

Machine Learning a Molecular Hamiltonian for Predicting Electron Dynamics

Harish S. Bhat · Karnamohit Ranka · Christine M. Isborn

Received: date / Accepted: date

Abstract We develop a computational method to learn a molecular Hamiltonian matrix from matrix-valued time series of the electron density. As we demonstrate for three small molecules, the resulting Hamiltonians can be used for electron density evolution, producing highly accurate results even when propagating 1000 time steps beyond the training data. As a more rigorous test, we use the learned Hamiltonians to simulate electron dynamics in the presence of an applied electric field, extrapolating to a problem that is beyond the field-free training data. We find that the resulting electron dynamics predicted by our learned Hamiltonian are in close quantitative agreement with the ground truth. Our method relies on combining a reduced-dimensional, linear statistical model of the Hamiltonian with a time-discretization of the quantum Liouville equation within time-dependent Hartree Fock theory. We train the model using a least-squares solver, avoiding numerous, CPU-intensive optimization steps. For both field-free and field-on problems, we quantify training and propagation errors, highlighting areas for future development.

Keywords electron dynamics · electron density · Hamiltonian · machine learning · system identification

1 Introduction

An intriguing new application of machine learning is to predict the dynamical electronic properties of a molecular system [8, 16, 19], which is essential to understanding phenomena such as charge transfer and response to an applied laser field. When discussing such electron dynamics predictions, we must start with the electronic time-dependent Schrödinger equation (TDSE):

$$i \frac{d\Psi(\mathbf{r}, t)}{dt} = \hat{H}(\mathbf{r}, t)\Psi(\mathbf{r}, t). \quad (1)$$

Here $\hat{H}(\mathbf{r}, t)$ is the electronic Hamiltonian operator that operates on the time-dependent many-body electronic wave function $\Psi(\mathbf{r}, t)$, where \mathbf{r} represents the spatial and spin coordinates of all electrons. One can derive from (1) an evolution equation for the time-dependent density operator. This operator equation can be represented in a finite-dimensional basis, yielding a matrix system of ordinary differential equations:

$$i \frac{d\mathbf{P}'(t)}{dt} = \left[\mathbf{H}'(t), \mathbf{P}'(t) \right]. \quad (2)$$

We call this the quantum Liouville-von Neumann equation. Boldface capital letters denote matrices, representations of operators in particular bases. Primes denote representations of operators in an orthonormal basis. Here $\mathbf{P}'(t)$ and $\mathbf{H}'(t)$ are time-dependent density and Hamiltonian matrices, respectively. The square brackets on the right-hand side denote a commutator; for matrices \mathbf{A} and \mathbf{B} , the commutator is $[\mathbf{A}, \mathbf{B}] = \mathbf{AB} - \mathbf{BA}$.

Harish S. Bhat
Applied Mathematics Department
University of California, Merced
5200 N. Lake Rd., Merced, CA 95343
E-mail: hbhat@ucmerced.edu

Karnamohit Ranka
Chemistry Department
University of California, Merced
5200 N. Lake Rd., Merced, CA 95343
E-mail: kranka@ucmerced.edu

Christine M. Isborn
Chemistry Department
University of California, Merced
5200 N. Lake Rd., Merced, CA 95343
E-mail: cisborn@ucmerced.edu

Inspecting a particular molecular system, one determines and writes the system’s Hamiltonian, a sum of kinetic and potential energy operators. As the Hamiltonian includes spatial derivatives within the kinetic energy operator, (1) will be a partial differential equation (PDE). For an N -electron system, ignoring spin, the PDE (1) will feature $3N$ spatial degrees of freedom. As N increases beyond $N = 1$, it becomes intractable to solve (1) directly for the time-dependent many-body wave function $\Psi(\mathbf{r}, t)$, even with modern numerical analysis and high-performance computing. For this reason, molecular electronic structure and dynamics calculations typically use simplified, mean-field approaches. One such approach is time-dependent Hartree-Fock (TDHF) [12, 29] theory, which solves (2) based on a simplified form of the wave function. In HF theory, we approximate the many-body wave function using a Slater determinant, an anti-symmetrized product of single-particle orbitals $\phi_i(\mathbf{r}, t)$, where \mathbf{r} now represents the spatial and spin coordinates of one electron. This approximation leads to a modified form of the Hamiltonian \hat{H} that appears in (1)—for details, see (3) below. Within HF theory, we then call (2) the TDHF equation.

Equation (2), used within TDHF theory or an alternative, similar approach called time-dependent density functional theory, is used in atomic, molecular, and materials calculations to simulate the dynamic electronic response to a perturbation, including predicting charge transfer and spectroscopic properties [13, 21, 25, 26, 32, 36, 50]. In these physical science settings, one starts with a system of interest, e.g., a molecule in an applied electric field. The system’s atomic configuration completely determines the Hamiltonian \mathbf{H}' and therefore the right-hand side of (2). Starting from an initial condition, the typical workflow is then to numerically solve (2) forward in time to generate simulations of interest, i.e., to generate $\mathbf{P}'(\mathbf{t})$ for $t > 0$ for a given perturbation.

Note that we write $\mathbf{H}'(t)$ to encapsulate two types of dependence on time t . First, \mathbf{H}' can depend explicitly on time—we will see this below when we consider \mathbf{V}_{ext} , an external, time-dependent potential. Second, within HF theory, even if \mathbf{H}' does not depend explicitly on time, it *is* in general a function of the density $\mathbf{P}'(t)$. In summary, $\mathbf{H}'(t)$ is shorthand for $\mathbf{H}'(t, \mathbf{P}'(t))$. This implies that (2) is in fact a *nonlinear* system.

Main Contribution. In this paper, we address a *system identification* problem for (2). Our main contribution is a computational method to estimate the molecular field-free matrix Hamiltonian $\mathbf{H}'(t)$ from time series observations of density matrices $\mathbf{P}'(t)$. By building a data-driven model of \mathbf{H}' , we identify the right-hand side of (2). We use a linear model for \mathbf{H}' , formulate a quadratic

loss function that stems from discretizing (2) in time, and eliminate unnecessary degrees of freedom. Thus we reduce model training to a least-squares problem. We demonstrate this method using training data consisting of density matrices $\mathbf{P}(t)$ for three small molecules.

Among other tests, we use the machine-learned (ML) Hamiltonian to propagate, i.e., to solve (2) forward in time. We find that using the ML Hamiltonian instead of the exact Hamiltonian results in a small, acceptable level of propagation error, even on a time interval that is twice the length of the time interval used for training. We then add a time-dependent external potential to our machine-learned, field-free Hamiltonian; we propagate forward in time using this augmented Hamiltonian. For each of the three molecules we consider, the resulting solutions are in close quantitative agreement with simulations that use the exact Hamiltonian. In short, our machine-learned Hamiltonian extrapolates well to a dynamical setting that differs from that of the training data.

To our knowledge, despite the surge of interest in applying machine learning to molecular simulation [1, 6, 7, 10, 17, 18, 23, 27, 31, 33–35, 37, 39–43, 47, 48], there are no other procedures in the literature to estimate molecular Hamiltonians from density matrix time series. Our work shares goals with other efforts to learn Hamiltonians, or energy functions and functionals that are ingredients in Hamiltonians. In this space, we primarily see efforts to learn classical Hamiltonians from time series [4, 5, 9, 22, 28, 30, 38, 46, 49] as well as efforts to learn quantum Hamiltonians or potentials for time-independent problems [2, 3, 15, 20, 24]. Recently, a neural network method to learn the exchange-correlation functional in time-dependent density functional theory has been developed [44]; solutions of the corresponding TDSE are used to train the networks.

In this paper, we consider small molecular systems modeled with a small basis set in order to focus on methodological development and careful analysis of errors. The present work forms a foundation on which we can build towards studying systems and theories (such as time-dependent density functional theory) in which the underlying potentials and functionals have yet to be completely determined. This is the overarching motivation for pursuing the present work.

Extrapolation. To clarify what we mean by extrapolation, let us consider a classical mass-spring system; one end of the spring is fixed and the other is connected to the mass. Let $F(t)$ be an external force applied to the mass, and let $x(t)$ denote the displacement from equilibrium of the mass at time t . We start with time series of $x(t)$ for a system where $x(0) \neq 0$ but $F(t) \equiv 0$, i.e.,

a field-free system. With this training data, suppose we seek a machine learning model that can predict the mass' position accurately, even when we switch on an applied forcing, e.g., $F(t) = A \cos(\omega t)$.

We view the task of (i) training with field-off data and (ii) predicting for field-on systems as an extrapolation task. For this classical example, one approach is to learn a model of $V(x)$, the spring's potential. Equipped with a sufficiently accurate model of V , the field-free dynamics can be predicted from Newton's second law: $m\ddot{x} = -V'(x)$. Again assuming that V is sufficiently accurate, we can predict the behavior of the mass-spring system accurately when the field is on, by solving $\ddot{x} = -V'(x) + F(t)$. Suitably expanded in size and complexity to the TDHF equation (2), this is essentially what we do in the present work.

2 Physical Considerations

2.1 Time-Dependent Hartree-Fock

In Section 1, we provided a highly summarized conceptual overview of deriving the TDHF equation from the TDSE. Here we expand on this overview and give more mathematical details. All equations use atomic units, with $e^2 = \hbar = m_e = 1$. An external perturbation, such as an applied electric field, within the Hamiltonian will give rise to the time-evolution of the wave function that dictates all properties of a quantum electronic system.

The molecules studied are closed-shell systems—all electrons in the system are spin-paired. Within HF theory, each pair of electrons with opposite spins can be described by the same spatial function ϕ_i , which is referred to as *restricted* HF theory. For closed-shell systems within such a formalism, the need to solve for N spatial orbitals occupied by N electrons reduces to solving for $(N/2)$ spatial orbitals, each of these doubly occupied to give a total of N electrons [45].

Given this choice, our Hamiltonian is then

$$\hat{H}(\mathbf{r}, t) = \sum_i^{N/2} \left[-\frac{\nabla^2}{2} - \sum_A \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} + \hat{V}_{\text{ext}}(\mathbf{r}, t) \right] + \sum_i^{N/2} \left[\sum_j^{N/2} 2 \times \int d\mathbf{r}' \frac{\phi_j^*(\mathbf{r}', t) \phi_j(\mathbf{r}', t)}{|\mathbf{r} - \mathbf{r}'|} - \int d\mathbf{r}' \frac{\phi_j^*(\mathbf{r}', t) \mathfrak{P}(\mathbf{r}, \mathbf{r}') \phi_j(\mathbf{r}', t)}{|\mathbf{r} - \mathbf{r}'|} \right], \quad (3)$$

where the first group includes one-electron terms summed over half the number of electrons (the total number of *spatially unique* electrons): the electron kinetic energy, the electron-nuclear attraction to all nuclei A with nuclear charge Z_A at fixed position \mathbf{R}_A , and the external

potential \hat{V}_{ext} . In this work, the external perturbation is an electric field treated classically within the dipole approximation $\hat{V}_{\text{ext}}(\mathbf{r}, t) = \mathbf{E}(t) \cdot \hat{\mu}(\mathbf{r})$. The first term in the second group is the electron-electron Coulomb repulsion. The operator $\mathfrak{P}(\mathbf{r}, \mathbf{r}')$ used in the last term denotes the operation of permutation between electrons represented by coordinate-variables \mathbf{r} and \mathbf{r}' . This term is known as the exchange operator and arises from the antisymmetry of the electronic wave function. The second group collectively represents the electron-electron interaction operator. This operator depends on the instantaneous charge distribution of all other electrons, resulting in an implicit time-dependence of the electronic Hamiltonian (in addition to the explicit time-dependence due to \hat{V}_{ext}).

We next define a (reduced one-body) density operator, $\hat{\rho}$, that allows us to represent the total density of electrons [11]:

$$\hat{\rho}(\mathbf{r}, t) = \sum_p f_p \phi_p(\mathbf{r}, t) \phi_p^*(\mathbf{r}, t) = \sum_p f_p |\phi_p\rangle \langle \phi_p|, \quad (4)$$

where f_p is the occupation of orbital ϕ_p : in a restricted, closed-shell system, $f_p = 2$ (if ϕ_p is occupied) or 0 (if ϕ_p is unoccupied). The corresponding density matrix (\mathbf{P}) is represented in the basis of $\{\phi_i\}$ as:

$$P_{ij}(t) = \int d\mathbf{r} \phi_i^*(\mathbf{r}, t) \hat{\rho}(\mathbf{r}, t) \phi_j(\mathbf{r}, t) = \langle \phi_i | \hat{\rho} | \phi_j \rangle. \quad (5)$$

We can now write down the Liouville-von Neumann equation in operator form:

$$i \frac{d\hat{\rho}(\mathbf{r}, t)}{dt} = [\hat{H}(\mathbf{r}, t), \hat{\rho}(\mathbf{r}, t)]. \quad (6)$$

This is an operator equation for the evolution of $\hat{\rho}$. The time-dependent molecular orbitals ϕ_i are often created from a linear combination of basis functions $\{\chi_\mu\}$, as $\phi_i = \sum_\mu c_{\mu,i}(t) \chi_\mu$, where $c_{\mu,i}(t)$ are the time-dependent coefficients. The elements of the density matrix \mathbf{P} are given in this basis by

$$P_{\mu\nu}(t) = \sum_p f_p c_{\mu,p}(t) c_{\nu,p}^*(t). \quad (7)$$

We transform \mathbf{P} to an orthonormal basis, yielding \mathbf{P}' (see Appendix). We then use the Liouville equation for the density operator to write the TDHF equation in matrix form

$$i \frac{d\mathbf{P}'(t)}{dt} = \left[\mathbf{H}'(t), \mathbf{P}'(t) \right], \quad (8)$$

where $\mathbf{H}'(t)$ is the Hamiltonian (or Fock) matrix that results from integrating (3) over \mathbf{r} in the orthonormal basis. In this work, primed notations (e.g., \mathbf{H}' , \mathbf{P}') are

used for matrices in the orthonormal basis and unprimed notations for matrices (*e.g.*, \mathbf{H} , \mathbf{P}) in the atomic orbital (AO) basis.

Although it is straightforward to write down the molecular Hamiltonian if the atomic positions are known, integration of the Hamiltonian within a given basis is more challenging and encodes ground and excited state information about the molecule within that basis. Learning the integrated form of the molecular matrix Hamiltonian is thus key to determining the electron dynamics.

2.2 Molecules and Exact Hamiltonian

Here we study three diatomic molecules: H_2 , HeH^+ , and LiH . The atoms in each of these diatomic systems are placed along the z -axis, equidistant from the origin. The interatomic separations for H_2 , HeH^+ and LiH are 0.74 Å, 0.772 Å, and 1.53 Å, respectively. These simple molecular systems increase in complexity, going from a symmetric two-electron homonuclear diatomic, to a two-electron heteronuclear diatomic, to a four-electron heteronuclear diatomic. The basis set used for these calculations is STO-3G, a minimal basis set made of s and p atomic orbitals. For H_2 and HeH^+ , this results in two basis functions (a 2×2 matrix for \mathbf{P} and \mathbf{H}), and for LiH this results in six basis functions (a 6×6 matrix for \mathbf{P} and \mathbf{H} , although some elements of the matrices are zero due to the linear symmetry of the molecule, as discussed later).

For each molecule, the electronic structure code provides the integrals that are the components of the exact Hamiltonian matrix \mathbf{H} , expressed in the same AO basis set as the density matrices. Specifically, we obtain real, symmetric, constant-in-time matrices for the kinetic energy and electron-nuclear potential energy. We also obtain a 4-index tensor of evaluated integrals, which we use together with the time-dependent density matrices $\mathbf{P}(t)$ to compute the electron-electron potential energy term. These ingredients allow us to compute, for each molecule, the exact Hamiltonian. Electron density propagation with this exact Hamiltonian, both within the electronic structure code and within our propagation code, is compared to that from our ML model Hamiltonian.

3 Electron Density Matrix Data

There are two steps involved in generating the training and test sets of the time series of density matrix data:

1. Generating an initial condition (the initial density matrix).
2. Generating a trajectory using the initial condition and the differential equation (2) for propagation.

For the first step, the HF stationary state solution is determined self-consistently within the electronic structure code. The density matrix corresponding to the alpha-spin part of the solution, represented in the AO basis, is used as the initial condition. The second step involves propagating the initial density matrix using the TDHF equation.

We performed each of these steps with the Gaussian electronic structure program [14], using a locally modified development version.

3.1 Initial Conditions

We have calculated initial density matrices for field-free and static field conditions. For the field-free calculations, we set the \mathbf{V}_{ext} term to 0. For the static field, $E_z = 0.05$ a.u. (atomic units). Applying a static field creates an initial electron density that is not a stationary state of the field-free Hamiltonian and is often referred to as a delta-kick perturbation.

3.2 Trajectory Data

The density matrix from the initial condition calculation is used as the starting point for generating the real-time TDHF electron dynamics trajectory, *i.e.* $\mathbf{P}(t)$.

For the field-free trajectories, we set \mathbf{V}_{ext} to zero during propagation; we use the density matrix with the delta kick perturbation as the initial condition. These trajectories serve as the training data for the ML Hamiltonian. There is a particular rationale behind choosing the delta kick perturbation. First, consider that a perturbation that is localized at one point in time is, by Fourier duality, maximally spread out in frequency space. Hence such a perturbation necessarily excites all modes of the system. Second, consider that if we generate a trajectory that does not excite a mode of a system, we cannot expect that a machine learning technique will be able to learn that that mode exists, let alone how to represent the unobserved mode accurately in either the potential or the full Hamiltonian. By choosing a delta kick trajectory that excites all modes, we ensure that it is at least possible in principle to learn the full potential/Hamiltonian.

For the field-on trajectories, the field-free initial density matrix is used and \mathbf{V}_{ext} takes the following form during propagation:

$$\mathbf{V}_{\text{ext}}(t) = \sum_{i \in \{x, y, z\}} E_i \sin(\omega t) \boldsymbol{\mu}_i = 0.05 \sin(0.0428t) \boldsymbol{\mu}_z, \quad (9)$$

where the time t , the field-intensity E_i along axis i , and the field-frequency ω are expressed in a.u. Here the field is applied only along the z-direction and μ_z is the z-component of the dipole moment matrix in the AO basis. The sinusoidal field is switched on for one full cycle (around 3.55 fs) starting at $t = 0$. These field-on trajectories test the ML Hamiltonian in a regime quite outside the field-free training regime.

Using a propagation step-size of 0.002 fs, the total length of each trajectory is 20000 time-steps (thus, each trajectory is 40 fs long). The real-time TDHF implementation in Gaussian uses as its propagation scheme the modified midpoint unitary transformation (MMUT) algorithm [25].

4 Learning the Molecular Hamiltonian

For a particular molecule, suppose we are given time series $\{\mathbf{P}'(t_j)\}_{j=0}^N$ sampled on an equispaced temporal grid $t_j = j\Delta t$. We assume that $\mathbf{P}'(t)$, the continuous-time trajectory corresponding to our time series, satisfies (2). Our goal is to learn the Hamiltonian \mathbf{H}' . Assume that the Hamiltonian contains no explicit time-dependence—this can be ensured by generating training data with no external potentials (e.g., no external forcing). Then \mathbf{H}' is a Hermitian matrix of functions of \mathbf{P}' , the density matrix. Our strategy therefore consists of three steps: (i) develop a model of \mathbf{H}' with a finite-dimensional set of parameters β , (ii) derive from (2) a statistical model, and (iii) use the model with available data to estimate β .

Note that in order to obtain \mathbf{P}' , \mathbf{H}' from \mathbf{P} , \mathbf{H} , we transform from the AO basis to its canonical orthogonalization [45]. We do this because the TDHF equation (2) holds in an orthonormal basis; the AO basis by itself is not orthonormal. We leave the details of this transformation to the Appendix.

Let us split \mathbf{H}' into its real and imaginary parts: $\mathbf{H}' = \mathbf{H}'_R + i\mathbf{H}'_I$. By Hermitian symmetry, \mathbf{H}' is determined completely by the upper-triangular component of \mathbf{H}'_R (including the diagonal) and by the upper-triangular component of \mathbf{H}'_I (not including the diagonal). If \mathbf{H}' has size $M \times M$, there are $M(M+1)/2$ elements of \mathbf{H}'_R and $M(M-1)/2$ elements of \mathbf{H}'_I that we must model. Hence there are a total of M^2 real degrees of freedom, which we can represent as an $M^2 \times 1$ vector \mathbf{h}' . Note that we can apply this same real and imaginary splitting to \mathbf{P}' ; since it is also Hermitian, it can also be determined completely by a real vector \mathbf{p}' of dimension $M^2 \times 1$. Then we formulate the following linear model for $\mathbf{h}'(\mathbf{p}')$ —in what follows, we use \sim to

denote either statistical models or their parameters:

$$\tilde{\mathbf{h}}' = \tilde{\beta}_0 + \tilde{\beta}_1 \mathbf{p}' \quad (10)$$

Here $\tilde{\beta}_0$ has size $M^2 \times 1$, while $\tilde{\beta}_1$ has maximal size $M^2 \times M^2$. For the smaller molecules in our study (H_2 and HeH^+), where the STO-3G basis set leads to a dimension of $M = 2$, we use (10) with no modifications.

For LiH, a larger molecule, to handle entries of \mathbf{p}' that are identically zero, and also to dramatically reduce the computational effort required for training, we modify the basic model (10). We can understand these modifications very succinctly by saying that we reduce both the number of columns of $\tilde{\beta}_1$ and the number of rows of all vectors in (10), namely $\tilde{\mathbf{h}}'$, $\tilde{\beta}_0$, and \mathbf{p}' . In more detail, what we do is first form a set Z consisting of indices of, separately, the real and imaginary parts of training data matrices $\{\mathbf{P}'_R(t_j)\}_{j=0}^N$ and $\{\mathbf{P}'_I(t_j)\}_{j=0}^N$ that are *not* identically zero. Let us use the notation $\tilde{\mathbf{H}}'$ to denote the $M \times M$ Hermitian matrix that corresponds to the real vector $\tilde{\mathbf{h}}'$. For both \mathbf{P}' and also for our model Hamiltonian $\tilde{\mathbf{H}}'$, we restrict attention to upper-triangular matrix indices that are in the set Z . To illustrate this concretely, LiH in the STO-3G basis has dimension $M = 6$, and so, at each instant of time, both \mathbf{P}' and $\tilde{\mathbf{H}}'$ are determined by $M^2 = 36$ entries, of which 21 correspond to real parts and 15 correspond to imaginary parts. Of these, only 10 real parts and 6 imaginary parts are not identically zero. In this way, we reduce the overall dimension of (10) from 36 down to $10 + 6 = 16$. We defer the formal details of this procedure to the Appendix.

Note that (10) is by no means the only possible model. We have explored higher-order polynomial models that, while remaining linear in the parameters $\tilde{\beta}$, allow $\tilde{\mathbf{h}}'$ to depend nonlinearly on \mathbf{p}' . We have also explored models in which $\tilde{\mathbf{h}}'$ is allowed to depend explicitly on time t , including through Fourier terms such as $\sin(\omega t)$ and $\cos(\omega t)$. None of these choices led to any improvement in validation or test error, so we focus on the linear model (10).

Now that we have (10), we turn our attention to (2). Then we use a centered-difference approximation to derive from (2) the statistical model

$$i \frac{\mathbf{P}'(t_{j+1}) - \mathbf{P}'(t_{j-1})}{2\Delta t} = \left[\tilde{\mathbf{H}}'(\mathbf{P}'(t_j)), \mathbf{P}'(t_j) \right] + \epsilon_j, \quad (11)$$

with ϵ_j denoting error. With $\|\mathbf{A}\|_F^2 = \sum_{i,j} A_{ij}^2$, the squared Frobenius norm, we form the sum of squared

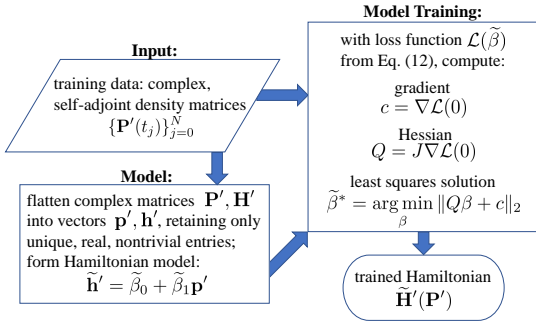


Fig. 1 Overall training procedure for learning the molecular, field-free Hamiltonian. In this paper, for each molecule, we train using time series with $N = 1000$. We use this field-free Hamiltonian to propagate for $2N = 2000$ steps; see Figures 2 and 3. We augment the learned Hamiltonian with an external potential (an electric field), yielding a field-on Hamiltonian that we use to propagate for $2N = 2000$ steps; see Figures 5 and 6.

errors loss function

$$\mathcal{L}(\tilde{\beta}) = \sum_{j=1}^{N-1} \left\| i \frac{\mathbf{P}'(t_{j+1}) - \mathbf{P}'(t_{j-1})}{2\Delta t} - \left[\tilde{\mathbf{H}}'(\mathbf{P}'(t_j)), \mathbf{P}'(t_j) \right] \right\|_F^2. \quad (12)$$

4.1 Reduction to Least Squares

The dependence of \mathcal{L} on $\tilde{\beta} = (\tilde{\beta}_0, \tilde{\beta}_1)$ is entirely through $\tilde{\mathbf{H}}'$. We estimate $\tilde{\beta}$ by solving the optimization problem $\tilde{\beta}^* = \arg \min_{\tilde{\beta}} \mathcal{L}(\tilde{\beta})$. Because (10) is linear in the parameters $\tilde{\beta}$, we observe that (12) must be quadratic in $\tilde{\beta}$. So, there exist constants Q (matrix), c (vector), and L_0 (scalar) such that

$$\mathcal{L}(\tilde{\beta}) = \frac{1}{2} \tilde{\beta}^T Q \tilde{\beta} + c^T \tilde{\beta} + \frac{L_0}{2}. \quad (13)$$

Here we can identify c as the gradient of \mathcal{L} with respect to $\tilde{\beta}$ evaluated at $\tilde{\beta} \equiv 0$, and Q as the Hessian of \mathcal{L} with respect to $\tilde{\beta}$. We compute this gradient and Hessian via automatic differentiation of \mathcal{L} . When Q is full rank, we have an exact minimizer $-Q^{-1}c$. As Q is typically rank deficient, we replace Q^{-1} with the Moore-Penrose pseudoinverse Q^\dagger :

$$\tilde{\beta}^* = -Q^\dagger c = \arg \min_{\tilde{\beta}} \|Q\tilde{\beta} + c\|_2. \quad (14)$$

When $(I - QQ^\dagger)c = 0$, the loss \mathcal{L} achieves its global minimum at $\tilde{\beta}^*$. For each of our molecules, we find that $\|(I - QQ^\dagger)c\|$ is small but non-zero. Still, we find empirically that (14) yields a nearly zero-norm gradient of

\mathcal{L} , as good as what can be achieved via other numerical optimization methods.

We have summarized the overall procedure in flowchart form in Figure 1. Equation (14) constitutes the end of the training procedure. In particular, we use a method in NumPy, `linalg.lstsq`, to compute (14), and so we avoid the full computation of Q^\dagger . Note that gradient-based optimization can also be used to minimize the loss (12), as we have verified. However, such a procedure requires millions of small steps, resulting in a training time (for LiH) that is 500-1000 times larger than the time required by a least-squares solver.

4.2 Error Metrics

Inserting (14) into (13) and using properties of the pseudoinverse, $(Q^\dagger)^T = (Q^T)^\dagger = Q^\dagger$ together with $Q^\dagger QQ^\dagger = Q^\dagger$, we obtain the training error

$$\mathcal{L}(\tilde{\beta}^*) = -\frac{1}{2} [c^T Q^\dagger c + L_0],$$

the value of the loss function at the optimal set of parameters. The training error measures a local-in-time error, essentially equivalent to starting at the training data point $\mathbf{P}'(t_j)$, propagating one step forward in time with our learned Hamiltonian (10) and comparing with the very next training data point $\mathbf{P}'(t_{j+1})$. Aggregating these *one-step errors*—squaring and summing their magnitudes—yields the training error $\mathcal{L}(\tilde{\beta}^*)$.

We contrast the training error with the *propagation error*. Once we have solved for the optimal parameter values $\tilde{\beta}^*$, the model Hamiltonian (10) is completely determined. Using this estimated Hamiltonian with the initial condition $\mathbf{P}'(0)$ from our training time series, we solve (2) forward in time using a Runge-Kutta scheme, generating our statistical estimates of $\tilde{\mathbf{P}}'(t_j)$ from $j = 1$ up to $j = 2N = 2000$, twice the length of the training data. For the Runge-Kutta integration, we set absolute and relative tolerances to 10^{-12} . We then define the propagation error to be

$$\mathcal{E} = \frac{1}{2N} \sum_{j=1}^{2N} \left\| \mathbf{P}'(t_j) - \tilde{\mathbf{P}}'(t_j) \right\|_F. \quad (15)$$

In contrast to the training error, (15) measures the divergence between two trajectories— \mathbf{P} (training) and $\tilde{\mathbf{P}}$ (propagation of ML Hamiltonian)—over *many* time steps. Both trajectories have exactly the same initial condition, and hence $j = 0$ is excluded from the sum. For $j > 0$, the two trajectories are computed using different numerical schemes (modified midpoint for the training data and Runge-Kutta for the ML Hamiltonian

	HeH ⁺	H ₂	LiH
$\mathcal{L}(\tilde{\beta}^*)$	4.75×10^{-6}	5.77×10^{-6}	2.30×10^{-6}
$\ \nabla\mathcal{L}(\tilde{\beta}^*)\ $	4.17×10^{-11}	3.44×10^{-11}	6.47×10^{-11}
\mathcal{E}	4.37×10^{-3}	4.89×10^{-3}	6.51×10^{-3}
\mathcal{E}_{Sch}	2.57×10^{-3}	2.50×10^{-3}	2.15×10^{-3}
\mathcal{E}_{Ham}	1.81×10^{-3}	2.40×10^{-3}	5.41×10^{-3}

Table 1 After training, we report the training loss and the norm of its gradient, along with three forms of propagation error. All results are for the field-free problem. Note that the training error is a sum of squared errors; for each molecule, if we divide by the training data length $N = 10^3$, we obtain mean-squared training errors that are all on the order of 10^{-9} , indicating approximately 4 decimal places of accuracy. The propagation errors show a roughly even breakdown into error due to different schemes versus error due to different Hamiltonians.

propagation) *and* different Hamiltonians. To control for scheme-related error, we compute

$$\mathcal{E}_{\text{Sch}} = \frac{1}{2N} \sum_{j=1}^{2N} \|\mathbf{P}(t_j) - \bar{\mathbf{P}}(t_j)\|_F, \quad (16)$$

where $\bar{\mathbf{P}}(t_j)$ is the result of propagating forward in time using the same Runge-Kutta scheme with the exact Hamiltonian \mathbf{H}' . This exact Hamiltonian is built by (i) extracting the Hamiltonian \mathbf{H} in the AO basis from the electronic structure output and then (ii) transforming \mathbf{H} to \mathbf{H}' using the procedure described in the Appendix. In (16), the two trajectories being compared have the same Hamiltonian and differ only in the numerical propagation schemes used to generate them. As a final error metric, we compute

$$\mathcal{E}_{\text{Ham}} = \frac{1}{2N} \sum_{j=1}^{2N} \|\tilde{\mathbf{P}}(t_j) - \bar{\mathbf{P}}(t_j)\|_F. \quad (17)$$

The two trajectories compared here are computed using the same Runge-Kutta scheme, but with different Hamiltonians. By the triangle inequality, we have $\mathcal{E} \leq \mathcal{E}_{\text{Sch}} + \mathcal{E}_{\text{Ham}}$. We may conceptualize this as breaking down the total error into the error due to different schemes (\mathcal{E}_{Sch}) and the error due to different Hamiltonians (\mathcal{E}_{Ham}).

5 Results

5.1 Training and Propagation Tests

We apply the procedure described in Section 4 to training time series of length $N = 1000$ for each of the three molecules HeH⁺, H₂, and LiH. See Section 3 for details on the generation of training data. The only additional preprocessing step here was to omit the first two time

steps, for each molecule, and to take the subsequent 1000 time steps as training data. This was carried out purely to avoid large numerical time-derivatives $d\mathbf{P}'/dt$ associated with the delta-kick perturbation at $t = 0$; these time-derivatives form a critical part of our loss function (12). We emphasize that these training trajectories were generated with no external potential/field, using delta-kick initial conditions described in Section 3.1.

We report the value of the loss and the norm of its gradient, after training, in the first two rows of Table 1. For each molecule, the training loss is of the order of 10^{-6} , which corresponds to an accuracy of roughly 4 decimal places. In order to visualize this accuracy, see Figures 2 and 3. For each molecule, we have plotted each of the non-zero real and imaginary components (note the y -axis labels) that fully determine the Hermitian density matrices $\mathbf{P}'(t_j)$ at each time step $t_j = j\Delta t$. In fact, in each panel, there are three curves: in black, we have plotted the actual training data produced by the electronic structure code; in blue, we have plotted $\bar{\mathbf{P}}(t_j)$, the result of propagating the exact Hamiltonian; and in red, we have plotted $\tilde{\mathbf{P}}(t_j)$, the result of propagating the ML Hamiltonian.

For HeH⁺ and H₂ (Fig. 2), the curves agree to a degree where they can hardly be distinguished. As we described above, due to the fact that in HF theory \mathbf{H}' depends on \mathbf{P}' , the TDHF equation (2) is nonlinear, and hence all of these oscillations are nonlinear oscillations. For LiH (Fig. 3), we can discern some divergence between the result of ML Hamiltonian propagation (red) and the other two curves, but only for those density matrix elements with relatively small variance. The sum of squares loss function (12) is biased in favor of fitting large-variance components; to avoid this, one could modify (12) to include weights that are inversely proportional to density element variances. The errors in Figure 3 consist primarily of oscillations about the black curve; the magnitudes of these oscillations are small and do not increase dramatically over time. Still, we should give the the machine-learned Hamiltonian credit for performing well when we use it to propagate for $2N = 2000$ steps, twice the length of the training data used. This hints at being able to use the machine-learned Hamiltonian to extrapolate beyond the field-free system used for training.

To understand more deeply the different sources of error, we refer to the final three rows of Table 1 together with the left panel of Figure 4. We think of \mathcal{E} as the overall RMS error between the training data \mathbf{P}' and our predicted trajectory $\tilde{\mathbf{P}}$, broken down into two components \mathcal{E}_{sch} and \mathcal{E}_{Ham} as explained above. If our goal is to track the training data, we incur errors of

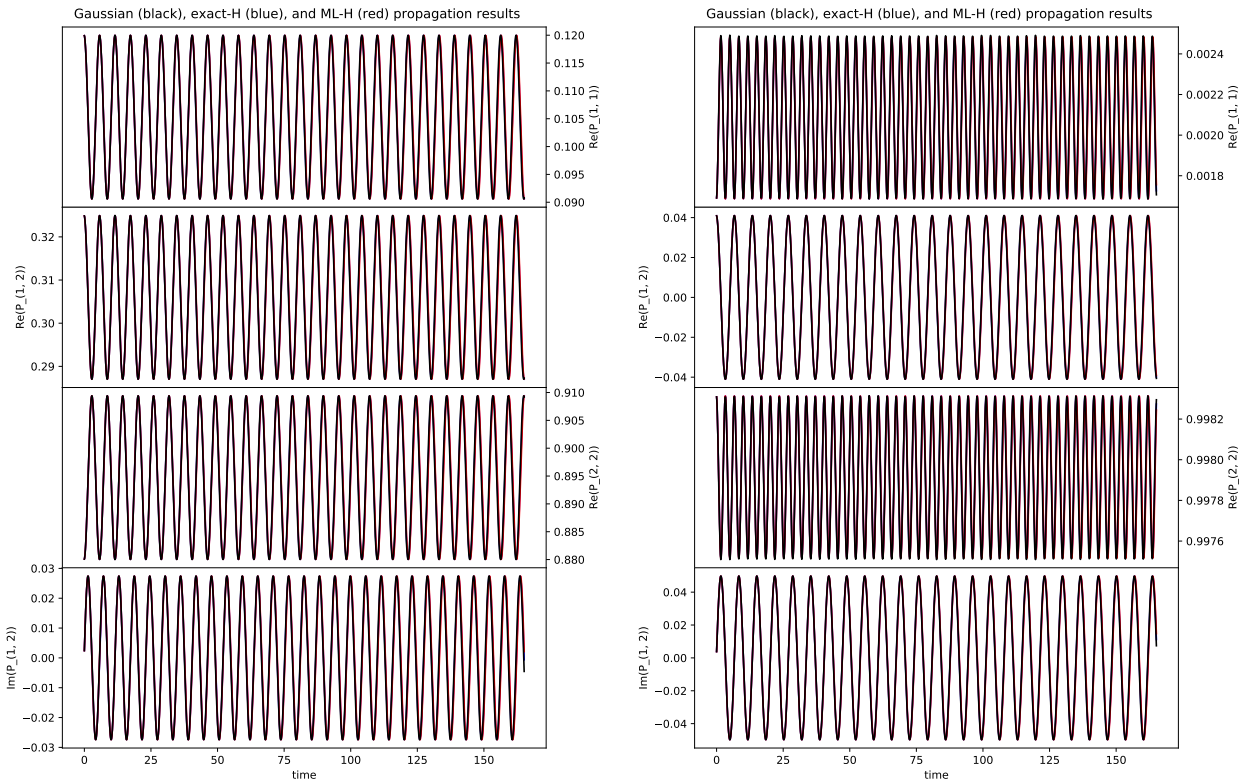


Fig. 2 HeH^+ (left) and H_2 (right) propagation with no field. For both molecules, we have plotted all unique real and imaginary parts of the time-dependent density matrices: actual training data (black), exact Hamiltonian propagation (blue), and ML Hamiltonian propagation (red). Note the close agreement of all three curves, on a time interval that is twice the length used for training.

	HeH^+	H_2	LiH
\mathcal{E}	3.59×10^{-4}	4.97×10^{-4}	4.86×10^{-3}
\mathcal{E}_{Sch}	2.94×10^{-4}	4.10×10^{-4}	4.87×10^{-3}
\mathcal{E}_{Ham}	7.22×10^{-5}	1.01×10^{-4}	1.33×10^{-4}

Table 2 For the field-on problem, we report three forms of propagation error corresponding to field-on versions of (15), (16), and (17). Here \mathcal{E} measures the difference between (i) propagation of the ML Hamiltonian plus \mathbf{V}_{ext} and (ii) the output of an electronic structure code for the field-on problem; \mathcal{E}_{Sch} measures the difference between (ii) and (iii) propagation of the exact Hamiltonian plus \mathbf{V}_{ext} . Finally, \mathcal{E}_{Ham} measures the difference between (i) and (iii). Overall, we find that the errors are lower than in Table 1, indicating that the ML Hamiltonian succeeds in solving the field-on problem.

the same order of magnitude when we use either the ML Hamiltonian or the exact Hamiltonian. Consistent with Fig. 3, we find the largest gap between exact and ML Hamiltonian propagation for LiH.

5.2 Electric Field Tests

After learning a field-free Hamiltonian for each of the three molecules, we compared the values of $\mathbf{H}'(t)$ and $\tilde{\mathbf{H}}'(t)$ along the training trajectories. We found that the ML Hamiltonian does not equal the exact Hamiltonian. This led us to question whether the ML Hamiltonian could solve a problem outside the training regime. We therefore augmented the ML Hamiltonian with an applied electric field, i.e., the time-dependent external potential \mathbf{V}_{ext} given in (9). Using the same Runge-Kutta scheme and tolerances described earlier, we propagated for $2N = 2000$ steps. We compared these results with test data produced by an electronic structure code, and also the results of propagating the exact Hamiltonian, augmented with \mathbf{V}_{ext} , via our Runge-Kutta method.

For a first view of the field-on results, see Table 2 and Figures 5 and 6. In particular, the top panels of Figure 5 show the applied electric field; note that it is switched off abruptly after one period. We can immediately discern that the applied field substantially

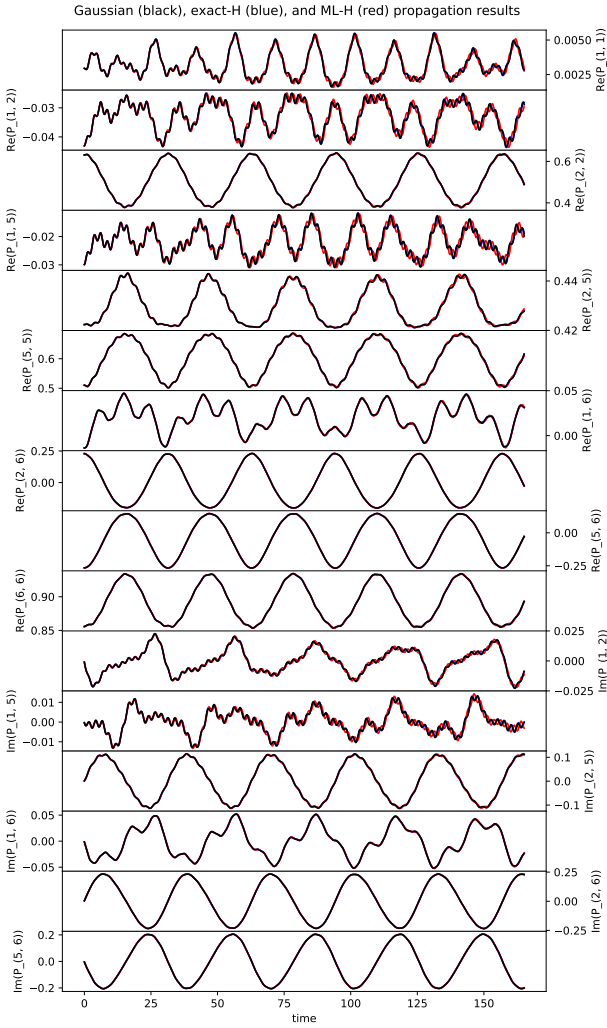


Fig. 3 LiH propagation with no field. We have plotted all unique real and imaginary parts of the time-dependent density matrices: actual training data (black), exact Hamiltonian propagation (blue), and ML Hamiltonian propagation (red). For density matrix elements with small variance, we discern slight disagreement especially at large times. For large-variance density matrix elements, the curves are in close agreement.

alters the electron density from the field-off case. Still, in each panel, we see excellent agreement between all three curves in each plot: the ground truth solution produced by an electronic structure code (black), the result of propagating the exact Hamiltonian plus \mathbf{V}_{ext} (blue), and the result of propagating the ML Hamiltonian plus \mathbf{V}_{ext} (red). Table 2, in which we find errors that are roughly an order of magnitude lower than those in Table 1, confirms that all computed densities are in close quantitative agreement. To repeat, all field-on results are for a time interval that is twice the length used for training, and training was conducted using field-off data only. Overall, we take these results to indicate that

the ML Hamiltonian can indeed extrapolate to problem settings beyond the one used for training.

For a deeper understanding of the field-on results, we focus on the right panel of Figure 4 and Figure 7. In the right panel of Figure 4, we compare (i) the result of propagating the ML Hamiltonian plus \mathbf{V}_{ext} against (ii) the ground truth solution, the output of the electronic structure code for the field-on problem. We also compare (ii) with (iii) the result of propagating the exact Hamiltonian plus \mathbf{V}_{ext} . The plots indicate that, for all three molecules and especially for LiH, the error between (i) and (ii) is almost identical to that between (ii) and (iii). This indicates that the bulk of the error is due to our use of a Runge-Kutta scheme instead of the MMUT scheme used in the electronic structure code. To confirm this, we consult Figure 7, in which we compare (i) and (iii) directly. All solutions here are computed using the same Runge-Kutta scheme. For each molecule, we see that the errors for the field-on problems are consistently smaller than those for the field-off problems. We conclude from these results that the ML Hamiltonian can be used to compute the electronic response to an applied electric field.

A short derivation will show that it is not automatic to expect the augmented ML Hamiltonian to propagate correctly. Let us work in continuous time, to eliminate error due to discrete-time propagation; in this idealized setting, we start with the statement that both of our field-free Hamiltonians, $\mathbf{H}'(t)$ (exact) and $\tilde{\mathbf{H}}'(t)$ (ML), satisfy the TDHF equation:

$$i \frac{d\mathbf{P}'(t)}{dt} = [\mathbf{H}'(t), \mathbf{P}'(t)]$$

$$i \frac{d\mathbf{P}'(t)}{dt} = [\tilde{\mathbf{H}}'(t), \mathbf{P}'(t)].$$

Subtracting these equations, and defining the error $\epsilon(t) = \mathbf{H}'(t) - \tilde{\mathbf{H}}'(t)$, we obtain

$$[\epsilon(t), \mathbf{P}'(t)] = 0. \quad (18)$$

Now we augment both Hamiltonians with an external field $\mathbf{V}_{\text{ext}}(t)$. Let $\mathbf{P}''(t)$ denote the true density for the problem with the external field. It must satisfy

$$i \frac{d\mathbf{P}''(t)}{dt} = [\mathbf{H}'(t) + \mathbf{V}_{\text{ext}}(t), \mathbf{P}''(t)].$$

Via $\mathbf{H}'(t) = \epsilon(t) + \tilde{\mathbf{H}}'(t)$, we obtain

$$i \frac{d\mathbf{P}''(t)}{dt} = [\tilde{\mathbf{H}}'(t) + \mathbf{V}_{\text{ext}}(t), \mathbf{P}''(t)] + \underbrace{[\epsilon(t), \mathbf{P}''(t)]}_{*}.$$

As (18) does not in general imply that the starred term vanishes, we cannot conclude that the true density $\mathbf{P}''(t)$ satisfies the TDHF equation with the augmented ML

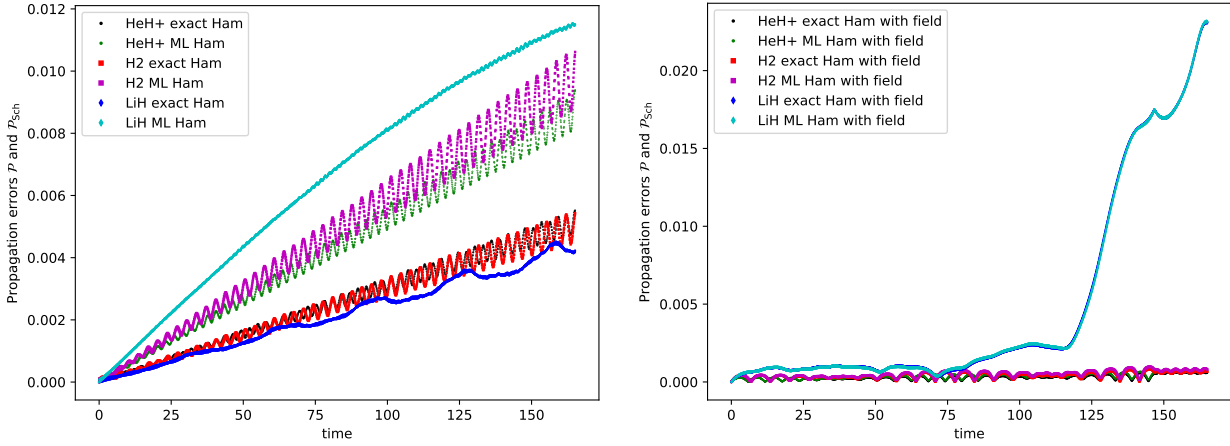


Fig. 4 Time-dependent propagation errors in which we compare the training data against either $\tilde{\mathbf{P}}$, the result of propagating the ML Hamiltonian, or \mathbf{P} , the result of propagating the exact Hamiltonian. All calculations on the left (respectively, right) are for the field-free (respectively, field-on) problem. For each molecule, the error incurred by propagating with the ML Hamiltonian is within a constant factor of the error incurred by propagating with the exact Hamiltonian. At the final time, all errors are on the order of 10^{-3} , except for the field-on calculations with LiH. The average values of these curves over all time correspond precisely to \mathcal{E} and \mathcal{E}_{Sch} —see (15), (16), and Table 1 for further details.

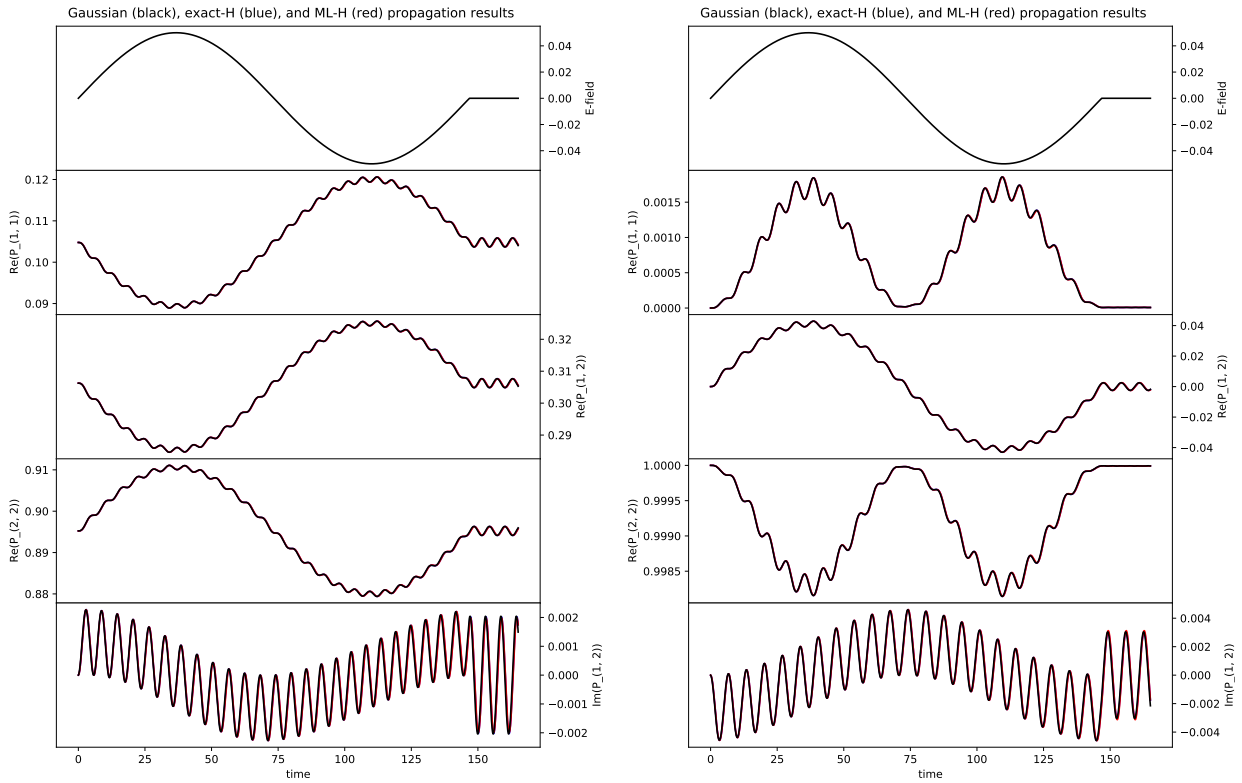


Fig. 5 HeH⁺ (left) and H₂ (right) propagation with field. The top panel of each plot gives the applied electric field (9). In subsequent panels, for both molecules, we plot all unique real and imaginary parts of the time-dependent density matrices: actual training data (black), exact Hamiltonian propagation (blue), and ML Hamiltonian propagation (red). By ML Hamiltonian, we mean the Hamiltonian trained on the field-free data plus \mathbf{V}_{ext} given by (9). Note the close agreement of all three curves, on a time interval that is twice the length used for training. This is a true test of whether the learned Hamiltonian can extrapolate to problem settings beyond the one used for training.

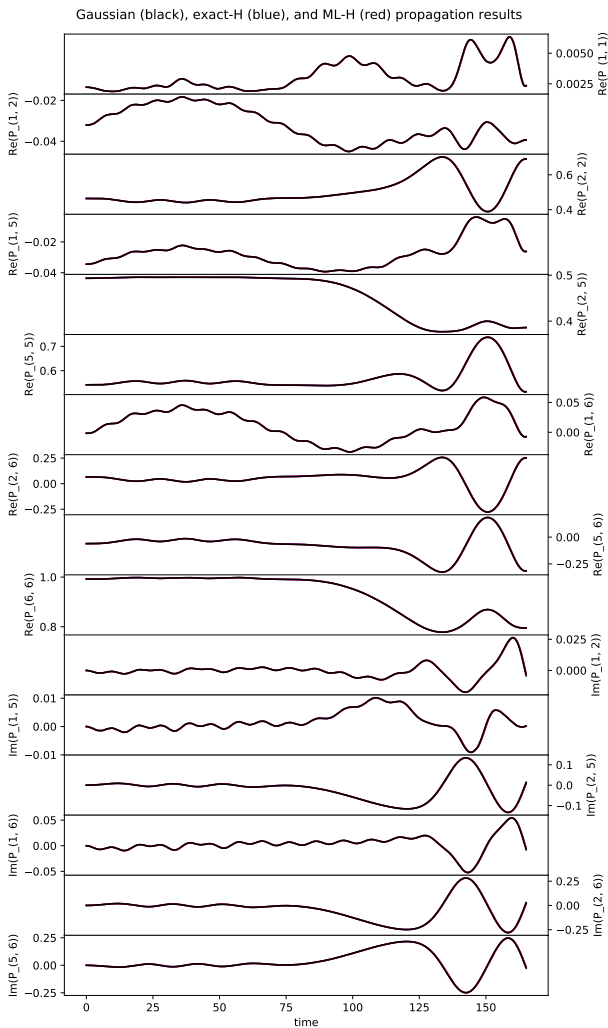


Fig. 6 LiH propagation with field. We plot all unique real and imaginary parts of the time-dependent density matrices: actual training data (black), exact Hamiltonian propagation (blue), and ML Hamiltonian propagation (red). By ML Hamiltonian, we mean the Hamiltonian trained on the field-free LiH data plus \mathbf{V}_{ext} given by (9). Note the close agreement of all curves, on a time interval that is twice the length used for training. This is a true test of whether the learned Hamiltonian can extrapolate to problem settings beyond the one used for training. We omit a plot of the electric field here—see the top panels of Figure 5.

Hamiltonian $\tilde{\mathbf{H}}'(t) + \mathbf{V}_{\text{ext}}(t)$. Based on the above derivation, if we solve the TDHF equation using the augmented ML Hamiltonian, we expect to obtain a time-dependent density that differs from $\mathbf{P}''(t)$. As we are able to use the ML Hamiltonian successfully on the problem with an applied electric field, we hypothesize that the error $\epsilon(t)$ is structured in such a way that enables us to extrapolate to new problems. We plan to test this hypothesis in future work.

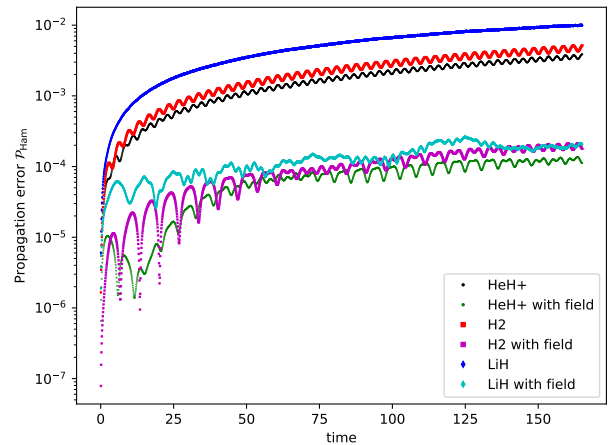


Fig. 7 Time-dependent propagation errors in which we compare \mathbf{P} , the result of propagating the ML Hamiltonian, with $\bar{\mathbf{P}}$, the result of propagating the exact Hamiltonian. All results were computed using the same Runge-Kutta scheme, isolating the error due to the different Hamiltonians. We include both field-free and field-on calculations. Note that all results are plotted on a log scale. The results show that when we propagate both the ML and exact Hamiltonians using the same scheme, the errors between the two resulting trajectories remain small even as we take hundreds of time steps. The average values of these curves over all time correspond precisely to \mathcal{E}_{Ham} —see (17) and Table 1 for further details.

5.3 Reproducibility

All code required to reproduce all training and test results (including plots) is available on GitHub at <https://github.com/hbhat4000/electrondynamics>. Training data is available from the authors upon request.

6 Discussion

Our current work demonstrates that, from a single time series consisting of time-dependent density matrices, we can effectively learn an integrated Hamiltonian matrix. This ML Hamiltonian can be used for propagation in both the field-off and field-on settings. Importantly, training with a single field-free trajectory, our ML Hamiltonian has the potential to predict the electronic response to a large variety of field pulse perturbations, opening the door to laser-field controlled chemistry. The present work leads to two main areas of future work. The first area concerns technical improvements to the procedure itself, including (i) to replace (12) with a weighted loss function, to account for density elements that oscillate on different vertical scales, (ii) to propagate our ML Hamiltonian using the MMUT scheme, thus eliminating the kind of error quantified by \mathcal{E}_{Sch} , and (iii) to further explore reducing the number of degrees of freedom in the ML Hamilto-

nian. The second area concerns improving our overall understanding of the procedure, and applying it to systems of greater chemical and physical interest. In this area, further work is needed to understand the difference between the exact and ML Hamiltonians, whether this difference can be decreased by training on multiple trajectories, and how far outside the training regime we can push the ML Hamiltonian. We can also seek to learn the \hat{H} operator rather than the \mathbf{H} matrix representation, which is of interest for determining the unknown exchange-correlation potential within time-dependent density functional theory. In this way, we can push this procedure beyond known physics (as explored here) to systems where the underlying potential energy terms are not known with sufficient accuracy or precision.

Acknowledgements This work was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Award Number DE-SC0020203. We acknowledge computational time on the MERCED cluster (funded by NSF ACI-1429783), and on the Nautilus cluster, which is supported by the Pacific Research Platform (NSF ACI-1541349), CHASE-CI (NSF CNS-1730158), and Towards a National Research Platform (NSF OAC-1826967). Additional funding for Nautilus has been supplied by the University of California Office of the President.

Appendix

Canonical Orthogonalization. Let \mathbf{S} be the overlap matrix with $S_{\mu\nu} = \langle \chi_\mu | \chi_\nu \rangle$. Because it is real and symmetric, we have $\mathbf{S} = \mathbf{U}\mathbf{s}\mathbf{U}^T$ where \mathbf{s} is diagonal and real, and \mathbf{U} is a real orthogonal matrix. Now we form $\mathbf{X} = \mathbf{U}\mathbf{s}^{-1/2}$. Then, we go from \mathbf{P} to \mathbf{P}' via

$$\mathbf{P}' = \mathbf{s}^{1/2}\mathbf{U}^T\mathbf{P}\mathbf{U}\mathbf{s}^{1/2}.$$

If \mathbf{H} is the Hamiltonian in the AO basis, the Hamiltonian in the orthogonalized basis is

$$\mathbf{H}' = \mathbf{s}^{-1/2}\mathbf{U}^T\mathbf{H}\mathbf{U}\mathbf{s}^{-1/2}.$$

Dimensionality Reduction. For LiH, certain elements of the density matrix $\mathbf{P}'(t)$ are identically zero for all t . We thus define a reduced state vector \mathbf{p}^\sharp that consists of the non-zero upper-triangular degrees of freedom, i.e., the $\widetilde{M} \leq M^2$ elements that are necessary to reconstruct all of \mathbf{P}' . Out of these \widetilde{M} elements, we take the first \widetilde{M}_R to correspond to elements of \mathbf{P}'_R and the remaining $\widetilde{M}_I = \widetilde{M} - \widetilde{M}_R$ to correspond to elements of \mathbf{P}'_I . Define Z by

$$Z = Z^R \cup Z^I \quad (19a)$$

$$Z^R = \{(i, j) \text{ s.t. } i \leq j \text{ and not } P_{ij}^R(t) \equiv 0\} \quad (19b)$$

$$Z^I = \{(i, j) \text{ s.t. } i < j \text{ and not } P_{ij}^I(t) \equiv 0\}. \quad (19c)$$

We form a mapping $\sigma : \{1, 2, \dots, \widetilde{M}\} \rightarrow Z$, whose purpose is to map one-dimensional indices of the reduced vector \mathbf{p}^\sharp to ordered pair indices of the full matrix \mathbf{P}' . We define σ implicitly through (19) and the following:

$$p_k^\sharp = \begin{cases} \mathbf{P}'_{R\sigma(k)} & k \leq \widetilde{M}_R \\ \mathbf{P}'_{I\sigma(k)} & k > \widetilde{M}_R. \end{cases} \quad (20)$$

Looping over the entries $k = 1, 2, \dots, \widetilde{M}$, this equation let us go back and forth from the full complex matrix \mathbf{P}' to the reduced real state vector \mathbf{p}^\sharp .

Importantly, we now follow precisely the same procedure, with the same mapping σ and set Z , to form a reduced Hamiltonian vector \mathbf{h}^\sharp . We then formulate a reduced-dimensionality version of (10):

$$\widetilde{\mathbf{h}}^\sharp = \widetilde{\beta}_0 + \widetilde{\beta}_1\mathbf{p}^\sharp. \quad (21)$$

The matrix $\widetilde{\beta}_1$ is now of dimension $\widetilde{M} \times \widetilde{M}$; all other objects in this equation are vectors of dimension $\widetilde{M} \times 1$. The training procedure then holds without further modifications.

References

- Bartók, A.P., De, S., Poelking, C., Bernstein, N., Kermode, J.R., Csányi, G., Ceriotti, M.: Machine learning unifies the modeling of materials and molecules. *Science Advances* **3**(12) (2017). DOI 10.1126/sciadv.1701816
- Behler, J.: Perspective: Machine learning potentials for atomistic simulations. *The Journal of Chemical Physics* **145**(17), 170901 (2016). DOI 10.1063/1.4966192
- Behler, J., Parrinello, M.: Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**(14), 146401 (2007)
- Bertalan, T., Dietrich, F., Mezić, I., Kevrekidis, I.G.: On learning Hamiltonian systems from data. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **29**(12), 121107 (2019). DOI 10.1063/1.5128231. URL <https://doi.org/10.1063/1.5128231>
- Bhat, H.S.: Learning and interpreting potentials for classical Hamiltonian systems. In: P. Cellier, K. Driessens (eds.) *Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2019. Springer, Cham* (2020). *Communications in Computer and Information Science* **1167**.
- Ceriotti, M.: Unsupervised machine learning in atomistic simulations, between predictions and understanding. *Journal of Chemical Physics* **150**(15) (2019). DOI 10.1063/1.5091842
- Chandrasekaran, A., Kamal, D., Batra, R., Kim, C., Chen, L., Ramprasad, R.: Solving the electronic structure problem with machine learning. *npj Computational Materials* **5**(1) (2019). DOI 10.1038/s41524-019-0162-7. URL <https://doi.org/10.1038/s41524-019-0162-7>
- Chen, W.K., Liu, X.Y., Fang, W.H., Dral, P.O., Cui, G.: Deep Learning for Nonadiabatic Excited-State Dynamics. *J. Phys. Chem. Lett.* **9**(23), 6702–6708 (2018)
- Chen, Z., Zhang, J., Arjovsky, M., Bottou, L.: Symplectic recurrent neural networks. In: 8th International Conference on Learning Representations, ICLR 2020 (2020). URL <https://openreview.net/forum?id=BkgYPREtPr>

10. Christensen, A.S., Faber, F.A., Von Lilienfeld, O.A.: Operators in quantum machine learning: Response properties in chemical space. *J. Chem. Phys.* **150**(6), 064105 (2019)
11. Dirac, P.A.M.: Note on Exchange Phenomena in the Thomas Atom. *Mathematical Proceedings of the Cambridge Philosophical Society* **26**(3), 376–385 (1930). DOI 10.1017/S0305004100016108
12. Dreuw, A., Head-Gordon, M.: Single-reference ab initio methods for the calculation of excited states of large molecules. *Chem. Rev.* **105**(11), 4009–4037 (2005). DOI 10.1021/cr0505627. URL <https://doi.org/10.1021/cr0505627>. PMID: 16277369
13. Eshuis, H., Balint-Kurti, G.G., Manby, F.R.: Dynamics of molecules in strong oscillating electric fields using time-dependent Hartree–Fock theory. *J. Chem. Phys.* **128**(11), 114113 (2008)
14. Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Scalmani, G., Barone, V., Petersson, G.A., Nakatsuji, H., Li, X., Caricato, M., Marenich, A.V., Bloino, J., Janesko, B.G., Gomperts, R., Mennucci, B., Hratchian, H.P., Ortiz, J.V., Izmaylov, A.F., Sonnenberg, J.L., Williams-Young, D., Ding, F., Lipparini, F., Egidi, F., Goings, J., Peng, B., Petrone, A., Henderson, T., Ranasinghe, D., Zakrzewski, V.G., Gao, J., Rega, N., Zheng, G., Liang, W., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Throssell, K., Montgomery Jr., J.A., Peralta, J.E., Ogliaro, F., Bearpark, M.J., Heyd, J.J., Brothers, E.N., Kudin, K.N., Staroverov, V.N., Keith, T.A., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A.P., Burant, J.C., Iyengar, S.S., Tomasi, J., Cossi, M., Millam, J.M., Klene, M., Adamo, C., Cammi, R., Ochterski, J.W., Martin, R.L., Morokuma, K., Farkas, O., Foresman, J.B., Fox, D.J.: *Gaussian Development Version Revision I.14+* (2018). Gaussian Inc. Wallingford CT
15. Fujita, H., Nakagawa, Y.O., Sugiura, S., Oshikawa, M.: Construction of Hamiltonians by supervised learning of energy and entanglement spectra. *Phys. Rev. B* **97**, 075114 (2018). DOI 10.1103/PhysRevB.97.075114. URL <https://link.aps.org/doi/10.1103/PhysRevB.97.075114>
16. Gastegger, M., Behler, J., Marquetand, P.: Machine learning molecular dynamics for the simulation of infrared spectra. *Chem. Sci.* **8**(10), 6924–6935 (2017)
17. Ghosh, K., Stuke, A., Todorović, M., Jørgensen, P.B., Schmidt, M.N., Vehtari, A., Rinke, P.: Deep Learning Spectroscopy: Neural Networks for Molecular Excitation Spectra. *Adv. Sci.* **6**(9), 1801367–1801374 (2019)
18. Grisafi, A., Wilkins, D.M., Csányi, G., Ceriotti, M.: Symmetry-Adapted Machine Learning for Tensorial Properties of Atomistic Systems. *Phys. Rev. Lett.* **120**(3), 36002 (2018)
19. Häse, F., Valteau, S., Pyzer-Knapp, E., Aspuru-Guzik, A.: Machine learning exciton dynamics. *Chem. Sci.* **7**(8), 5139–5147 (2016)
20. Innocenti, L., Banchi, L., Ferraro, A., Bose, S., Paternostro, M.: Supervised learning of time-independent Hamiltonians for gate design. *New Journal of Physics* **22**(6), 065001 (2020). DOI 10.1088/1367-2630/ab8aaf
21. Isborn, C.M., Li, X., Tully, J.C.: TDDFT Ehrenfest dynamics: Collisions between atomic oxygen and graphite clusters. *J. Chem. Phys.* **126**, 134307 (2007)
22. Jin, P., Zhang, Z., Zhu, A., Tang, Y., Karniadakis, G.E.: Sympnets: Intrinsic structure-preserving symplectic networks for identifying Hamiltonian systems (2020). ArXiv:2001.03750
23. Jørgensen, M.S., Mortensen, H.L., Meldgaard, S.A., Kolsbjerg, E.L., Jacobsen, T.L., Sørensen, K.H., Hammer, B.: Atomistic structure learning. *Journal of Chemical Physics* (2019). DOI 10.1063/1.5108871
24. Li, H., Collins, C., Tanha, M., Gordon, G.J., Yaron, D.J.: A density functional tight binding layer for deep learning of chemical Hamiltonians. *Journal of Chemical Theory and Computation* **14**(11), 5764–5776 (2018). DOI 10.1021/acs.jctc.8b00873. URL <https://doi.org/10.1021/acs.jctc.8b00873>. PMID: 30351008
25. Li, X., Smith, S.M., Markevitch, A.N., Romanov, D.A., Levis, R.J., Schlegel, H.B.: A time-dependent Hartree-Fock approach for studying the electronic optical response of molecules in intense fields. *Physical Chemistry Chemical Physics* **7**(2), 233–239 (2005). DOI 10.1039/b415849k
26. Lopata, K., Govind, N.: Modeling fast electron dynamics with real-time time-dependent density functional theory: Application to small molecules and chromophores. *J. Chem. Theory Comput.* **7**(5), 1344–1355 (2011). DOI 10.1021/ct200137z
27. Lu, C., Liu, Q., Sun, Q., Hsieh, C.Y., Zhang, S., Shi, L., Lee, C.K.: Deep Learning for Optoelectronic Properties of Organic Semiconductors. *J. Phys. Chem. C* **124**, 7048–7060 (2020)
28. Mattheakis, M., Sondak, D., Dogra, A.S., Protopapas, P.: Hamiltonian Neural Networks for solving differential equations (2020). ArXiv:2001.11107
29. Micha, D.A., Runge, K.: Time-dependent many-electron approach to slow ion-atom collisions: The coupling of electronic and nuclear motions. *Phys. Rev. A* **50**, 322–336 (1994). DOI 10.1103/PhysRevA.50.322. URL <https://link.aps.org/doi/10.1103/PhysRevA.50.322>
30. Miller, S.T., Lindner, J.F., Choudhary, A., Sinha, S., Ditto, W.L.: Mastering high-dimensional dynamics with Hamiltonian neural networks (2020). ArXiv:2008.04214
31. Montavon, G., Rupp, M., Gobre, V., Vazquez-Mayagoitia, A., Hansen, K., Tkatchenko, A., Müller, K.R., Anatole Von Lilienfeld, O.: Machine learning of molecular electronic properties in chemical compound space. *New J. Phys.* **15**, 095003 (2013)
32. Nascimento, D.R., DePrince III, A.E.: Linear absorption spectra from explicitly time-dependent equation-of-motion coupled-cluster theory. *J. Chem. Theory Comput.* **12**(12), 5834–5840 (2016). DOI 10.1021/acs.jctc.6b00796
33. Nebgen, B., Lubbers, N., Smith, J.S., Sifain, A.E., Lokhov, A., Isayev, O., Roitberg, A.E., Barros, K., Tretiak, S.: Transferable Dynamic Molecular Charge Assignment Using Deep Neural Networks. *J. Chem. Theory Comput.* **14**(9), 4687–4698 (2018)
34. Paruzzo, F.M., Hofstetter, A., Musil, F., De, S., Ceriotti, M., Emsley, L.: Chemical shifts in molecular solids by machine learning. *Nat. Commun.* **9**, 4501 (2018)
35. Pronobis, W., Schütt, K.T., Tkatchenko, A., Müller, K.R.: Capturing intensive and extensive DFT/TDDFT molecular properties with machine learning. *Eur. Phys. J. B* **91**(8), 178–184 (2018)
36. Provorse, M.R., Isborn, C.M.: Electron dynamics with real-time time-dependent density functional theory. *Int. J. Quant. Chem.* **116**(10), 739–749 (2016). DOI 10.1002/qua.25096
37. Ramakrishnan, R., Hartmann, M., Tapavicza, E., von Lilienfeld, O.A.: Electronic spectra from TDDFT and machine learning in chemical space. *J. Chem. Phys.* **143**, 084111 (2015)

38. Rezende, D.J., Racanière, S., Higgins, I., Toth, P.: Equivariant Hamiltonian flows (2019). ArXiv:1909.13739
39. Rodríguez, M., Kramer, T.: Machine learning of two-dimensional spectroscopic data. *Chem. Phys.* **520**(December 2018), 52–60 (2019)
40. Schleder, G.R., Padilha, A.C.M., Acosta, C.M., Costa, M., Fazzio, A.: From DFT to machine learning: recent approaches to materials science—a review. *Journal of Physics: Materials* **2**(3), 032001 (2019). DOI 10.1088/2515-7639/ab084b. URL <https://doi.org/10.1088/2515-7639/ab084b>
41. Sifain, A.E., Lubbers, N., Nebgen, B.T., Smith, J.S., Lokhov, A.Y., Isayev, O., Roitberg, A.E., Barros, K., Tretiak, S.: Discovering a Transferable Charge Assignment Model Using Machine Learning. *J. Phys. Chem. Lett* **9**(16), 4495–4501 (2018)
42. Smith, J.S., Nebgen, B.T., Zubatyuk, R., Lubbers, N., Devereux, C., Barros, K., Tretiak, S., Isayev, O., Roitberg, A.E.: Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. *Nature Communications* (2019). DOI 10.1038/s41467-019-10827-4
43. Snyder, J.C., Rupp, M., Hansen, K., Müller, K.R., Burke, K.: Finding density functionals with machine learning. *Physical Review Letters* **108**(25), 1–5 (2012). DOI 10.1103/PhysRevLett.108.253002
44. Suzuki, Y., Nagai, R., Haruyama, J.: Machine learning exchange-correlation potential in time-dependent density-functional theory. *Phys. Rev. A* **101**, 050501 (2020). DOI 10.1103/PhysRevA.101.050501
45. Szabo, A., Ostlund, N.S.: *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, first edn. Dover Publications, Inc., Mineola (1996)
46. Toth, P., Rezende, D.J., Jaegle, A., Racanière, S., Botev, A., Higgins, I.: Hamiltonian generative networks. In: 8th International Conference on Learning Representations, ICLR 2020 (2020). URL <https://openreview.net/forum?id=HJenn6VFvB>
47. Wilkins, D.M., Grisafi, A., Yang, Y., Lao, K.U., DiStasio, R.A., Ceriotti, M.: Accurate molecular polarizabilities with coupled cluster theory and machine learning. *Proc. Natl. Acad. Sci. U.S.A.* **116**(9), 3401–3406 (2019)
48. Ye, S., Hu, W., Li, X., Zhang, J., Zhong, K., Zhang, G., Luo, Y., Mukamel, S., Jiang, J.: A neural network protocol for electronic excitations of N-methylacetamide. *Proc. Natl. Acad. Sci. U.S.A.* **116**(24), 11612–11617 (2019)
49. Zhong, Y.D., Dey, B., Chakraborty, A.: Symplectic ODE-Net: Learning Hamiltonian Dynamics with Control. In: 8th International Conference on Learning Representations, ICLR 2020 (2020). URL <https://openreview.net/pdf?id=ryxmb1rKDS>
50. Zhu, Y., Herbert, J.M.: Self-consistent predictor/corrector algorithms for stable and efficient integration of the time-dependent Kohn-Sham equation. *J. Chem. Phys.* **148**(4), 044117 (2018)