Equivariant neural network for Green's functions of molecules and materials

Xinyang Dong [1,*] Emanuel Gull [2] and Lei Wang[3,4]

[1]*AI for Science Institute, Beijing 100080, China*
[2]*Department of Physics, University of Michigan, Ann Arbor, Michigan 48109, USA*
[3]*Beijing National Laboratory for Condensed Matter Physics and Institute of Physics,*
Chinese Academy of Sciences, Beijing 100190, China
[4]*Songshan Lake Materials Laboratory, Dongguan, Guangdong 523808, China*

The many-body Green's function provides access to electronic properties beyond density-functional theory level in *ab inito* calculations. In this manuscript, we propose a deep learning framework for predicting the finite-temperature Green's function in atomic orbital space, aiming to achieve a balance between accuracy and efficiency. By predicting the self-energy matrices in Lehmann representation using an equivariant message-passing neural network, our method respects its analytical property and the $E(3)$ equivariance. The Green's function is obtained from the predicted self-energy through the Dyson equation with target total number of electrons. We present proof-of-concept benchmark results for both molecules and simple periodic systems, showing that our method is able to provide an accurate estimate of physical observables such as energy and density of states based on the predicted Green's function.

I. INTRODUCTION

The single-particle Green's function plays a fundamental role in the computational study of quantum field theories in condensed matter physics, quantum chemistry, and material science. It provides in particular information about the single-particle excitation spectrum, which can be compared to scanning tunneling microscopy and angle-resolved photoemission spectroscopy experiments. In recent years, rapid development in *ab initio* theory [1–5] and the numerical implementation [6–15] of Green's function methods has enabled systematic calculations of interacting quantum many-body systems.

The field theory formulation provides a theoretically rigorous view of the finite-temperature physics of interacting quantum systems, complementary to results from ground-state methodologies such as the density-functional theory (DFT). However, computing Green's functions is in general orders of magnitude more expensive than a DFT calculation, limiting the methodology to small systems.

This motivates research into the application of data-driven machine learning approaches to quantum field theories. Such methods balance accuracy with efficiency, and prior work has shown considerable success. For instance, Refs. [16–18] developed machine learning models to predict the Green's function of the single-site Anderson impurity model, serving as impurity solvers for the dynamical mean-field theory (DMFT). Reference [19] employed kernel ridge regression to predict self-energies and spectral functions of realistic systems starting from a mean-field Hartree-Fock solution. Still,

the power of state-of-the-art deep learning models of finite-temperature field theories when applied to realistic systems has yet to be demonstrated.

In this manuscript, we propose a deep learning framework for predicting the many-body Green's function and self-energy for both molecules and periodic systems in atomic orbital space. We employ a message-passing neural network [20,21] that maps atomic configurations to matrices, in analogy to a framework that successfully predicts DFT Hamiltonians [22–27]. To achieve better accuracy and data efficiency in the training process, we use an equivariant setup as in Refs. [24–32]. By constructing the fundamental self-energy matrices in the Lehmann representation using the equivariant features, we ensure the fulfillment of their analytical properties by construction.

This paper is organized as follows: In Sec. II, we introduce the theory of finite-temperature Green's function (Sec. II A), the equivariant message passing neural network (Sec. II B), and how we employ the neural network to predict self-energies and Green's functions (Secs. II C and II D). In Sec. III, we present results of proof-of-concept benchmarks for both molecules (Sec. III A) and periodic systems (Sec. III B). Section IV provides the conclusion and outlook.

II. METHOD

A. Green's function formalism

Within the Born-Oppenheimer approximation and in the absence of relativistic effects, the second-quantized Hamiltonian of realistic systems can be written as [33,34]

$$H = \sum_{ij} \sum_{\sigma} h_{ij} c_{i\sigma}^{\dagger} c_{j\sigma} + \frac{1}{2} \sum_{ijkl} \sum_{\sigma\sigma'} U_{ijkl} c_{i\sigma}^{\dagger} c_{k\sigma'}^{\dagger} c_{l\sigma'} c_{j\sigma}, \quad (1)$$

*xinyang.dongxy@gmail.com

075112-1

where $c_{i\sigma}^{\dagger}$ ($c_{i\sigma}$) are the creation (annihilation) operators with orbital index i and spin index σ, h_{ij} is the one-electron kinetic and electron-nuclei integral, and U_{ijkl} the Coulomb repulsion integral. The atomic orbitals $g_i(\mathbf{r})$ considered in this work may be nonorthogonal, defining an overlap matrix in orbital space [34]:

$$S_{ij} = \int d\mathbf{r} g_i^*(\mathbf{r}) g_j(\mathbf{r}). \tag{2}$$

We will use i as spin-orbital index to omit the explicit spin index σ and use bold symbols for matrices in spin-orbital space in the rest of this paper.

To compute physical properties of an electron system, we introduce the single-particle finite-temperature Green's function [33,35]

$$G_{ij}(\tau) = -\langle \mathcal{T} c_i(\tau) c_j^{\dagger}(0) \rangle, \tag{3}$$

where $\tau \in [0, \beta]$ is the imaginary time, $\beta = \frac{1}{k_B T}$ the inverse temperature, and $\mathcal{T}$ the time-ordering operator [33]. The imaginary time Green's function $\mathbf{G}(\tau)$ corresponds to a frequency space or Matsubara Green's function with the transform

$$\mathbf{G}(i\omega_n) = \int_0^{\beta} d\tau \, \mathbf{G}(\tau) e^{i\omega_n \tau}, \tag{4}$$

where the fermionic Matsubara frequencies ω_n are defined as $\omega_n = (2n+1)\pi/\beta$, $n \in \mathbb{Z}$. The connection between the noninteracting Green's function $\mathbf{G}_0$ and the full Green's function $\mathbf{G}$ is given by the Dyson equation

$$\mathbf{G}(i\omega_n) = \mathbf{G}_0(i\omega_n) + \mathbf{G}_0(i\omega_n)\mathbf{\Sigma}(i\omega_n)\mathbf{G}(i\omega_n), \tag{5}$$

where $\mathbf{G}_0(i\omega_n) = [(i\omega_n + \mu)\mathbf{S} - \mathbf{h}]^{-1}$, with μ the chemical potential, $\mathbf{h}$ as in Eq. (1), and $\mathbf{\Sigma}(i\omega_n)$ the Matsubara frequency self-energy which is a function of the full Green's function $\mathbf{\Sigma} \equiv \mathbf{\Sigma}[\mathbf{G}]$. The self-energy can be split into two parts

$$\mathbf{\Sigma}[\mathbf{G}](i\omega_n) = \mathbf{\Sigma}^{(\mathrm{HF})}[\mathbf{G}] + \tilde{\mathbf{\Sigma}}[\mathbf{G}](i\omega_n), \tag{6}$$

where $\mathbf{\Sigma}^{(\mathrm{HF})}[\mathbf{G}]$ is the static Hartree-Fock (HF) self-energy and $\tilde{\mathbf{\Sigma}}[\mathbf{G}](i\omega_n)$ denotes the frequency-dependent dynamical self-energy. Usually $\mathbf{\Sigma}^{(\mathrm{HF})}[\mathbf{G}]$ is combined with the one-body integral $\mathbf{h}$ into the so-called Fock matrix $\mathbf{F} = \mathbf{h} + \mathbf{\Sigma}^{(\mathrm{HF})}$. The total energy of the system is given by

$$E_{\mathrm{tot}} = \tfrac{1}{2}\mathrm{Tr}[\mathbf{P}(\mathbf{h} + \mathbf{F})] + \tfrac{1}{2}\mathrm{Tr}[\tilde{\mathbf{\Sigma}} * \mathbf{G}] + E_{nn}, \tag{7}$$

where E_{nn} is the nuclei-nuclei Coulomb energy, $\mathbf{P} = \mathbf{G}(\tau = 0^-)$ is the density matrix, $*$ is the imaginary time convolution operator, and the trace is defined as $\mathrm{Tr}[\mathbf{A}] = -\sum_i A_{ii}(\beta^-)$.

The Green's function can be analytically continued from the Matsubara frequencies to the entirety of the complex plane, and will be analytic in the upper half of the complex plane. The limit of the Green's function taken towards the real frequency axis corresponds to the so-called retarded Green's function. It can be obtained from Matsubara data with numerical analytical continuation and directly yields the spectral function (DOS) of the system

$$A(\omega) = -\frac{1}{\pi}\mathrm{Im}(\mathrm{Tr}[\mathbf{S}\mathbf{G}(\omega)]). \tag{8}$$

Refs. [8,15,36–41] contain further references and detailed explanations of the finite-temperature formalism and its numerical implementation.

B. Equivariant message-passing neural networks

In graph neural networks (GNN) or message-passing neural networks (MPNN), atomic structures of isolated molecules or periodic solids are represented by nodes and edges, where nodes indicate atoms and edges demonstrate the connection between atom pairs [20,21]. In the "message passing" process, starting from the initial element-based embedding such as the one-hot encoding of nuclear charge, the feature vector $\mathbf{f}_i$ associated with node i is iteratively updated through convolutions with its neighbors based on their features $\mathbf{f}_j$ and distances $\mathbf{r}_{ij}$.

The node features of the graph can be used to construct desired physical quantities such as the interatomic potential [31,32] or DFT (tight-binding) Hamiltonian [22–27]. The GNN (MPNN) setup can be either rotational invariant or equivariant. Since the objects we are interested in are matrices in atomic orbital space which are equivariant under rotations, we choose to use the equivariant message-passing neural network.

The core operation in equivariant neural network architecture is the tensor product operation that couples two representations in an equivariant way [30]

$$\mathbf{z}^{(l_3)} = \mathbf{x}^{(l_1)} \otimes \mathbf{y}^{(l_2)}, \tag{9}$$

$$z_{m_3}^{l_3} = \sum_{m_1=-l_1}^{l_1} \sum_{m_2=-l_2}^{l_2} C_{m_3,m_2,m_1}^{l_3,l_2,l_1} x_{m_1}^{l_1} y_{m_2}^{l_2}, \tag{10}$$

where C denotes the Clebsch-Gordan (CG) coefficients, $l \in \mathbb{N}$ are angular-momentum quantum numbers, and m are magnetic quantum numbers. l_3 satisfies the relation $|l_1 - l_2| \leqslant l_3 \leqslant l_1 + l_2$, and the parity of $\mathbf{z}$ is given by $p(\mathbf{z}) = p(\mathbf{x})p(\mathbf{y})$. When building neural networks, the tensor product operation is usually supplemented with an equivariant linear operation to mix channels of each irreducible representation

$$\mathbf{z}_c^{(l_3)} = \sum_{c'} W_{cc'} \mathbf{z}_{c'}^{(l_3)}, \tag{11}$$

where W is a trainable weight matrix. We will use the $\otimes$ operator to denote the weighted tensor product operation in the rest of this paper for better readability.

The network structure we use in this work is similar to the tensor field network [28] and NequIP [31]. In each message-passing layer, the features on each node are updated by collecting information from all its neighbors:

$$\mathbf{f}_i' = \sum_{j \in \mathrm{neigh}(i)} \mathbf{f}_j \otimes \mathbf{e}_{ij}, \tag{12}$$

$$\mathbf{e}_{ij} = R(B(\|\mathbf{x}_{ij}\|))Y(\mathbf{x}_{ij}/\|\mathbf{x}_{ij}\|). \tag{13}$$

Here $Y(\mathbf{x}_{ij}/\|\mathbf{x}_{ij}\|)$ denotes spherical expansion of the direction of distances between different nodes, R denotes a multilayer perceptron, and B is a trainable edge-length embedding layer as described in Ref. [31]

$$B(x) = \frac{2}{x_c}\frac{\sin\left(\frac{b\pi}{x_c}x\right)}{x}f(x, x_c), \tag{14}$$

where b is a trainable parameter, x_c is a given cutoff length, and f is a polynomial envelope function defined in Ref. [42]. An equivariant nonlinear activation function [30] is applied to all the node features after the updates. Note that the spherical expansion of interatomic distances fits into the construction of atomic basis functions such as Gaussian-type orbitals or linear combination of atomic orbitals (LCAO), whose angular components are spherical harmonics. Since Eq. (13) takes relative distances between atom pairs, this type of construction is equivariant with respect to the $E(3)$ group which comprises translations, rotations, and reflections [30].

The full matrix in atomic orbital space is constructed in a block-wise manner where each block corresponds to the interaction between two atoms. To obtain these pairwise features, we use a pair interaction layer to get diagonal and off-diagonal features as in PhiSNet [24] and QHNet [27]:

$$\mathbf{f}_{ii} = \tilde{\mathbf{f}}_i + \text{ResBlock}(\tilde{\mathbf{f}}_i \otimes \tilde{\mathbf{f}}_i), \quad (15a)$$

$$\mathbf{f}_{ij} = \tilde{\mathbf{f}}_i + \text{ResBlock}(\tilde{\mathbf{f}}_i \otimes \mathbf{e}_{ij} \otimes \tilde{\mathbf{f}}_j). \quad (15b)$$

The features $\tilde{\mathbf{f}}$ are computed from the node features $\mathbf{f}$ via a ResBlock that contains equivariant linear and activation functions

$$\tilde{\mathbf{f}} = \text{ResBlock}(\mathbf{f}) = \text{Linear}(\mathbf{f} + \text{Linear}(\text{Activation}(\mathbf{f}))). \quad (16)$$

Each matrix block in the full matrix is constructed using the inverse operation of the tensor product

$$\mathbf{M}_{ij}^{l_1, l_2} = \sum_{l_3=|l_2-l_1|}^{l_2+l_1} \overset{l_1, l_2}{\otimes} \mathbf{f}_{ij}^{(l_3)}. \quad (17)$$

C. Predicting Green's functions and self-energies with equivariant neural networks

We aim to predict both the static quantity $\mathbf{\Sigma}^{(\text{HF})}$ (or the Fock matrix $\mathbf{F}$) and the dynamical quantities $\mathbf{G}$ and $\tilde{\mathbf{\Sigma}}$ using a neural network. In finite-temperature theories, $\mathbf{G}(i\omega_n)$ and $\tilde{\mathbf{\Sigma}}(i\omega_n)$ are functions of imaginary time or Matsubara frequency. Both functions are strongly constrained by their analytical properties, and respecting these properties guarantees, among others, causality and the conservation of probability density [43–45].

To construct such functions, we start from a Lehmann representation

$$G_{ij}(z) = \frac{1}{Z} \sum_{mn} \frac{\langle n|c_i|m\rangle \langle m|c_j^\dagger|n\rangle}{z + E_n - E_m} (e^{-\beta E_n} + e^{-\beta E_m}), \quad (18)$$

$$\tilde{\Sigma}_{ij}(z) = \sum_s \frac{t_{is} t_{js}^*}{z - t_{ss}}, \quad (19)$$

where z is a complex frequency value defined on the upper half plane $z \in \mathcal{C}^+$, and $Z = \sum_n e^{-\beta E_n}$ is the partition function. In the Lehmann representation of $\tilde{\Sigma}$, s represents virtual orbitals in addition to the physical orbitals, and t_{xy} are terms in a corresponding effective Hamiltonian. See Ref. [46] for a detailed derivation.

The Lehmann representation implies that both $\mathbf{G}(z)$ and $\tilde{\mathbf{\Sigma}}(z)$ are Carathéodory functions up to a conventional factor of the imaginary unit i [44]. This mathematical property constrains the values that the functions can assume in the complex

plane: Given a set of frequency-dependent values $\mathbf{G}(z)$ or $\tilde{\mathbf{\Sigma}}(z)$, a generalized Pick criterion states that the so-called generalized Pick matrix should be positive semidefinite [44]. Additionally the behavior of $\mathbf{G}(z)$ and $\tilde{\mathbf{\Sigma}}(z)$ for $z \to i\infty$ is constrained by the short-time evolution of the Hamiltonian and, via the Hamburger moment problem [47], defines the moments of the spectral function. Values of $\mathbf{G}$ or $\tilde{\mathbf{\Sigma}}$ in the complex plane, and in particular on the real or the imaginary axis, can therefore not be considered as independent quantities and should not be predicted independently.

The issue can be circumvented by using the Lehmann representations of $\mathbf{G}$ and $\tilde{\mathbf{\Sigma}}$ [Eqs. (18) and (19)] directly, which share a general form [44,46,48]

$$\mathbf{Y}(z) = \sum_l \frac{\mathbf{X}_l}{z - \lambda_l}, \quad (20)$$

with $\mathbf{X}_l$ positive semidefinite (PSD) matrices and λ_l real numbers. Predicting PSD matrices $\mathbf{X}_l$ associated with real frequency sampling points λ_l ensures that the resulting Green's function and self-energy fulfill the Carathéodory constraint by construction.

To obtain universal real frequency grids that are applicable to all systems and that scale well as temperature is lowered, we employ the discrete Lehmann representation (DLR) [49,50], which is derived from the truncated spectral Lehmann representation of the imaginary time Green's function

$$\mathbf{G}(\tau) = \int_{-\Lambda}^{\Lambda} K(\tau, \omega) \rho(\omega) d\omega. \quad (21)$$

Here $\Lambda = \beta \omega_{\text{max}}$ is a finite truncation parameter, $\rho(\omega)$ is the spectral density, and the analytical continuation kernel is defined as

$$K(\tau, \omega) = -\frac{e^{-\omega\tau}}{1 + e^{-\beta\omega}}. \quad (22)$$

The DLR frequencies ω_k are chosen based on the discretization of $K(\tau, \omega)$, such that within given accuracy,

$$\mathbf{G}(\tau) \approx \sum_{l=1}^N K(\tau, \omega_l) \mathbf{g}_l, \quad (23)$$

and the corresponding spectral function is given by

$$\rho(\omega) = \sum_{l=1}^N \mathbf{g}_l \delta(\omega - \omega_l). \quad (24)$$

The dynamical self-energy $\tilde{\mathbf{\Sigma}}$ follows similar expression as $\mathbf{G}$. See Ref. [49] for further derivations and additional references.

D. Work flow

Figure 1 demonstrates the general workflow of our method. In finite-temperature grand canonical ensemble calculations, the values of the Green's function are subject to a strict constraint in order to give the correct total number of electrons of the system $N_e = \text{Tr}[\mathbf{PS}]$. Therefore, we choose to use the neural network as a self-energy solver to predict $\mathbf{\Sigma}^{(\text{HF})}$ and $\tilde{\mathbf{\Sigma}}(i\omega_n)$ instead of predicting $\mathbf{G}(i\omega_n)$ directly. All Matsubara frequency-dependent quantities are sampled on the sparse
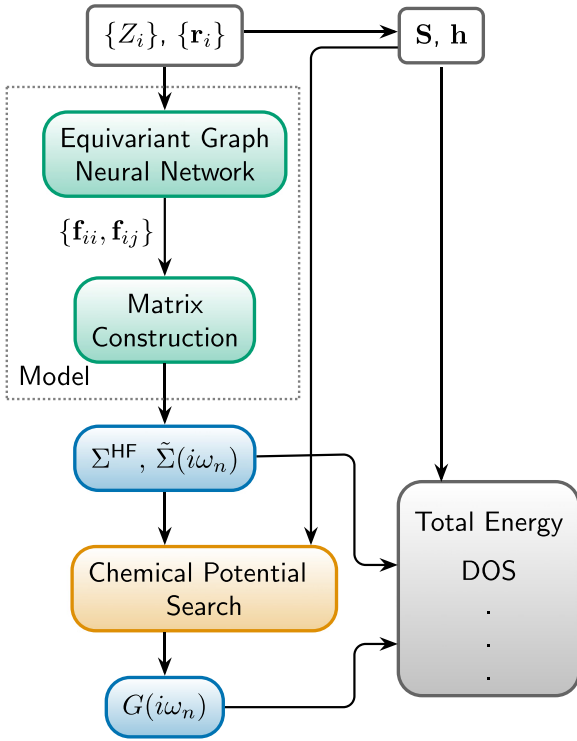
FIG. 1. Flow chart of predicting the Green's function and downstream physical observables from atomic charges $\{Z_i\}$ and positions $\{\mathbf{r}_i\}$. The green boxes indicate the neural network model with trainable parameters. The yellow box specifies the μ-search procedure in Eq. (30). Various physical observables that can be computed from the predicted self-energy and Green's function (blue boxes) are listed in the gray box.

sampling grid [12] to assure accurate and efficient transformations between the time and frequency domain.

Starting from the nuclear charges Z and positions $\mathbf{r}$ of atoms, we construct $N + 1$ matrices $\mathbf{M}$ using the equivariant message-passing neural network explained in Sec. II B, where N is the number of DLR frequencies used for constructing $\tilde{\mathbf{\Sigma}}(i\omega_n)$. The predicted matrices are symmetrized to ensure the Hermitian and PSD properties needed for constructing $\mathbf{\Sigma}^{(\mathrm{HF})}$ and $\tilde{\mathbf{\Sigma}}(i\omega_n)$:

$$\mathbf{M}^{(\mathrm{Herm})} = \mathbf{M} + \mathbf{M}^\dagger, \tag{25}$$

$$\mathbf{M}^{(\mathrm{PSD})} = \mathbf{M}\mathbf{M}^\dagger. \tag{26}$$

During the training process of the neural network, the loss is computed by summing over the square of Frobnius norms of matrix differences in $\mathbf{\Sigma}^{(\mathrm{HF})}$ and $\tilde{\mathbf{\Sigma}}(i\omega_n)$ on all frequency points of all data points x:

$$l_{\mathrm{tot}} = \sum_x l_{x,1} + l_{x,2}, \tag{27}$$

$$l_{x,1} = \left|\left| \mathbf{\Sigma}^{(\mathrm{HF})}_{x,\mathrm{pred}} - \mathbf{\Sigma}^{(\mathrm{HF})}_{x,\mathrm{label}} \right|\right|^2_F, \tag{28}$$

$$l_{x,2} = \sum_n ||\tilde{\mathbf{\Sigma}}_{x,\mathrm{pred}}(i\omega_n) - \tilde{\mathbf{\Sigma}}_{x,\mathrm{label}}(i\omega_n)||^2_F. \tag{29}$$

With the predicted self-energies, the Green's function can be computed using the Dyson equation [see Eq. (5)]

$$\mathbf{G}(i\omega_n) = \frac{1}{(i\omega_n + \mu)\mathbf{S} - \mathbf{h} - \mathbf{\Sigma}^{(\mathrm{HF})} - \tilde{\mathbf{\Sigma}}(i\omega_n)}, \tag{30}$$

with the chemical potential μ determined through a chemical potential search procedure such that the total number of electrons matches the target value. The overlap matrix $\mathbf{S}$ and the one-body integral $\mathbf{h}$ are easy to compute with given atomic structures and basis functions so we treat them as input.

With the predicted $\mathbf{\Sigma}^{(\mathrm{HF})}$, $\tilde{\mathbf{\Sigma}}(i\omega_n)$, $\mathbf{G}(i\omega_n)$, we have full access to the one-particle properties of the given electron system. We will show and compare the predicted total energies [Eq. (7)] and spectral functions (DOS) [Eq. (8)] of different systems in this manuscript.

III. RESULTS

To demonstrate that our method is general, we benchmark the neural network self-energy solver using both small molecules and periodic systems with different temperatures, Matsubara frequency grids, and self-energy approximations. The results for molecules are obtained at relatively high temperature using the self-consistent second-order Green's function perturbation theory (GF2) [1,39–41,51] with a Chebyshev sparse sampling grid [7,22]. The convergence to the zero-temperature limit is achieved by requiring that the total energy differences between the finite temperature and ground-state Hartree Fock (HF) calculations are below 10^{-10} hartree. For periodic systems the calculations are performed at lower temperature using the self-consistent GW approximation [15] with the intermediate representation (IR) sparse sampling grid [6,22,52,53], which has a better scaling as a function of temperature. The GF2 calculations are performed with the full interaction tensor, while the GW calculations are performed with the decomposed interaction using def2-svp-ri auxiliary basis. See Appendix for explicit equations for computing the GF2 and GW self-energy. The interaction tensor, overlap matrix $\mathbf{S}$, and one-body integral $\mathbf{h}$ are all generated using the PySCF [54,55] package. The DLR frequencies are generated using LibDLR [50] with $\omega_{\max} = 100$ for all temperatures. Diagrammatic calculations are performed using Green [56] open source software package.

In the graph neural network setup, molecules are treated as fully connected graphs, i.e., all atom pairs are connected by edges. A cutoff radius $r_{\max}$ is set for periodic systems such that each atom is only connected to all other atoms within this range, see Sec. III B for detailed explanations. We use a three-layer message-passing network for all systems.

In the post-processing procedure, $A(\omega)$ is computed from $\mathbf{G}(i\omega_n)$ using the Nevanlinna analytical continuation method [43,56] which is good at resolving sharp peaks around the Fermi level. Each diagonal component of $\mathbf{G}(i\omega_n)$ in orthogonal basis is continued separately with a broadening parameter η, and $A(\omega)$ is scaled with $1/(\pi\eta)$ in all plots. We use hartree as energy unit through out the paper with all other quantities presented in units that are in correspondence with the energy unit.
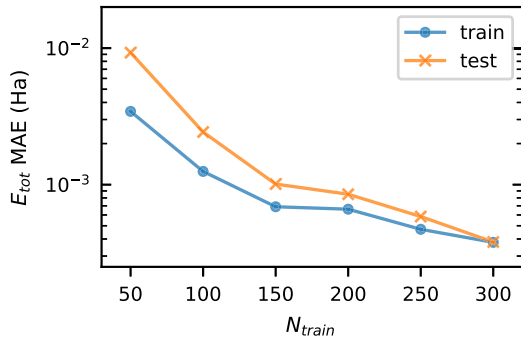
FIG. 2. Learning curve for a single water molecule with STO-3G basis in terms of the MAE of total energy.

A. Molecules

1. Single water molecule

As a first proof-of-concept application of the neural network self-energy solver, we compute the Green's functions, total energies, and spectral functions of a single water molecule using both the STO-3G and cc-pVDZ basis. All results are obtained at $\beta = 100$ with 320 Chebyshev sparse sampling points. The configurations are taken from the data set used by Refs. [57,58], which are sampled from *ab initio* molecular dynamics trajectories [59]. 100 randomly selected configurations are used as test data set, and the rest are used as training data set.

Figure 2 shows the improvement of accuracy in terms of the mean absolute error (MAE) of total energy in STO-3G basis as we increase the training set size. With 100 training data, both the training and testing errors are below 1 mHa per atom, and the train-test gap vanishes with 300 training data.

The testing MAE of various predicted quantities with 300 training data are summarized in Table I. The values of HOMO, LUMO, and gap are obtained from the Nevanlinna analytical continuation with $\eta = 0.01$ and a resolution of 10^{-4}. Figure 3 shows the comparisons of $\text{Tr}[\mathbf{SG}(i\omega_n)]$ and DOS for data points with the largest MAE in $\mathbf{G}(i\omega_n)$ in the test data sets. The results presented in Table I and Fig. 3 demonstrate that our method consistently gives accurate estimations for both basis sets.

2. Small organic molecules

We further benchmark our method with two small organic molecules, benzene and ethanol, taken from the original MD-17 [60] data set. We trained our model with 800 randomly selected configurations and used an additional 100 randomly
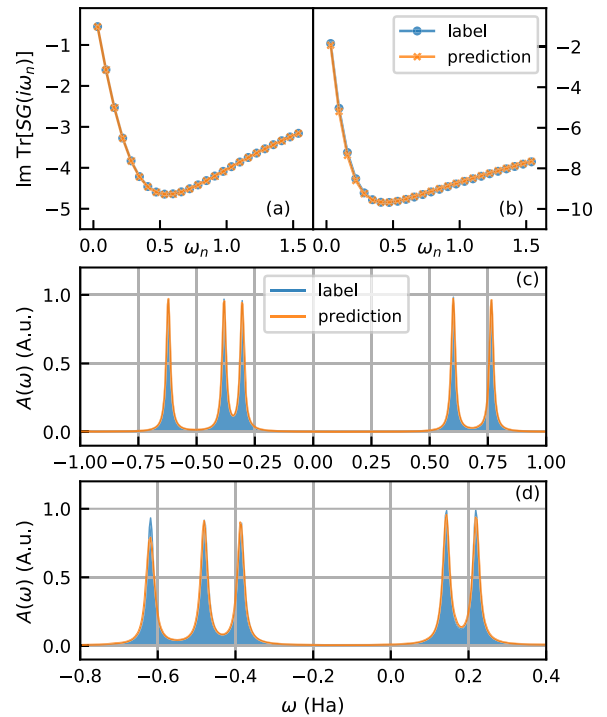

FIG. 3. Comparisons of the label and predicted Green's function [(a), (b)] and DOS [(c), (d)] of single water molecule. (a), (c): Data point with largest MAE in $\mathbf{G}(i\omega_n)$ with STO-3G basis. (b), (d): Data point with largest MAE in $\mathbf{G}(i\omega_n)$ with cc-pVDZ basis.

selected configurations as test data set for each molecule. All calculations are performed at $\beta = 100$ with 320 Chebyshev sparse sampling points using the STO-3G basis.

Table II summarizes the prediction errors of the two molecules. The values of HOMO, LUMO, and gap are obtained from the Nevanlinna analytical continuation with $\eta = 0.01$ and a resolution of 10^{-3}. Comparing the results in Tables I and II, we see that the prediction errors of benzene and ethanol are larger than those of a single water molecule. This can be primarily attributed to the more complicated atomic configurations of these organic molecules, and the prediction accuracy of benzene is slightly better than ethanol due to its more rigid structure. Analogous occurrences have been observed in other machine learning models, such as in Refs. [27,57,60]. Moreover, as the required matrices become larger in orbital space, the prediction task becomes more difficult. As we are only predicting the self-energies, $\mathbf{G}(i\omega_n)$ and E_{tot} are derived properties that do not factor into the supervised learning procedure. With the current setup, we

TABLE I. MAE of 100 testing data for single water molecule.

$\mathbf{\Sigma}^{(\mathrm{HF})}$	$\tilde{\mathbf{\Sigma}}(i\omega_n)$	$\mathbf{G}(i\omega_n)$	E_{tot}	HOMO	LUMO	$\times 10^{-3}$ gap
STO-3G						
0.0312	0.00335	0.0971	0.380	0.163	0.251	0.361
cc-pVDZ						
0.0371	0.00320	0.190	0.819	0.162	0.265	0.253

TABLE II. MAE of 100 testing data for benzene and ethanol.

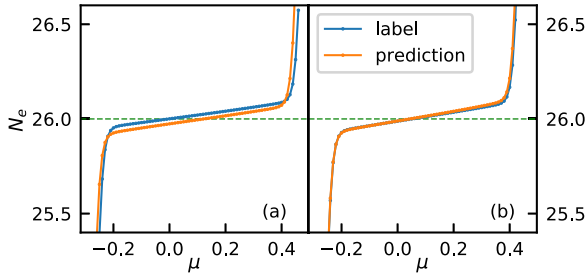| $\mathbf{\Sigma}^{(\mathrm{HF})}$	$\tilde{\mathbf{\Sigma}}(i\omega_n)$	$\mathbf{G}(i\omega_n)$	E_{tot}	HOMO	LUMO	$\times 10^{-3}$ gap
Benzene						
0.158	0.00926	0.225	7.23	1.97	1.12	2.11
Ethanol						
0.612	0.0622	3.19	9.54	3.35	6.01	7.96

FIG. 4. $\mu - N_e$ curve of ethanol with (a) largest MAE in $\mathbf{G}(i\omega_n)$, (b) smallest MAE in $\mathbf{G}(i\omega_n)$.

manage to control the MAE of total energy to around 1 mHa per atom, and the errors of HOMO, LUMO, and gap are also at the order of 10^{-3}. For applications demanding higher energy accuracy, it is possible to supplement the current workflow with a fine-tune procedure utilizing energy as the learning target for better results.

Part of the error amplification from $\mathbf{\Sigma}$ to $\mathbf{G}$ and E_{tot} comes from the chemical potential search procedure introduced in Sec. II D. Figure 4 shows the $\mu - N_e$ curve of data points with the largest and smallest MAE in $\mathbf{G}(i\omega_n)$ for ethanol. Since the $\mu - N_e$ curve is relatively flat around the target electron number, small errors in $\mathbf{\Sigma}^{\text{HF}}$ and $\tilde{\mathbf{\Sigma}}(i\omega_n)$ might cause a non-negligible μ-shift that propagates to $\mathbf{G}(i\omega_n)$. The comparison of $\text{Tr}[\mathbf{SG}(i\omega_n)]$ and DOS of these two data points are shown in Fig. 5. For data point with the largest $\mathbf{G}$ error, the shift in μ causes an obvious shift in the low-frequency part of $\mathbf{G}(i\omega_n)$ as shown in panel (a). However, this shift would not significantly
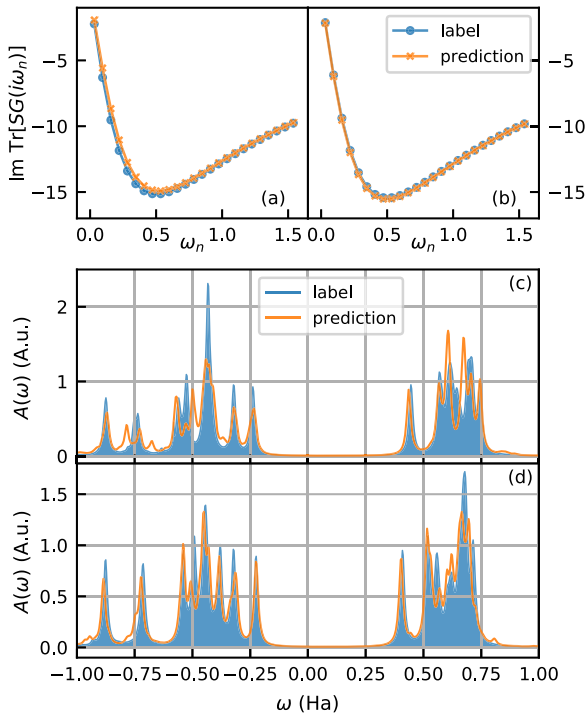

FIG. 5. Comparisons of the label and predicted Green's function [(a), (b)] and DOS [(c), (d)] of ethanol. (a), (c): Data point with largest MAE in $\mathbf{G}(i\omega_n)$. (b), (d): Data point with smallest MAE in $\mathbf{G}(i\omega_n)$.
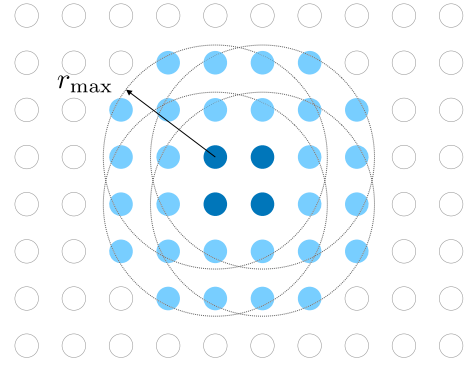

FIG. 6. Schematic plot for periodic system set up. Dark blue dots represent atoms in the center unit cell, dotted lines indicate the cutoff range, and light blue dots are the considered images of atoms.

affect the band gap or HOMO LUMO as shown in panel (c) since the frequency grid is shifted by μ in the mean time.

B. Periodic system

The neural network self-energy solver can be applied to periodic systems in a similar way as isolated molecules. For periodic system calculations carried out in k space, the matrix elements are determined by summing the contributions from all unit cells, each labeled by $\mathbf{R}$, within the periodic lattice

$$\mathbf{M}(\mathbf{k}) = \sum_{\mathbf{R}} e^{-i\mathbf{k}\cdot\mathbf{R}} \mathbf{M}(\mathbf{R}). \tag{31}$$

In the corresponding graph neural network setup, each atom in the center unit cell is connected to all atom images within a given cutoff radius r_{max}. This setup employs a local approximation to address the periodicity of the system [22,26,31,61]. See Fig. 6 for a schematic plot. Unlike constructing the real-space matrices in LCAO which naturally fits into this type of tight-binding setup [22,25,26], recovering k-space matrices in Gaussian basis requires summing over the features from different unit cells with k-dependent phase factors given in Eq. (31).

As an example, we apply our method to Gamma-point calculations of diamond and silicon using their conventional cell with lattice parameters 3.57 Å and 5.43 Å. The configurations are taken from the data set of Ref. [62]. 100 (200) randomly drawn configurations are used as training data for diamond (silicon) the models are tested with another 100 randomly selected configurations. r_{max} is set to be 6 Å for diamond and 10 Å for silicon. To improve prediction accuracy, we trained two neural networks separately for $\mathbf{\Sigma}^{(\text{HF})}$ and $\tilde{\mathbf{\Sigma}}(i\omega_n)$ in this example. All calculations are performed at $\beta = 500$ with 136 IR sparse sampling points generated with $\Lambda = 10^5$ using the STO-3G basis.

The prediction MAE are summarized in Table III with the Nevanlinna continuation performed with $\eta = 0.005$ and a resolution of 10^{-3}. As illustrated in the table, our method demonstrates the ability to predict the total energies and spectral properties of periodic systems with an accuracy comparable to that achieved for molecules, indicating that this method is promising for more complicated applications in real material calculations.

TABLE III. MAE of 100 testing data for diamond and silicon.

$\Sigma^{(HF)}$	$\tilde{\Sigma}(i\omega_n)$	$G(i\omega_n)$	E_{tot}	HOMO	LUMO	gap
						$\times 10^{-3}$
Diamond						
0.0808	0.0118	0.507	3.36	1.09	1.21	1.55
Silicon						
0.0933	0.00371	0.528	8.01	0.463	1.67	1.72

IV. CONCLUSION AND OUTLOOK

In this manuscript, we introduce a general framework for predicting the finite-temperature self-energy and Green's function using equivariant neural network. The proof of concept examples demonstrate that from the predicted self-energy and Green's function, we are able to obtain fairly accurate energy and band gap values for both molecules and periodic systems.

The inference of the neural network scales quadratically with the number of atoms, which is much more efficient than performing actual many-body calculations. Therefore, a trained model could be used for rapid preliminary calculations of electron systems to identify desired properties. On the other hand, the predicted self-energy and Green's function can also serve as a reasonable initial guess of corresponding self-consistent diagrammatic method which accelerates the convergence. For future developments, including specific observables such as energy in the loss function, Eq. (29) is expected to give improved accuracy when accurate results are required. Integrating recent developments of equivariant graph neural network [27,32,63] into our model is anticipated to further improve the prediction accuracy and efficiency. Besides observables that are directly related to the Green's function and self-energy, force and other response properties would also be accessible via automatic differentiation through the trained model.

In summary, the equivariant neural network self-energy solver provides a new opportunity to leverage the rapid development of geometric deep learning to fast and accurate prediction of molecular and material properties at many-body level.

ACKNOWLEDGMENTS

APPENDIX: SELF-CONSISTENT DIAGRAMMATIC METHODS

This Appendix provides the equations for approximating the self-energy within self-consistent diagrammatic methods. The orbital index i and spin index σ will be written separately for clarity.

The HF self-energy is static (frequency independent) and only depends on the density matrix $\mathbf{P}$:

$$\Sigma^{(HF)}_{ij,\sigma} = \sum_{kl}\sum_{\sigma'}(U_{ijkl} - U_{ilkj}\delta_{\sigma\sigma'})P_{kl,\sigma'}, \quad (A1)$$

$$P_{kl,\sigma} = G_{lk,\sigma}(\tau = 0^-). \quad (A2)$$

In the GF2 approximation, the dynamical part of Σ is approximated with two second-order bold self-energy diagrams. The corresponding second order self-energy is given by

$$\Sigma^{(GF2)}_{ij,\sigma}(\tau) = -\sum_{klmnpq} U_{ilnp}G_{lk,\sigma}(\tau)$$
$$\times \sum_{\sigma'} G_{pq,\sigma'}(\tau)G_{mn,\sigma'}(-\tau)(U_{kjqm} - U_{qjkm}\delta_{\sigma\sigma'}). \quad (A3)$$

Within the self-consistent GW approximation, the dynamical part of the self-energy consists of an infinite series of RPA-like bubble diagrams. The self-energy $\Sigma^{(GW)}$ reads as

$$\Sigma^{(GW)}_{ij,\sigma}(\tau) = -\sum_{kl} \tilde{W}_{ilkj}(\tau)G_{lk,\sigma}(\tau), \quad (A4)$$

where $\tilde{\mathbf{W}}$ is the effective screened interaction. The GW self-energy is usually computed using decomposed interaction to get a better scaling, i.e., the interaction tensor is written in a decomposed form

$$U_{ijkl} = \sum_Q V_{ij}(Q)V_{kl}(Q), \quad (A5)$$

and the effective screened interaction is given by

$$\tilde{W}_{ijkl}(i\Omega_n) = \sum_{QQ'} V_{ij}(Q)P_{QQ'}(i\Omega_n)V_{kl}(Q'), \quad (A6)$$

with $\Omega_n = 2n\pi/\beta$, $n \in \mathbb{Z}$ the bosonic Matsubara frequencies. $\mathbf{P}$ is an auxiliary function given by

$$\mathbf{P}(i\Omega_n) = [\mathbf{I} - \mathbf{P}_0(i\Omega_n)]^{-1}\mathbf{P}_0(i\Omega_n), \quad (A7)$$

$$\mathbf{P}_{0,QQ'}(\tau) = -\sum_{\sigma\sigma'}\sum_{abcd} V_{da}(Q)$$
$$\times G_{c\sigma',d\sigma}(-\tau)G_{a\sigma,b\sigma'}(\tau)V_{bc}(Q'). \quad (A8)$$

[1] N. E. Dahlen and R. van Leeuwen, Self-consistent solution of the Dyson equation for atoms and molecules within a conserving approximation, J. Chem. Phys. **122**, 164102 (2005).

[2] D. Zgid and G. K.-L. Chan, Dynamical mean-field theory from a quantum chemical perspective, J. Chem. Phys. **134**, 094115 (2011).

[3] T. N. Lan, A. A. Kananenka, and D. Zgid, Communication: Towards ab initio self-energy embedding theory in quantum chemistry, J. Chem. Phys. **143**, 241102 (2015).

[4] A. A. Rusakov, S. Iskakov, L. N. Tran, and D. Zgid, Self-energy embedding theory (seet) for periodic systems, J. Chem. Theory Comput. **15**, 229 (2019).

[5] T. Zhu, Z.-H. Cui, and G. K.-L. Chan, Efficient formulation of *ab initio* quantum embedding in periodic systems: Dynamical mean-field theory, J. Chem. Theory Comput. **16**, 141 (2020).

[6] H. Shinaoka, J. Otsuki, M. Ohzeki, and K. Yoshimi, Compressing Green's function using intermediate representation between imaginary-time and real-frequency domains, Phys. Rev. B **96**, 035147 (2017).

[7] E. Gull, S. Iskakov, I. Krivenko, A. A. Rusakov, and D. Zgid, Chebyshev polynomial representation of imaginary-time response functions, Phys. Rev. B **98**, 075127 (2018).

[8] S. Iskakov, A. A. Rusakov, D. Zgid, and E. Gull, Effect of propagator renormalization on the band gap of insulating solids, Phys. Rev. B **100**, 085112 (2019).

[9] A. Shee and D. Zgid, Coupled cluster as an impurity solver for Green's function embedding methods, J. Chem. Theory Comput. **15**, 6010 (2019).

[10] X. Dong, D. Zgid, E. Gull, and H. U. R. Strand, Legendre-spectral Dyson equation solver with super-exponential convergence, J. Chem. Phys. **152**, 134107 (2020).

[11] O. J. Backhouse and G. H. Booth, Efficient excitations and spectra within a perturbative renormalization approach, J. Chem. Theory Comput. **16**, 6294 (2020).

[12] J. Li, M. Wallerberger, N. Chikano, C.-N. Yeh, E. Gull, and H. Shinaoka, Sparse sampling approach to efficient *ab initio* calculations at finite temperature, Phys. Rev. B **101**, 035144 (2020).

[13] S. J. Bintrim and T. C. Berkelbach, Full-frequency GW without frequency, J. Chem. Phys. **154**, 041101 (2021).

[14] X. Dong, E. Gull, and H. U. R. Strand, Excitations and spectra from equilibrium real-time Green's functions, Phys. Rev. B **106**, 125153 (2022).

[15] C.-N. Yeh, S. Iskakov, D. Zgid, and E. Gull, Fully self-consistent finite-temperature GW in Gaussian Bloch orbitals for solids, Phys. Rev. B **106**, 235104 (2022).

[16] L.-F. Arsenault, A. Lopez-Bezanilla, O. A. von Lilienfeld, and A. J. Millis, Machine learning for many-body physics: The case of the Anderson impurity model, Phys. Rev. B **90**, 155136 (2014).

[17] E. Sheridan, C. Rhodes, F. Jamet, I. Rungger, and C. Weber, Data-driven dynamical mean-field theory: An error-correction approach to solve the quantum many-body problem using machine learning, Phys. Rev. B **104**, 205120 (2021).

[18] E. J. Sturm, M. R. Carbone, D. Lu, A. Weichselbaum, and R. M. Konik, Predicting impurity spectral functions using machine learning, Phys. Rev. B **103**, 245118 (2021).

[19] C. Venturella, C. Hillenbrand, J. Li, and T. Zhu, Machine learning many-body Green's functions for molecular excitation spectra, J. Chem. Theory Comput. **20**, 143 (2024).

[20] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, Neural message passing for quantum chemistry, in *Proceedings of the 34th International Conference on Machine Learning* (PMLR, Sydney, Australia, 2017), Vol. 70, pp. 1263–1272.

[21] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, R. Faulkner, C. Gulcehre, F. Song, A. Ballard, J. Gilmer, G. Dahl, A. Vaswani, K. Allen, C. Nash, V. Langston, C. Dyer *et al.*, Relational inductive biases, deep learning, and graph networks, arXiv:1806.01261 [cs.LG].

[22] H. Li, Z. Wang, N. Zou, M. Ye, R. Xu, X. Gong, W. Duan, and Y. Xu, Deep-learning density functional theory Hamiltonian for efficient *ab initio* electronic-structure calculation, Nat. Comput. Sci. **2**, 367 (2022).

[23] M. Su, J.-H. Yang, H.-J. Xiang, and X.-G. Gong, Efficient determination of the Hamiltonian and electronic properties using graph neural network with complete local coordinates, Mach. Learn.: Sci. Tech. **4**, 035010 (2023).

[24] O. T. Unke, M. Bogojeski, M. Gastegger, M. Geiger, T. E. Smidt, and K.-R. Müller, Se(3)-equivariant prediction of molecular wavefunctions and electronic densities, in *Neural Information Processing Systems* (Curran Associates Inc., Red Hook, NY, US, 2021).

[25] X. Gong, H. Li, N. Zou, R. Xu, W. Duan, and Y. Xu, General framework for e(3)-equivariant neural network representation of density functional theory Hamiltonian, Nat. Commun. **14**, 2848 (2023).

[26] Y. Zhong, H. Yu, M. Su, X. Gong, and H. Xiang, Transferable equivariant graph neural networks for the Hamiltonians of molecules and solids, npj Comput. Mater. **9**, 182 (2023).

[27] H. Yu, Z. Xu, X. Qian, X. Qian, and S. Ji, Efficient and equivariant graph networks for predicting quantum Hamiltonian, in *Proceedings of the 40th International Conference on Machine Learning, ICML'23* (2023), pp. 40412–40424.

[28] N. Thomas, T. Smidt, S. Kearnes, L. Yang, L. Li, K. Kohlhoff, and P. Riley, Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds, arXiv:1802.08219 [cs.LG].

[29] J. Brandstetter, R. Hesselink, E. van der Pol, E. J. Bekkers, and M. Welling, Geometric and physical quantities improve E(3) equivariant message passing, in *International Conference on Learning Representations* (2022).

[30] M. Geiger and T. Smidt, e3nn: Euclidean neural networks, arXiv:2207.09453 [cs.LG].

[31] S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky, E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials, Nat. Commun. **13**, 2453 (2022).

[32] A. Musaelian, S. Batzner, A. Johansson, L. Sun, C. J. Owen, M. Kornbluth, and B. Kozinsky, Learning local equivariant representations for large-scale atomistic dynamics, Nat. Commun. **14**, 579 (2023).

[33] J. W. Negele and H. Orland, *Quantum Many-particle Systems* (Addison-Wesley, Boca Raton, 1988).

[34] A. Szabo and N. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Dover Books on Chemistry (Dover Publications, New York, NY, 1996).

[35] G. D. Mahan, *Many-Particle Physics*, 3rd ed. (Springer, New York, NY, 2012).

[36] G. Stefanucci and R. van Leeuwen, *Nonequilibrium Many-Body Theory of Quantum Systems: A Modern Introduction* (Cambridge University Press, Cambridge, UK, 2013).

[37] J. M. Luttinger and J. C. Ward, Ground-state energy of a many-fermion system. II, Phys. Rev. **118**, 1417 (1960).

[38] N. E. Dahlen, R. van Leeuwen, and U. von Barth, Variational energy functionals of the green function and of the density tested on molecules, Phys. Rev. A **73**, 012511 (2006).

[39] J. J. Phillips and D. Zgid, Communication: The description of strong correlation within self-consistent Green's function second-order perturbation theory, J. Chem. Phys. **140**, 241101 (2014).

[40] A. A. Rusakov and D. Zgid, Self-consistent second-order Green's function perturbation theory for periodic systems, J. Chem. Phys. **144**, 054106 (2016).

[41] A. R. Welden, A. A. Rusakov, and D. Zgid, Exploring connections between statistical mechanics and Green's functions for realistic systems: Temperature dependent electronic entropy and internal energy from a self-consistent second-order green's function, J. Chem. Phys. **145**, 204106 (2016).

[42] J. Gasteiger, J. Groß, and S. Günnemann, Directional message passing for molecular graphs, in *International Conference on Learning Representations* (2020).

[43] J. Fei, C.-N. Yeh, and E. Gull, Nevanlinna analytical continuation, Phys. Rev. Lett. **126**, 056402 (2021).

[44] J. Fei, C.-N. Yeh, D. Zgid, and E. Gull, Analytical continuation of matrix-valued functions: Carathéodory formalism, Phys. Rev. B **104**, 165111 (2021).

[45] B. Farid, On the Luttinger-Ward functional and the convergence of skeleton diagrammatic series expansion of the self-energy for hubbard-like models, arXiv:2108.10903 [cond-mat.str-el].

[46] C. Gramsch and M. Potthoff, Lehmann representation of the nonequilibrium self-energy, Phys. Rev. B **92**, 235135 (2015).

[47] N. Akhiezer, *The Classical Moment Problem: And Some Related Questions in Analysis*, University Mathematical Monographs (Oliver and Boyd, Philadelphia, PA, 1965).

[48] Z. Huang, E. Gull, and L. Lin, Robust analytic continuation of Green's functions via projection, pole estimation, and semidefinite relaxation, Phys. Rev. B **107**, 075151 (2023).

[49] J. Kaye, K. Chen, and O. Parcollet, Discrete Lehmann representation of imaginary time Green's functions, Phys. Rev. B **105**, 235115 (2022).

[50] J. Kaye, K. Chen, and H. U. Strand, libdlr: Efficient imaginary time calculations using the discrete Lehmann representation, Comput. Phys. Commun. **280**, 108458 (2022).

[51] J. J. Phillips, A. A. Kananenka, and D. Zgid, Fractional charge and spin errors in self-consistent Green's function theory, J. Chem. Phys. **142**, 194108 (2015).

[52] N. Chikano, J. Otsuki, and H. Shinaoka, Performance analysis of a physically constructed orthogonal representation of imaginary-time Green's function, Phys. Rev. B **98**, 035104 (2018).

[53] N. Chikano, K. Yoshimi, J. Otsuki, and H. Shinaoka, ir-basis: Open-source database and software for intermediate-representation basis functions of imaginary-time Green's function, Comput. Phys. Commun. **240**, 181 (2019).

[54] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K.-L. Chan, Pyscf: the Python-based simulations of chemistry framework, WIREs Comput. Mol. Sci. **8**, e1340 (2018).

[55] Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui, J. J. Eriksen, Y. Gao, S. Guo, J. Hermann, M. R. Hermes, K. Koh, P. Koval, S. Lehtola, Z. Li, J. Liu *et al.*, Recent developments in the PySCF program package, J. Chem. Phys. **153**, 024109 (2020).

[56] Green, green-phys.org.

[57] M. Welborn, L. Cheng, and T. F. Miller III, Transferability in machine learning for electronic structure via the molecular orbital basis, J. Chem. Theory Comput. **14**, 4772 (2018).

[58] L. Cheng, M. Welborn, A. S. Christensen, and T. F. Miller III, A universal density matrix functional from molecular orbital-based machine learning: Transferability across organic molecules, J. Chem. Phys. **150**, 131103 (2019).

[59] L. Cheng, M. Welborn, A. S. Christensen, and T. F. Miller, Thermalized (350K) QM7b, GDB-13, water, and short alkane quantum chemistry dataset including MOB-ML features (1.1) [Data set], CaltechDATA, https://doi.org/10.22002/D1.1177 (2019).

[60] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K.-R. Müller, Machine learning of accurate energy-conserving molecular force fields, Sci. Adv. **3**, e1603015 (2017).

[61] T. Xie and J. C. Grossman, Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties, Phys. Rev. Lett. **120**, 145301 (2018).

[62] Q. Gu, Z. Zhouyin, S. K. Pandey, P. Zhang, L. Zhang, and W. E, Deeptb: A deep learning-based tight-binding approach with *ab initio* accuracy, arXiv:2307.04638 [cond-mat.mtrl-sci].

[63] S. Passaro and C. L. Zitnick, Reducing SO(3) convolutions to SO(2) for efficient equivariant GNNs, in *Proceedings of the 40th International Conference on Machine Learning, ICML'23* (2023), pp. 27420–27438.