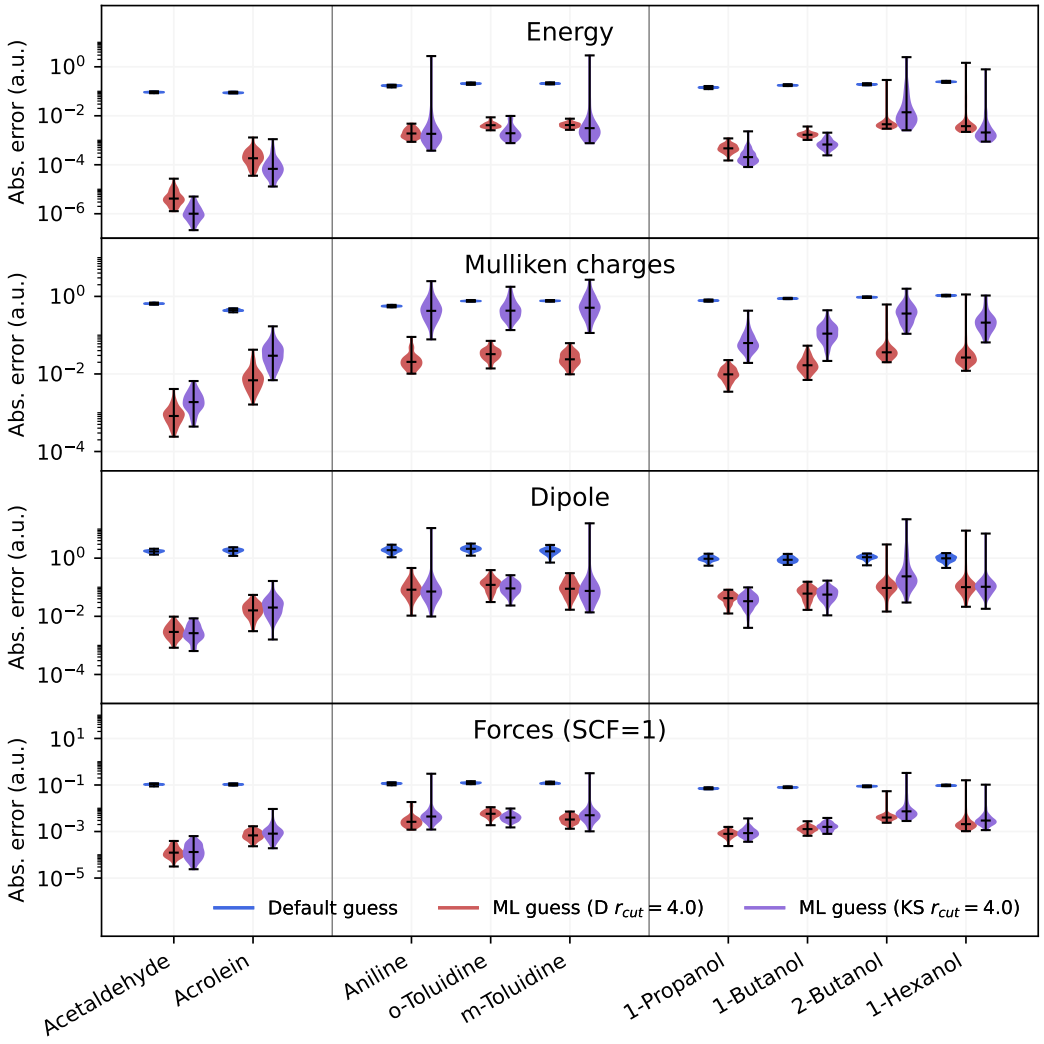

Figure S1: Plot of the error (in logarithmic scale) for energy, Mulliken charges, dipole moment and the error for forces after a single SCF cycle. The blue line represents the default guess provided by Gaussian, and red and violet lines refer to the ML guess obtained with models trained on the density matrices and the KS matrices, respectively (r_{cut} = 4.0 Å).

S5 Effect of basis sets

In this section, we apply the proposed method for learning the density matrix to the dataset obtained with a level of theory of DFT ωB96X-D/6-311G(d,p), that is, the one with a larger basis set compared to what was reported in Section 3 of the main text.

We present the results of molecule-specific models for m-toluidine, a system that shows longer-range π conjugation and $n-\pi$ interactions. The size of the density matrices for the m-toluidine molecules is 198 for ωB96X-D/6-311G(d,p) and 130 for ωB96X-D/6-31G(d), respectively. Although the sizes of the density matrices differ under the two bases are different, their corresponding subblocks share the same equivariance. We can therefore follow exactly the same protocol illustrated in Section 3.1: the specific models are trained with 1 500 training samples evenly sampled from the first 3 000 frames of m-toluidine, and are then tested on the last 7 000 frames. We used models with $\nu = 2,\ 4 \leq d_{\max} \leq 8$, and $\nu = 3,\ 4 \leq d_{\max} \leq 6$, which we rationalize below. For the local truncation, we tested the cutoffs of 4.0 Å and 6.5 Å, respectively. The distribution of RMSE values as a function of the degree $d_{\max}$ used to generate the ACE bases, for the two choices of the correlation order ν, is shown in Figure S2.

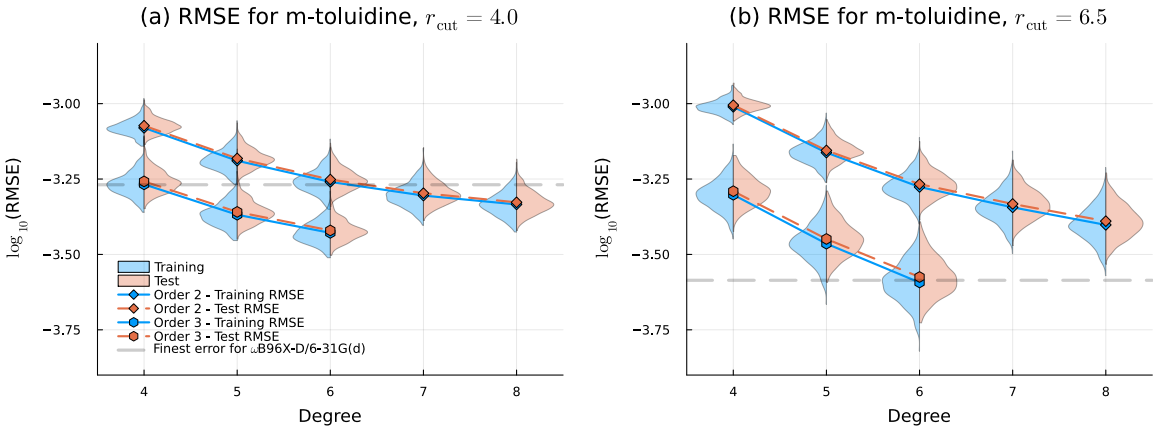

Figure S2: The RMSEs of density matrices for m-toluidine at a level of theory of DFT ωB96X-D/6-311G(d,p): (a) with $r_{\text{cut}} = 4.0$ Å, (b) with $r_{\text{cut}} = 6.5$ Å, obtained by the corresponding specific models with respect to different degrees $d_{\max}$ and for various orders ν. The solid and dashed colored lines refer to the average training and test set errors, and the shaded areas show the distribution of the errors for the corresponding models, while the gray dashed lines stand for the finest RMSEs obtained in the smaller basis set case.

As can be seen in Figure S2, a convergence trend similar to that reported for the smaller basis set case is observed, where the training and test set RMSEs not only align nicely with each other, but are also distributed nearly normally. In particular, the test set RMSE decreases monotonically with increasing ACE basis size, as previously observed. This implies that smaller errors can be expected by continuously increasing the model parameters. However, we chose to terminate at an earlier stage where the models achieved the same, or even better, level of accuracy than the finest RMSEs obtained in the smaller basis set case (shown as the gray dashed lines in Figure S2).

Remarkably, the smaller cutoff radius of 4.0 Å appears to be more preferable in this test, unlike with the smaller basis set scenario, where larger cutoffs always result in strictly smaller errors. Here, the 4.0 Å models maintain similar accuracy to the corresponding 6.5 Å models most of the time and yield a more concentrated error distribution, which is a sign of better transferability.

While it cannot be ruled out that the model parameters of the proposed approach may require fine-tuning in certain scenarios, the similar convergence trend and the even lower RMSEs obtained with the larger basis set suggest that the proposed approach is not greatly affected by the basis set

selection.

S6 Additional tables

Molecule	**N frames**
Acetaldehyde	10000
Acrolein	10000
Aniline	10000
o-Toluidine	10000
m-Toluidine	10000
Benzene	100
Toluene	100
Phenol	100
Benzaldehyde	100
p-Toluidine	100
1-Propanol	10000
1-Butanol	10000
2-Butanol	10000
1-Hexanol	10000
Ethanol	100
2-Propanol	100
2-Hexanol	100
1-Heptanol	100

Table S2: Datasets used in this work. Different molecules are grouped according to their chemical class: aldehydes, aromatics, alcohols. The level of theory is DFT ωB97X-D/6-31G(d).

Molecule	Specific Model				Unified Model				Unified Model-A			Unified-Alcohol-Model		
	training frames	test frames	r_{cut} =4.0	r_{cut} =6.5	training frames	test frames	r_{cut} =4.0	r_{cut} =6.5	training frames	test frames	r_{cut} =4.0	training frames	test frames	r_{cut} =4.0
Acetaldehyde	0:2999	3000:9999	$4.416{\cdot}10^{-5}$	$4.240{\cdot}10^{-5}$	0:30:8999	9000:9999	$3.278{\cdot}10^{-4}$	$3.861{\cdot}10^{-4}$	0:30:8999	9000:9999	$3.299{\cdot}10^{-4}$	-	-	-
Acrolein	0:2:5999	6000:9999	$2.514{\cdot}10^{-4}$	$1.409{\cdot}10^{-4}$	0:30:8999	9000:9999	$5.028{\cdot}10^{-4}$	$4.277{\cdot}10^{-4}$	0:30:8999	9000:9999	$5.081{\cdot}10^{-4}$	-	-	-
Aniline	0:2:2999	3000:9999	$4.300{\cdot}10^{-4}$	$1.634{\cdot}10^{-4}$	0:30:8999	9000:9999	$4.868{\cdot}10^{-4}$	$3.177{\cdot}10^{-4}$	0:30:8999	9000:9999	$4.876{\cdot}10^{-4}$	-	-	-
o-Toluidine	0:2:2999	3000:9999	$5.430{\cdot}10^{-4}$	$2.738{\cdot}10^{-4}$	0:30:8999	9000:9999	$5.962{\cdot}10^{-4}$	$3.584{\cdot}10^{-4}$	0:30:8999	9000:9999	$5.940{\cdot}10^{-4}$	-	-	-
m-Toluidine	0:2:2999	3000:9999	$5.384{\cdot}10^{-4}$	$2.596{\cdot}10^{-4}$	0:30:8999	9000:9999	$5.824{\cdot}10^{-4}$	$3.727{\cdot}10^{-4}$	0:30:8999	9000:9999	$5.822{\cdot}10^{-4}$	-	-	-
Benzene*	-	-	-	-	-	0:99	$4.058{\cdot}10^{-4}$	$8.638{\cdot}10^{-3}$	-	0:99	$3.646{\cdot}10^{-4}$	-	-	-
Toluene*	-	-	-	-	-	0:99	$6.369{\cdot}10^{-4}$	$7.674{\cdot}10^{-3}$	-	0:99	$5.980{\cdot}10^{-4}$	-	-	-
Phenol**	-	-	-	-	-	0:99	$4.809{\cdot}10^{-3}$	$1.795{\cdot}10^{-2}$	0:10:89	others	$6.770{\cdot}10^{-4}$	-	-	-
Benzaldehyde**	-	-	-	-	-	0:99	$4.129{\cdot}10^{-3}$	$1.997{\cdot}10^{-2}$	0:10:89	others	$1.201{\cdot}10^{-3}$	-	-	-
p-Toluidine**	-	-	-	-	-	0:99	$2.840{\cdot}10^{-3}$	$1.907{\cdot}10^{-3}$	0:10:89	others	$6.293{\cdot}10^{-4}$	-	-	-
Propanol	0:2999	3000:9999	$3.049{\cdot}10^{-4}$	$2.421{\cdot}10^{-4}$	0:30:8999	9000:9999	$4.427{\cdot}10^{-4}$	$4.080{\cdot}10^{-4}$	0:30:8999	9000:9999	$4.444{\cdot}10^{-4}$	0:24:8999	9000:9999	$3.932{\cdot}10^{-4}$
Butanol	0:3:8999	9000:9999	$4.510{\cdot}10^{-4}$	$7.643{\cdot}10^{-4}$	0:30:8999	9000:9999	$4.921{\cdot}10^{-4}$	$1.273{\cdot}10^{-3}$	0:30:8999	9000:9999	$4.934{\cdot}10^{-4}$	0:24:8999	9000:9999	$4.477{\cdot}10^{-4}$
2-Butanol	0:2999	3000:9999	$9.173{\cdot}10^{-4}$	$7.603{\cdot}10^{-4}$	0:30:8999	9000:9999	$1.494{\cdot}10^{-3}$	$6.581{\cdot}10^{-4}$	0:30:8999	9000:9999	$1.509{\cdot}10^{-3}$	0:24:8999	9000:9999	$7.598{\cdot}10^{-4}$
Hexanol	0:2:2999	3000:9999	$1.031{\cdot}10^{-3}$	$7.021{\cdot}10^{-4}$	0:30:8999	9000:9999	$5.423{\cdot}10^{-4}$	$9.614{\cdot}10^{-4}$	0:30:8999	9000:9999	$5.314{\cdot}10^{-4}$	0:24:8999	9000:9999	$4.602{\cdot}10^{-4}$
Ethanol*	-	-	-	-	-	0:99	$9.644{\cdot}10^{-4}$	$1.021{\cdot}10^{-3}$	-	0:99	$8.999{\cdot}10^{-4}$	-	0:99	$5.556{\cdot}10^{-4}$
2-Propanol*	-	-	-	-	-	0:99	$7.701{\cdot}10^{-4}$	$7.573{\cdot}10^{-4}$	-	0:99	$7.480{\cdot}10^{-4}$	-	0:99	$4.686{\cdot}10^{-4}$
2-Hexanol*	-	-	-	-	-	0:99	$7.384{\cdot}10^{-4}$	$4.853{\cdot}10^{-3}$	-	0:99	$7.353{\cdot}10^{-4}$	-	0:99	$5.485{\cdot}10^{-4}$
Heptanol*	-	-	-	-	-	0:99	$8.896{\cdot}10^{-4}$	$1.538{\cdot}10^{-3}$	-	0:99	$8.980{\cdot}10^{-4}$	-	0:99	$7.048{\cdot}10^{-4}$

Table S3: test set RMSEs for different molecules obtained by (3,8)-models trained with different datasets. In the table, when "-" appears, it means that the corresponding molecule is not included in the training process, and hence there is no dedicated model for these molecules.

Molecule	Tol.	Default guess	Specific Model	
			$r_{\text{cut}} = 4.0$	$r_{\text{cut}} = 6.5$
Acetaldehyde	10^{-6}	9.4 ± 0.1	5.2 ± 0.1 (∼ 44%)	5.1 ± 0.1 (∼ 46%)
	10^{-7}	12.4 ± 0.2	7.8 ± 0.1 (∼ 37%)	7.8 ± 0.1 (∼ 37%)
	10^{-8}	14.9 ± 0.1	9.8 ± 0.1 (∼ 34%)	9.7 ± 0.1 (∼ 35%)
Acrolein	10^{-6}	10.5 ± 0.1	7.5 ± 0.2 (∼ 29%)	6.3 ± 0.2 (∼ 40%)
	10^{-7}	13.4 ± 0.1	9.8 ± 0.1 (∼ 26%)	9.1 ± 0.1 (∼ 32%)
	10^{-8}	16.0 ± 0.2	11.5 ± 0.1 (∼ 28%)	11.2 ± 0.1 (∼ 30%)
Aniline	10^{-6}	9.9 ± 0.1	7.2 ± 0.1 (∼ 28%)	6.5 ± 0.1 (∼ 35%)
	10^{-7}	12.5 ± 0.1	9.7 ± 0.1 (∼ 23%)	8.7 ± 0.1 (∼ 31%)
	10^{-8}	14.7 ± 0.1	11.9 ± 0.1 (∼ 19%)	11.1 ± 0.1 (∼ 24%)
o-Toluidine	10^{-6}	10.0 ± 0.0	7.6 ± 0.1 (∼ 24%)	7.2 ± 0.1 (∼ 28%)
	10^{-7}	12.8 ± 0.1	9.9 ± 0.1 (∼ 22%)	9.7 ± 0.1 (∼ 24%)
	10^{-8}	14.9 ± 0.0	12.3 ± 0.1 (∼ 18%)	11.7 ± 0.1 (∼ 22%)
m-Toluidine	10^{-6}	10.0 ± 0.0	7.3 ± 0.1 (∼ 27%)	7.1 ± 0.1 (∼ 29%)
	10^{-7}	12.7 ± 0.1	9.8 ± 0.1 (∼ 23%)	9.5 ± 0.1 (∼ 25%)
	10^{-8}	14.8 ± 0.1	11.9 ± 0.1 (∼ 19%)	11.7 ± 0.1 (∼ 21%)
1-Propanol	10^{-6}	9.0 ± 0.0	6.0 ± 0.1 (∼ 33%)	6.0 ± 0.1 (∼ 33%)
	10^{-7}	11.0 ± 0.1	8.0 ± 0.0 (∼ 27%)	8.0 ± 0.0 (∼ 27%)
	10^{-8}	13.2 ± 0.1	9.8 ± 0.1 (∼ 25%)	9.8 ± 0.1 (∼ 26%)
1-Butanol	10^{-6}	9.0 ± 0.0	6.4 ± 0.1 (∼ 29%)	6.4 ± 0.1 (∼ 29%)
	10^{-7}	11.0 ± 0.0	8.1 ± 0.1 (∼ 27%)	8.2 ± 0.1 (∼ 26%)
	10^{-8}	13.2 ± 0.1	10.1 ± 0.1 (∼ 24%)	10.2 ± 0.1 (∼ 23%)
2-Butanol	10^{-6}	9.0 ± 0.0	7.2 ± 0.1 (∼ 20%)	7.5 ± 0.1 (∼ 16%)
	10^{-7}	11.1 ± 0.1	9.0 ± 0.1 (∼ 19%)	9.3 ± 0.1 (∼ 16%)
	10^{-8}	13.2 ± 0.1	10.9 ± 0.1 (∼ 18%)	11.0 ± 0.1 (∼ 17%)
1-Hexanol	10^{-6}	9.0 ± 0.0	6.4 ± 0.1 (∼ 29%)	6.6 ± 0.1 (∼ 27%)
	10^{-7}	10.9 ± 0.1	8.2 ± 0.1 (∼ 24%)	8.3 ± 0.1 (∼ 23%)
	10^{-8}	13.0 ± 0.0	10.3 ± 0.1 (∼ 21%)	10.4 ± 0.1 (∼ 20%)

Table S4: Average of the number of iterations obtained using the predicted density matrix as guess for the calculation on the same geometry. The results refer to the (3,8)-models trained targeting the density matrix. The first column specifies the molecule, the second column indicates the convergence tolerance for the SCF procedure, the third column reports the average of the number of iterations obtained using the default guess available in the Gaussian program and in the last two columns we report the average number of iterations obtained with our predicted density matrix as guess, using the specific model for each molecule and with 4.0 and 6.5 radius of cutoff, respectively. The values reported in parentheses indicate the percentage of reduction with respect to the iterations required by the default guess to converge.

Molecule	Tol.	Default guess	Specific Model	
			$r_{\text{cut}} = 4.0$	$r_{\text{cut}} = 6.5$
Acetaldehyde	10^{-6}	9.4 ± 0.1	4.8 ± 0.1 (∼ 49%)	4.7 ± 0.1 (∼ 50%)
	10^{-7}	12.4 ± 0.2	7.3 ± 0.2 (∼ 41%)	7.3 ± 0.2 (∼ 41%)
	10^{-8}	14.9 ± 0.1	9.3 ± 0.1 (∼ 38%)	9.2 ± 0.1 (∼ 38%)
Acrolein	10^{-6}	10.5 ± 0.1	7.2 ± 0.2 (∼ 32%)	5.9 ± 0.2 (∼ 44%)
	10^{-7}	13.4 ± 0.1	9.6 ± 0.1 (∼ 29%)	8.6 ± 0.1 (∼ 36%)
	10^{-8}	16.0 ± 0.2	11.2 ± 0.1 (∼ 30%)	10.8 ± 0.1 (∼ 32%)
Aniline	10^{-6}	9.9 ± 0.1	7.0 ± 0.1 (∼ 29%)	6.3 ± 0.1 (∼ 36%)
	10^{-7}	12.5 ± 0.1	9.4 ± 0.2 (∼ 25%)	8.8 ± 0.2 (∼ 30%)
	10^{-8}	14.7 ± 0.1	11.7 ± 0.2 (∼ 20%)	11.0 ± 0.2 (∼ 25%)
o-Toluidine	10^{-6}	10.0 ± 0.0	7.0 ± 0.1 (∼ 30%)	6.9 ± 0.1 (∼ 31%)
	10^{-7}	12.8 ± 0.1	9.6 ± 0.1 (∼ 25%)	9.4 ± 0.1 (∼ 27%)
	10^{-8}	14.9 ± 0.0	11.7 ± 0.1 (∼ 21%)	11.4 ± 0.1 (∼ 24%)
m-Toluidine	10^{-6}	10.0 ± 0.0	7.1 ± 0.2 (∼ 29%)	6.9 ± 0.2 (∼ 31%)
	10^{-7}	12.7 ± 0.1	9.7 ± 0.2 (∼ 23%)	9.4 ± 0.2 (∼ 25%)
	10^{-8}	14.8 ± 0.1	11.9 ± 0.2 (∼ 20%)	11.6 ± 0.2 (∼ 22%)
1-Propanol	10^{-6}	9.0 ± 0.0	6.0 ± 0.0 (∼ 33%)	5.9 ± 0.0 (∼ 34%)
	10^{-7}	11.0 ± 0.1	8.0 ± 0.1 (∼ 27%)	7.8 ± 0.1 (∼ 29%)
	10^{-8}	13.2 ± 0.1	9.9 ± 0.1 (∼ 25%)	9.7 ± 0.1 (∼ 27%)
1-Butanol	10^{-6}	9.0 ± 0.0	6.2 ± 0.1 (∼ 31%)	6.2 ± 0.1 (∼ 32%)
	10^{-7}	11.0 ± 0.0	8.1 ± 0.1 (∼ 27%)	8.1 ± 0.1 (∼ 27%)
	10^{-8}	13.2 ± 0.1	10.1 ± 0.1 (∼ 24%)	10.0 ± 0.1 (∼ 24%)
2-Butanol	10^{-6}	9.0 ± 0.0	7.3 ± 0.1 (∼ 18%)	7.1 ± 0.1 (∼ 21%)
	10^{-7}	11.1 ± 0.1	9.3 ± 0.1 (∼ 17%)	9.1 ± 0.1 (∼ 18%)
	10^{-8}	13.2 ± 0.1	11.1 ± 0.1 (∼ 16%)	11.0 ± 0.1 (∼ 17%)
1-Hexanol	10^{-6}	9.0 ± 0.0	6.6 ± 0.1 (∼ 27%)	6.7 ± 0.1 (∼ 25%)
	10^{-7}	10.9 ± 0.1	8.3 ± 0.1 (∼ 24%)	8.6 ± 0.1 (∼ 21%)
	10^{-8}	13.0 ± 0.0	10.2 ± 0.1 (∼ 22%)	10.4 ± 0.1 (∼ 20%)

Table S5: Average of the number of iterations obtained using the predicted density matrix as guess for the calculation on the same geometry. The results refer to the (3,8)-models trained targeting the Kohn-Sham matrix. The first column specifies the molecule, the second column indicates the convergence tolerance for the SCF procedure, the third column reports the average of the number of iterations obtained using the default guess available in the Gaussian program and in the last two columns we report the average number of iterations obtained with our predicted density matrix as guess, using the specific model for each molecule and with 4.0 and 6.5 radius of cutoff, respectively. The values reported in parentheses indicate the percentage of reduction with respect to the iterations required by the default guess to converge.

Molecule	Tol.	Default guess	Specific Model	Unified Model	Unified Model-A
Acetaldehyde	10^{-6}	9.4 ± 0.1	5.2 ± 0.1 (∼ 44%)	7.5 ± 0.1 (∼ 20%)	7.5 ± 0.1 (∼ 20%)
	10^{-7}	12.4 ± 0.2	7.8 ± 0.1 (∼ 37%)	9.7 ± 0.1 (∼ 22%)	9.7 ± 0.1 (∼ 22%)
	10^{-8}	14.9 ± 0.1	9.8 ± 0.1 (∼ 34%)	11.1 ± 0.1 (∼ 26%)	11.1 ± 0.1 (∼ 25%)
Acrolein	10^{-6}	10.5 ± 0.1	7.5 ± 0.2 (∼ 29%)	8.3 ± 0.2 (∼ 21%)	8.3 ± 0.2 (∼ 21%)
	10^{-7}	13.4 ± 0.1	9.8 ± 0.1 (∼ 26%)	10.5 ± 0.1 (∼ 22%)	10.5 ± 0.1 (∼ 21%)
	10^{-8}	16.0 ± 0.2	11.5 ± 0.1 (∼ 28%)	12.0 ± 0.1 (∼ 25%)	12.1 ± 0.1 (∼ 24%)
Aniline	10^{-6}	9.9 ± 0.1	7.2 ± 0.1 (∼ 28%)	7.5 ± 0.1 (∼ 24%)	7.5 ± 0.1 (∼ 24%)
	10^{-7}	12.5 ± 0.1	9.7 ± 0.1 (∼ 23%)	9.9 ± 0.1 (∼ 21%)	9.9 ± 0.1 (∼ 21%)
	10^{-8}	14.7 ± 0.1	11.9 ± 0.1 (∼ 19%)	12.1 ± 0.1 (∼ 18%)	12.1 ± 0.1 (∼ 18%)
o-Toluidine	10^{-6}	10.0 ± 0.0	7.6 ± 0.1 (∼ 24%)	7.8 ± 0.1 (∼ 22%)	7.8 ± 0.1 (∼ 22%)
	10^{-7}	12.8 ± 0.1	9.9 ± 0.1 (∼ 22%)	10.1 ± 0.1 (∼ 21%)	10.1 ± 0.1 (∼ 21%)
	10^{-8}	14.9 ± 0.0	12.3 ± 0.1 (∼ 18%)	12.5 ± 0.1 (∼ 16%)	12.5 ± 0.1 (∼ 16%)
m-Toluidine	10^{-6}	10.0 ± 0.0	7.3 ± 0.1 (∼ 27%)	7.6 ± 0.1 (∼ 24%)	7.6 ± 0.1 (∼ 24%)
	10^{-7}	12.7 ± 0.1	9.8 ± 0.1 (∼ 23%)	9.9 ± 0.1 (∼ 22%)	9.9 ± 0.1 (∼ 22%)
	10^{-8}	14.8 ± 0.1	11.9 ± 0.1 (∼ 19%)	12.1 ± 0.1 (∼ 18%)	12.1 ± 0.1 (∼ 18%)
Benzene*	10^{-6}	9.0 ± 0.0	-	7.4 ± 0.1 (∼ 18%)	7.4 ± 0.1 (∼ 18%)
	10^{-7}	11.3 ± 0.1	-	9.7 ± 0.1 (∼ 14%)	9.7 ± 0.1 (∼ 14%)
	10^{-8}	13.5 ± 0.1	-	11.7 ± 0.1 (∼ 13%)	11.7 ± 0.1 (∼ 13%)
Toluene*	10^{-6}	9.0 ± 0.0	-	8.0 ± 0.0 (∼ 11%)	7.9 ± 0.1 (∼ 12%)
	10^{-7}	11.6 ± 0.1	-	10.1 ± 0.1 (∼ 13%)	10.0 ± 0.0 (∼ 14%)
	10^{-8}	13.9 ± 0.1	-	12.4 ± 0.1 (∼ 11%)	12.3 ± 0.1 (∼ 12%)
Phenol**	10^{-6}	9.9 ± 0.1	-	10.0 ± 0.1 (∼ -1%)	8.1 ± 0.1 (∼ 18%)
	10^{-7}	12.2 ± 0.1	-	12.2 ± 0.1 (∼ 1%)	10.5 ± 0.1 (∼ 14%)
	10^{-8}	14.6 ± 0.1	-	14.4 ± 0.1 (∼ 2%)	12.3 ± 0.1 (∼ 15%)
Benzaldehyde**	10^{-6}	10.7 ± 0.1	-	10.5 ± 0.1 (∼ 2%)	9.0 ± 0.0 (∼ 16%)
	10^{-7}	13.2 ± 0.1	-	12.9 ± 0.0 (∼ 2%)	11.5 ± 0.1 (∼ 14%)
	10^{-8}	15.7 ± 0.1	-	15.1 ± 0.1 (∼ 4%)	14.2 ± 0.1 (∼ 10%)
p-Toluidine**	10^{-6}	10.0 ± 0.0	-	8.3 ± 0.1 (∼ 16%)	7.8 ± 0.1 (∼ 21%)
	10^{-7}	12.6 ± 0.1	-	11.1 ± 0.1 (∼ 12%)	10.2 ± 0.1 (∼ 20%)
	10^{-8}	14.9 ± 0.1	-	13.3 ± 0.1 (∼ 11%)	12.6 ± 0.1 (∼ 15%)

Continued on the next page

1-Propanol	10^{-6}	9.0 ± 0.0	6.0 ± 0.1 (∼ 33%)	6.5 ± 0.1 (∼ 27%)	6.5 ± 0.1 (∼ 28%)
	10^{-7}	11.0 ± 0.1	8.0 ± 0.0 (∼ 27%)	8.1 ± 0.1 (∼ 26%)	8.1 ± 0.1 (∼ 26%)
	10^{-8}	13.2 ± 0.1	9.8 ± 0.1 (∼ 25%)	10.1 ± 0.1 (∼ 23%)	10.1 ± 0.1 (∼ 23%)
1-Butanol	10^{-6}	9.0 ± 0.0	6.4 ± 0.1 (∼ 29%)	6.6 ± 0.1 (∼ 27%)	6.5 ± 0.1 (∼ 27%)
	10^{-7}	11.0 ± 0.0	8.1 ± 0.1 (∼ 27%)	8.2 ± 0.1 (∼ 26%)	8.2 ± 0.1 (∼ 26%)
	10^{-8}	13.2 ± 0.1	10.1 ± 0.1 (∼ 24%)	10.2 ± 0.1 (∼ 23%)	10.2 ± 0.1 (∼ 22%)
2-Butanol	10^{-6}	9.0 ± 0.0	7.2 ± 0.1 (∼ 20%)	7.0 ± 0.1 (∼ 22%)	7.1 ± 0.1 (∼ 22%)
	10^{-7}	11.1 ± 0.1	9.0 ± 0.1 (∼ 19%)	8.8 ± 0.1 (∼ 21%)	8.8 ± 0.1 (∼ 21%)
	10^{-8}	13.2 ± 0.1	10.9 ± 0.1 (∼ 18%)	10.8 ± 0.1 (∼ 19%)	10.8 ± 0.1 (∼ 19%)
1-Hexanol	10^{-6}	9.0 ± 0.0	6.4 ± 0.1 (∼ 29%)	6.5 ± 0.1 (∼ 28%)	6.5 ± 0.1 (∼ 28%)
	10^{-7}	10.9 ± 0.1	8.2 ± 0.1 (∼ 24%)	8.1 ± 0.1 (∼ 25%)	8.2 ± 0.1 (∼ 25%)
	10^{-8}	13.0 ± 0.0	10.3 ± 0.1 (∼ 21%)	10.2 ± 0.1 (∼ 21%)	10.2 ± 0.1 (∼ 21%)
Ethanol*	10^{-6}	9.1 ± 0.0	-	7.2 ± 0.1 (∼ 21%)	7.1 ± 0.1 (∼ 21%)
	10^{-7}	11.3 ± 0.1	-	9.0 ± 0.0 (∼ 20%)	9.0 ± 0.0 (∼ 20%)
	10^{-8}	13.7 ± 0.1	-	10.9 ± 0.0 (∼ 20%)	10.8 ± 0.1 (∼ 21%)
2-Propanol*	10^{-6}	9.0 ± 0.0	-	7.1 ± 0.1 (∼ 21%)	7.0 ± 0.0 (∼ 22%)
	10^{-7}	11.3 ± 0.1	-	9.1 ± 0.1 (∼ 20%)	9.0 ± 0.1 (∼ 20%)
	10^{-8}	13.7 ± 0.1	-	11.0 ± 0.0 (∼ 19%)	10.9 ± 0.0 (∼ 20%)
2-Hexanol*	10^{-6}	9.0 ± 0.0	-	7.0 ± 0.1 (∼ 23%)	7.0 ± 0.1 (∼ 22%)
	10^{-7}	11.0 ± 0.0	-	8.6 ± 0.1 (∼ 22%)	8.6 ± 0.1 (∼ 22%)
	10^{-8}	13.0 ± 0.0	-	10.5 ± 0.1 (∼ 19%)	10.5 ± 0.1 (∼ 19%)
1-Heptanol*	10^{-6}	9.0 ± 0.0	-	6.6 ± 0.1 (∼ 27%)	6.6 ± 0.1 (∼ 27%)
	10^{-7}	10.9 ± 0.1	-	8.3 ± 0.1 (∼ 24%)	8.2 ± 0.1 (∼ 24%)
	10^{-8}	13.0 ± 0.0	-	10.2 ± 0.1 (∼ 21%)	10.3 ± 0.1 (∼ 21%)

Table S6: Average of the number of iterations obtained using the predicted density matrix as guess for the calculation on the same geometry. The results refer to the (3,8)-models trained targeting the density matrix ($r_{\text{cut}} = 4.0$). The first column specifies the molecule, the second column indicates the convergence tolerance for the SCF procedure, the third column reports the average of the number of iterations obtained using the default guess available in the Gaussian program and in the last three columns we report the average number of iterations obtained with our predicted density matrix as guess, using the specific models, the unified model and the Unified Model-A, respectively. The values reported in parentheses indicate the percentage of reduction with respect to the iterations required by the default guess to converge. The test molecules with a superscript $*$ are not included in the training process at all, and those with $**$ are involved in the training of Unified Model-A, with only 10 frames each included.

Molecule	Tol.	Default guess	Specific Model	Unified Alcohols Model
1-Propanol	10^{-6}	9.0 ± 0.0	6.0 ± 0.1 (∼ 33%)	6.4 ± 0.1 (∼ 29%)
	10^{-7}	11.0 ± 0.1	8.0 ± 0.0 (∼ 27%)	8.1 ± 0.1 (∼ 27%)
	10^{-8}	13.2 ± 0.1	9.8 ± 0.1 (∼ 25%)	10.0 ± 0.1 (∼ 24%)
1-Butanol	10^{-6}	9.0 ± 0.0	6.4 ± 0.1 (∼ 29%)	6.4 ± 0.1 (∼ 29%)
	10^{-7}	11.0 ± 0.0	8.1 ± 0.1 (∼ 27%)	8.1 ± 0.1 (∼ 26%)
	10^{-8}	13.2 ± 0.1	10.1 ± 0.1 (∼ 24%)	10.2 ± 0.1 (∼ 23%)
2-Butanol	10^{-6}	9.0 ± 0.0	7.2 ± 0.1 (∼ 20%)	6.9 ± 0.1 (∼ 24%)
	10^{-7}	11.1 ± 0.1	9.0 ± 0.1 (∼ 19%)	8.4 ± 0.1 (∼ 24%)
	10^{-8}	13.2 ± 0.1	10.9 ± 0.1 (∼ 18%)	10.5 ± 0.1 (∼ 21%)
1-Hexanol	10^{-6}	9.0 ± 0.0	6.4 ± 0.1 (∼ 29%)	6.2 ± 0.1 (∼ 31%)
	10^{-7}	10.9 ± 0.1	8.2 ± 0.1 (∼ 24%)	8.1 ± 0.1 (∼ 26%)
	10^{-8}	13.0 ± 0.0	10.3 ± 0.1 (∼ 21%)	10.1 ± 0.1 (∼ 22%)
Ethanol*	10^{-6}	9.1 ± 0.0	-	7.0 ± 0.0 (∼ 23%)
	10^{-7}	11.3 ± 0.1	-	8.4 ± 0.1 (∼ 26%)
	10^{-8}	13.7 ± 0.1	-	10.4 ± 0.1 (∼ 24%)
2-Propanol*	10^{-6}	9.0 ± 0.0	-	7.0 ± 0.0 (∼ 22%)
	10^{-7}	11.3 ± 0.1	-	8.4 ± 0.1 (∼ 25%)
	10^{-8}	13.7 ± 0.1	-	10.5 ± 0.1 (∼ 23%)
2-Hexanol*	10^{-6}	9.0 ± 0.0	-	6.7 ± 0.1 (∼ 26%)
	10^{-7}	11.0 ± 0.0	-	8.4 ± 0.1 (∼ 24%)
	10^{-8}	13.0 ± 0.0	-	10.5 ± 0.1 (∼ 20%)
1-Heptanol*	10^{-6}	9.0 ± 0.0	-	6.4 ± 0.1 (∼ 29%)
	10^{-7}	10.9 ± 0.1	-	8.1 ± 0.1 (∼ 26%)
	10^{-8}	13.0 ± 0.0	-	10.2 ± 0.1 (∼ 22%)

Table S7: Average of the number of iterations obtained using the predicted density matrix as guess for the calculation on the same geometry. The results refer to the (3,8)-models trained targeting the density matrix. The first column specifies the molecule, the second column indicates the convergence tolerance for the SCF procedure, the third column reports the average of the number of iterations obtained using the default guess available in the Gaussian program and in the last two columns we report the average number of iterations obtained with our predicted density matrix as guess, using the specific models and the unified model trained on alcohols, respectively. The values reported in parentheses indicate the percentage of reduction with respect to the iterations required by the default guess to converge. The test molecules with a superscript $*$ are not included in the training process at all.

Molecule	RMSE
Acetaldehyde	$1.7 \cdot 10^{-5}$
Acrolein	$3.1 \cdot 10^{-5}$
Aniline	$3.0 \cdot 10^{-5}$
o-Toluidine	$4.1 \cdot 10^{-4}$
m-Toluidine	$4.2 \cdot 10^{-4}$
1-Propanol	$2.4 \cdot 10^{-4}$
1-Butanol	$3.0 \cdot 10^{-4}$
2-Butanol	$2.3 \cdot 10^{-4}$
1-Hexanol	$5.7 \cdot 10^{-4}$

Table S8: RMSEs obtained by the (3,8)-models trained on different datasets ($r_{\text{cut}} = 4.0$). Errors are evaluated on the isolated geometries optimized at the ωB97X-D/6-31G(d) level of theory.

S7 Additional plots

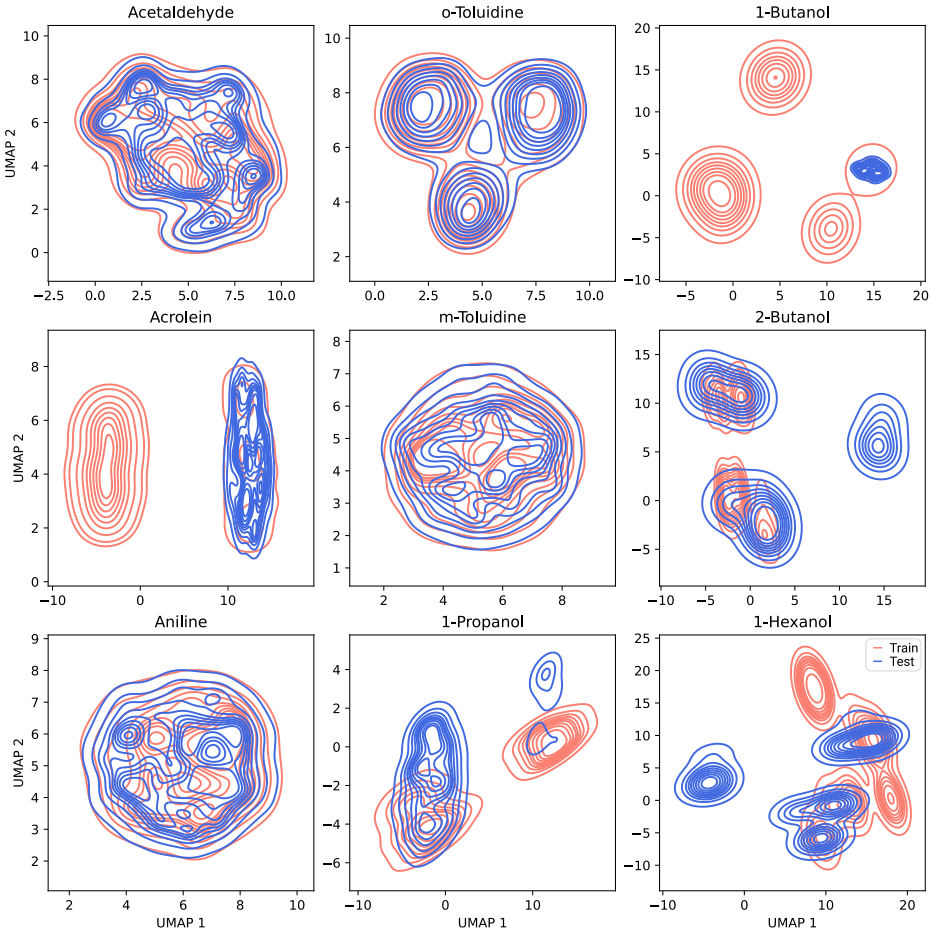

Figure S3: Visualization of the training (pink) and test (blue) sets for the specific models. Each plot is generated by computing the Coulomb matrix as input to a UMAP, and plotting the first two components.

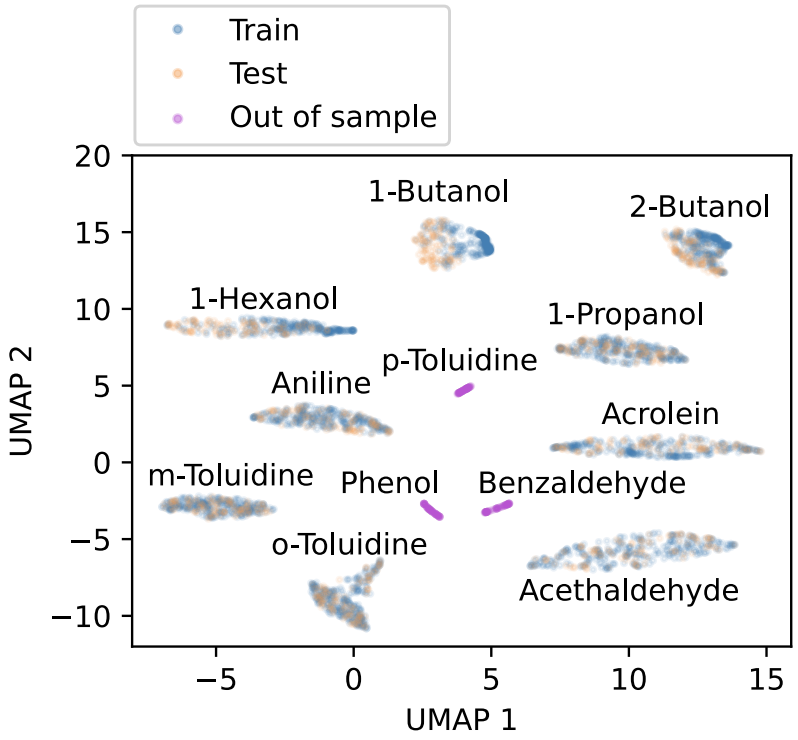

Figure S4: Visualization of the training (blue) and test (yellow) sets for the Unified Model, along with phenol, p-toluidine, and benzaldehyde (purple), which the model struggles to predict. For each molecule, the eigenvalues of the Coulomb matrix (zero-padded and sorted in decreasing order) were provided as input to a single UMAP computed on the full dataset, and the first two components are plotted.

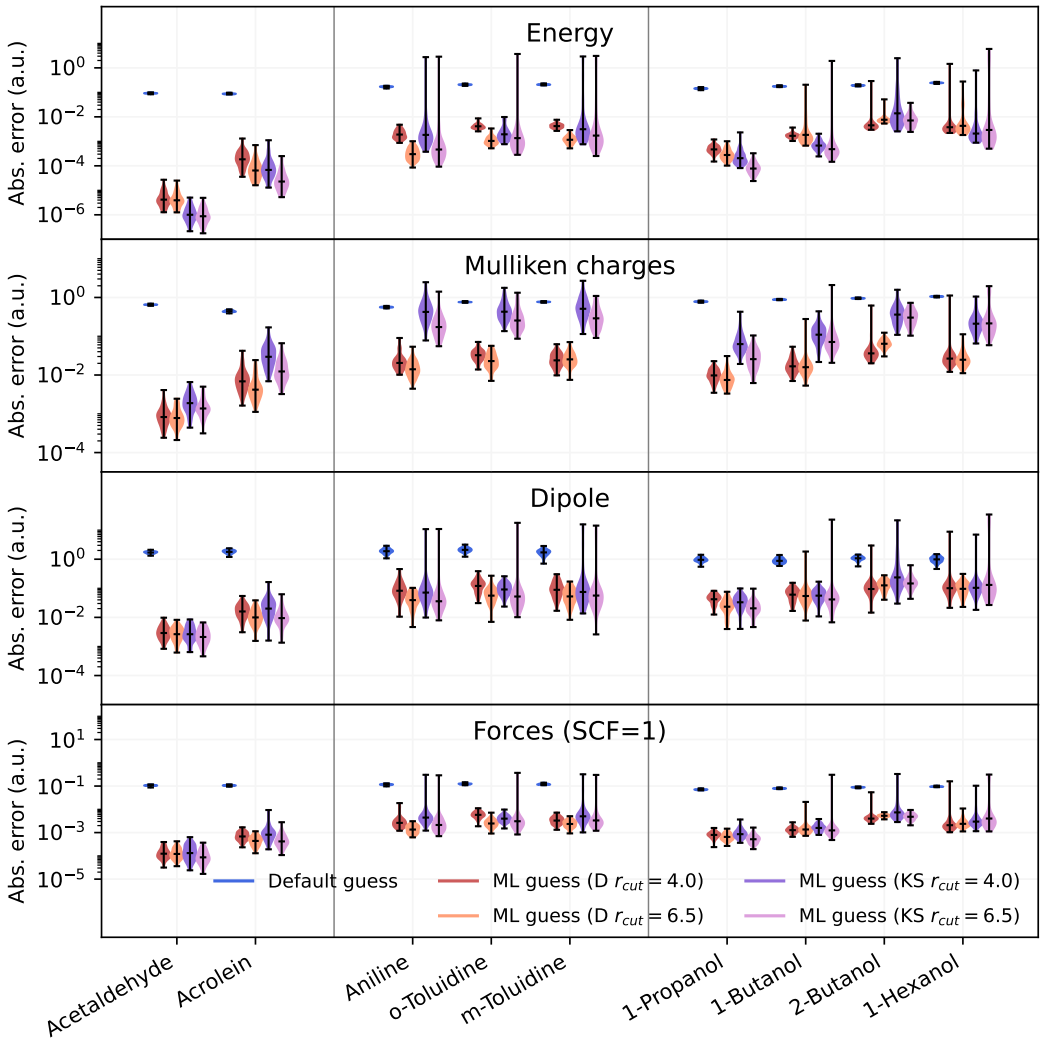

Figure S5: Plot of the error (in logarithmic scale) for energy, Mulliken charges, dipole moment and the error for forces after a single SCF cycle. The blue line represents the default guess provided by Gaussian, red and orange lines refer to models trained on density matrices with cutoff of 4.0 and 6.5 Å, respectively, and violet and pink lines refer to models trained on Kohn-Sham matrices with cutoff of 4.0 and 6.5 Å, respectively.

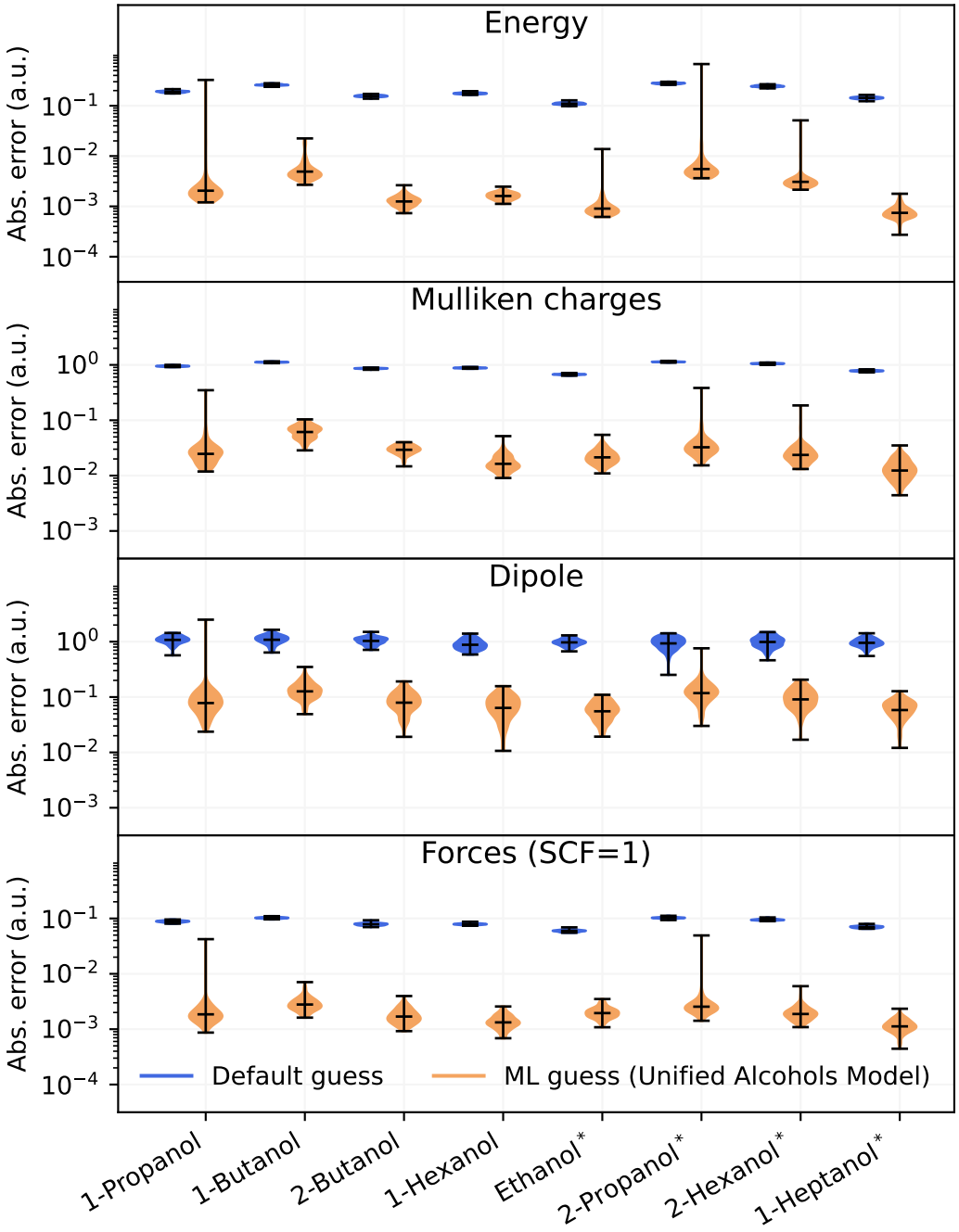

Figure S6: Plot of the error (in logarithmic scale) for energy, Mulliken charges, dipole moment and the error for forces after a single SCF cycle. The blue line represents the default guess provided by Gaussian, while the orange line corresponds to the density matrix predicted using the unified alcohols model. The test molecules with a superscript $*$ are not included in the training process at all.

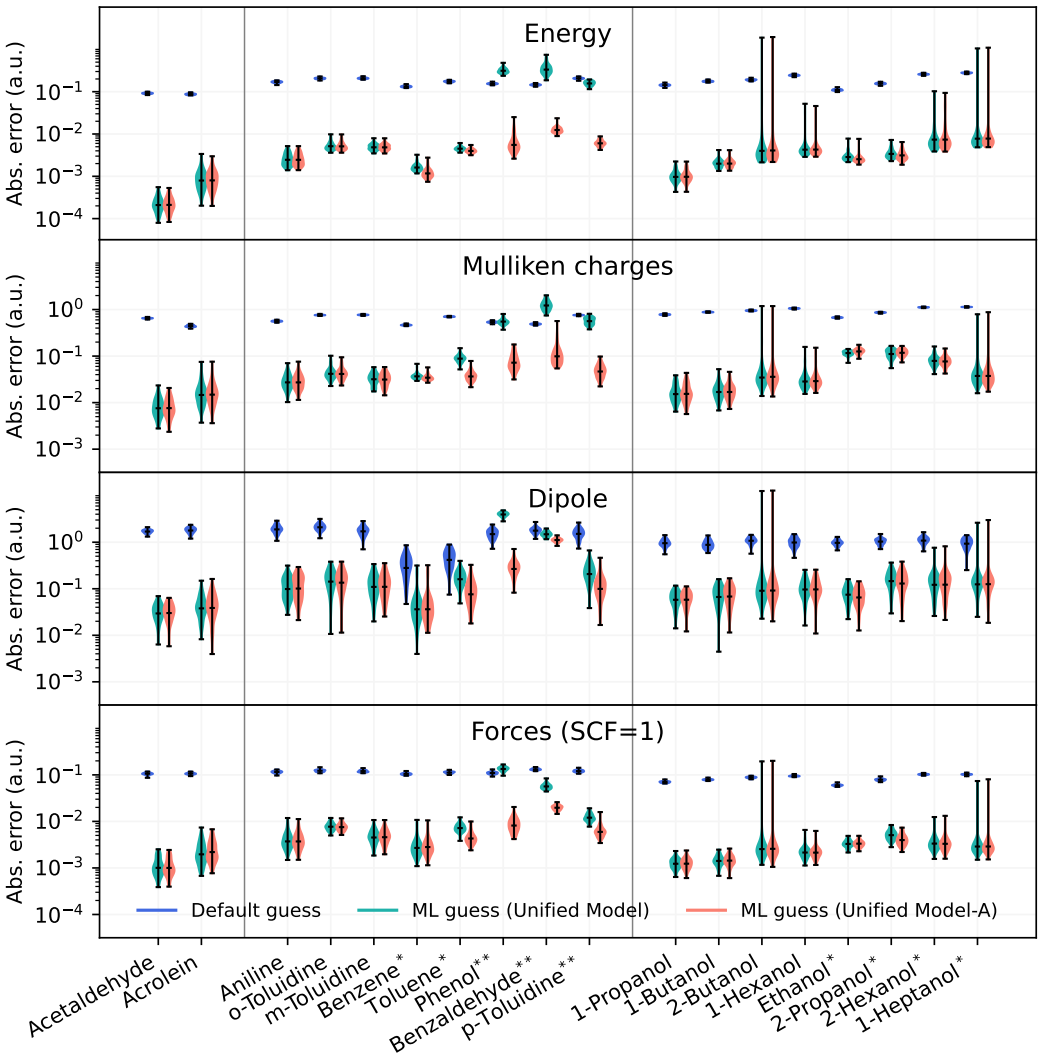

Figure S7: Plot of the error (in logarithmic scale) for energy, Mulliken charges, dipole moment and the error for forces after a single SCF cycle. The blue line represents the default guess provided by Gaussian, while the sea-green and pink lines correspond to the density matrix predicted using the unified model and Unified Model-A, respectively. The test molecules with a superscript $*$ are not included in the training process at all, and those with $**$ are involved in the training of Unified Model-A, with only 10 frames each included.

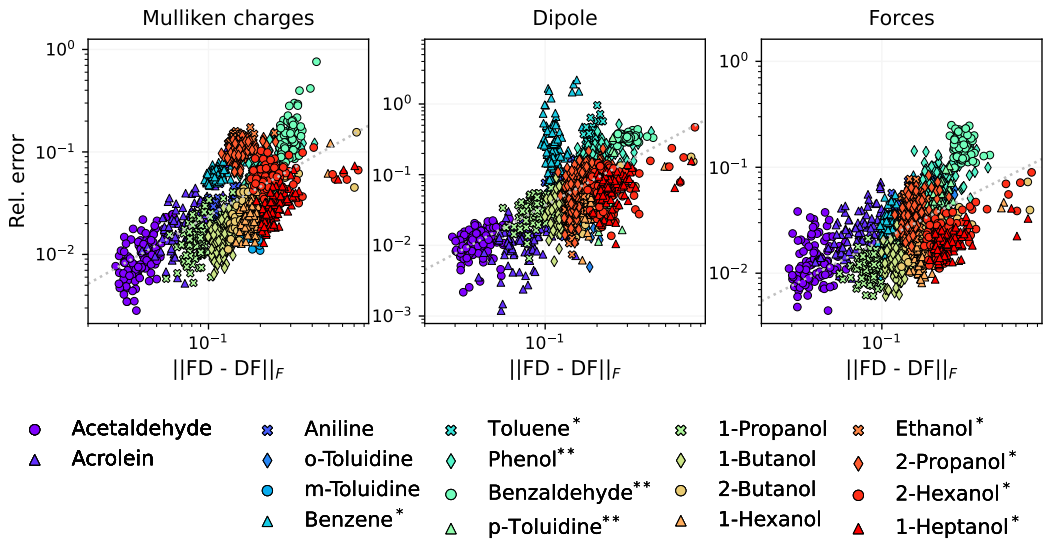

Figure S8: Plot of the Frobenius norm of the commutator between F and D versus the relative error in Mulliken charges, dipole moment and atomic forces, using the density matrix predicted by Unified Model-A as a guess. The test molecules with a superscript $*$ are not included in the training process at all, and those with $**$ are involved in the training of Unified Model-A, with only 10 frames each included. As expected, for the dipole moment of benzene, which is close to zero, a clear correlation is not observed.

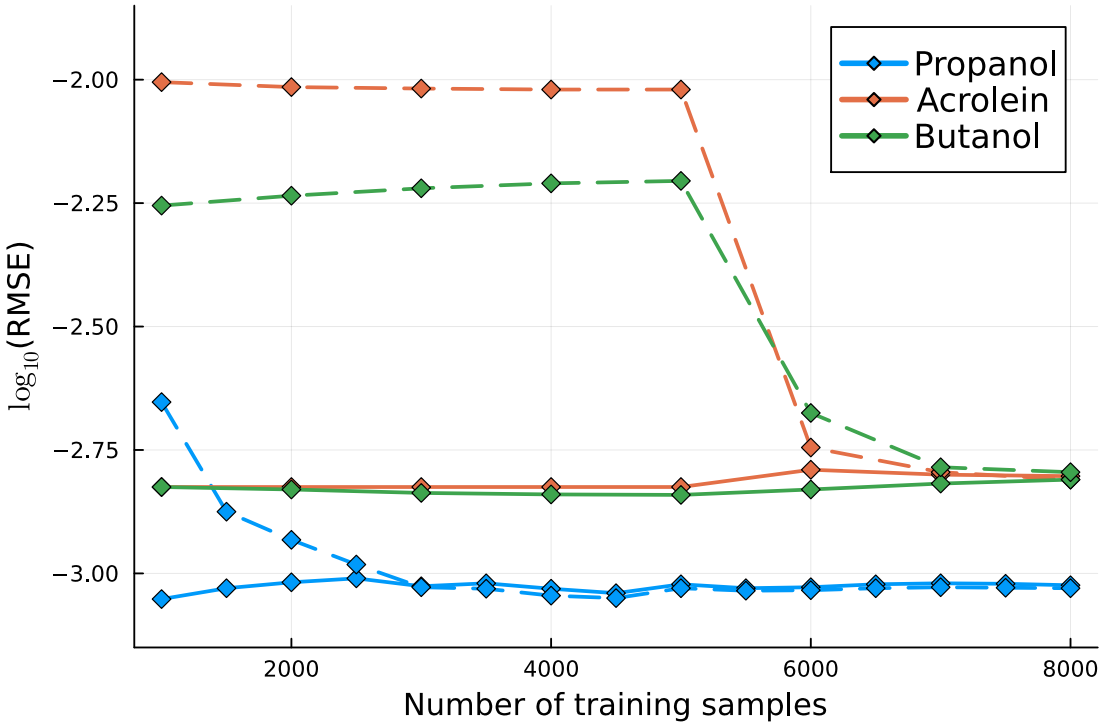

Figure S9: Plot of the relationship between the RMSEs and the number of training configurations for different molecules obtained with the model of correlation order $\nu = 2$ and maximum polynomial degree $d_{\max} = 6$. The solid and dashed lines refer to training set and test set RMSEs, respectively. We pick the amount of training data points for the specific models for each molecule by finding a balance between the training and test set errors.

References

[1] Liwei Zhang et al. "Equivariant analytical mapping of first principles Hamiltonians to accurate and transferable materials models". In: *npj Computational Materials* 8.1 (July 2022). ISSN: 2057-3960. DOI: 10.1038/s41524-022-00843-2.