

Language-inspired machine learning approach for solving strongly correlated problems with dynamical mean-field theory

Hovan Lee^{1,*}, Zelong Zhao,² George H. Booth², Weifeng Ge,^{3,†} and Cedric Weber²

¹*Department of Physics, Royal Holloway, University of London, Egham TW20 0EX, United Kingdom*

²*King's College London, Theory and Simulation of Condensed Matter (TSCM), The Strand, London WC2R 2LS, United Kingdom*

³*Nebula AI Group, School of Computer Science, Fudan University, Shanghai, China*



(Received 2 April 2025; revised 20 June 2025; accepted 9 July 2025; published 24 July 2025)

We present SCALINN—strongly correlated approach with language-inspired neural network—as a method for solving the Anderson impurity model and reducing the computational cost of dynamical mean-field theory calculations. Inspired by the success of generative Transformer networks in natural language processing, SCALINN utilizes an in-house modified Transformer network in order to learn correlated Matsubara Green's functions, which act as solutions to the impurity model. This is achieved by providing the network with low-cost Matsubara Green's functions, thereby overcoming the computational cost of high accuracy solutions. Across different temperatures and interaction strengths, the performance of SCALINN is demonstrated in both physical observables (spectral function, Matsubara Green's functions, quasiparticle weight), and the mean squared error cost values of the neural network, showcasing the network's ability to accelerate Green's function based calculations of correlated materials by predicting DMFT solver solutions, which—given a sufficiently comprehensive training database—can be applied across materials and regimes, enabling *ab initio* DFT+DMFT calculations without prior knowledge of the material properties.

DOI: [10.1103/bk5q-pfb2](https://doi.org/10.1103/bk5q-pfb2)

I. INTRODUCTION

Widespread interest has been devoted in the last three decades to strongly correlated materials, which are being used in emerging technologies, such as spintronics, quantum computing and high temperature superconductivity. They are characterized by strong electronic interactions between their *d*- or *f*-band valence electrons, where the interplay between electron itinerancy (quasiparticle behavior) and electron-electron Coulomb interactions (Mott physics) provides a challenge for standard electronic theories, such as density function theories. The energy scales for the two phenomena overlap for strongly correlated systems. As such, the full characterization of these materials cannot be simplified as the perturbative expansion of one phenomenon against the static backdrop of the another; The competing effects of itinerancy and interaction-induced localization necessitates treating these properties on equal footing.

Through the use of dynamical mean-field theory (DMFT) [1], significant developments have been made towards the understanding of the underlying physics of strong electron correlations in recent years. In particular, the harnessing

of DMFT to colloquial materials modeling methods—such as density functional theory (DFT) with the DFT+DMFT combined approach [2]—has yielded progress into high temperature superconductivity [3–5], cold-atom optical trapping [6–8], and topologically ordered phases [9–11] among other research topics [12]. In the investigations of these materials, the DFT+DMFT approach provides structural and energetic calculations. These calculations are essential for materials research, particularly when sampling the large phase spaces of material structures with structural relaxation.

Despite these formidable achievements, practical challenges remain with the DFT+DMFT approach. At the heart of the theory is the choice of the quantum engine that solves the many-body Anderson impurity model (AIM) [13]. This AIM serves as an underlying model that provides the local Green's function—the descriptor of creation, propagation, and subsequent annihilation of an electron or hole—of the material of interest via a self-consistent mapping. Various methods have been developed to calculate the Green's functions of the AIM within the self-consistent cycle of a DMFT calculation. However, each of these solvers have their own limitations. Examples of these solvers include the Hubbard-I (HI) solver [14], which assumes no electron itinerancy, and is therefore an approximation that is only reasonable for highly localized systems.

The iterative perturbation theory (IPT) solver [1] and its third-order extension, which contains all first-, second- (and optionally third) order irreducible diagrams in the proper self-energy. This is an accurate solver in the low correlation regime, but does not generalize well beyond systems that are half-filled and relatively low interaction strength.

*Contact author: hovan.lee@rhul.ac.uk

†Contact author: wfge@fudan.edu.cn

Continuous time quantum Monte Carlo (CTQMC) methods [15–18] splits the Hamiltonian of the system into two parts (generally hybridization or interaction), and expands the full partition function as powers of one such part, where these powers are stochastically sampled. While formally exact, this method is nonetheless burdened with the sign problem, random errors, and the requirement to analytically continue the resulting Green’s function.

Finally, the exact diagonalization (ED) solver [19–21] which computes the eigenvalues of the AIM Hamiltonian directly, and without approximations in the physics. However, this approach becomes exponentially prohibitive as an increasing number of impurity or bath electron orbitals of the AIM are taken into consideration. This therefore limits the number of orbitals in the calculation, and introduces a finite bath discretization error of the hybridization into the resulting Green’s function.

While this represents an incomplete list of all DMFT solvers that have been considered to date, it is clear that deficiencies remain, and research into improved AIM solvers is still a highly active area of research.

A potential solution could emerge from another discipline; machine learning. Due to advancements in hardware accelerators, improvements in memory capacity, and access to increasingly large databases, deep learning has dominated the field of machine learning in the last decade [22–24]. These developments have given hope to materials scientists; methods to overcome long standing bottlenecks in condensed matter may be in reach [25,26]. A far-from-complete list of these machine learning driven efforts include: predictions of protein structures [27], the learning of exchange-correlation functionals for density functional theory [28], the challenge of ill-conditioned analytic continuation of dynamical quantities [29], or previous efforts aimed at tackling the same problem of DMFT solvers [30–32].

In this work, we develop and demonstrate the capabilities of SCALINN—strongly correlated approach with language-inspired neural network—which is based on the Transformer architecture [24]. SCALINN predicts the Green’s functions of strongly correlated systems as ordered sequences in the Matsubara domain, this is done in order to reliably solve single-impurity Anderson models (SIAM) within self-consistent DMFT calculations. Here, prediction refers to the machine learning terminology of the output of a neural network that has been trained on a dataset.

Rather than encoding the SIAM Hamiltonian directly into the network as input parameters, the characteristics of the SIAM are instead encoded as (potentially multiple) computationally cheap input Green’s functions. Optionally, additional characteristic system parameters are also supplemented to the network input, in order to generate more accurate estimates of the target Green’s function from the decoder of the transformer.

The Transformer framework was chosen due to (1) its potentially infinite memory span—Green’s functions as continuous sequential frequency data exhibit nonlocal dependence between different frequency points, and therefore the procedural generation of these points require the memory provided by the Transformer network, (2) the parallelism enabled by the Transformer formalism, and (3) in the absence of word

embedding, the query, key and value method provides a rich representation between Green’s function frequency sequence entries. (4) The independence of the model on the number of Matsubara points in training, here the number of points is not case specific, in comparison to previously attempted methods such as fully connected deep neural networks. In contrast, our approach provides predictions of varying lengths without the need of retraining the model, allowing for the learning of abstract information from the dataset of, and the prediction on, materials with a wide range of characteristic energy scales.

On the topic of computational and memory efficiency, compact representations [33–35] have enabled efficient storage of Green’s functions by expressing them in terms of carefully chosen basis functions, potentially allowing for the learning of compact representation coefficients through AIM or Hubbard model parameters. In particular, Arsenault *et al.* [36] successfully solved the DMFT equations for the Hubbard model defined on a three-dimensional cubic lattice over a broad range of parameters (including first and second-neighbor hoppings, chemical potential, and interaction strength) with high accuracy. Their method employed a compact representation of the Green’s function using Legendre polynomials. Our method attempts to take another approach altogether by leveraging the advantages provided by the transformer network in order to learn the mapping between computationally cheap, low-level SIAM input Green’s functions, and the fully correlated output Green’s functions. This enables reliable predictions within DMFT iterations in a nonperturbative manner. Moreover, once the machine learning model has been trained to a satisfactory accuracy (to be defined in Sec. II B), the run-time computational cost is independent of the complexity of the problem.

In comparison to other neural networks, a fully connected network or a kernel ridge regression model will only be able to predict series of fixed length. Our model, in contrast, can provide predictions of varying lengths without the need of retraining the model, allowing for the learning of abstract information from the dataset of, and the prediction on, materials with a wide range of characteristic energy scales.

In our present work, we present a general method to predict DMFT solver solutions. Provided that the training database is sufficient, this method could be applied to any materials in any regime, allowing for *ab initio* DFT+DMFT calculations, where the properties of the material are not known *a priori*.

The performance of SCALINN at different values of temperature and interaction strengths is showcased in this work, along with predictions of spectral functions, quasiparticle weights, and Matsubara Green’s functions. More technical parameters, including training and testing errors of the network across different hyperparameters are given in Ref. [37]. We report improvements over the team’s previous machine learning DMFT solver [31]. In particular, SCALINN is able to predict the Green’s function of the Hubbard model at inverse temperature $\beta = 100$ (the energy range at which low-energy excitation occurs) across various values of Coulomb interaction U with a mean squared error (MSE) of $\sim 10^{-7}$, enabling the prediction of the self-energy and spectral function of the system. This is a significant improvement from the previous attempt which was able to reach a validation loss of $\sim 10^{-3}$ at $\beta = 50$, which was shown to be able to calculate the

quasiparticle weight of the system, but was unable to provide predictions on the self-energy and spectral function.

We attribute this improvement of performance to a more intuitive formulation of the problem: the previous work was carried out with fully connected networks, where the only sourcing of improvement of network prediction accuracy was to increase the number of layers in the network, leading to predictions that are biased towards the training dataset. In this work, the input data is provided to the network as Green's function sequential tokens, onto which we can apply the capabilities provided by the attention mechanism.

II. THEORY

A. Dynamical mean-field theory

The SIAM is defined as

$$\begin{aligned} \hat{H}_{\text{SIAM}} = & \underbrace{\sum_{i,\sigma} (\varepsilon_{i\sigma} - \mu) \hat{a}_{i\sigma}^\dagger \hat{a}_{i\sigma}}_{\hat{H}_{\text{bath}}} + \underbrace{\sum_{i,\sigma} (V_{i\sigma} \hat{c}_\sigma^\dagger \hat{a}_{i\sigma} + \text{H.c.})}_{\hat{H}_{\text{hyb}}} \\ & + \underbrace{\sum_{\sigma} (\varepsilon_d - \mu) \hat{c}_\sigma^\dagger \hat{c}_\sigma + U \hat{c}_\uparrow^\dagger \hat{c}_\uparrow \hat{c}_\downarrow^\dagger \hat{c}_\downarrow}_{\hat{H}_{\text{imp}}}, \end{aligned} \quad (1)$$

where the first term $\hat{H}_{\text{bath}}$ describes the dispersion of the uncorrelated bath electrons with spin state σ and site index i , their corresponding annihilation (creation) operators are $\hat{a}_{i,\sigma}^{(\dagger)}$ and the chemical potential is denoted by μ . The second term $\hat{H}_{\text{hyb}}$ determines the hybridization between the impurity and the bath electrons, parameterized by the hybridization strength $V_{i\sigma}$. The last term $\hat{H}_{\text{imp}}$ characterizes the impurity electrons at the single impurity energy level ε_d , with creation (annihilation) operators $\hat{c}_\sigma^{(\dagger)}$ and on-site screened Coulomb interaction strength U .

In the absence of interactions, this Hamiltonian can be written in the following block matrix form:

$$\underline{\hat{H}}_{\text{SIAM}}^0 = \begin{bmatrix} \varepsilon_d - \mu & \underline{V} \\ \underline{V}^\dagger & \underline{\varepsilon} - \mu \underline{\mathbb{I}} \end{bmatrix}. \quad (2)$$

Due to the absence of interactions, there are no terms that explicitly requires the consideration of different spins, and as such the spin index σ is dropped. $\underline{V}$ is a vector of the hybridization strengths V_i , and $\underline{\varepsilon}$ is a diagonal matrix containing the dispersion energies of the uncorrelated bath electrons ε_i .

The corresponding noninteracting, total (encapsulating both the impurity and bath parts of the system) Green's function of the system is defined as

$$\begin{aligned} (i\omega_n \underline{\mathbb{I}} - \underline{\hat{H}}_{\text{SIAM}}^0) \underline{G}_{\text{tot}}^0(i\omega_n) &= \underline{\mathbb{I}}, \\ \underline{G}_{\text{tot}}^0(i\omega_n) &= \begin{bmatrix} i\omega_n + \mu - \varepsilon_d & -\underline{V} \\ -\underline{V}^\dagger & i\omega_n + \mu - \underline{\varepsilon} \end{bmatrix}^{-1}, \end{aligned} \quad (3)$$

where $i\omega_n$ are the discrete Matsubara frequencies.

As a side note, although the DMFT procedure can be carried out in real frequency ω , we choose to operate in Matsubara frequency as this approach does not require artificial broadening and does not introduce sharp features in the Green's functions that destabilize convergence. This is in line with methods that interface DMFT with DFT such

as TRIQS [38], ABINIT [39], and CASTEP [40]. Most of these implementations are performed in the CTQMC Matsubara frequency framework, as such these solvers scale well with the number of orbitals. On this note, numerical renormalization group (NRG) provides a unique toolset to analyze the spectral features of the single band Hubbard model with a remarkable accuracy. As such, NRG has suggested as a solver for DMFT through various machine learning approaches [41–43]. However, it is a notorious challenge to extend NRG to multi-orbital systems, and a port of call for DFT+DMFT charge self-consistent approaches lies instead with Matsubara based solvers such as quantum Monte Carlo.

To concentrate on the strongly correlated portion of the system, the characterization of the impurity electrons does not require a complete description of the bath electron dispersions or the hybridization strengths. As such, the bath information can be condensed; taking only the top left block of $G_{\text{tot}}^0(i\omega_n)$, which is denoted as the noninteracting impurity Green's function $G_{\text{imp}}^0(i\omega_n)$, this $G_{\text{imp}}^0(i\omega_n)$ can be analytically calculated via blockwise inversion of Eq. (3):

$$G_{\text{imp}}^0(i\omega_n) = \frac{1}{i\omega_n + \mu - \varepsilon_d - \Delta(i\omega_n)}, \quad (4)$$

$$\Delta(i\omega_n) = \sum_i \frac{|V_i|^2}{i\omega_n - \varepsilon_i}, \quad (5)$$

where $\Delta(i\omega_n)$ is known as the hybridization function or the dynamical mean field.

In ED, this $\Delta(i\omega_n)$ is approximated with a finite number of bath orbitals:

$$\Delta(i\omega_n) \approx \Delta^{\text{bath}}(i\omega_n) = \sum_{p=1}^{N_b} \frac{V_p^2}{i\omega_n - \varepsilon_p}, \quad (6)$$

with number of bath sites N_b and bath site index p . Each bath site has hybridization strength V_p and energy ε_p . These parameters are typically fit via numerical techniques [44].

The effects of interactions is then supplemented to $G_{\text{imp}}^0(i\omega_n)$ via the Dyson equation to obtain the (interacting) impurity Green's function $G_{\text{imp}}(i\omega_n)$:

$$[G_{\text{imp}}(i\omega_n)]^{-1} = [G_{\text{imp}}^0(i\omega_n)]^{-1} - \Sigma(i\omega_n), \quad (7)$$

where $\Sigma(i\omega_n)$ is the self-energy term given by the DMFT solver.

The resulting $G_{\text{imp}}(i\omega_n)$ is approximated as the wave vectors $\underline{k}$ averaged Green's function of the Hubbard model:

$$\begin{aligned} \hat{H}_{\text{Hubbard}} = & -t \sum_{\langle i,j \rangle, \sigma} c_{j,\sigma}^\dagger c_{i,\sigma} + U \sum_i c_{i,\uparrow}^\dagger c_{i,\uparrow} c_{i,\downarrow}^\dagger c_{i,\downarrow} \\ & - \mu \sum_{i,\sigma} c_{i,\sigma}^\dagger c_{i,\sigma}, \end{aligned} \quad (8)$$

where $c_{i,\sigma}^{(\dagger)}$ denotes the annihilation (creation) operator on the site index i of spin state σ , t is the hopping energy between neighboring sites, U is the on-site screened Coulomb interaction strength, and μ is the chemical potential.

With periodic boundary conditions, the Green's function of the Hubbard model is can be presented in terms of wave vectors $\underline{k}$, which approximates to the SIAM interacting impurity

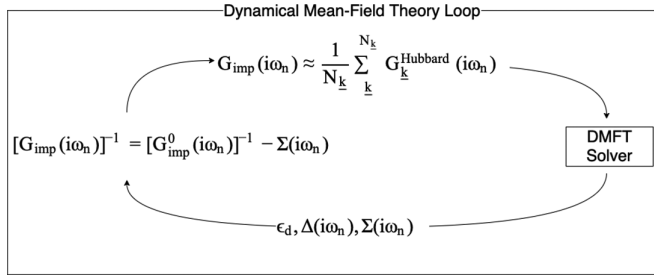


FIG. 1. Schematic diagram of dynamical mean-field theory (DMFT) iterative loop. To initiate the DMFT calculation, a set of initial parameters are chosen for the impurity energy level ϵ_d , the hybridization function $\Delta(i\omega_n)$, and the self-energy $\Sigma(i\omega_n)$ (as denoted at the bottom of the figure). The noninteracting impurity Green's function $G_{\text{imp}}^0(i\omega_n)$ is evaluated via Eq. (4), which is then used to calculate the impurity Green's function $G_{\text{imp}}(i\omega_n)$ with Eq. (7) (left of figure). Next, the local Green's function of the Hubbard model $\frac{1}{N_k} \sum_k G_k^{\text{Hubbard}}(i\omega_n)$ is approximated with $G_{\text{imp}}(i\omega_n)$ (top of figure). The DMFT solver is applied to provide a new set of ϵ_d , $\Delta(i\omega_n)$, and $\Sigma(i\omega_n)$ parameters to be used in the next iteration of the DMFT loop. This process is carried out until convergence.

Green's function via

$$G_{\text{imp}}(i\omega_n) \approx \frac{1}{N_k} \sum_k G_k^{\text{Hubbard}}(i\omega_n), \quad (9)$$

where $G_k^{\text{Hubbard}}(i\omega_n)$ is the k dependent Hubbard model Green's function, and N_k is the number of k points taken into consideration.

In short, Eqs. (1) to (9) provides a pathway to approximating the Green's function of the Hubbard model—a model that is colloquially used to describe the transition between the insulating and conducting phenomena of strongly correlated materials—with the simpler SIAM, providing that the parameters of the SIAM are well chosen.

In DMFT, the choice of the parameters of the SIAM are facilitated through the use of the DMFT solver, and the DMFT self-consistency cycle: a set of initial SIAM parameters are chosen, calculations of $G_{\text{imp}}(i\omega_n)$ are performed, this Green's function is then approximated to the Hubbard model Green's function, which is then fed into the DMFT solver to provide a new updated set of parameters for the SIAM, as shown in Fig. 1.

Updating this self-consistent loop until convergence is a well defined but demanding task, and is where the overhead of computational cost sits. Conventionally this task is carried out by the DMFT solver, here SCALINN is applied to approximate the iterative update of $G_{\text{imp}}(i\omega_n)$ with near instantaneous overhead.

B. SCALINN solver

1. Modes of operation

We considered two different modes of operation for the SCALINN solver as shown in Fig. 2. The first corresponds to the potential to overcome bath discretization error in ED solvers; due to the growth in the dimensions of the Hamiltonian as the number of bath sites increases, we

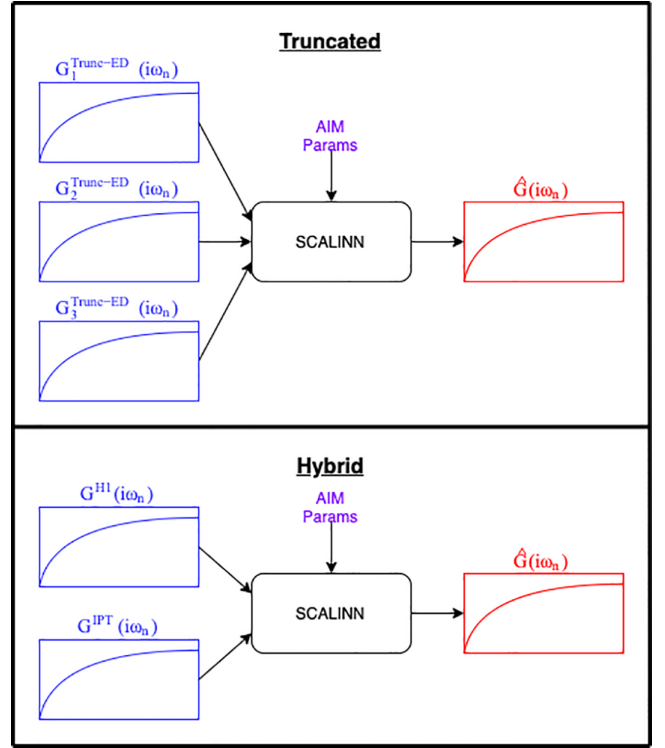


FIG. 2. Illustration of the modes of operation in SCALINN. Input (output) Green's functions are depicted in blue (red). (Top) Truncated mode: a set of three Green's functions calculated from the exact-diagonalization solver with a number of bath sites discarded are used as inputs. (Bottom) Hybrid mode: input corresponds to Green's functions calculated from the Hubbard I and the iterative perturbation theory solvers. Both modes of operation also include Anderson impurity model parameters as inputs, and result in a prediction Green's function $\hat{G}(i\omega_n)$.

attempted to train networks that would reconstruct the Green's function of N_b bath site calculations by providing the network with a set of ED solver calculated Green's functions in which a number of the N_b bath sites are discarded. We call this the 'truncated' mode of operation.

While we have trained models from various different numbers of bath sites. Due to the computational cost to generate the training datasets, we report our results from a model trained on seven-bath sites in the main text, with details and results of the other models in Ref. [37]. We report that as the number of bath sites used to train the model increases, the accuracy of predictions increases.

Given the exponential scaling in calculations with respect to bath size, this method would provide substantial benefit in computational cost. Three of these truncated systems were created, such that all of the seven bath orbitals from the full system is present in at least one of the truncated systems. This is a similar motivation to the 'distributed exact diagonalization' approach [45]. The mapping between the Green's functions of these truncated systems and the full seven-bath site Green's function is then learned by the network. We report that this approach produces Green's functions that are in agreement with the exact calculations obtained from CTQMC.

The second approach is for the network to learn G^{ED} from input Green's functions of complementary characteristics. In our attempts, the HI and the IPT solvers were used. These two input models, while computationally cheap, allow for a description of different extremes in correlated physics. Specifically, HI is superior in the atomic limit, while IPT describes the low- U itinerant physics as a perturbative expansion in powers of U . We will denote this approach the “hybrid” method.

One should note that the choice of the HI and IPT solvers at half-filling are due to the computation requirement to generate a large database of solutions in order to train the network, it is commonly known, for example, that the IPT solver cannot handle systems away from half-filling. This choice was made for the purposes of showcasing the ability of a trained transformer network to predict Green's functions that matches classically computed results. We foresee that away from half-filling, the network will need to be trained with solutions from other DMFT solvers in order to predict physically reasonable results. The key point is that once such networks are trained to satisfaction, the computational cost of using the network to predict a Green's function solution is independent on the complexity of the system under investigation.

2. Encoder-decoder structure and autoregression

As colloquially used in natural language processing, the transformer architecture [24] constitutes an encoder-decoder structure to generate output sequences from input prompts; The encoder maps the input prompt sequences to abstract sets of numerical sequences, whereas the decoder takes and converts the abstract sequences into output tokens. This is illustrated in Fig. 3.

In the case of the SCALINN solver with the “hybrid” approach, the encoder takes input Green's function sequences of N decreasing Matsubara frequencies: the Green's function calculated with the Hubbard-I solver $G^{HI}(i\omega_n) = (G_{N-1}^{HI}, G_{N-2}^{HI}, \dots, G_0^{HI})$ —where G_{N-1}^{HI} represents the Green's function value at the highest Matsubara frequency considered, G_{N-2}^{HI} is the value at the next highest Matsubara frequency, etc.—and $G^{IPT}(i\omega_n)$, a Green's function calculated from the iterative perturbation theory (IPT) solver with the same decreasing frequency order as $G^{HI}(i\omega_n)$. The network then converts these sequences into abstract sequences $A(N) = (a_{N-1}, a_{N-2}, \dots, a_0)$.

Similarly for the “truncated” approach, instead of $G^{HI}(i\omega_n)$ and $G^{IPT}(i\omega_n)$ input Green's functions, we supply the encoder with Green's functions calculated from the truncated systems. Apart from this difference the methodology remains the same, in order to avoid the explanations of the same concepts twice, the remainder of this section will continue with the explanation of the hybrid mode of operation.

Next, an input corresponding to the guess of the Green's function value at the highest frequency considered $G_{N-1}^{input} = \frac{1}{i\omega_{n=N-1}}$ is supplied to the decoder. This input, along with the abstract sequences $A(N)$ (which are themselves mapped onto a set of attention vectors as described below) are then mapped by the decoder into the predicted token corresponding to the Green's function value at the next Matsubara frequency $\hat{G}_{N-2}$.

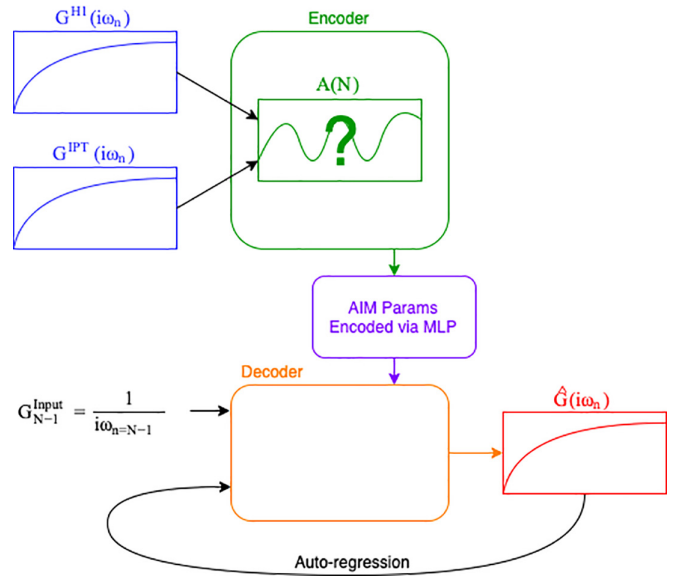


FIG. 3. Illustration of the transformer framework in SCALINN: input Green's functions (blue) are mapped to some abstract sequence $A(N)$ by the encoder (green). The decoder (orange) takes the input $G_{N-1}^{input} = \frac{1}{i\omega_{n=N-1}}$, together with the encoder output, to autoregressively generate prediction (red) tokens $\hat{G}$ that form the prediction Green's function $\hat{G}(i\omega_n)$. Anderson impurity model parameters are included within the encoder output via multilayer perceptron blocks.

The decoder then generates the next predicted token $\hat{G}_{N-3}$ by taking $A(N)$ (in their attention vector form), G_{N-1}^{input} , and $\hat{G}_{N-2}$ as inputs. The decoder carries out this process autoregressively, generating $\hat{G}$ tokens from G_{N-1}^{input} , $A(N)$ and all previous output tokens, until the prediction at the lowest frequency $\hat{G}_0$ is reached.

The significance of the use of autoregression with descending frequency is that the network begins its prediction at the limit of high frequency. At this limit, the Green's function describes single particle excitations. When expanded in this high-frequency limit, the leading term of physical Green's functions scales as $1/i\omega_n$, which is the value we make use of as the input token G_{N-1}^{input} .

As the network iterates towards predictions at lower and lower frequencies, the referencing of the $G^{HI}(i\omega_n)$ and $G^{IPT}(i\omega_n)$ sequences, as well as the training of the network, becomes more and more consequential. This provides a systematic approach from the single electron ionization physics, towards the intermediate and the low-frequency regimes that describe phenomena such as the Mott transition and Fermi liquid behavior.

3. Self-attention mechanism

In order to map and manipulate the abstract sequences, the transformer architecture makes use of the self-attention mechanism—a method which splits each token in a sequence (in this case these tokens would be taken from the abstractly coded sequences of $G^{HI}(i\omega_n)$, $G^{IPT}(i\omega_n)$, G_{N-1}^{input} and the sequence of generated $\hat{G}$ tokens) into three learned vectors: the query, key and value vectors.

Whilst the naming of the query, key and value vectors is not immediately important in this work, a conceptual understanding of these vectors can be grasped in the context of information retrieval and database query processing. In these domains, queries represents user input statements of some desired information. However, a query does not uniquely identify a single result in the database, and is instead referenced with the key descriptors of the objects in the database in order to return a ranked list of results based on the similarities between the query and the keys. The value of the objects correspond to the data content of each object.

For physical Green's functions, we know that successive points in $G(i\omega_n)$ are complementary to one another— $G(i\omega_n)$ are smooth and continuous, reflecting the fact that the system's excitations (e.g., electrons, holes, or quasiparticles) are not isolated events but are part of a continuous spectrum of states. Moreover, emergent effects (e.g., Mott transitions or Kondo physics) introduce correlations that spans all frequencies scales. For example, the formation of a Mott gap at low frequencies is influenced by the high-frequency structure of the self-energy. We make use of the attention mechanism to handle these dependencies between Matsubara frequency points.

In the case of SCALINN, a set of query, key, and value vectors are evaluated for each token based on the training of the network. The length d_k query vector of a specific token is to be multiplied with the length d_k key vectors of all the tokens in the sequence via dot product. This is done in order to calculate the similarities of the query vector of the specific token with the key vectors of the tokens in the sequence. The output of the self-attention mechanism is then a weighted sum where the value vectors of the token are weighted by the dot product between the query and key vectors.

When vectorized across a length n' input sequence, this dot product would be expressed as $\underline{Q} \cdot \underline{K}^T$, where $\underline{Q}$ is a $n' \times d_k$ matrix in which the n' rows correspond to the length d_k query vectors of the tokens, $\underline{K}^T$ is a $d_k \times n'$ matrix where the n' columns correspond to the length d_k key vectors of the tokens. This $n' \times n'$ matrix $\underline{Q} \cdot \underline{K}^T$ is then scaled by $1/\sqrt{d_k}$, and subsequently applied to a softmax function. This softmax function normalizes the rows of $\underline{Q} \cdot \underline{K}^T / \sqrt{d_k}$, such that the n_i th row correspond to a normalized set of n' metrics based on the similarities between the query vector of the n_i th token and the key vectors of the tokens across the whole sequence: this set of metrics are the attention weights, corresponding to how much attention should be placed on the each of the n' tokens given the query of the n_i th token.

A dot product is then applied between this attention weight matrix and the $\underline{V}$ matrix of size $n' \times d_v$, such that the attention values are applied as the weights of a weighted sum to the value vectors of the sequence.

The full operation is expressed as

$$\text{Attention}(\underline{Q}, \underline{K}, \underline{V}) = \text{softmax} \left(\frac{\underline{M} + \underline{Q} \cdot \underline{K}^T}{\sqrt{d_k}} \right) \cdot \underline{V}, \quad (10)$$

where $\underline{M}$ is a $n' \times n'$ matrix of mask values, and is described in Sec. [II B 4](#).

Throughout our tests (please refer to Sec. 3 of the Supplemental Material [\[37\]](#)), we found that the most accurate models are trained by splitting the AIM parameters into impurity parameters (ϵ_d, U), and hybridization parameters (ϵ, V), and encoding these two sets of parameters through multilayer perceptron blocks (successive layers of fully connected networks) into the encoder-decoder self-attention mechanism.

4. Masking and Mastubara encoding

Despite handling the input prompts as sequences, the transformer network does not parse inputs and previously predicted tokens sequentially. Instead this process is carried out in parallel: each head of the multihead attention blocks employed in the network applies its own version of the attention mechanism of Eq. (10).

Generally, with a higher number of heads, the richness of information that can be learned in these multihead self-attention blocks increases. For example, a number of heads may be dedicated to capture the local (in terms of energy) smooth and continuous nature of Green's functions, whilst other heads may capture many-body effects that spans the whole frequency spectrum. The hyperparameter corresponding to the number of heads is varied, and the performances of models with different number of heads, and different number of decoder blocks (n_{Decoder}) are listed in Table A1 of Supplemental Material [\[37\]](#).

As such, the order of the input sequences in descending Matsubara frequency must be encoded into the tokens of the sequence explicitly. Our attempts at various different Matsubara encoding schemes are detailed in Ref. [\[37\]](#).

This nonsequential in manner of sequence-handling however, would mean that during training, the self-attention heads of the network are able to reference full length N target Green's function sequences $G^{\text{Training}}(i\omega_n)$ obtained from the exact diagonalization solver for seven-bath sites. These $G^{\text{Training}}(i\omega_n)$ correspond to the values that a fully trained network would be able to predict (i.e., the predicted Green's function $\hat{G} = (\hat{G}_{N-1}, \hat{G}_{N-2}, \dots, \hat{G}_0)$). Whereas, the network should be trained such that it cannot reference $\hat{G}$ tokens of lower frequencies that it has not generated yet.

As such, mask matrices $\underline{M}$ are applied to the self-attention blocks in the decoder, in order to mask these lower frequency tokens such that the network will not be able to reference them during training. For details on masking, please refer to Ref. [\[37\]](#). An illustration of masking and the attention mechanism is shown in Fig. [4](#).

5. Training versus run time

As alluded to, there is a slight difference of the network during training and during run time. In training the network, the decoder is sequentially provided with the seven bath site ED Green's function values starting at the highest frequency considered. This is done in order for the network to generate a prediction token for the Green's function value at the next highest frequency.

The cost function to be minimized is defined as the mean squared error to the seven-bath ED solution at a set of training

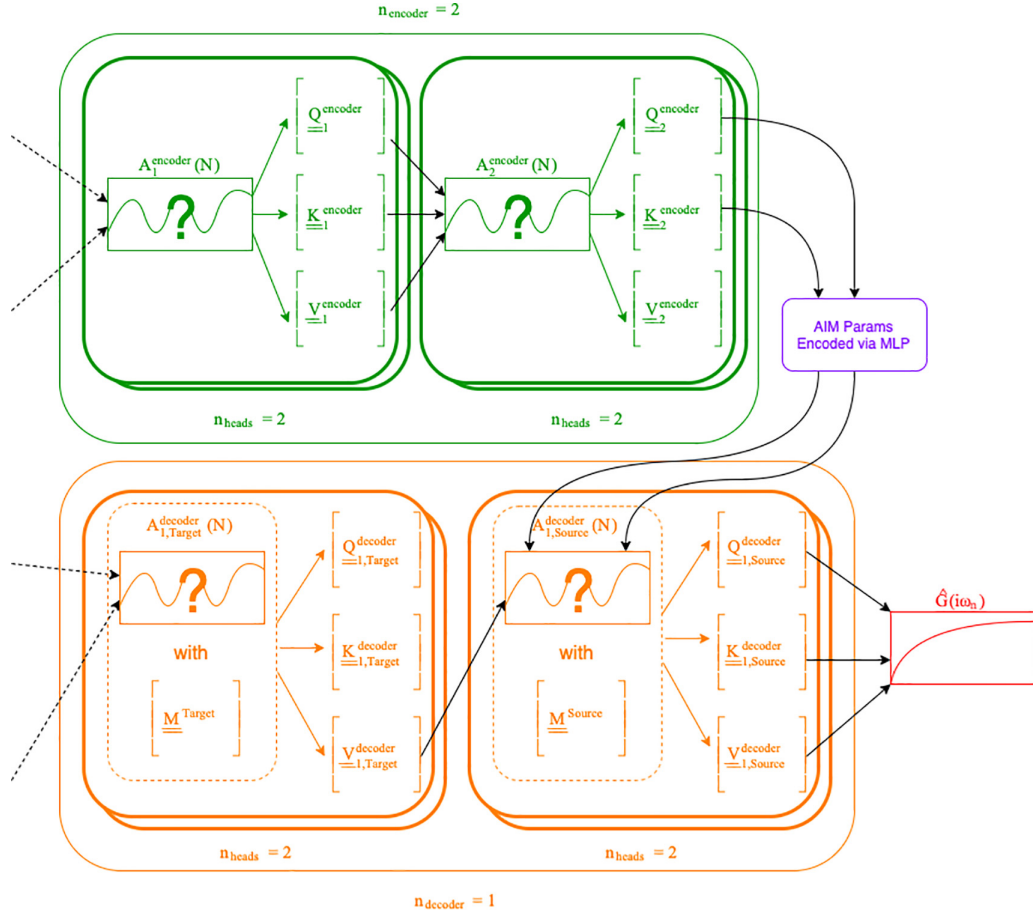


FIG. 4. Illustration of masking and the self-attention mechanism: dotted lines denote the inputs to the encoder (green) and the decoder (orange). The inputs of the encoder are mapped into the abstract sequence $A_1^{\text{encoder}}(N)$. In this example, the hyperparameter $n_{\text{encoder}} = 2$ was chosen, giving rise to two successive sets of self-attention query, key and value vectors in the encoder, denoted as $\underline{Q}_i^{\text{encoder}}, \underline{K}_i^{\text{encoder}}, \underline{V}_i^{\text{encoder}}$ respectively, for $i = 1$ and 2 . The query, key, and value vectors of the first encoder block is then mapped to the abstract sequence for the next encoder block $A_2^{\text{encoder}}(N)$. From $A_2^{\text{encoder}}(N)$, the second encoder set of query, key and value vectors are obtained. Likewise, the input to the decoder are mapped to $A_1^{\text{decoder}}(N)$, onto which a target mask $\underline{M}^{\text{Target}}$ is applied in order to train the mapping of the decoder target query, key and value vectors via Eq. (10). The decoder source block then takes the last set of encoder query and key vectors $\underline{Q}_2^{\text{encoder}}, \underline{K}_2^{\text{encoder}}$, along with the target decoder value vectors $\underline{V}_{1,\text{Target}}^{\text{decoder}}$ in order to obtain the abstract decoder source sequence $A_{1,\text{Source}}^{\text{decoder}}(N)$. From $A_{1,\text{Source}}^{\text{decoder}}(N)$ the source decoder query, key, and value vectors are calculated in order to predict the Green's function $\hat{G}(i\omega_n)$. All encoder and decoder subblocks are depicted with $n_{\text{heads}} = 2$ heads to showcase the parallelism of the method. The residual connection, feed forward, and add and norm blocks of the traditional transformer network are also employed within SCALINN, these blocks are not shown in this figure for brevity.

points,

$$J(\hat{G}(i\omega_n), G^{\text{ED}}(i\omega_n)) = \frac{1}{N} \sum_{n=0}^{N-2} (\hat{G}(i\omega_n) - G^{\text{ED}}(i\omega_n))^2, \quad (11)$$

where $\hat{G}(i\omega_n)$ is the indexed set of predicted tokens $\hat{G}$, $G^{\text{ED}}(i\omega_n)$ is the seven-bath ED solution of the SIAM, and the sum is performed over Matsubara frequency points of the generated tokens. N can be increased such that the model is trained to generate as many Matsubara frequency points as required. In practice, increasing the number of Matsubara points considered beyond $N = 32$ in this mean squared error evaluation does not dramatically increase the accuracy of the prediction, as will be shown below, the results obtained from the model is within the statistical error bars of the CTQMC result.

During run-time however, only the highest frequency point is provided to the decoder with $G_{N-1}^{\text{input}} = \frac{1}{i\omega_{n=N-1}}$ as mentioned above. To continue past the first prediction, the subsequent generated tokens are provided back to the decoder autoregressively.

III. RESULTS

A. Computational details

To setup the training data of the model, three $G^{\text{ED}}(i\omega_n)$ data sets at different inverse temperatures of $\beta = 10, 50$, and 100 were generated, each consisting of 14 000 different seven-bath SIAM Hamiltonians. The ED Green's functions were generated using the ED-KCL package [46], with DMFT functionality where required enabled via interface to the TRIQS package [38]. Implementation of the Transformer model made

use of the PYTORCH package [47]. The training of our base model, detailed in Table A1 of Supplemental Material [37], was conducted on a single NVIDIA A100 GPU, where each model was trained on 80% of the data for 350 000 steps, which took approximately 9 hours per model, with the rest of the data used for evaluation of the model accuracy. During run-time, each DMFT loop takes $\sim 10^3$ seconds on the A100 GPU. It should be noted that this GPU computational cost is independent on the number of bath sites of the training dataset. In contrast, a traditional seven-bath site ED calculation on the same GPU takes ~ 1 minute, the size of the Green's function matrix to be inverted (as shown in Fig. 1) scales as 2^N for a paramagnetic calculation, and as such the computational cost of each ED DMFT loop scales accordingly.

In order to generate a physically relevant dataset, the bath energy values ϵ , and hybridization parameters V are randomly sampled, with boundaries that are consistent compared to those of the Bethe lattice. For further details, please refer to Ref. [37].

B. Truncated Hamiltonian

We first consider the “truncated” mode of operation of the SCALINN model, whereby predictions of the seven-bath SIAM models are created from ED Green's functions with only three bath orbitals, mitigating the exponential increase in cost of ED with respect to bath size. The three-bath SIAM Hamiltonians were created from their seven-bath counterparts by inclusion of the $\epsilon = 0W$ bath orbital, and then selection of one particle-hole symmetric pair of bath orbitals at finite frequency. This allows us to create three three-bath SIAM approximations to each SIAM of interest. These are solved with ED to provide the input to SCALINN, as described in Sec. II B 1.

An example of the effect that this bath dropout has on Green's functions is shown in Fig. 5(a), where these truncated bath models differ significantly from the desired seven-bath solution. Unless otherwise specified, $G(i\omega_n)$ indicates only the imaginary component of the Green's functions in all the figures shown below.

SCALINN nevertheless predicts the seven-bath ED solution with high accuracy after training of the model. SCALINN models are trained at three different inverse temperatures, with the average training error reduced to below $J(\hat{G}(i\omega_n), G^{\text{ED}}(i\omega_n)) < 10^{-5}$ in all cases. We can analytically continue the predicted Matsubara Green's functions onto the real-frequency axis via Padé approximants, to consider their accuracy of the real-frequency spectrum, shown in Fig. 5(b) for a representative test-set prediction, with three values of inverse temperature β and three values of interaction U .

The SCALINN predictions agree very well with the ED ground truth for these systems, even on the real axis, with only small deviations are observed in Fig. 5(b). As the interaction increases, classic hallmarks of correlated materials emerge, with low U describing metals with a single quasiparticle peak at zero energy. As temperature increases (or as β decreases), so does the rate of scattering between electrons, leading to a broadening of these peaks. At intermediate interaction strength $U = 3W$, lower and upper Hubbard bands at $\pm U/2W$ emerge around the quasiparticle peaks at $\omega = 0$. Once again, due to increasing rates of scattering, these three

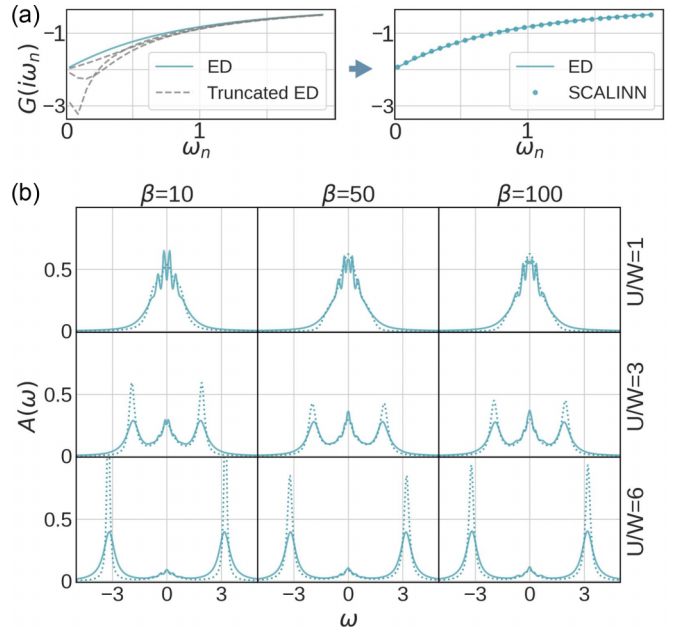


FIG. 5. Representative SCALINN predictions for a test-set SIAM with the truncated bath scheme. In a), the desired (seven-bath) ED solution $G^{\text{ED}}(i\omega_n)$ at $\beta = 100$, $U = 1$ is plotted as a blue line, and the three truncated (three-bath) $G^{\text{Trunc}}(i\omega_n)$ are plotted in dashed gray. These $G^{\text{Trunc}}(i\omega_n)$ act as inputs to the Transformer model to produce the SCALINN predictions $\hat{G}(i\omega_n)$, which are plotted as blue dots. In (b), Padé analytic continuation obtains the predicted spectral function $A(\omega)$, shown here for three different values of β and U . Once again, the ED solution is plotted as blue lines, and the SCALINN predictions are plotted as blue dots.

sets of peaks are broadened as temperature increases. Lastly, at higher interaction strength still $U = 6W$, the magnitude of the quasiparticle peaks are greatly reduced in favor of the Hubbard bands. However, due to the lack of DMFT self-consistency in this case, the fully insulating Mott solution was not recovered.

C. Hybrid approximation solver

In addition to the truncated bath approach, as discussed in Sec. II B 1, we also consider a Hubbard-I+IPT hybrid scheme. In this hybrid method, models were trained from a combination of Hubbard-I and IPT approximate solutions to the target SIAM, and used as inputs to the transformer model to predict target ED-quality outputs. Once again, three separate inverse temperature models were trained to reach errors of $J(\hat{G}(i\omega_n), G^{\text{ED}}(i\omega_n)) < 10^{-5}$ across the entire training set.

In Fig. 6, the various Matsubara Green's functions are presented, including the Hubbard-I and IPT inputs, the ED ground truth and SCALINN prediction. As expected, the Hubbard-I inputs are all insulating solutions, as observed with $G^{\text{HI}}(i\omega_n)$ tending towards zero as $i\omega_n \rightarrow 0$, consistent with its description of the atomic solution. This is increasingly erroneous for lower temperatures and interaction strengths, where delocalized solutions should be found. In contrast, the IPT input favors delocalized descriptions, where all solutions reach a maximum absolute value as $i\omega_n \rightarrow 0$, which is in error particularly for higher interactions in the nonperturbative U/W limit. In contrast to these computationally cheap input

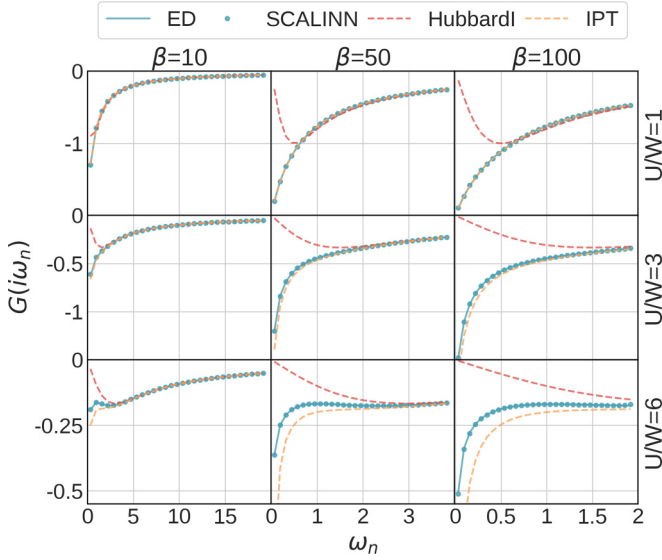


FIG. 6. Representative SCALINN predictions for a test-set SIAM with the HI+IPT hybrid scheme. The Hubbard-I solution $G^{\text{HI}}(i\omega_n)$ (dashed red), the IPT solver solution $G^{\text{IPT}}(i\omega_n)$ (dashed yellow), and the ED target $G^{\text{ED}}(i\omega_n)$ (solid blue lines) are plotted alongside the SCALINN predictions from these input Green's functions (blue dots) at $\beta = 10, 50, 100$ and at $U/W = 1, 3, 6$.

models, the SCALINN predictions match the ED results to remarkably high accuracy from these inputs.

Finally, we consider the utility of the scheme as a solver within a fully self-consistent DMFT calculation. In this, the IPT and Hubbard-I approximations can be found in the absence of bath discretization error of the hybridization. Therefore, while the training of the SCALINN model is performed in the presence of the finite bath approximation to the hybridization, its use within a DMFT scheme can aim to eliminate both the bath discretization error, as well as approximations to the correlated effects of approximate solvers. In order to benchmark the accuracy of the approach, we therefore turn to comparison with a CTQMC solver which can obtain correlated Green's functions in the limit of a continuous hybridization via Monte Carlo sampling.

In Fig. 7, we consider DMFT on the continuous hybridization of the Bethe lattice at $\beta = 10$ compared to CTQMC. Note that the entire DMFT calculation was performed via inference on the trained neural network, without any need for extra calculations or retraining, at zero added computational cost. At each DMFT iteration we only need to calculate the approximated Green's functions: $G^{\text{HI}}(i\omega_n)$, $G^{\text{IPT}}(i\omega_n)$, these Green's functions are given to the network as inputs for the model to predict an updated $\hat{G}(i\omega_n)$. This process is performed iteratively until convergence was reached. For both the final self-consistent Matsubara Greens function, and the quasiparticle weight, Z . This quasiparticle weight can be computed as

$$Z = \frac{\text{Im} \{ \Sigma(i\omega_n) \}}{\omega_n} \Big|_{\omega_n \rightarrow 0}, \quad (12)$$

from the converged DMFT self-energy. Values close to unity indicate a metallic solution, with lower values describing the increased effective mass towards a Mott insulating solution.

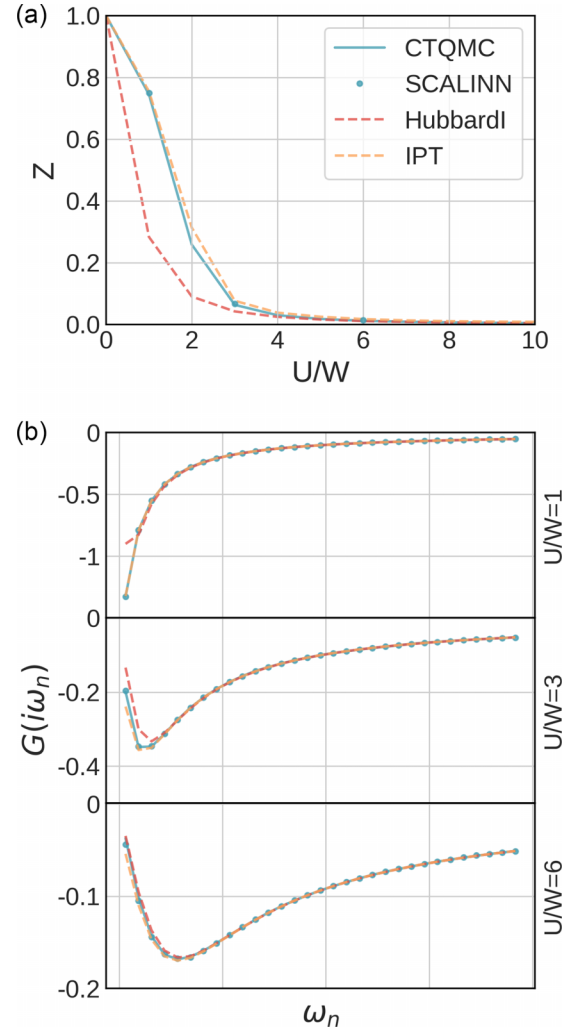


FIG. 7. Converged self-consistent DMFT results on the Bethe Hubbard lattice for different solvers: IPT, Hubbard-I, SCALINN, and CTQMC. (a) Quasiparticle weight. (b) Imaginary part of Matsubara Green's functions. DMFT is converged at $\beta = 10$ for the continuous hybridization of the Bethe lattice without bath discretization error, with the IPT and Hubbard-I solutions at each iteration used as input for the SCALINN solver approach.

This is observed from the results, with DMFT+SCALINN agreeing almost perfectly with the DMFT+CTQMC renormalization factor, with the DMFT+IPT biasing towards the metallic phase, and DMFT+Hubbard-I the atomic Mott phase.

Moreover, the self-consistent DMFT Matsubara Green's functions with these various solvers are plotted in Fig. 7(b), where as interaction strength increases, the solutions become increasingly insulating, as can be observed from the $i\omega_n \rightarrow 0$ trend of $G(i\omega_n)$, with once again the discrepancy between SCALINN and CTQMC solvers is indistinguishable on the scale of the plot. It should be noted that these insulating solutions differ from the single-shot SIAM solutions of Fig. 6, due to the self-consistent update of the continuous hybridization in the full DMFT scheme.

Despite the agreement reached between the predicted and the CTQMC calculated Green's functions, a stable

convergence at $\beta = 50$ and 100 in the strongly correlated regime has not been achieved. We attribute this to the increased size of the database required to train networks of systems at lower temperatures. Note that one advantage of our purposed method is that it is scalable for large databases, in contrast to alternatives such as Kernel ridge or Gaussian methods.

IV. CONCLUSIONS

With the insights gained from drawing comparisons to natural language processing problems, we developed a novel and promising approach to predict Green's function sequences in the Matsubara domain via modifications to a Transformer model. These predicted sequences exhibit levels of accuracy that were previously restricted to comparatively high computational costs of exact diagonalization. We considered approaches to both predict these Green's functions of general SIAM models from inputs based on results from computationally accessible lower levels of theory, as well as an approach to mitigate the bath discretization error in describing SIAM's with a continuous hybridization spectrum. Finally, we combined these developments in a fully self-consistent DMFT scheme to solve the Bethe lattice Hubbard model with results indistinguishable from exact CTQMC benchmarks.

However, while the approach showcases much potential, there exist remaining challenges to overcome. Firstly, the approach was restricted to a relatively narrow class of SIAM models from a training dataset of a maximum of seven fermionic levels. This set of training data can still be improved upon to reach competitive levels as a DMFT solver. As such, the extension to multi-impurity, matrix-valued Green's functions and a wider set of representative hybridizations is required. The advantage gained from the use of this model is that once a model is trained from this set of larger number of fermionic levels, wider hybridization representation extended dataset, the run-time GPU computational cost does not increase accordingly. Note however that one of the key take-aways of our work is that sampling the entire space of AIM (those represented by seven-bath sites and all those that are not) is not required to achieve excellent accuracy for the machine learning approach. Essentially our work demonstrates that training the network on discretized AIMs provides a solid

network that can apply to any AIMs. One remarkable demonstration is that comparison to the exact (within statistical error bars) CTQMC results for DMFT. Extending the database with AIM with larger bath sites has been shown in our work to only provide minor improvements, see for instance the analysis in Eq. (6) in Ref. [37] where we review the dependence on number of bath sites. Adaptations of the model to *enforce* desirable features such as causality of the output Green's functions or symmetries would be beneficial, and help avoid convergence issues in the self-consistent DMFT loops which manifested at times for low temperatures or quantum phase transitions.

Moreover, alternative methods to provide training data of exact Green's functions would allow for an extension to overcome the bath discretization which is manifest in the training of the model. Nevertheless, our findings highlight the power and adaptability of the Transformer model within the field of correlated materials and its potential for pushing the frontiers of computational problem solving in this domain.

Finally, additional networks can also be trained with solutions from other DMFT solvers, such as the publicly available TRIQS-CTSEG solver [48].

ACKNOWLEDGMENTS

G.H.B. gratefully acknowledges support from the Air Force Office of Scientific Research under award number FA8655-22-1-7011. We are also grateful to the King's Computational Research, Engineering and Technology Environment (CREATE) and UK Materials and Molecular Modelling Hub for computational resources, which is partially funded by EPSRC (EP/T022213/1, EP/W032260/1, and EP/P020194/1).

C.W., G.H.B, W.G., and Z.Z. conceived the project. C.W. and Z.Z. build the database. W.G., H.L. and Z.Z. developed the Machine Learning model and trained the Transformer model. H.L. and Z.Z. performed data analysis. All wrote the manuscript.

The authors declare no conflict of interest.

DATA AVAILABILITY

The data and code that support the findings of this article are openly available [49,50].

-
- [1] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, Exact functionals, effective actions and (dynamical) mean-field theories: Some remarks, *Rev. Mod. Phys.* **68**, 13 (1996).
 - [2] G. Kotliar, S. Y. Savrasov, K. Haule, V. S. Oudovenko, O. Parcollet, and C. A. Marianetti, Electronic structure calculations with dynamical mean-field theory, *Rev. Mod. Phys.* **78**, 865 (2006).
 - [3] H.-C. Jiang and T. P. Devereaux, Superconductivity in the doped Hubbard model and its interplay with next-nearest hopping t' , *Science* **365**, 1424 (2019).
 - [4] M. Kheirkhah, Z. Yan, Y. Nagai, and F. Marsiglio, First- and second-order topological superconductivity and temperature-driven topological phase transitions in the extended Hubbard

model with spin-orbit coupling, *Phys. Rev. Lett.* **125**, 017001 (2020).

- [5] A. Foley, S. Verret, A.-M. S. Tremblay, and D. Sénéchal, Coexistence of superconductivity and antiferromagnetism in the Hubbard model for cuprates, *Phys. Rev. B* **99**, 184510 (2019).
- [6] J. Koepsell, J. Vijayan, P. Sompet, F. Grusdt, T. A. Hilker, E. Demler, G. Salomon, I. Bloch, and C. Gross, Imaging magnetic polarons in the doped Fermi-Hubbard model, *Nature (London)* **572**, 358 (2019).
- [7] P. T. Brown, D. Mitra, E. Guardado-Sanchez, R. Nourafkan, A. Reymbaut, C.-D. Hébert, S. Bergeron, A.-M. Tremblay, J. Kokalj, D. A. Huse *et al.*, Bad metallic transport in a cold atom Fermi-Hubbard system, *Science* **363**, 379 (2019).

- [8] B. Yang, H. Sun, R. Ott, H.-Y. Wang, T. V. Zache, J. C. Halimeh, Z.-S. Yuan, P. Hauke, and J.-W. Pan, Observation of gauge invariance in a 71-site Bose–Hubbard quantum simulator, *Nature (London)* **587**, 392 (2020).
- [9] C. Shao, E. V. Castro, S. Hu, and R. Mondaini, Interplay of local order and topology in the extended Haldane–Hubbard model, *Phys. Rev. B* **103**, 035125 (2021).
- [10] J. Fraxanet, D. González-Cuadra, T. Pfau, M. Lewenstein, T. Langen, and L. Barbiero, Topological quantum critical points in the extended Bose–Hubbard model, *Phys. Rev. Lett.* **128**, 043402 (2022).
- [11] D. González-Cuadra, A. Dauphin, P. R. Grzybowski, P. Wójcik, M. Lewenstein, and A. Bermudez, Symmetry-breaking topological insulators in the Bose–Hubbard model, *Phys. Rev. B* **99**, 045139 (2019).
- [12] C. Weber, D. D. O’Regan, N. D. M. Hine, P. B. Littlewood, G. Kotliar, and M. C. Payne, Importance of many-body effects in the kernel of hemoglobin for ligand binding, *Phys. Rev. Lett.* **110**, 106402 (2013).
- [13] P. W. Anderson, Localized magnetic states in metals, *Phys. Rev.* **124**, 41 (1961).
- [14] J. Hubbard, Electron correlations in narrow energy bands, *Proc. Math. Phys. Eng. Sci.* **276**, 238 (1963).
- [15] E. Gull, P. Werner, A. Millis, and M. Troyer, Performance analysis of continuous-time solvers for quantum impurity models, *Phys. Rev. B* **76**, 235123 (2007).
- [16] P. Werner and A. J. Millis, Hybridization expansion impurity solver: General formulation and application to Kondo lattice and two-orbital models, *Phys. Rev. B* **74**, 155107 (2006).
- [17] P. Werner, A. Comanac, L. de’Medici, M. Troyer, and A. J. Millis, Continuous-time solver for quantum impurity models, *Phys. Rev. Lett.* **97**, 076405 (2006).
- [18] A. N. Rubtsov, V. V. Savkin, and A. I. Lichtenstein, Continuous-time quantum Monte Carlo method for fermions, *Phys. Rev. B* **72**, 035122 (2005).
- [19] M. Caffarel and W. Krauth, Exact diagonalization approach to correlated fermions in infinite dimensions: Mott transition and superconductivity, *Phys. Rev. Lett.* **72**, 1545 (1994).
- [20] M. Capone, L. de’Medici, and A. Georges, Solving the dynamical mean-field theory at very low temperatures using the Lanczos exact diagonalization, *Phys. Rev. B* **76**, 245116 (2007).
- [21] A. Liebsch and H. Ishida, Temperature and bath size in exact diagonalization dynamical mean field theory, *J. Phys.: Condens. Matter* **24**, 053201 (2012).
- [22] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, *arXiv:1412.6980*.
- [23] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, *Nature (London)* **521**, 436 (2015).
- [24] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, Attention is all you need, in *Advances in Neural Information Processing Systems*, edited by I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates Inc., 2017), Vol. 30.
- [25] V. Dunjko and H. J. Briegel, Machine learning and artificial intelligence in the quantum domain: A review of recent progress, *Rep. Prog. Phys.* **81**, 074001 (2018).
- [26] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences*, *Rev. Mod. Phys.* **91**, 045002 (2019).
- [27] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko *et al.*, Highly accurate protein structure prediction with AlphaFold, *Nature (London)* **596**, 583 (2021).
- [28] J. Kirkpatrick, B. McMorrow, D. H. Turban, A. L. Gaunt, J. S. Spencer, A. G. Matthews, A. Obika, L. Thiry, M. Fortunato, D. Pfau *et al.*, Pushing the frontiers of density functionals by solving the fractional electron problem, *Science* **374**, 1385 (2021).
- [29] R. Fournier, L. Wang, O. V. Yazyev, and Q. Wu, Artificial neural network approach to the analytic continuation problem, *Phys. Rev. Lett.* **124**, 056401 (2020).
- [30] 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).
- [31] 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).
- [32] J. Rogers, T.-H. Lee, S. Pakdel, W. Xu, V. Dobrosavljevic, Y.-X. Yao, O. Christiansen, and N. Lanatà, Bypassing the computational bottleneck of quantum-embedding theories for strong electron correlations with machine learning, *Phys. Rev. Res.* **3**, 013101 (2021).
- [33] L. Boehnke, H. Hafermann, M. Ferrero, F. Lechermann, and O. Parcollet, Orthogonal polynomial representation of imaginary-time Green’s functions, *Phys. Rev. B* **84**, 075145 (2011).
- [34] 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).
- [35] J. Kaye, K. Chen, and O. Parcollet, Discrete Lehmann representation of imaginary time Green’s functions, *Phys. Rev. B* **105**, 235115 (2022).
- [36] L. Arsenault, O. Von Lilienfeld, and A. Millis, Machine learning for many-body physics: Efficient solution of dynamical mean-field theory, *arXiv:1506.08858*.
- [37] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/bk5q-pfb2> for taking note of the different positional encoding techniques in the concatenation and addition methods of the Matsubara frequencies.
- [38] O. Parcollet, M. Ferrero, T. Ayril, H. Hafermann, I. Krivenko, L. Messio, and P. Seth, TRIQS: A toolbox for research on interacting quantum systems, *Comput. Phys. Commun.* **196**, 398 (2015).
- [39] A. Romero, D. Allan, B. Amadon, G. Antonius, T. Applencourt, L. Baguet, J. Bieder, F. Bottin, J. Bouchet, E. Bousquet, and Others, ABINIT: Overview and focus on selected capabilities, *J. Chem. Phys.* **152**, 124102 (2020).
- [40] E. Plekhanov, P. Hasnip, V. Sacksteder, M. Probert, S. J. Clark, K. Refson, and C. Weber, Many-body renormalization of forces in *f*-electron materials, *Phys. Rev. B* **98**, 075129 (2018).
- [41] 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).
- [42] H. Kim, D. Kim, and D. Kim, Minimal neural network to learn the metal-insulator transition in the dynamical mean-field theory, *New Physics: Sae Mulli* **72**, 487 (2022).

- [43] J. B. Rigo and A. K. Mitchell, Machine learning effective models for quantum systems, *Phys. Rev. B* **101**, 241105(R) (2020).
- [44] C. Mejuto-Zaera, L. Zepeda-Núñez, M. Lindsey, N. Tubman, B. Whaley, and L. Lin, Efficient hybridization fitting for dynamical mean-field theory via semi-definite relaxation, *Phys. Rev. B* **101**, 035143 (2020).
- [45] M. Granath and H. U. R. Strand, Distributional exact diagonalization formalism for quantum impurity models, *Phys. Rev. B* **86**, 115111 (2012).
- [46] C. Weber, A. Amaricci, M. Capone, and P. B. Littlewood, Augmented hybrid exact-diagonalization solver for dynamical mean field theory, *Phys. Rev. B* **86**, 115136 (2012).
- [47] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, Automatic differentiation in PyTorch, Openreview (2017), <https://openreview.net/forum?id=BJJsrnfCZ>.
- [48] N. Kavokine, H. Lu, T. Ayrál, M. Ferrero, N. Wentzell, and O. Parcollet, CTSEG: A segment picture quantum impurity solver based on TRIQS, *JOSS* **10**, 7425 (2025).
- [49] Z. Zhao and H. Lee, Example database (2023), https://figshare.com/articles/dataset/Example_database/23144474.
- [50] The code for SCALINN is available at <https://github.com/zelong-zhao/SCALINN>.