Sparse sampling approach to efficient ab initio calculations at finite temperature

Jia Li, Markus Wallerberger, Naoya Chikano, Chia-Nan Yeh, Emanuel Gull, and Hiroshi Shinaoka

Sparse sampling approach to efficient *ab initio* calculations at finite temperature

Jia Li,[1] Markus Wallerberger,[1] Naoya Chikano,[2] Chia-Nan Yeh,[1] Emanuel Gull,[1, 3] and Hiroshi Shinaoka[2]

[1] *Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA*
[2] *Department of Physics, Saitama University, Saitama 338-8570, Japan*
[3] *Center for Computational Quantum Physics, The Flatiron Institute, New York, New York, 10010, USA*

Efficient *ab initio* calculations of correlated materials at finite temperature require compact representations of the Green's functions both in imaginary time and Matsubara frequency. In this paper, we introduce a general procedure which generates sparse sampling points in time and frequency from compact orthogonal basis representations, such as Chebyshev polynomials and intermediate representation (IR) basis functions. These sampling points accurately resolve the information contained in the Green's function, and efficient transforms between different representations are formulated with minimal loss of information. As a demonstration, we apply the sparse sampling scheme to diagrammatic GW and GF2 calculations of a hydrogen chain, of noble gas atoms and of a silicon crystal.

I. INTRODUCTION

The finite temperature Green's function formalism [1–4] is a standard approach to study equilibrium properties of correlated quantum systems at finite temperature. Many-body theories, approximations, and methods based on this formalism are widely used in condensed matter physics, quantum chemistry, and material science. Applications to low-energy effective model Hamiltonians include lattice Monte Carlo [5], dynamical mean-field theory [6] with its cluster [7], multi-orbital extensions [8, 9], and diagrammatic extensions [10–12], and diagrammatic or continuous-time quantum Monte Carlo methods [13, 14]. In the context of *ab initio* calculations of correlated materials, examples include the GW method [15–23], the self-consistent second order approximation (GF2) [24–31], variants of the dynamical mean field theory [8, 32–36], and the self-energy embedding theory [37–43].

The fundamental object of these theories are numerically computed one- and two-particle Green's functions and derived quantities such as self-energies and vertex functions. These quantities are known on a finite Matsubara frequency or imaginary time grid.

In *ab initio* calculations, the difference between the scales of the bare Hamiltonian (which sets a frequency range), temperature (which dictates the frequency resolution), and the energy scales for competing quantum phenomena spans many orders of magnitude. As a consequence, a naive representation of the Green's function requires an imaginary-time grid too large to store in memory, and solving equations such as the Dyson equation to the required accuracy becomes prohibitively expensive. This issue becomes even more pronounced for two-particle response functions, which are generically a function of multiple time and orbital indices.

Compact representations of Green's functions are crucial to address this problem. Representations based on power meshes [44, 45], Legendre polynomials [27, 46], Chebyshev polynomials [47, 48], intermediate numerical representations (IR) [49–51], quadrature rules [52, 53], and spline interpolations [28] have been proposed, as well as high frequency tail expansions [54–56].

Green's function representations, in addition to being compact, also need to enable efficient calculations. The two main stages in most methods are the evaluation of self-energy diagrams (usually best done in imaginary time, as the interaction is instantaneous) and the solution of the Dyson equation (usually best done in Matsubara space, where the equation is diagonal in frequency). Some of the representations mentioned above are only compact in either time or frequency, and transforming between those domains is expensive or involves a loss of accuracy. Others, such as the orthogonal polynomial bases, can efficiently and accurately be transformed between coefficients and imaginary time sampling points, but frequency transforms result in a loss of compactness.

It is therefore natural to ask if there is a set of 'sparse' sampling points in both frequency and time such that, if the Green's function is evaluated on these points, one may reconstruct the continuous Green's functions in both time and frequency to high precision. This will then allow to perform diagram calculations in time, Dyson equation solutions in frequency, and transformations in between with minimal loss of accuracy.

This paper presents such a compact representation in both time and frequency by proposing a sparse sampling scheme for finite temperature Green's functions, following prior efforts using frequency interpolation [28, 48] and MP2 quadratures [53]. We illustrate our scheme at the example of Chebyshev [47] and IR basis functions [49]. Our method aims to accurately resolve all the information contained within finite temperature Green's functions in a compact set of sampling points, and enables efficient and accurate transforms between imaginary time and Matsubara frequency. The sparsity of the sampling points directly corresponds to the compactness of the basis representation, which leads to system-independent time and frequency grids with few control parameters.

The paper will proceed as follows. In Sec. II we will introduce and derive the sparse sampling method. Sec. III will discuss the application to low-order diagrammatic methods such as GW and GF2. Sec. IV will show appli-

cations to well-known systems in order to illustrate the power of the method.

II.   SPARSE SAMPLING METHOD

A.   General description and notation

The main object of this paper is the Green's function G^α, which we assume is expanded into a compact representation in terms of N basis functions, such that in imaginary time and Matsubara frequencies

$$G^\alpha(\tau) = \sum_{l=0}^{N-1} G_l^\alpha F_l^\alpha(\tau), \tag{1}$$

$$\hat{G}^\alpha(i\omega_n^\alpha) = \sum_{l=0}^{N-1} G_l^\alpha \hat{F}_l^\alpha(i\omega_n^\alpha) \tag{2}$$

$$\hat{F}_l^\alpha(i\omega_n^\alpha) = \int_0^\beta d\tau F_l^\alpha(\tau) e^{i\omega_n^\alpha \tau}, \tag{3}$$

where G_l^α are expansion coefficients, $F_l^\alpha(\tau)$ imaginary time basis functions with Fourier transform $\hat{F}_l^\alpha(i\omega_n^\alpha)$ (the 'hat' denoting frequency representations), $\omega_n^\alpha = \pi(2n + \delta_{\alpha,F})/\beta$ Matsubara frequencies, and α denotes the statistics (F for fermions and B for bosons).

In the Chebyshev representation [47]

$$F_l^\alpha(\tau) \equiv T_l[x(\tau)], \tag{4}$$

where $T_l(x)$ are Chebyshev polynomials of the first kind and $x(\tau) = 2\tau/\beta - 1$. In the IR basis [51]

$$F_l^\alpha(\tau) \equiv U_l^\alpha(\tau) \tag{5}$$

where $U_l^\alpha(\tau)$ depend on the statistics and a dimensionless parameter $\Lambda = \beta\omega_{\max}$ with a cutoff frequency $\omega_{\max}$. Appendix A summarizes notations for Chebyshev and IR as well as their practical differences.

To determine G_l^α from $G^\alpha(\tau)$, we choose a finite set of M sampling points $\bar{\tau}_0^\alpha, \ldots, \bar{\tau}_{M-1}^\alpha \in [0, \beta]$ ($M \geq N$). If $\{\bar{\tau}_k^\alpha\}$ are chosen such that the discretized basis vectors $\{F_0^\alpha(\bar{\tau}_k^\alpha)\}, \ldots, \{F_{N-1}^\alpha(\bar{\tau}_k^\alpha)\}$ are linearly independent, the exact values of G_l^α can be computed (transformed) from sampled values of $G^\alpha(\tau)$. Similarly, if a subset of Matsubara frequencies $\{i\bar{\omega}_k^\alpha\}$ is chosen such that the basis functions are linearly independent, G_l^α can be obtained from $\hat{G}^\alpha(i\bar{\omega}_k^\alpha)$.

If as many sampling points M are chosen as imaginary time points N, these transformations are

$$G_l^\alpha = \sum_{k=0}^{N-1} [\mathbf{F}_\alpha^{-1}]_{lk} G^\alpha(\bar{\tau}_k^\alpha) \tag{6}$$

$$= \sum_{k=0}^{N-1} [\hat{\mathbf{F}}_\alpha^{-1}]_{lk} \hat{G}^\alpha(i\bar{\omega}_k^\alpha), \tag{7}$$
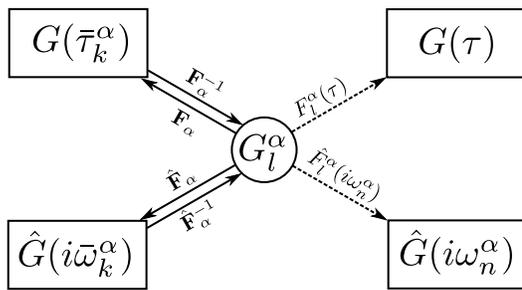
FIG. 1. Schematic illustration of relations between different representations. Solid lines denote transformations between the basis representation coefficients G_l^α (center) and Green's functions evaluated at imaginary time or frequency sampling points via transformation matrices. Dashed lines represent basis expansions of G_l^α to arbitrary imaginary time or frequency points.

where $\mathbf{F}_\alpha$ and $\hat{\mathbf{F}}_\alpha$ are $N \times N$ matrices:

$$[\mathbf{F}_\alpha]_{kl} = F_l^\alpha(\bar{\tau}_k^\alpha) \tag{8}$$

$$[\hat{\mathbf{F}}_\alpha]_{kl} = \hat{F}_l^\alpha(i\bar{\omega}_k^\alpha). \tag{9}$$

This procedure only requires evaluating the Green's function on N sampling points, and linear transforms between the time or frequency domain and the basis representation G_l^α become invertible. G_l^α can thus serve as a proxy to transform between imaginary time and frequency sampling points, as well as evaluation at arbitrary τ and $i\omega_n$ values, as illustrated in Fig. 1.

For $M > N$ (more sampling points than basis coefficients), the inverses in Eqs. (6) and (7) are replaced by the corresponding pseudoinverses $\mathbf{F}^+ \equiv (\mathbf{F}^\dagger \mathbf{F})^{-1} \mathbf{F}^\dagger$, and the exact transform is replaced by a least squares fitting procedure.

In practical calculations, different choices of basis functions and sampling points lead to differently conditioned equation systems. A naive choice of sampling points, such as uniformly distributed time or frequency grids, results in almost linearly dependent basis vectors and ill-conditioned transforms, which improve very slowly when additional grid points are added. For an efficient method, a minimal set of sampling points that generates well-conditioned transformation matrices is desired in order to minimize the number of function evaluations and the loss of accuracy during transforms.

In the remainder of this section, we show that such a set with $M = N$ can be generated according to the distribution of the roots of the basis functions.

B.   Imaginary time sampling

Chebyshev.– For a truncated Chebyshev basis of size N, the N sampling points in τ are naturally given by the roots of the $(N+1)$-th basis function $T_N(\tau)$ as

$$\bar{\tau}_k^\alpha \equiv \tau \left(\cos \left(\pi \frac{2k+1}{2N} \right) \right), \tag{10}$$

for $k = 0, \ldots, N - 1$ with the mapping $\tau(x) = \beta(x + 1)/2$. These sampling points lead to very well conditioned transformation matrices due to the discrete orthogonality of Chebyshev polynomials, and the condition number of $\mathbf{F}_\alpha$ (defined as $\|\mathbf{F}_\alpha\|_2\|\mathbf{F}_\alpha^{-1}\|_2$) takes the constant value of $\sqrt{2}$.

IR basis.– For the IR basis with N basis functions and given β and $\omega_{\max}$, we choose the sampling points $\{\bar{\tau}_k^\alpha\}$ to be the midpoints of the grid composed of the $N - 1$ roots of the highest order basis function $U_{N-1}^\alpha(\tau)$ and the boundary points 0 and β. We choose not to use the roots of the next basis function $U_N^\alpha(\tau)$ like in the Chebyshev case due to the fact that the IR basis is a numerical basis, and thus it is more convenient to determine sampling points from the available basis functions.

C. Matsubara frequency sampling

We generate sampling points in Matsubara frequencies following an algorithm similar to the imaginary time cases, with two additional considerations. First, function values should only be evaluated on the discrete Matsubara frequencies. Second, fermionic and bosonic Matsubara frequencies have to be treated separately, and the zero bosonic frequency (which represents static physics) has to be considered explicitly.

Chebyshev.– In the Chebyshev representation, we follow the same heuristics as in the τ sampling by finding or approximating zeros of the next basis function $\hat{T}_N^\alpha(i\omega_n^\alpha)$ defined in Matsubara frequency.

For fermions and even N, when continued to continuous Matsubara frequency space, $\hat{T}_N^F(i\omega_n^\alpha)$ has N roots on the imaginary axis $(-i\infty, i\infty)$. We take N Matsubara frequencies closest to these roots as sampling points.

For bosons and odd N, $\hat{T}_N^B(i\omega_n^\alpha)$ has $N - 1$ roots. We define $N - 1$ sampling points as the Matsubara frequencies closest to the roots. We take the zero bosonic frequency $i\omega_n^B = 0$ as the last sampling point. The zero bosonic frequency, which corresponds to a constant offset in τ and often has to be treated separately, serves as a natural complement.

The requirement that even N should be used for fermions and odd N for bosons is necessary because the other cases (odd N for fermions or even N for bosons) will not yield adequate number of sampling points due to the analytic structure of $\hat{T}_N^\alpha(i\omega_n^\alpha)$. See Appendix B for a discussion.

IR basis.– For the IR basis, the procedure for getting frequency sampling points is more empirical due to the numerical nature of the basis function. We partition all Matsubara frequencies into contiguous groups such that the highest order basis function $\hat{U}_{l_{\max}}^\alpha(i\omega_n^\alpha)$ has the same sign within each group ($l_{\max} \equiv N - 1$). $\hat{U}_{l_{\max}}^\alpha(i\omega_n^\alpha)$ is either purely real (for even $l_{\max}$) or purely imaginary (for odd $l_{\max}$). We therefore use the sign of the corresponding real or imaginary part as the sign of $\hat{U}_{l_{\max}}^\alpha(i\omega_n^\alpha)$. The
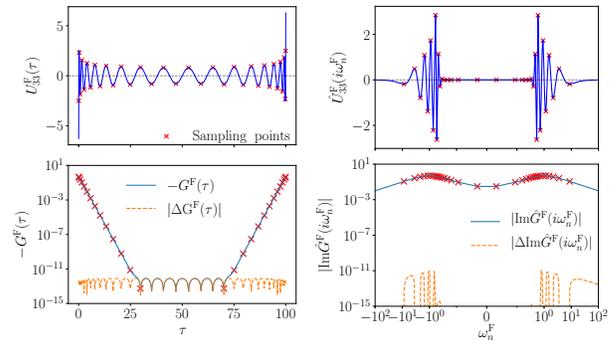
FIG. 2. Distribution of sampling points and results transformed from imaginary time (left panels) and Matsubara frequency (right panels) for the IR basis by sparse sampling. We consider a model of a semicircular density of states of half width 1 at $\beta = 100$ defined in Eq. (11). We take $\omega_{\max} = 1$ for the IR basis. The sampling points are denoted by crosses. Top row: The basis functions used to generate sampling points ($l = 33$). Bottom row: Comparison of the reconstructed Green's function to exact results.

sampling points $i\bar{\omega}_k^\alpha$ are chosen to be those that maximize $|\hat{U}_{l_{\max}}^\alpha(i\omega_n^\alpha)|$ in each group.

By checking the resulting sampling points numerically, we conclude that by requiring N to be even ($l_{\max}$ odd) for fermionic basis and odd ($l_{\max}$ even) for bosonic basis, the number of sampling points is exactly N, and the bosonic sampling points naturally include zero.

D. Numerical demonstration

The transformation defined in Eqs. (6) and (7) is exact if the Green's function is a linear combination of a finite set of basis functions (1). With physical Green's functions, this is seldom the case, and any finite expansion incurs a truncation error. Fortunately, in both the IR and the Chebyshev expansion, the truncation error is controlled: the analyticity of the finite-temperature Green's function in $(0, \beta)$ guarantees exponential convergence of the Chebyshev expansion, and the construction of the IR basis from analytic continuation guarantees the same thing for the IR expansion [49].

To demonstrate the behavior of the sparse sampling scheme when applied to physical Green's functions, we consider a model with semicircular density of states (full bandwidth is 2) for the IR basis in Fig. 2:

$$\rho(\omega) = \frac{2}{\pi}\sqrt{1 - \omega^2}. \qquad (11)$$

As an example, the top left panel shows U_{34}^F as a function of imaginary time (blue lines) and illustrates the $N = 34$ sampling points for the fermionic basis of $\beta = 100$ and $\omega_{\max} = 1$ (red crosses). The sampling points cluster near $\tau = 0$ and β, where this basis function is rapidly oscillating.

The top right panel of Fig. 2 shows the distribution of the Matsubara frequency sampling points generated for the same basis for $N = 34$. The Fourier transformed basis function $\hat{U}_{33}^{\mathrm{F}}(i\omega_n)$ (blue lines) exhibits $N - 1 = 33$ sign changes, which define $N = 34$ sampling points (red crosses). The sampling points are distributed almost logarithmically, which allows us to capture all the features of $\hat{U}_{l_{\max}}^{\mathrm{F}}(i\omega_n)$ from low to high frequency.

In the bottom row of Fig. 2, the left and right panels illustrate the sampling of $G(\tau)$ and $\hat{G}(i\omega_n)$, respectively. The sampling points capture relevant features of the Green's function in both cases. We compare the interpolated and extrapolated results with the numerically exact values. For imaginary time, one can see agreement at the level of $\sim 10^{-12}$ in the whole interval of $[0, \beta]$, which matches the singular value cutoff we used (10^{-12}). For Matsubara frequency, the coefficients obtained by the sparse sampling not only interpolate $\hat{G}(i\omega_n)$ but also extrapolate it precisely beyond the highest sampling frequency.

E. Technical details

In practical applications, it is advisable to precompute the sampling points and transformation matrices for the basis functions employed, to avoid unnecessary evaluations of $\hat{T}_l^\alpha(i\omega_n^\alpha)$ and $\hat{U}_l^\alpha(i\omega_n^\alpha)$.

The `irbasis` library [51] provides numerical data of the IR basis functions in the dimensionless form for selected values of Λ from $\Lambda = 10$ up to $\Lambda = 10^7$. The numerical evaluation of the basis functions in Matsubara frequency is also implemented.

The procedures we have presented are not unique, and we do not claim that they are optimal definitions of sampling points. One may design other choices with similar or, potentially, better numerical performance. The number of sampling points may also exceed the basis size N, as long as inversions in Eqs. (6) and (7) are replaced with a pseudoinverse, and the resulting transformations are well-conditioned. Nevertheless, the algorithms introduced in this section provide a systematic and unambiguous way to obtain the minimum sets of sampling points which yield well-conditioned numerical transforms and high accuracy. Appendix B includes numerical investigations of the transformation matrices.

III. SPARSE SAMPLING APPROACH TO SOLVING DIAGRAMMATIC EQUATIONS

Self-consistent second order Green's function theory (GF2) is a second order perturbation theory which renormalizes the Green's function [24–31]. Self-consistent GW [15, 17, 21–23] further renormalizes the interaction.

GF2 and GW calculations involve evaluations of diagrams in τ and solving convolution equations in Matsubara frequencies. Quantities such as the Green's function

and the screened interaction are repeatedly converted back and forth between time and frequency. These calculations therefore serve as a good test platform to demonstrate the numerical stability and computational advantage of the sparse sampling scheme.

In *ab initio* calculations of molecules, a finite spatial basis set is chosen for each atom. Since the energy scale $\omega_{\max}$ in IR basis is only well defined for Green's functions under orthogonal spatial orbitals, for better estimations of $\omega_{\max}$ or Λ we perform canonical orthogonalizations [57] on the atomic orbitals for a chosen basis set in each system, and use the resulting orthogonal spatial basis in all calculations. The Hamiltonian reads

$$H = \sum_{ij\sigma} h_{ij} c_{i\sigma}^\dagger c_{j\sigma} + \frac{1}{2} \sum_{ijkl} \sum_{\sigma\sigma'} V_{ijkl} c_{i\sigma}^\dagger c_{k\sigma'}^\dagger c_{l\sigma'} c_{j\sigma} \quad (12)$$

where h is the single-particle Hamiltonian, and V is the Coulomb integrals defined on the orthogonalized orbitals.

The self-consistency in GF2 and GW is defined by the Dyson equation

$$\hat{G}(i\omega_n^{\mathrm{F}}) = [(i\omega_n^{\mathrm{F}} + \mu)I - F - \hat{\tilde{\Sigma}}(i\omega_n^{\mathrm{F}})]^{-1}. \quad (13)$$

The Fock matrix $F = h + \Sigma^{\mathrm{HF}}$ includes the frequency independent Hartree-Fock contribution

$$\Sigma_{ij}^{\mathrm{HF}} = (2V_{ijkl} - V_{ilkj})\rho_{kl}, \quad (14)$$

where i, j, k, l are orbital indices. GF2 approximates the frequency-dependent self-energy $\tilde{\Sigma}$ as

$$\begin{aligned} \tilde{\Sigma}_{ij}(\tau) = &-G_{kl}(\tau)G_{qm}(\tau)G_{np}(-\tau)\times \\ &\times V_{ikpq}(2V_{ljmn} - V_{mjln}), \end{aligned} \quad (15)$$

while GW approximates $\tilde{\Sigma}$ as

$$\tilde{\Sigma}_{ij}(\tau) = -G_{lk}(\tau)\tilde{W}_{ilkj}(\tau), \quad (16)$$

where $\tilde{W} = W - V$ is the frequency-dependent part of the screened interaction W. W is calculated by the random phase approximation (RPA) [58] as

$$\hat{W}_{ijkl}(i\omega_n^{\mathrm{B}}) = V_{ijkl} + V_{ijpq}\hat{P}_{qpsr}(i\omega_n^{\mathrm{B}})\hat{W}_{rskl}(i\omega_n^{\mathrm{B}}), \quad (17)$$

where the bare polarization is given by

$$P_{ijkl}(\tau) = -G_{il}(\tau)G_{jk}(-\tau). \quad (18)$$

At self-consistency, physical properties are evaluated from G and Σ. For example, density matrix ρ is given by $\rho_{ij} = G_{ji}(0^-) = -G_{ji}(\beta)$, and the total electronic energy is

$$\begin{aligned} E = &\operatorname{Tr}[\rho H_0] + \frac{1}{2}\operatorname{Tr}[\rho\Sigma_{\mathrm{HF}}] \\ &+ \frac{1}{2\beta}\sum_n \operatorname{Tr}[\hat{\tilde{\Sigma}}(i\omega_n^{\mathrm{F}})\hat{G}(i\omega_n^{\mathrm{F}})]. \end{aligned} \quad (19)$$
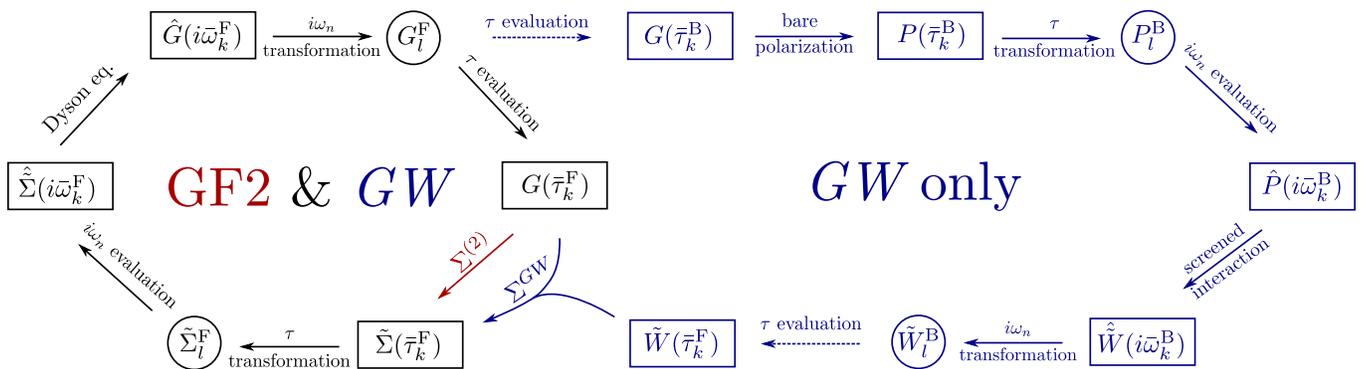
FIG. 3. Illustration of GF2 and GW procedures using the sparse sampling scheme. The red and blue lines denote GF2-only and GW-only steps, respectively. Dashed arrows indicate evaluations that change the statistics of the representation.

Fig. 3 illustrates how the self-consistent calculations can be performed by sparse sampling. In a GF2 calculation, we first evaluate Green's function at sampling points $\bar{\tau}_k^{\mathrm{F}}$, and construct the self-energy $\tilde{\Sigma}(\bar{\tau}_k^{\mathrm{F}})$ following second-order approximation (15). The self-energy is then transformed to the basis representation $\tilde{\Sigma}_l^{\mathrm{F}}$ following Eq. (6), which is then evaluated on the frequency sampling points to get $\hat{\tilde{\Sigma}}(i\bar{\omega}_k^{\mathrm{F}})$. The Dyson equation (13) is solved for each $i\bar{\omega}_k^{\mathrm{F}}$ to obtain $\hat{G}(i\bar{\omega}_k^{\mathrm{F}})$. We then transform the Green's function to its basis representation G_l^{F} following Eq. (7). The updated Green's function in τ is recovered by evaluating G_l^{F} on sampling points $\bar{\tau}_k^{\mathrm{F}}$. The procedure is repeated using the updated Green's function until self-consistency, which corresponds to the inner loop in Fig. 3. We can see that the compact basis representations G_l^{F} and $\tilde{\Sigma}_l^{\mathrm{F}}$ serve as proxies to convert back and forth between τ and frequency domains, all evaluated on corresponding sampling points.

While all quantities involved in GF2 are fermionic, in GW we have to switch between fermionic quantities (G and Σ) and bosonic quantities (P and W). This is achieved again by using compact basis representations as proxies: when calculating the polarization P, we evaluate G_l^{F} on the *bosonic* sampling points $\bar{\tau}_k^{\mathrm{B}}$ and assemble $P(\bar{\tau}_k^{\mathrm{B}})$ following Eq. (18). We then carry out the calculation of $\tilde{W}$ on the frequency sampling points $i\bar{\omega}_k^{\mathrm{B}}$, and obtain the compact basis representation $\tilde{W}_l^{\mathrm{B}}$. Finally we evaluate $\tilde{W}_l^{\mathrm{B}}$ back on the *fermionic* sampling points $\bar{\tau}_k^{\mathrm{F}}$, and compute self-energy using the GW approximation (16). The rest of the procedure is identical to GF2.

Collective physical observables such as the total energy and the density matrix can be evaluated accurately from the compact basis representation of G and Σ. Efficient evaluation of the density is beneficial in calculations with a fixed number of electrons, where one has to adjust the chemical potential in each self-consistent iteration [25]. Appendix C discusses these technical aspects in detail.

Note that since most basis functions, including Chebyshev and IR, cannot capture constant shifts in frequency (which correspond to a delta function at $\tau = 0$), it is im-

portant that one only expands the frequency-dependent components such as $\tilde{\Sigma}$ and $\tilde{W}$ using compact basis representations. One also needs to be careful when zero-energy mode exists in a bosonic quantity, in which case the IR basis function cannot describe the constant shift in imaginary time [50].

IV. RESULTS

A. Hydrogen chain

We first apply our sparse sampling scheme to GF2 and GW caculations of a system composed of 10 hydrogen atoms placed on a straight line with equal spacing r. The hydrogen chain, due to its simplicity, serves as a benchmark platform for testing numerical methods of correlated electrons. Reference data for the hydrogen chain were carefully compared and analyzed in Ref. 59 with many methods including GF2 and GW. It is therefore convenient to use this system to analyze the sparse sampling scheme.

We perform GF2 and GW calculations for H_{10} with $r = 1\,a_0$ and $\beta = 1000\,E_{\mathrm{h}}^{-1}$ ($T \sim 315.8\,\mathrm{K}$). We use the minimal basis set STO-6g with only one $1s$ orbital per atom. Hartree-Fock calculations show that the difference between the highest and the lowest Hartree-Fock energy levels is about $\Delta E \approx 5.76\,E_{\mathrm{h}}$ ($\sim 156\,\mathrm{eV}$). The dimensionless parameter for the IR basis can thus be estimated by taking Λ to bound the value $\beta\Delta E \approx 5.76 \times 10^3$. We take $\Lambda = 10^5$ in all our calculations.

Both Chebyshev and IR basis functions are used together with the sparse sampling scheme. To demonstrate convergence, we examine a series of calculations with different sizes of basis functions. Typically we choose N to be an even number, which is then used as the size of the fermionic basis. For corresponding bosonic basis functions in GW, we used the closest odd number $N - 1$ as the basis size. The Python library `irbasis` version 2.0.0b1 [51] is used for calculating IR basis functions.

With each basis of fixed size N, we perform self-

consistent GF2 and GW calculations following the procedures described in the previous section. The initial guess of the Green's function is constructed from Hartree-Fock calculations using the PySCF library [60]. Self-consistent calculations, as illustrated in Fig. 3, are then executed repeatedly until the energy difference is below the convergence tolerance $E_{\text{tol}} = 10^{-8} E_{\text{h}}$ between two consecutive iterations. Energy values converged with respect to basis size N are cross-checked with reference data to ensure correctness. All results are then compared to converged values to assess errors in energy and density matrix as a function of basis size N.

In the left panel of Fig. 4, we show the convergence of errors in total energy with respect to N. GF2 and GW share similar convergence behavior, in that the error decreases almost exponentially with respect to N in either basis representation. In the Chebyshev representation, we obtain convergence of the total energy below the tolerance E_{tol} with around 350 Chebyshev polynomials. With IR basis, convergence is reached with less than 100 basis functions. Similarly, the right panel of Fig. 4 illustrates the convergence of the maximum error in density matrix, exhibiting an exponential decay of errors. This indicates that with a reasonable number of sampling points, we can reach very high precision in both global observables such as the energy and local properties like the density matrix. The observation that GW shows a convergence behavior similar to GF2 indicates that no substantial additional errors are introduced during the frequent switching between fermionic and bosonic representations.

The sparse sampling scheme is stable thanks to the well-conditioned transformation matrices generated from the sampling points. We demonstrate this in Fig. 5, which shows the relative magnitude of basis expansion coefficients for the converged solution. N is chosen large enough to ensure that all quantities are well approximated by the basis representations, with $N = 600$ for Chebyshev and $N = 130$ for IR with $\Lambda = 10^5$. Even after several iterations with multiple transforms forward and backward between different types of sampling points, we see that for all quantities, expansion coefficients decay at least exponentially as we expect from the properties of the basis, down to a relative size below 10^{-12}. The truncation error due to the finite basis expansion is therefore controlled, and no amplification of error is observed during the self-consistent iterations.

The sparse sampling scheme ensures that the number of τ grid points and the number of Matsubara frequencies is the same as the basis size N. We reach a precision of 8 digits in total energy with only hundreds of τ points and Matsubara frequencies, a significant improvement from the conventional approach used in Ref. 59, where $\sim 10^4$ Matsubara frequencies were used for higher temperature ($\beta \sim 100 E_{\text{h}}^{-1}$) and bigger convergence threshold ($E_{\text{tol}} \sim 10^{-6} E_{\text{h}}$ or $10^{-7} E_{\text{h}}$). This greatly reduces the computational cost and memory requirement in all parts of the calculations while still being accurate. The sparse sam-
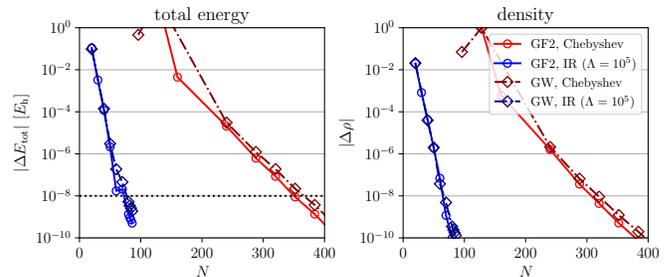
FIG. 4. Error in total energy and density from converged GF2 and GW calculations for H_{10} in minimal basis at $\beta = 10^3 E_{\text{h}}^{-1}$. The left panel shows the convergence of the total energy with the dashed horizontal line representing the convergence threshold of $10^{-8} E_{\text{h}}$. The right panel shows the convergence of the density.
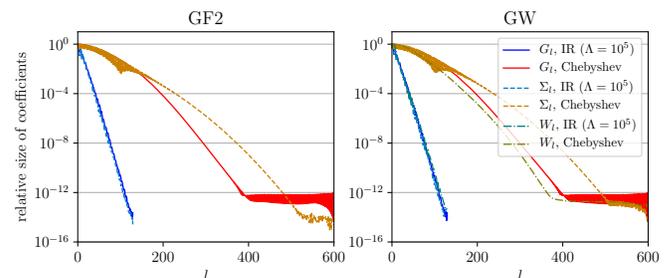
FIG. 5. Relative size of basis expansion coefficients with converged GF2 (left panel) and GW (right panel) calculations for H_{10} in STO-6g at $r = 1.0$ a.u., $\beta = 1000 E_{\text{h}}^{-1}$.

pling scheme thus allows to tackle problems that were too costly to calculate, especially low temperature calculations of systems with large energy scales.

B. Noble gas atoms

Noble gas atoms such as Kr have deep core states, if no pseudopotentials are employed. Due to the large energy scale caused by the core states, it is computationally very demanding for conventional methods to resolve sharp features close to $\tau = 0$ or β in $G(\tau)$ or the slow decay of $\hat{G}(i\omega_n)$ at high frequency. Even with a compact polynomial basis such as Chebyshev, thousands of basis functions are required to represent the Green's functions [47], and effective core potentials (ECP), which absorb electron in inner orbitals to the ionic potential, have to be employed in most practical calculations. We choose this problem to demonstrate the power of the sparse sampling method when dealing with large energy scales, while avoiding additional physical or technical difficulties.

The IR basis is a natural choice for systems with large energy scales. As long as the spectral cutoff ω_{max} is chosen to include all energy scales in the system, exponential convergence of the coefficients is guaranteed by construction, usually with no more than a couple hundred basis

functions [49]. The full potential of IR basis is realized when combined with the sparse sampling scheme developed in this paper. Numerical difficulties in either τ or frequency domain are reduced to a single issue: whether the basis functions can capture all relevant quantities well. Therefore, by using a sparse sampling method with the IR basis, we are able to treat noble gas atoms efficiently even in all-electron calculations.

We choose the all-electron correlation consistent basis set cc-pVDZ [61], and perform GF2 and GW calculations of four noble gas atoms: He, Ne, Ar, and Kr, at $\beta = 1000\, E_{\mathrm{h}}^{-1}$. Similar to the case for H_{10}, we estimate the dimensionless parameter Λ according to the Hartree-Fock energy spectrum for each individual atom: $\Lambda = 10^4$ for He, $\Lambda = 10^5$ for Ne, and $\Lambda = 10^6$ for Ar and Kr. The sparse sampling scheme is used in all calculations, and we vary the basis size N to explore the convergence behavior. The energy convergence threshold is set to $E_{\mathrm{tol}} = 10^{-8}\, E_{\mathrm{h}}$, which is much smaller than the energy scale of Kr ($\sim 10^3\, E_{\mathrm{h}}$).

Fig. 6 shows the energy convergence of all four atoms with respect to basis size N with GF2 (left column) and GW (right column). The upper panels indicate that the basis converges for all atoms, with absolute difference in energy dropping in an exponential trend below the convergence tolerance E_{tol}. The lower panels put the convergence in a relative scale, where all atoms in both GF2 and GW can reach $\sim 10^{-10}$ relative convergence.

Our results show that we can obtain fast basis convergence with around 100 basis functions for all systems studied. This is consistent with the property of IR basis that the basis size N scales only logarithmically with Λ [50].

C. Application to solids

One can apply the sparse sampling technique to periodic systems the same way as in molecules. Additional momentum dependence can be viewed as an extra "orbital index". As an example, we apply our method to the GW calculation of a silicon crystal, using a $4 \times 4 \times 4$ momentum mesh, inverse temperature $\beta = 1000\, E_{\mathrm{h}}^{-1}$, and energy self-consistency threshold $E_{\mathrm{tol}} = 10^{-6}\, E_{\mathrm{h}}$. We choose the GTH-DZVP basis of Gaussian orbitals [62], which has 13 orbitals per Si atom. The initial Green's function is constructed from the LDA solution in this basis set with a GTH-LDA pseudopotential. We orthogonalize the spatial basis set in all GW calculations. The LDA spectrum shows a maximum energy scale of $\Delta E \approx 3\, E_{\mathrm{h}}$, thus we choose $\Lambda = 10^4$ and 10^5 for IR basis to bound the value $\beta \Delta E \approx 3 \times 10^3$.

Table I shows the total energy convergence with respect to the size of basis representations N. With both Chebyshev and IR, the total energy per unit cell converges below the self-consistency threshold of $10^{-6}\, E_{\mathrm{h}}$ to the same value $E_{\mathrm{tot}} = -7.880\,430\, E_{\mathrm{h}}$ as N increases. The Chebyshev basis converges at around $N \approx 300$, and the
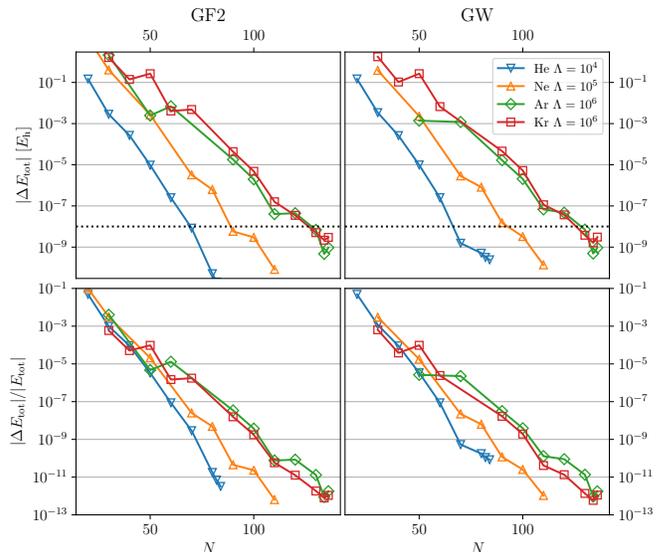
FIG. 6. Total energy convergence in GF2 and GW calculations of noble gas atoms with respect to the number of IR basis functions N. Left and right columns show GF2 and GW results respectively. Top row: absolute differences in total energy to the converged value. Dashed black horizontal line illustrates the energy convergence threshold of $10^{-8}\, E_{\mathrm{h}}$. Bottom row: relative differences in total energy.

TABLE I. GW total energy of the silicon crystal per unit cell at $\beta = 1000\, E_{\mathrm{h}}^{-1}$, with self-consistency convergence threshold $E_{\mathrm{tol}} = 10^{-6}\, E_{\mathrm{h}}$. Note that the energy values reflect the pseudopotential approximation in our calculations and are only used to assess basis convergence.

Chebyshev		IR $\Lambda = 10^4$		IR $\Lambda = 10^5$	
N	$E_{\mathrm{tot}}\ [E_{\mathrm{h}}]$	N	$E_{\mathrm{tot}}\ [E_{\mathrm{h}}]$	N	$E_{\mathrm{tot}}\ [E_{\mathrm{h}}]$
100	-8.0270874	80	-7.8804300	110	-7.8804300
150	-7.8861508	82	-7.8804300	112	-7.8804300
200	-7.8806642	84	-7.8804300		
250	-7.8804379	86	-7.8804300		
300	-7.8804302				
350	-7.8804298				

IR basis at $N \approx 80$ for $\Lambda = 10^4$, $N \approx 110$ for $\Lambda = 10^5$. In contrast, conventional approaches using uniform Matsubara frequency grid and high frequency tail expansions require $\sim 10^5$ frequency points.

The sparse sampling method thus allows us to obtain numerically precise values for physical quantities with much less computational cost. Figure 7 shows the momentum-resolved spectral functions calculated from the converged GW Green's function via analytical continuation to the real frequency axis using the maximum entropy (MaxEnt) method [63–65]. The indirect band gap of the silicon crystal between Γ and X points is
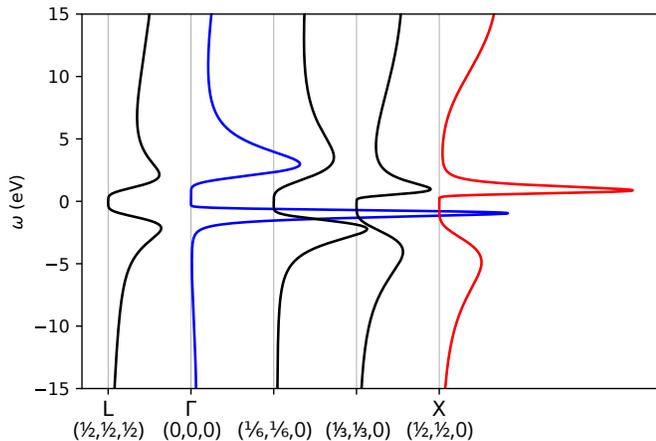
FIG. 7. Momentum resolved spectral function of silicon at $\beta = 1000\, E_{\mathrm{h}}^{-1}$, calculated from converged GW Green's functions. Blue and red lines highlight spectra at Γ and X points, which are the locations of the indirect band gap.

well captured, with an estimated value of 1.84 eV. The band gap from our fully self-consistent GW shows decent agreement with previous GW results obtained with quasi-particle approximations [21, 23].

V. CONCLUSIONS

We have introduced a sparse sampling scheme for imaginary-time Green's functions which greatly reduces the size of imaginary time and frequency grids in realistic calculations. We have developed general procedures to generate imaginary time and frequency sampling points from basis representations such as Chebyshev and IR, with efficient transformations between time and frequency domain. We have applied the procedure to low-order diagrammatic methods and demonstrated the numerical performance in realistic molecules and solids. With no more than a few hundred sampling points, the sparse sampling scheme can accurately capture physics in realistic systems with large energy scales and at low temperatures.

The sparse sampling reveals the full power of compact basis representations of the Green's function as the compactness is maintained in all stages of a calculation. The number of sampling points necessary for the required precision scales slowly as temperature is lowered and energy scales are increased. Our work therefore allows finite-temperature many-body methods to access physics of correlated electrons at larger energy scale and lower temperature.

Potential applications of the sparse sampling scheme include state-of-the-art methods for strongly correlated materials such as GW+DMFT [32–36] and the self-energy embedding theory (SEET) [41]. Quantitative estimates of superconducting temperature T_{c} based on the Migdal-Eliashberg theory in *ab initio* calculations may

benefit from the sparse sampling scheme so that the constant density of states approximation is no longer necessary [66].

In addition, two-particle quantities play a key role in many diagrammatic calculations such as computing lattice susceptibility in DMFT [67] as well as diagrammatic extensions of DMFT [10, 68]. Two-particle quantities are difficult to handle even in effective model calculations due to the multiple indices for frequencies and spin/orbital degrees of freedom. The IR basis has recently been extended to two-particle quantities [69]. The application of the sparse sampling scheme to two-particle quantities is an interesting topic for future research.

Note added. After the submission of our manuscript, the authors became aware of another independent development of similar techniques. We direct the reader to Ref. 70 in case of interest.

Appendix A: Compact orthogonal representation of Green's function

1. Chebyshev basis

The Chebyshev polynomials of the first kind $T_l(x)$ form an orthogonal system in the interval $[-1, 1]$, which can be mapped in to the interval $[0, \beta]$ via

$$x(\tau) = \frac{2\tau}{\beta} - 1, \quad \tau(x) = \frac{\beta(x+1)}{2} \qquad \text{(A1)}$$

such that $F_l^{\alpha}(\tau) = T_l[x(\tau)]$. We use the notation $T_l(\tau)$ to represent the order l Chebyshev polynomial mapped onto the interval $[0, \beta]$.

Approximating an analytical function with the first N Chebyshev polynomials is convenient due to the discrete orthogonality on the roots of the $(N+1)$-th Chebyshev polynomial

$$\frac{1}{N} \sum_{k=0}^{N-1} T_i(x_k) T_j(x_k) = \frac{1 + \delta_{i,0}}{2} \delta_{ij} \qquad \text{(A2)}$$

where x_k are the roots of $T_N(x)$. The Chebyshev coefficients are therefore well approximated by Clenshaw–Curtis quadrature:

$$G_l^\alpha = \frac{2}{N(1+\delta_{0,l})} \sum_{k=1}^{N-1} G^\alpha(\tau_k) T_l(\tau_k) + \mathcal{O}(2^{-N}) \quad \text{(A3)}$$

where $\tau_k = \tau(x_k)$ defined in (A1).

With the Chebyshev coefficients G_l^α, one can perform fast interpolation of $G^\alpha(\tau)$ at any $\tau \in [0, \beta]$ using recursion relations. Fourier transforms of the basis function $\hat{T}_l^\alpha(i\omega_n^\alpha)$ can also be computed, see Ref. 47.

2. IR basis

The IR basis introduced in Refs. 49 is designed to better capture properties of Green's functions in physical systems rather than arbitrary analytic functions. The IR basis has been applied to numerical analytic continuation [71] and DMFT calculations [72]. This section provides a brief description of the IR basis following the notation used in Ref. 51. The IR basis originates from the Lehmann representation of the single-particle Green's function

$$G^\alpha(\tau) = -\int_{-\omega_{\max}}^{\omega_{\max}} d\omega K^\alpha(\tau, \omega) \rho^\alpha(\omega), \quad \text{(A4)}$$

where the spectrum $\rho^\alpha(\omega)$ is bounded in the interval $[-\omega_{\max}, \omega_{\max}]$ ($\omega_{\max}$ is a cutoff frequency). The kernel $K^\alpha(\tau, \omega)$ reads

$$K^\alpha(\tau, \omega) \equiv \omega^{\delta_{\alpha,B}} \frac{e^{-\tau\omega}}{1 \pm e^{-\beta\omega}} \quad \text{(A5)}$$

for $\tau \in [0, \beta]$, where the $+$ and $-$ signs are used for fermions and bosons, respectively. The extra ω factor for bosons in Eq. (A5) is introduced in order to avoid a singularity of the kernel at $\omega = 0$.

For a fixed value of β and $\omega_{\max}$, the IR basis functions are defined through the singular value decomposition (SVD)

$$K^\alpha(\tau, \omega) = \sum_{l=0}^{\infty} S_l^\alpha U_l^\alpha(\tau) V_l^\alpha(\omega) \quad \text{(A6)}$$

where one observes an exponential decay of the singular values S_l^α (> 0) with increasing l. $U_l(\tau)$ and $V_l(\omega)$ form an orthonormal system for $\tau \in [0, \beta]$ and $y \in [-\omega_{\max}, \omega_{\max}]$, respectively.

A Green's function can be expanded as

$$G^\alpha(\tau) = \sum_{l=0}^{\infty} G_l^\alpha U_l^\alpha(\tau), \quad \text{(A7)}$$

$$G_l^\alpha = -S_l^\alpha \rho_l^\alpha, \quad \text{(A8)}$$

where

$$\rho_l^\alpha \equiv \int_{-\omega_{\max}}^{\omega_{\max}} d\omega \rho^\alpha(\omega) V_l^\alpha(\omega). \quad \text{(A9)}$$

If $|\rho_l^\alpha|$ does not grow, the exponential decay of S_l^α ensures exponential convergence of G_l^α. The accuracy of the expansion can be controlled by applying a cut-off on the singular values.

In calculations of realistic systems, one can set $\omega_{\max}$ large enough to capture the expected spectral width. The basis functions change their shapes through the change of the dimensionless quantity $\Lambda = \beta\omega_{\max}$ as temperature is lowered. This leads to a logarithmic growth of the basis size with respect to β.

The IR basis $\hat{U}_l^\alpha(i\omega_n^\alpha)$ does not compactly describe a constant shift in Matsubara frequency which corresponds to an unbounded spectrum. Thus, any constant term must be subtracted beforehand particularly when expanding the self-energy. Also, $U_l^B(\tau)$ does not describe a constant shift in imaginary time, corresponding to a zero-energy mode. Such terms must be treated separately as well [50].

The dimensionless form of the IR basis is defined as

$$U_l^\alpha(\tau) = \sqrt{\frac{2}{\beta}} u_l^\alpha(x(\tau)), \quad \text{(A10)}$$

$$\hat{U}_l^\alpha(i\omega_n^\alpha) = \sqrt{\beta} u_{ln}^\alpha$$

$$= \sqrt{\frac{\beta}{2}} \int_{-1}^{1} dx e^{i\pi\{n+(1/2)\delta_{\alpha,F}\}(x+1)} u_l^\alpha(x), \quad \text{(A11)}$$

where $u_l^\alpha(x)$ form an orthonormalized basis for $x \in [-1, 1]$.

3. Practice differences of Chebychev and IR bases

In practice, the choice of the basis representation (Chebyshev or IR) is largely application-dependent. Compared to the numerical evaluated IR basis functions with an additional control parameter Λ, the Chebyshev polynomials are easy to compute, and have only one precision-control parameter, i.e. the number of polynomials N, due to the exponential convergence guaranteed by the approximation theory [47]. However, the number of the Chebyshev polynomials required to approximate an imaginary-time Green's function to desired precision increases as $\mathcal{O}(\beta^{1/2})$ at low T, faster than the logarithmic scaling of the IR basis [47, 50]. Therefore, for systems with large energy scales or at low temperature, with a reasonable choice of Λ the IR basis provides more compact representation of the Green's function, while at smaller energy scale or higher temperature the Chebyshev may be favorable due to its simplicity.

Appendix B: Sampling points and condition numbers

1. Matsubara sampling points for Chebyshev

In the Chebyshev representation, we follow the same heuristics as in the τ sampling by finding or approximating zeros of the next basis function $\hat{T}_N^\alpha(i\omega_n^\alpha)$. The following properties can be shown using the recursion relation developed in Ref. 47.

- $\hat{T}_l^\alpha(i\omega_n^\alpha)$ can be written as a polynomial $I_l^\alpha(z_n) = \hat{T}_l^\alpha(z_n^{-1})$ with respect to the inverse frequency $z_n = 1/\omega_n^\alpha$.

- In the fermionic case, $I_l^F(z)$ has order $l+1$. It has l roots away from $z=0$ when l is even, and $l-1$ when l is odd.

- In the bosonic case, $I_l^B(z)$ has order l. It has $l-1$ roots away from $z=0$ when l is odd, and $l-2$ when l is even except for $l=0$, where there is no root.

Since $z_n \to 0$ implies ω_n^α going to infinity, we cannot use $z=0$ to define our sampling points. In order to take full advantage of the root structure of $I_l^\alpha(z)$ while being able to have at least N sampling points, we put the following restrictions on the basis size N, based on whether fermionic or bosonic statistics is concerned.

In the fermionic case, we require that basis size N to be even, such that $I_N^F(z)$ has exactly N roots $\{z_k\}$ away from $z=0$. In general, z_k^{-1} do not coincide exactly with Matsubara frequencies. We therefore approximate them by choosing the sampling points $\bar{\omega}_k^F$ to be the closest Matsubara frequency to z_k^{-1}.

In the bosonic case, even if we require N to be odd, there still are only $N-1$ roots away from $z=0$ from $I_N^B(z)$. The zero bosonic frequency $i\omega_n^B = 0$, which corresponds to a constant shift in τ, serves as a natural complement. The sampling points $\bar{\omega}_k^B$ is therefore composed of $N-1$ non-zero frequencies from approximating roots of $I_l^B(z)$ and one zero-frequency point.

Following this algorithm, we are able to get exactly N sampling points for either fermionic or bosonic statistics.

2. Condition numbers of transformation matrices

Every time Eqs. (6) and (7) are evaluated, numerical errors, such as round-off error in floating point operations, may be amplified due to the (pseudo-)inversion process. This error amplification can be quantified by the condition number of the transformation matrices $\mathbf{F}_\alpha$ and $\hat{\mathbf{F}}_\alpha$, defined as the product of the 2-norms of the matrix and its inverse. In Fig. 8 we show the behaviors of the condition numbers as a function of the basis size N for Chebyshev and IR (left panel), and as a function of Λ for IR basis (right panel). We can see that up
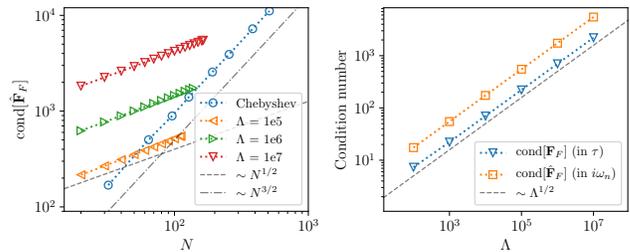
FIG. 8. Condition number of the transformation matrices. Left panel shows the condition number of frequency transformation matrices as a function of basis size N, for both Chebyshev and IR. Right panel shows the condition number of both τ and $i\omega_n$ transformation matrices with respect to Λ for the fermionic IR basis. N is chosen to be the maximum number of coefficients with the same cutoff in singular, provided in the `irbasis` library.

to a significant number of basis functions, the condition numbers are $< 10^4$, which indicates well-conditioned inversion problems. Since the values of N and Λ shown in Fig. 8 cover most values used in the calculations for this paper, the sparse sampling scheme guarantees stable numerical routines to get accurate results.

We observe that the condition number scales as $\mathcal{O}(N^{3/2})$ for Chebyshev, and $\mathcal{O}(N^{1/2})$ for IR. It also shows $\mathcal{O}(\Lambda^{1/2})$ scaling for IR. Although the origin of the condition number still requires further analysis, we believe that the algorithms we propose in this paper yield stable and predictable numerical procedures.

Appendix C: Technical details of self-consistent calculations using the sparse sampling method

1. Transforming between fermionic and bosonic statistics

Besides the transformation matrices defined in Eqs. (8) and (9), two additional matrices may be precomputed to allow fast switching between fermionic and bosonic representations in GW

$$[\mathbf{F}^{F\to B}]_{kl} = F_l^F(\bar{\tau}_k^B) \tag{C1}$$

$$[\mathbf{F}^{B\to F}]_{kl} = F_l^B(\bar{\tau}_k^F). \tag{C2}$$

Note that the inverse transform of those matrices are not well defined in general, since sampling points generated for one type of statistics usually do not serve as good sampling points for the other.

2. Evaluation of total energy and density matrix

In the total energy evaluation (19), the frequency summation term can be rewritten using an auxiliary scalar

quantity S such that

$$\frac{1}{2\beta} \sum_n \text{Tr}\,[\hat{\bar{\Sigma}}(i\omega_n^{\text{F}})\hat{G}(i\omega_n^{\text{F}})] = \frac{1}{2\beta} \sum_n \hat{S}(i\omega_n^{\text{F}})$$
$$= \frac{1}{2}S(0^-) = -\frac{1}{2}S(\beta) \quad \text{(C3)}$$

where

$$\hat{S}(i\omega_n^{\text{F}}) = \text{Tr}\,[\hat{\bar{\Sigma}}(i\omega_n^{\text{F}})\hat{G}(i\omega_n^{\text{F}})]. \tag{C4}$$

We first evaluate $\hat{S}(i\bar{\omega}_k^{\text{F}})$ on the frequency sampling points $i\bar{\omega}_k^{\text{F}}$, which is then transformed to the basis representation S_l^{F}. The value of $S(\beta)$ is now a straightforward

basis expansion at $\tau = \beta$

$$S(\beta) = \sum_{l=0}^{N-1} S_l^{\text{F}} F_l^{\text{F}}(\beta). \tag{C5}$$

For the Chebyshev basis we simply have $T_l(\beta) = 1$. In the case of IR basis, it is desirable to also tabulate the values $U_l^{\text{F}}(\beta)$ along with the transformation matrices for efficient evaluations of relevant quantities.

Similarly, the calculation of the density matrix ρ is a straightforward evaluation at $\tau = \beta$ from G_l^{F}. In calculations where the number of electrons is fixed, the chemical potential μ needs to be adjusted in each self-consistent iteration through a root finding procedure to conserve particle number [25]. This step involves repeated density evaluations and solutions of the Dyson equation using the frequency-dependent self-energy, and sometimes becomes the bottleneck of the calculation. The sparse sampling scheme massively reduces the number of frequency points needed in this process, which leads to a significant speedup over traditional approaches.

[1] T. Matsubara, A New Approach to Quantum-Statistical Mechanics, Progress of Theoretical Physics **14**, 351 (1955).

[2] A. A. Abrikosov, L. P. Gorkov, and I. E. Dzyaloshinski, *Methods of Quantum Field Theory in Statistical Physics* (Dover Publications, New York, 1975).

[3] J. W. Negele and H. Orland, *Quantum many-particle systems (Advanced Book classics)* (Westview Press, 1998).

[4] J. M. Luttinger and J. C. Ward, Ground-State Energy of a Many-Fermion System. II, Physical Review **118**, 1417 (1960).

[5] D. J. Scalapino and R. L. Sugar, Monte Carlo calculations of coupled boson-fermion systems. II, Physical Review B **24**, 4295 (1981).

[6] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions, Reviews of Modern Physics **68**, 13 (1996).

[7] T. Maier, M. Jarrell, T. Pruschke, and M. H. Hettler, Quantum cluster theories, Reviews of Modern Physics **77**, 1027 (2005).

[8] G. Kotliar, S. Y. Savrasov, K. Haule, V. S. Oudovenko, O. Parcollet, and C. A. Marianetti, Electronic structure calculations with dynamical mean-field theory, Reviews of Modern Physics **78**, 865 (2006).

[9] K. Held, I. A. Nekrasov, G. Keller, V. Eyert, N. Blümer, A. K. McMahan, R. T. Scalettar, T. Pruschke, V. I. Anisimov, and D. Vollhardt, Realistic investigations of correlated electron systems with LDA + DMFT, physica status solidi (b) **243**, 2599 (2006).

[10] H. Hafermann, S. Brener, A. N. Rubtsov, M. I. Katsnelson, and A. I. Lichtenstein, Cluster dual fermion approach to nonlocal correlations, JETP Letters **86**, 677 (2008).

[11] A. Toschi, A. A. Katanin, and K. Held, Dynamical vertex approximation: A step beyond dynamical mean-field

theory, Physical Review B **75**, 045118 (2007).

[12] G. Rohringer, H. Hafermann, A. Toschi, A. A. Katanin, A. E. Antipov, M. I. Katsnelson, A. I. Lichtenstein, A. N. Rubtsov, and K. Held, Diagrammatic routes to nonlocal correlations beyond dynamical mean field theory, Reviews of Modern Physics **90**, 25003 (2018).

[13] N. V. Prokof'ev and B. V. Svistunov, Polaron problem by diagrammatic quantum monte carlo, Physical Review Letters **81**, 2514 (1998).

[14] E. Gull, A. J. Millis, A. I. Lichtenstein, A. N. Rubtsov, M. Troyer, and P. Werner, Continuous-time Monte Carlo methods for quantum impurity models, Reviews of Modern Physics **83**, 349 (2011).

[15] L. Hedin, New Method for Calculating the One-Particle Green's Function with Application to the Electron-Gas Problem, Physical Review **139**, A796 (1965).

[16] F. Aryasetiawan and O. Gunnarsson, The GW method, Reports on Progress in Physics **61**, 237 (1998).

[17] A. Stan, N. E. Dahlen, and R. van Leeuwen, Levels of self-consistency in the GW approximation, The Journal of Chemical Physics **130**, 114105 (2009).

[18] A. Kutepov, S. Y. Savrasov, and G. Kotliar, Ground-state properties of simple elements from GW calculations, Physical Review B **80**, 041103(R) (2009).

[19] M. J. Van Setten, F. Caruso, S. Sharifzadeh, X. Ren, M. Scheffler, F. Liu, J. Lischner, L. Lin, J. R. Deslippe, S. G. Louie, C. Yang, F. Weigend, J. B. Neaton, F. Evers, and P. Rinke, GW100: Benchmarking G_0W_0 for Molecular Systems, Journal of Chemical Theory and Computation **11**, 5665 (2015).

[20] E. Maggio, P. Liu, M. J. van Setten, and G. Kresse, GW 100: A Plane Wave Perspective for Small Molecules, Journal of Chemical Theory and Computation **13**, 635 (2017).

[21] M. Grumet, P. Liu, M. Kaltak, J. Klimeš, and G. Kresse, Beyond the quasiparticle approximation: Fully self-

consistent *GW* calculations, Physical Review B **98**, 155143 (2018).

[22] A. L. Kutepov, Electronic structure of Na, K, Si, and LiF from self-consistent solution of Hedin's equations including vertex corrections, Physical Review B **94**, 155101 (2016).

[23] A. L. Kutepov, Self-consistent solution of Hedin's equations: Semiconductors and insulators, Physical Review B **95**, 195120 (2017).

[24] N. E. Dahlen and R. Van Leeuwen, Self-consistent solution of the Dyson equation for atoms and molecules within a conserving approximation, Journal of Chemical Physics **122**, 164102 (2005).

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

[26] J. J. Phillips, A. A. Kananenka, and D. Zgid, Fractional charge and spin errors in self-consistent Green's function theory, Journal of Chemical Physics **142**, 194108 (2015).

[27] A. A. Kananenka, J. J. Phillips, and D. Zgid, Efficient Temperature-Dependent Green's Functions Methods for Realistic Systems: Compact Grids for Orthogonal Polynomial Transforms, Journal of Chemical Theory and Computation **12**, 564 (2016).

[28] A. A. Kananenka, A. R. Welden, T. N. Lan, E. Gull, and D. Zgid, Efficient Temperature-Dependent Green's Function Methods for Realistic Systems: Using Cubic Spline Interpolation to Approximate Matsubara Green's Functions, Journal of Chemical Theory and Computation **12**, 2250 (2016).

[29] A. A. Rusakov and D. Zgid, Self-consistent second-order Green's function perturbation theory for periodic systems, Journal of Chemical Physics **144**, 54106 (2016).

[30] 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, The Journal of Chemical Physics **145**, 204106 (2016).

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

[32] P. Sun and G. Kotliar, Extended dynamical mean-field theory and *GW* method, Physical Review B **66**, 085120 (2002).

[33] S. Biermann, F. Aryasetiawan, and A. Georges, First-Principles Approach to the Electronic Structure of Strongly Correlated Systems: Combining the *GW* Approximation and Dynamical Mean-Field Theory, Physical Review Letters **90**, 086402 (2003).

[34] J. M. Tomczak, M. Casula, T. Miyake, F. Aryasetiawan, and S. Biermann, Combined GW and dynamical mean-field theory: Dynamical screening effects in transition metal oxides, EPL (Europhysics Letters) **100**, 67001 (2012).

[35] J. M. Tomczak, P. Liu, A. Toschi, G. Kresse, and K. Held, Merging GW with DMFT and non-local correlations beyond, European Physical Journal: Special Topics **226**, 2565 (2017).

[36] P. Werner and M. Casula, Dynamical screening in correlated electron systems—from lattice models to realistic materials, Journal of Physics: Condensed Matter **28**, 383001 (2016).

[37] A. A. Kananenka, E. Gull, and D. Zgid, Systematically improvable multiscale solver for correlated electron systems, Physical Review B **91**, 121111(R) (2015).

[38] T. N. Lan, A. A. Kananenka, and D. Zgid, Communication: Towards ab initio self-energy embedding theory in quantum chemistry, The Journal of Chemical Physics **143**, 241102 (2015).

[39] D. Zgid and E. Gull, Finite temperature quantum embedding theories for correlated systems, New Journal of Physics **19**, 023047 (2017).

[40] T. N. Lan and D. Zgid, Generalized Self-Energy Embedding Theory, The Journal of Physical Chemistry Letters **8**, 2200 (2017).

[41] T. N. Lan, A. Shee, J. Li, E. Gull, and D. Zgid, Testing self-energy embedding theory in combination with GW, Physical Review B **96**, 155106 (2017).

[42] L. N. Tran, S. Iskakov, and D. Zgid, Spin-Unrestricted Self-Energy Embedding Theory, The Journal of Physical Chemistry Letters **9**, 4444 (2018).

[43] A. A. Rusakov, S. Iskakov, L. N. Tran, and D. Zgid, Self-Energy Embedding Theory (SEET) for Periodic Systems, Journal of Chemical Theory and Computation **15**, 229 (2019).

[44] W. Ku, *Electronic Excitations in Metals and Semiconductors: Ab Initio Studies of Realistic Many-Particle Systems*, Ph.D. thesis, University of Tennessee (2000).

[45] W. Ku and A. G. Eguiluz, Band-Gap Problem in Semiconductors Revisited: Effects of Core States and Many-Body Self-Consistency, Physical Review Letters **89**, 126401 (2002).

[46] L. Boehnke, H. Hafermann, M. Ferrero, F. Lechermann, and O. Parcollet, Orthogonal polynomial representation of imaginary-time Green's functions, Physical Review B **84**, 075145 (2011).

[47] E. Gull, S. Iskakov, I. Krivenko, A. A. Rusakov, and D. Zgid, Chebyshev polynomial representation of imaginary-time response functions, Physical Review B **98**, 75127 (2018).

[48] A. Kutepov, K. Haule, S. Y. Savrasov, and G. Kotliar, Electronic structure of Pu and Am metals by self-consistent relativistic *GW* method, Physical Review B **85**, 155129 (2012).

[49] H. Shinaoka, J. Otsuki, M. Ohzeki, and K. Yoshimi, Compressing Green's function using intermediate representation between imaginary-time and real-frequency domains, Physical Review B **96**, 35147 (2017).

[50] N. Chikano, J. Otsuki, and H. Shinaoka, Performance analysis of a physically constructed orthogonal representation of imaginary-time Green's function, Physical Review B **98**, 35104 (2018).

[51] N. Chikano, K. Yoshimi, J. Otsuki, and H. Shinaoka, irbasis: Open-source database and software for intermediate-representation basis functions of imaginary-time Green's function, Computer Physics Communications **240**, 181 (2019).

[52] L. Steinbeck, A. Rubio, L. Reining, M. Torrent, I. White, and R. Godby, Enhancements to the GW space-time method, Computer Physics Communications **125**, 105 (2000).

[53] M. Kaltak, J. Klimeš, and G. Kresse, Low scaling algorithms for the random phase approximation: Imaginary time and laplace transformations, Journal of Chemical Theory and Computation **10**, 2498 (2014).

[54] A. A. Rusakov, J. J. Phillips, and D. Zgid, Local Hamilto-

nians for quantitative Green's function embedding methods, Journal of Chemical Physics **141**, 194105 (2014).

[55] A.-B. Comanac, *Dynamical Mean Field Theroy of Correlated Electron Systems: New Algorithms and Applications to Local Observables*, Ph.D. thesis, Columbia University (2007).

[56] N. Blümer, *Mott-Hubbard Metal-Insulator Transition and Optical Conductivity in High Dimensions*, Ph.D. thesis, Universität Augsburg (2002).

[57] A. Szabo and N. S. Ostlund, *McGraw-Hili, New York* (Dover Publications Inc New edition edn, 1989).

[58] D. Bohm and D. Pines, A Collective Description of Electron Interactions: III. Coulomb Interactions in a Degenerate Electron Gas, Physical Review **92**, 609 (1953).

[59] M. Motta, D. M. Ceperley, G. K.-L. Chan, J. A. Gomez, E. Gull, S. Guo, C. A. Jiménez-Hoyos, T. N. Lan, J. Li, F. Ma, A. J. Millis, N. V. Prokof'ev, U. Ray, G. E. Scuseria, S. Sorella, E. M. Stoudenmire, Q. Sun, I. S. Tupitsyn, S. R. White, D. Zgid, and S. Zhang, Towards the Solution of the Many-Electron Problem in Real Materials: Equation of State of the Hydrogen Chain with State-of-the-Art Many-Body Methods, Physical Review X **7**, 031059 (2017).

[60] 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, Wiley Interdisciplinary Reviews: Computational Molecular Science **8**, e1340 (2018).

[61] T. H. Dunning, Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen, The Journal of Chemical Physics **90**, 1007 (1989).

[62] S. Goedecker, M. Teter, and J. Hutter, Separable dual-space Gaussian pseudopotentials, Physical Review B **54**, 1703 (1996).

[63] R. K. Bryan, Maximum entropy analysis of oversampled data problems, European Biophysics Journal **18**, 165 (1990).

[64] M. Jarrell and J. E. Gubernatis, Bayesian inference and the analytic continuation of imaginary-time quantum Monte Carlo data, Physics Reports 10.1016/0370-1573(95)00074-7 (1996).

[65] R. Levy, J. LeBlanc, and E. Gull, Implementation of the maximum entropy method for analytic continuation, Computer Physics Communications **215**, 149 (2017).

[66] W. Sano, T. Koretsune, T. Tadano, R. Akashi, and R. Arita, Effect of Van Hove singularities on high-T_c superconductivity in H_3S, Physical Review B **93**, 094525 (2016).

[67] M. Jarrell, Hubbard model in infinite dimensions: A quantum Monte Carlo study, Physical Review Letters **69**, 168 (1992).

[68] K. Held, A. A. Katanin, and A. Toschi, Dynamical Vertex Approximation, Progress of Theoretical Physics Supplement **176**, 117 (2009).

[69] H. Shinaoka, J. Otsuki, K. Haule, M. Wallerberger, E. Gull, K. Yoshimi, and M. Ohzeki, Overcomplete compact representation of two-particle Green's functions, Physical Review B **97**, 205111 (2018).

[70] M. Kaltak and G. Kresse, Minimax isometry method, arXiv preprint arXiv:1909.01740 (2019).

[71] J. Otsuki, M. Ohzeki, H. Shinaoka, and K. Yoshimi, Sparse modeling approach to analytical continuation of imaginary-time quantum Monte Carlo data, Physical Review E **95**, 061302(R) (2017).

[72] Y. Nagai and H. Shinaoka, Smooth Self-energy in the Exact-diagonalization-based Dynamical Mean-field Theory: Intermediate-representation Filtering Approach, Journal of the Physical Society of Japan **88**, 064004 (2019).