# Reduced basis emulation of pairing in finite systems

V. V. Baran[1,2,3,*] and D. R. Nichita[2,3]

[1]*Research Institute of the University of Bucharest (ICUB), 050107 Bucharest, Romania*

[2]*Faculty of Physics, University of Bucharest, 405 Atomiştilor, RO-077125, Bucharest-Măgurele, Romania*

[3]*"Horia Hulubei" National Institute of Physics and Nuclear Engineering, 30 Reactorului, RO-077125, Bucharest-Măgurele, Romania*

(Received 29 December 2022; revised 25 March 2023; accepted 28 March 2023; published 7 April 2023)

In recent years, reduced-basis methods (RBMs) have been adapted to the many-body eigenvalue problem and they have been used, largely in nuclear physics, as fast emulators able to bypass expensive direct computations while still providing highly accurate results. This work is meant to show that the RBM is an efficient and accurate emulator for the strong correlations induced by the pairing interaction in a variety of finite systems like ultrasmall superconducting grains, interacting topological superfluids, and mesoscopic hybrid superconductor-semiconductor devices, all of which require an expensive, beyond-mean-field, particle-number conserving description. These systems are modeled by the number-conserving Richardson pairing Hamiltonian and its appropriate generalizations. Their ground state is solved for exactly using the density matrix renormalization group. The reduced basis is assembled iteratively from a small number of exact ground-state vectors, well-chosen from across the relevant parameter space using a fast estimate of the emulation error and a greedy local optimization algorithm. The RBM emulation is found to accurately describe the weak-to-strong pairing crossover in small grains, the third-order topological phase transition of the interacting Richardson-Kitaev chain, and the complex charge stability diagram of a hybrid quantum dot–superconductor device. For all considered systems the emulation error decreases exponentially with the reduced-basis dimensionality. The number of basis vectors necessary to reach a fixed emulation accuracy only shows a modest linear growth with increasing system size. RBMs are confirmed to be cheap and accurate emulators for the widely encountered superconducting phenomena. Capable of providing orders of magnitude computational speedup with respect to approaches based only on traditional many-body solvers, they open new possibilities in building and solving models of interacting many-body systems and in better interfacing them with experimental design and data analysis.

DOI: 10.1103/PhysRevB.107.144503

## I. INTRODUCTION

Strongly correlated systems are ubiquitous in several fields of fundamental physics including condensed matter [1,2] and nuclear physics [3]. Across these fields, the quantum-mechanical treatment of many interacting particles is extremely challenging both formally and computationally, which has led to an increased interest in building efficient emulators whose predictions go beyond what is possible with direct calculations and make many-query computations more affordable [4–6].

Recently, the eigenvector continuation (EC) approach was introduced in Ref. [7] as an emulator capable of efficient interpolations and extrapolations for the extremal eigenstates of a Hamiltonian defined by one or more variable parameters. Nuclear physicists have been using it for the emulation of scattering and reactions [8–14] and shell model calculations [15], as a perturbation theory resummation tool [16–18], and for sensitivity analysis and uncertainty quantification [19–21]. Furthermore, it has been shown to provide a universal set of states capable of reproducing the entire single-particle spectrum across the nuclear chart with remarkable accuracy [22].

The success of EC relies on the smooth eigenstate behavior over the parameter manifold being effectively limited to a very low-dimensional subspace (compared to the full Hilbert space dimensionality). In practice, this means that the EC ansatz may be computed as a linear combination of (a small number of) "exact" solutions of the considered problem, by solving a generalized eigenvalue problem in the restricted subspace. The "exact" or "truth" direct solutions are to be obtained using any high-fidelity (albeit usually computationally expensive) numerical many-body method.

More recently, EC has naturally been recognized [23,24] as being an instance of a larger class of reduced-basis methods (RBMs) [25–27], part of the general framework of reduced-order models [24,28,29]. Simultaneously, the idea of using RBMs in the context of eigenvalue problems has been independently rediscovered and applied to build surrogate models (emulators) for the efficient determination of phase diagrams in quantum spin models in Ref. [30].

Given that RBMs have only been used in recent years for the study of quantum many-body systems, their applicability remains to be tested in various scenarios. It is the purpose of this work to show that the strong correlations induced by the pairing interaction in the ground state of finite systems may be successfully emulated with RBMs.

*virgil.v.baran@unibuc.ro

 

In Sec. II, we shortly review the reduced-basis (rb) methodology in the context of eigenvalue problems. Section III is devoted to the RBM treatment of the generic pairing Hamiltonian, also known as the Richardson Hamiltonian. It used to describe pairing between protons or between neutrons in atomic nuclei [3] and electron pairing in ultrasmall superconducting grains [1,31] or small superconducting islands [32–36] in condensed matter physics. In Sec. IV we obtain the first RBM description of a topological phase transition [37], and in Sec. V we show that the RBM is able to successfully emulate the complex charging patterns of an interacting quantum dot in contact with a superconducting island [32]. Finally, in Sec. VI we draw conclusions.

## II. REDUCED-BASIS METHODOLOGY

We present here a short review of the general reduced-basis methodology employed in this work, for self-consistency. For a more detailed treatment, we refer the reader to Refs. [24,30].

All systems under consideration are modeled by Hamiltonians that may be written in a so-called affine decomposition (see also the specific examples in the next sections),

$$\mathcal{H}(\boldsymbol{\xi}) = \sum_{p=0}^{N_{\mathcal{H}}} f_p(\boldsymbol{\xi}) H_p, \tag{1}$$

where $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_n)$ is a set of $n$ control parameters in the domain of interest. Here, $f_p(\boldsymbol{\xi})$ are a set of $N_{\mathcal{H}} + 1$ functions dictating the full parameter dependence of the system's Hamiltonian $\mathcal{H}(\boldsymbol{\xi})$, with the corresponding $H_p'$s being parameter-independent. By convention, we denote the parameter-independent part of $\mathcal{H}(\boldsymbol{\xi})$ by $H_0$, and thus we take $f_0(\boldsymbol{\xi}) = 1$.

The core RBM idea is to construct a low-dimensional effective representation $\psi^{(\mathrm{rb})}(\boldsymbol{\xi})$ for the ground state of $\mathcal{H}(\boldsymbol{\xi})$ that is sufficiently accurate over the parameter domain of interest. In doing so, we only employ a relatively small number $N_{\mathrm{rb}}$ of sample or training points $\boldsymbol{\xi}_k$ where "exact" solutions $\psi_k^{(\mathrm{ex})}$ of the Hamiltonian problem are actually computed, i.e.,

$$|\psi^{(\mathrm{rb})}(\boldsymbol{\xi})\rangle = \sum_{k=1}^{N_{\mathrm{rb}}} c_k(\boldsymbol{\xi}) \left|\psi_k^{(\mathrm{ex})}\right\rangle, \tag{2}$$

with $N_{\mathrm{rb}}$ being much smaller than the dimension of the full many-body Hilbert space. Here, the expansion coefficients $c_k(\boldsymbol{\xi})$ for any parameter values may be obtained upon solving the effective generalized eigenvalue problem for $h_{k\ell}(\boldsymbol{\xi}) = \langle\psi_k^{(\mathrm{ex})}|\mathcal{H}(\boldsymbol{\xi})|\psi_\ell^{(\mathrm{ex})}\rangle$ in the (nonorthogonal) reduced basis of $\psi^{(\mathrm{ex})}$'s,

$$h\,c = E^{(\mathrm{rb})}\,S\,c, \tag{3}$$

involving the overlap matrix

$$S_{k\ell} \equiv \left\langle\psi_k^{(\mathrm{ex})}\middle|\psi_\ell^{(\mathrm{ex})}\right\rangle. \tag{4}$$

Regarding the choice of sampling points, one efficient way to sample the parameter space is through an active learning protocol which combines a fast estimate of the emulation error and a greedy optimization algorithm that becomes progressively more accurate [30,38]. In this so-called "offline" emulation phase, the reduced basis is constructed iteratively by repeating two main steps.

The first step is an evaluation over the relevant parameter space of the residual

$$\mathrm{Res}(\boldsymbol{\xi}) \equiv \|\mathcal{H}(\boldsymbol{\xi})|\psi^{(\mathrm{rb})}(\boldsymbol{\xi})\rangle - E^{(\mathrm{rb})}(\boldsymbol{\xi})|\psi^{(\mathrm{rb})}(\boldsymbol{\xi})\rangle\|, \tag{5}$$

which quantifies the degree to which the exact Schrödinger equation is satisfied by the current reduced-basis approximate solution $\psi^{(\mathrm{rb})}$. Note that only the emulated information is used in this step and there is no need to perform the error evaluation using new costly "exact" solutions (which would defeat the purpose of the emulation).

The location in the parameter domain of the maximum residual indicates where the emulation is least accurate and would benefit the most from a new sampling point (see also Appendix A for more details on this point). It is at the residual's maximum location that, in the second step, we run the "exact" solver and add the resulting vector to the reduced basis, thus increasing $N_{\mathrm{rb}}$ by one unit.

These steps are repeated until the emulation error (as quantified by the above defined residual) decreases below a suitably chosen threshold. On the one hand, the success of the RBM in reaching rapidly a good enough emulation accuracy within a low-dimensional subspace is naturally related to the high linear dependence of the basis of training vectors $\psi_k^{(\mathrm{ex})}$ [23,26].

On the other hand, this may lead to a numerically singular overlap matrix $S$ and thus to numerical instabilities in the solution to Eq. (3). It is customary to improve upon this by orthogonalizing the $\psi_k^{(\mathrm{ex})}$'s, leading to a standard eigenvalue problem. In practice, we perform this orthogonalization by solving the eigenvalue problem for the Hermitian overlap matrix $S\widetilde{c}_a = s_a\widetilde{c}_a$ and using its orthogonal eigenvectors $\widetilde{c}_a$ to build a new and numerically stable reduced basis.

Within this proper orthogonal decomposition (POD) approach [23,26,30] we have the option of further compressing our basis by discarding those $\widetilde{c}_a$ vectors corresponding to eigenvalues $s_a$ lower than a certain tolerance value, thus capturing an effective low-dimensional representation of the full set of "exact" ground-state solutions $\psi_k^{(\mathrm{ex})}$. The largest-$s$ $\widetilde{c}$-vector can be seen as the "average" ground state, while the small-$s$ $\widetilde{c}$-subspace accommodates fine ground-state variations over the sampled parameter domain. That the small-$s$ contributions are mostly decoupled from the low-energy subspace may be seen also from the equivalent isospectral form of Eq. (3) involving $S^{-1/2}hS^{-1/2}$. We quantitatively discuss the effects of this lossy basis compression on the emulation performance in the next section.

The reduced basis assembled during the emulator's training in the offline phase is guaranteed to approximate sufficiently well our system's "exact" ground state over the entire relevant parameter domain. This basis is then used in the final "online" emulation phase to evaluate efficiently the behavior of any affinely decomposable observable of interest

$\mathcal{A}(\boldsymbol{\xi}) = \sum_{p=1}^{N_\mathcal{A}} g_p(\boldsymbol{\xi}) A_p$ using the reduced-basis representation

$$A^{(\mathrm{rb})}(\boldsymbol{\xi}) = \langle \psi^{(\mathrm{rb})}(\boldsymbol{\xi}) | \mathcal{A}(\boldsymbol{\xi}) | \psi^{(\mathrm{rb})}(\boldsymbol{\xi}) \rangle = \sum_{p=1}^{N_\mathcal{A}} \sum_{k,\ell=1}^{N_{\mathrm{rb}}} g_p(\boldsymbol{\xi}) c_k^*(\boldsymbol{\xi}) c_\ell(\boldsymbol{\xi}) \langle \psi_k^{(\mathrm{ex})} | A_p | \psi_\ell^{(\mathrm{ex})} \rangle, \tag{6}$$

where the parameter-independent matrix elements of the various $A_p'$s in the $\psi_k^{(\mathrm{ex})}$ basis need only be evaluated once using "exact" methods. If employing the above POD approach, then any operator matrix element in the compressed basis may be trivially computed from the matrix elements in the original $\psi^{(\mathrm{ex})}$ basis. In the case of observables with nonaffine parameter dependencies, techniques such as the Empirical Interpolation Method may be used to construct an affine approximation [39,40].

We finally note that the RBM framework is blind to the actual method employed for obtaining the "exact" ground-state solutions: the only input needed for a minimal RBM emulator consists of the matrix elements $\langle \psi_k^{(\mathrm{ex})} | H_p | \psi_\ell^{(\mathrm{ex})} \rangle$ of the Hamiltonian affine components $H_p$ together with the overlaps $\langle \psi_k^{(\mathrm{ex})} | \psi_\ell^{(\mathrm{ex})} \rangle$ used to construct the effective eigenvalue problem of Eq. (3). Additionally, the matrix elements $\langle \psi_k^{(\mathrm{ex})} | H_p H_{p'} | \psi_\ell^{(\mathrm{ex})} \rangle$ must be provided to the RBM emulator for evaluating the residual in Eq. (5) during the offline phase, and the matrix elements of any operator of interest $\langle \psi_k^{(\mathrm{ex})} | A_p | \psi_\ell^{(\mathrm{ex})} \rangle$ for the online phase.

All "exact" solutions are obtained here by the density matrix renormalization group (DMRG) technique [41] in the matrix product state (MPS) formulation [42], using the Itensor library [43,44] for the numerical implementation. The latter allows for complete control over the numerical accuracy of the obtained solutions, while providing MPS tools for evaluating efficiently the matrix elements mentioned in the previous paragraph. We employed a $10^{-9}$ energy convergence tolerance for the DMRG sweeps in the presence of a maximum bond dimension of 2000 and utilized the ITensor "noise" term to ensure that the energy global minimum is reached.

## III. EMULATING THE RICHARDSON PAIRING HAMILTONIAN

### A. Exact ground-state structure

We first discuss qualitatively the structure of the ground state for the particle-number conserving generic $s$-wave singlet-pairing Hamiltonian of spin-1/2 fermions

$$H(G) \equiv \sum_{i=1}^{L} \sum_{\sigma=\uparrow,\downarrow} \epsilon_i \, c_{i,\sigma}^\dagger c_{i,\sigma} - G \sum_{i,j=1}^{L} c_{i,\uparrow}^\dagger c_{i,\downarrow}^\dagger c_{j,\downarrow} c_{j,\uparrow} \equiv \sum_{i=1}^{L} \epsilon_i N_i - G \sum_{i,j=1}^{L} P_i^\dagger P_j, \tag{7}$$

where $i$ indicates one of the $L$ pairs of conjugated degenerate single particle levels with energy $\epsilon_i = \epsilon_{i,\uparrow} = \epsilon_{i,\downarrow}$ and $G$ denotes the pairing interaction strength. In the following, we will take the single particle levels to be uniformly distributed with energy spacing $\epsilon = \epsilon_{i+1} - \epsilon_i$.

For simplicity, we restrict ourselves to even particle-number systems with no unpaired particles. Then, in the absence of the interaction term the ground state is given by the Hartree-Fock product state

$$|\mathrm{HF}\rangle = \prod_{i=1}^{N_p} P_i^\dagger \, |0\rangle \quad (G = 0), \tag{8}$$

where $N_p$ is the number of pairs in the system. The weak pairing regime $G \ll \epsilon$ may be described to a very good approximation by coherent pair-excitations through a coupled-cluster doubles variational ansatz [45]

$$|\psi\rangle \simeq \exp\left[ \sum_{ij} z_{ij} P_i^\dagger P_j \right] |\mathrm{HF}\rangle \quad (G \ll \epsilon). \tag{9}$$

The opposite limit of very strong pairing with $G \lll \epsilon$ or equivalently $\epsilon = 0$ is known as the seniority model [46] in nuclear physics and as the zero bandwidth model [47] in the condensed matter physics community, and has a product of identical collective pairs as the exact ground state,

$$|\psi\rangle = \left( \sum_{i=1}^{L} P_i^\dagger \right)^{N_p} |0\rangle \quad (G \lll \epsilon). \tag{10}$$

The crossover regime between weak and strong pairing (where the exact ground-state structure changes qualitatively from coherent pair-excitations to a collective pair condensate) is situated on an energy scale comparable to the critical BCS pairing strength $G_{\mathrm{cr}}$ [45]. At this value of the pairing strength, the $U(1)$-particle-number-breaking BCS solution collapses and gives zero pairing correlations. This effect may be understood from the BCS gap equation $\Delta = G/2 \sum_{i=1}^{L} \Delta / \sqrt{(\epsilon_i - \mu)^2 + \Delta^2}$ (involving the chemical potential $\mu$), which admits a nontrivial solution only for

$$G > \frac{2}{\sum_{i=1}^{L} |\epsilon_i - \mu|^{-1}} \equiv G_{\mathrm{cr}}. \tag{11}$$

This is contrary to the expectation for macroscopic systems with a finite-level density at the Fermi surface $\rho_F$, and thus with infinitely many available levels in a finite energy window $\omega_D$ around the Fermi surface. In this case $G_{\mathrm{cr}}$ vanishes and the gap is related to the pairing strength by the well-known BCS relation $\Delta \simeq 2\omega_D \exp[-1/(G\rho_F)]$, displaying explicitly the finite-level density $\rho_F$.

For the range of system sizes considered in this work, the critical pairing strength varies from $G_{\mathrm{cr}} \simeq 0.32\epsilon$ for $L = 2N_p = 12$ to $G_{\mathrm{cr}} \simeq 0.16\epsilon$ for $L = 2N_p = 200$.

Note that while the previous considerations apply also to more general pairing Hamiltonians involving nonconstant matrix elements of the pairing potential, the specific Hamiltonian of Eq. (7) actually admits an exact solution originally discovered by Richardson [48]. The Richardson solution provides a unified description of all pairing regimes in terms of a product of distinct collective pair operators

$$|\psi\rangle = \prod_{\alpha=1}^{N_p} \left( \sum_{i=1}^{L} \frac{P_i^\dagger}{2\epsilon_i - e_\alpha} \right) |0\rangle \quad (\text{any } G), \tag{12}$$

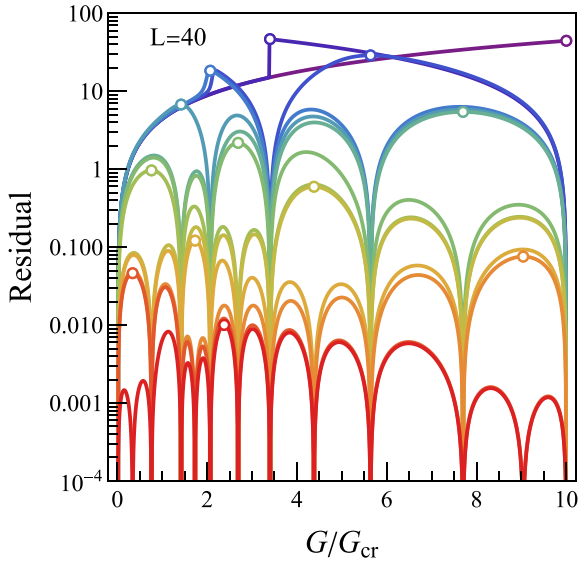


FIG. 1. Evolution of the residual (5), in units of the level spacing $\epsilon$, during the offline emulation phase of the pairing Hamiltonian (7) for a system of $N_p = 20$ pairs distributed over $L = 40$ levels. The sampling interval for the pairing strength $G$ is $0 \leqslant G \leqslant 10G_{cr}$, with the corresponding $G_{cr} \simeq 0.22\epsilon$ for $L = 2N_p = 40$. The color spectrum (violet, blue, green, orange, red) correlates with the increasing iteration number of the greedy self-learning algorithm (and thus also with the increasing reduced-basis dimension $N_{rb}$). The position of the residual's maximum at each iteration is indicated by an open circle of the appropriate color.

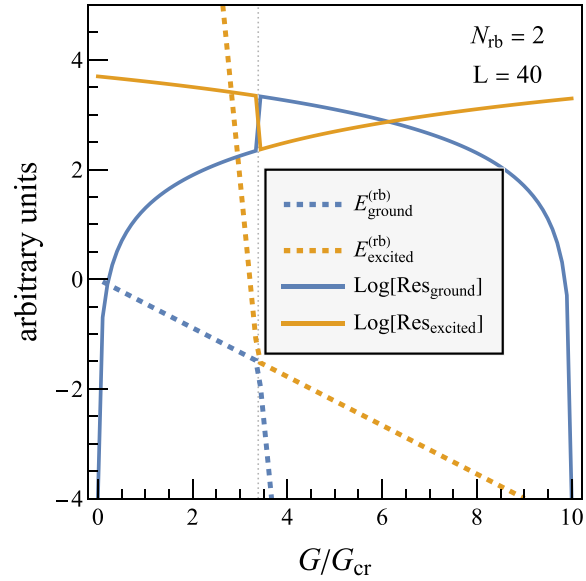


FIG. 2. Energies $E^{(rb)}$ (dashed lines) and residuals (5) (solid lines) corresponding the ground and excited states in a two-dimensional reduced subspace $N_{rb} = 2$ with sampling vectors situated at the ends of the sampling interval $0 \leqslant G \leqslant 10G_{cr}$. For the chosen system size $L = 2N_p = 40$ the crossing occurs around $G \simeq 3.4G_{cr}$ (vertical dotted gray line).

each collective pair being defined by a different pair energy $e_m$. The pair energies may be found for any value of the pairing strength $G$ by solving a set of coupled nonlinear equations [2].

### B. RBM results

We are now in a position to benchmark the RBM capabilities in emulating the weak to strong pairing crossover. For this purpose, we consider the Richardson pairing Hamiltonian of Eq. (7) with the pairing strength $G$ as the control parameter, $\xi = G$, and we take the level spacing $\epsilon = 1$ as the unit of energy. We sample the interval $0 \leqslant G \leqslant 10G_{cr}$ using the greedy algorithm described in the previous section, leading to the residual evolution over 14 iterations illustrated in Fig. 1 for $N_p = 20$ pairs distributed on $L = 40$ levels.

We take the first sampling point at $G = 0$ and thus $|\psi_1^{(ex)}\rangle = |\mathrm{HF}\rangle$ state as the initial reduced-basis vector. In this first iteration the emulation accuracy naturally degrades when moving away from the weak pairing regime. The residual thus reaches its maximum at the other end of the interval ($G = 10G_{cr}$), where the next basis vector $\psi_2^{(ex)}$ is computed by a DMRG evaluation.

In the presence of sampling points situated at both ends of our parameter interval the new residual maximum is now more centrally located around $G = 3.4G_{cr}$, where we compute with DRMG a new solution $\psi_3^{(ex)}$. This procedure is repeated until achieving the desired emulation accuracy, as given by the global maximum of the residual.

As also remarked in Ref. [30] and contrary to a naive expectation, the decrease of the residual's maximum from one iteration of the greedy sampling algorithm to the next is not always monotonic. Additionally, more or less pronounced discontinuities are present in the residual profile in Fig. 1 during the early iterations.

To better understand this behavior, we give in Fig. 2 more details on the second iterative step $N_{rb} = 2$ involving sampling points at both interval endpoints. Figure 2 shows how the $\psi_1^{(ex)}$-dominated vector and the $\psi_2^{(ex)}$-dominated vector (having continuous energies and residuals over the entire parameter range) exchange roles as ground and excited states around $G = 3.4G_{cr}$. At this point, the $\psi_1^{(ex)}$-dominated ground state suddenly becomes $\psi_2^{(ex)}$-dominated and experiences a residual jump. For denser $N_{rb} > 2$ samplings, the residual mismatch for the various $\psi_k^{(ex)}$-dominated vectors is naturally lower. As a consequence the jumps in the ground-state residual become less pronounced, as also seen in Fig. 1.

The residual jump of the ground state at $N_{rb} = 2$ is amplified when increasing the system's size, as shown in Fig. 3. Nevertheless, in all the tested examples, ranging from small $L = 2N_p = 12$ to large $L = 2N_p = 200$ systems, the residual maximum decays exponentially beyond $N_{rb} = 2$. This leads to an approximately linear increase with the system's size $L$ of the number of basis states $N_{rb}$ needed to reach our chosen threshold for the emulation error of $\mathrm{Res}/\epsilon = 10^{-2}$, of about one additional state per 10 extra levels. Note also that we choose to scale up the system while keeping intact the sampling interval $0 \leqslant G \leqslant 10G_{cr}$ to emulate the weak to strong pairing crossover under controlled conditions. The reduction of the critical pairing strength $G_{cr}$ by a factor of two from $L = 2N_p = 12$ ($G_{cr} \simeq 0.32\epsilon$) to $L = 2N_p = 200$

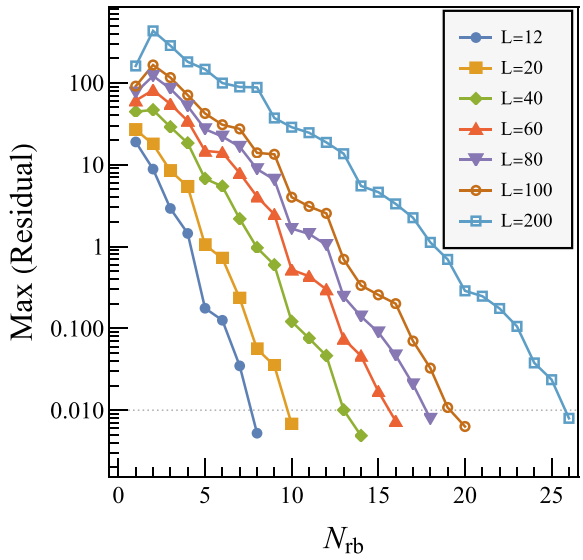


FIG. 3. Evolution of the maximum of the residual (5), in units of the level spacing $\epsilon$, versus the reduced-basis dimension $N_{\rm rb}$ during the offline emulation phase of the pairing Hamiltonian (7) for various systems sizes, at half filling $L = 2N_p$, for the sampling interval $0 \leqslant G \leqslant 10G_{\rm cr}$. The $L = 40$ values correspond to the open circles in Fig. 1.

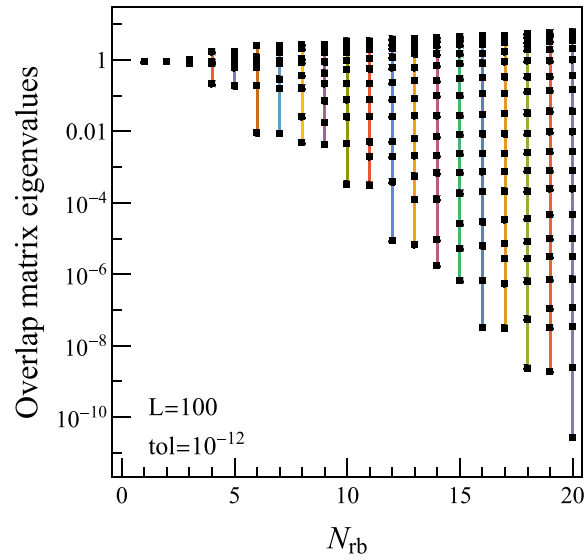


FIG. 4. Evolution of the eigenvalues of the overlap matrix (4) with increasing reduced-basis dimension $N_{\rm rb}$ during the offline emulation phase of the pairing Hamiltonian (7) for the sampling interval $0 \leqslant G \leqslant 10G_{\rm cr}$. The systems size is $L = 2N_p = 100$, corresponding to $G_{\rm cr} \simeq 0.18\epsilon$. The chosen POD basis compression tolerance for the overlap matrix eigenvalues is tol $= 10^{-12}$ such that none of the latter are discarded.

($G_{\rm cr} \simeq 0.16\epsilon$) would then imply at most a doubling of the reduced-basis growth rate if keeping the absolute $G/\epsilon$ interval constant during system scaling.

The exponentially fast decay of the residual during the greedy sampling algorithm is related to high linear dependence of the basis vectors (see the discussion in Sec. II). This in turn implies an exponential suppression of new overlap-matrix eigenvalues during the reduced-basis construction (offline) phase, which is confirmed in Fig. 4.

We note however that the POD basis compression losses may have a large impact on the emulator's ability to actually reach the target accuracy. We explore this in Fig. 5 where the evolution of the residual's maximum is shown for various tolerances of the overlap matrix eigenvalues. If the tolerance is not low enough (and too many vectors are discarded from the reduced basis), then the remaining subspace may not be large enough to accommodate all relevant ground-state variations over the parameter domain. This leads to a saturation in the residual's decay beyond a basis dimensionality for which any new overlap-matrix eigenvalue will not exceed the tolerance value. For a fixed choice of the residual function, a global RBM description with arbitrarily high accuracy may thus be numerically inaccessible in extreme cases where the required POD tolerance becomes comparable to the numerical tolerance of the chosen diagonalization routine. In such situations, one could avoid the inherent numerical instabilities by breaking the sampling interval into smaller subintervals and turning to a piece-wise RBM description using multiple reduced bases, assembled independently on each subinterval. Throughout this section, all results except for those presented in Fig. 5 are obtained without any POD basis compression.

We conclude this section by assessing the quality of the greedy algorithm in converging to the exact solution of the Richardson pairing Hamiltonian. As a first convergence metric, we use the correlation energy

$$E_{\rm corr}(G) \equiv E_{\rm HF}(G) - E_{\rm ground}(G), \tag{13}$$

where $E_{\rm HF}(G) = \langle {\rm HF}|H(G)|{\rm HF}\rangle = \sum_{i=1}^{N_p}(2\epsilon_i - G)$ is the energy of the Hartree-Fock state. $E_{\rm corr}$ measures the amount of correlations captured by the wave function under

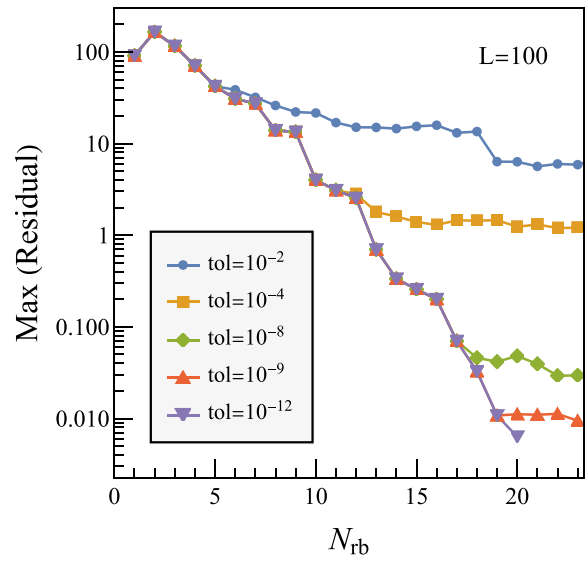


FIG. 5. Evolution of the residual's maximum, in units of the level spacing $\epsilon$, versus the reduced-basis dimension $N_{\rm rb}$ during the offline emulation phase of the pairing Hamiltonian (7), for various POD compression tolerances. The sampling interval is $0 \leqslant G \leqslant 10G_{\rm cr}$ and the systems size is $L = 2N_p = 100$, corresponding to $G_{\rm cr} \simeq 0.18\epsilon$.

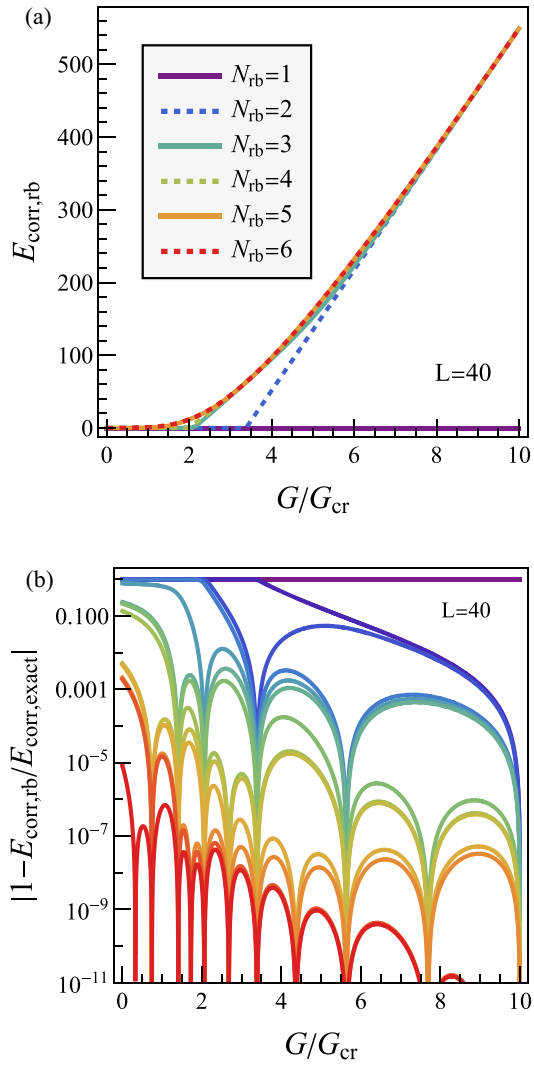


FIG. 6. Evolution of the reduced-basis correlation energy $E_{\mathrm{corr}}$ (13) (a), in units of the level spacing $\epsilon$, and the corresponding error relative to its exact value (b) during the offline emulation phase. The sampling interval for the pairing strength $G$ is $0 \leqslant G \leqslant 10G_{\mathrm{cr}}$, with the corresponding $G_{\mathrm{cr}} \simeq 0.22\epsilon$ for the chosen system size $L = 2N_p = 40$.

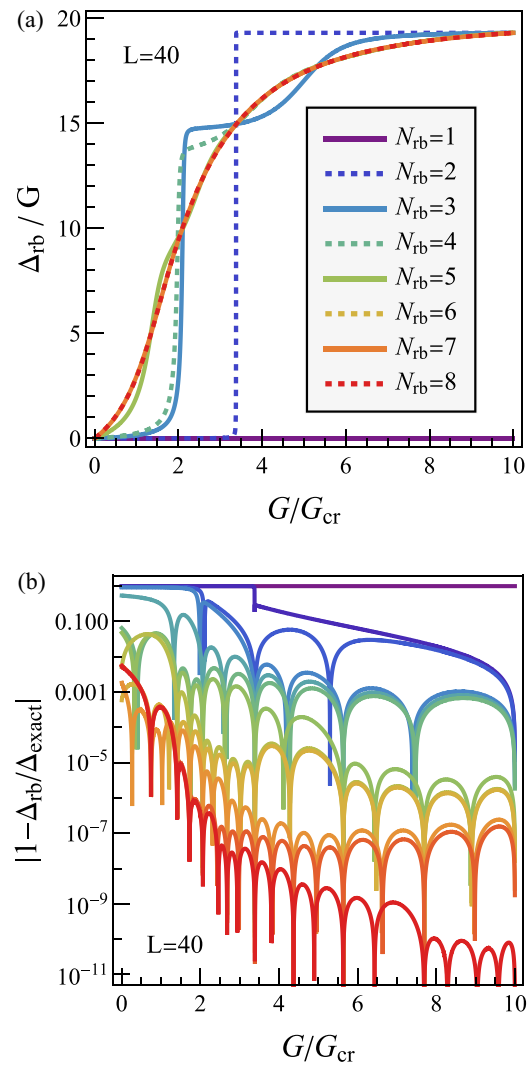


FIG. 7. Evolution of the reduced-basis canonical gap $\Delta$ (14) (a), and the corresponding error relative to its exact value (b), during the offline emulation phase. The sampling interval for the pairing strength $G$ is $0 \leqslant G \leqslant 10G_{\mathrm{cr}}$, with the corresponding $G_{\mathrm{cr}} \simeq 0.22\epsilon$ for the chosen system size $L = 2N_p = 40$.

consideration beyond the Hartree-Fock state and is thus sensitive on the wave function's structure especially in the weak pairing regime, as can be seen in Fig. 6(b). Even with the $|\mathrm{HF}\rangle$ state as one of the sampling points, in the extremely weak pairing regime $G \simeq 0$ the error in the correlation energy (relative to its exact value) remains finite due to the vanishing of $E_{\mathrm{corr}}(G = 0)$. This gives an indication on the rate at which the Hartree-Fock state is approached by our approximate solution, which turns out to be similar to that obtained using a more complex variational ansatz [45] (for the chosen residual accuracy goal of $10^{-2}\epsilon$).

As a second convergence metric, we consider the canonical gap

$$\Delta = G \sum_{i=1}^{L} \sqrt{n_i(1-n_i)}, \tag{14}$$

where $n_i = \langle N_i \rangle / 2$ indicates the occupation probability of each level $i$. Due to its dependence on the occupation probabilities, this quantity exhibits a more pronounced sensitivity to the structure of the wave function than the correlation energy. We show in Fig. 7(a) the profile of $\Delta/G$ obtained during the first iterations of the greedy algorithm, which converges rapidly towards the exact curve in an oscillatory manner. This leads to an increased number of zeros in the gap error presented in Fig. 7(b), corresponding both to the sampling points and to the intersections of the oscillating approximate profile with the exact curve. Similarly to the case of the correlation energy, the smallness of the gap in the weak pairing regime leads to the errors being largest there. However, the oscillatory behavior of the reduced-basis gap approximation leads to a nonmonotonic error decrease at $G \simeq 0$.

The success of the RBM emulation of the generic Richardson pairing Hamiltonian raises the question of whether RBMs can provide benefits in modeling more complex

condensed matter systems like topological superfluids or hybrid superconductor-semiconductor nano-devices, which we investigate in the following.

## IV. EMULATING FINITE TOPOLOGICAL SUPERFLUIDS

### A. The Richardson-Gaudin-Kitaev chain

As a generalization of the Richardson pairing Hamiltonian, we consider the Richardson-Gaudin-Kitaev (RGK) chain introduced in Ref. [37], a key example of an interacting, particle-conserving, fermionic superfluid in one spatial dimension displaying a topologically nontrivial superfluid phase. We note that, in the context of topologically protected quantum computation, determining the consequences of particle-number conservation on the properties of Majorana modes and on their braiding statistics (usually discussed within a mean-field treatment) has attracted a lot of attention in recent years and is an open problem at the time of this writing; see Ref. [49] and references therein.

Being formulated within the number-conserving setting, the one-dimensional RGK Hamiltonian of Ref. [37] includes long-range interactions to evade the Mermin-Wagner-Hohenberg (or Coleman in field theory) theorem and to display a true gap in the thermodynamic limit. Its explicit form in the momentum representation is given by

$$H_{\mathrm{RGK}} = \sum_{k\in\mathcal{S}_k^\phi} \varepsilon_k c_k^\dagger c_k - 8G \sum_{k,k'\in\mathcal{S}_{k+}^\phi} \eta_k \eta_{k'} c_k^\dagger c_{-k}^\dagger c_{-k'} c_{k'} \qquad (15)$$

for spinless fermions $c_k^\dagger$, with momentum-dependent single-particle spectrum

$$\varepsilon_k = -2t_1 \cos k - 2t_2 \cos 2k, \qquad (16)$$

where $t_1$ and $t_2$ are the nearest- and next-nearest-neighbor hopping amplitudes, and a momentum-dependent modulation of the interaction strength

$$\eta_k = \sin(k/2)\sqrt{t_1 + 4t_2 \cos^2(k/2)}, \qquad (17)$$

which is odd in $k$, as is characteristic of $p$-wave superconductivity. The specific values of the allowed momenta in the cases of interest below, i.e., for periodic ($\phi = 0$) and antiperiodic ($\phi = 2\pi$) boundary conditions on a chain of $L$ sites, are $\mathcal{S}_k^0 = \mathcal{S}_{k+}^0 \oplus \mathcal{S}_{k-}^0 \oplus \{0, -\pi\}$ and $\mathcal{S}_k^{2\pi} = \mathcal{S}_{k+}^{2\pi} \oplus \mathcal{S}_{k-}^{2\pi}$, with $\mathcal{S}_{k\pm}^0 = L^{-1}\{\pm 2\pi, \pm 4\pi, \cdots, \pm(\pi L - 2\pi)\}$ and $\mathcal{S}_{k\pm}^{2\pi} = L^{-1}\{\pm\pi, \pm 3\pi, \cdots, \pm(\pi L - \pi)\}$. As in Ref. [37], we assumed here that the boundary conditions are controlled by varying the enclosed flux (in ring geometry) from $\phi = 0$ to $\phi = 2\pi$ (in units of $\Phi_0 = h/2e$).

Physically, the factorized form of the interaction term ensures that the pairing model has a sensible form also in real space, where it describes long-range pair hopping. Computationally, the separable form allows for the RGK Hamiltonian to be represented exactly as a matrix product operator (MPO) of bond dimension 4 and thus to be efficiently solved with DMRG even when the exact solvability condition $4\eta_k^2 = \epsilon_k + 2t_1 + 2t_2$ is not fulfilled.

If instead the latter condition is realized, then the RGK Hamiltonian assumes the standard form of a hyperbolic Richardson-Gaudin integrable model [2,50]. Its eigenstates involving $N_p$ fermion pairs and $N_\nu = 0$ or 1 unpaired fermions are then given by

$$|\psi\rangle = \prod_{\alpha=1}^{N_p} \left( \sum_{k\in\mathcal{S}_{k+}^\phi} \frac{\eta_k}{\eta_k^2 - e_\alpha} c_k^\dagger c_{-k}^\dagger \right) |\nu\rangle \quad (\text{any } G), \qquad (18)$$

where, similar to Eq. (12), the spectral parameters $e_\alpha$ are determined by solving a set of nonlinear equations.

While the RGK Hamiltonian (15) conserves the total number of particles in the system, its mean-field approximation involves pairing terms of the form $\Delta cc$ and $\Delta^* c^\dagger c^\dagger$ and only conserves the fermion parity (fermion number modulo 2). The superconducting correlations do not couple opposite-parity states and thus the latter's energy levels are allowed to cross under symmetry-preserving deformations of the Hamiltonian. The ground-state fermion parity acts as a topological invariant, and its switch signaled by a level crossing is a topological transition [51]. For instance, one could also evolve a Kitaev chain in ring geometry from periodic to antiperiodic boundary conditions by adiabatically changing the hopping parameter for just one of the chain's bonds $t \to t(1 - 2\lambda)$, $\lambda \in [0, 1]$ [52]. If the ground-state fermion parity is the same/different at $\lambda = 0$ (periodic) than at $\lambda = 1$ (antiperiodic), then an even/odd number of level crossings has occurred during this process. If the number of level crossings is odd, then one of them always occurs at $\lambda = 1/2$; here one the bonds is effectively cut ($t = 0$) and the ring becomes an open chain. In this case two unpaired Majorana zero modes are found at the chain's boundaries, and the system is said to be in the topologically nontrivial phase. For an overview of the ongoing search for Majorana zero modes in superconductors, see Ref. [53].

The criterion introduced in Ref. [37] for establishing the emergence of topological superfluidity in the beyond mean-field (particle-number conserving) description of many-fermion systems uses the ground-state energies for even $N = 2M$ and odd $N = 2M \pm 1$ number of particles (for both periodic and antiperiodic boundary conditions) to identify the relevant fermion parity switches. To do so, one computes the ground-state energy $E_{\mathrm{gs}}^\phi(N)$ for a number $N$ of fermions in the chain of $L$ sites, imposing periodic ($\phi = 0$) or antiperiodic ($\phi = 2\pi$) boundary conditions, and compares

$$E_{\mathrm{odd}}^\phi = \left[E_{\mathrm{gs}}^\phi(N+1) + E_{\mathrm{gs}}^\phi(N-1)\right]/2 \text{ and}$$
$$E_{\mathrm{even}}^\phi = E_{\mathrm{gs}}^\phi(N), \qquad (19)$$

assuming $N$ is even. Their difference determines the fermion parity, which has opposite sign at $\phi = 0$ and $\phi = 2\pi$ in the topologically nontrivial phase.

### B. RGK chain: RBM results

To benchmark the RBM capabilities in emulating the RGK chain across its various phases, we follow Ref. [37] and take $t_1 = 1$ as the unit of energy and set $t_2 = 0$; furthermore, we consider the chain at quarter filling $N/L = 1/4$. For this specific value, the critical reduced strength

$$g = GL/2 \qquad (20)$$

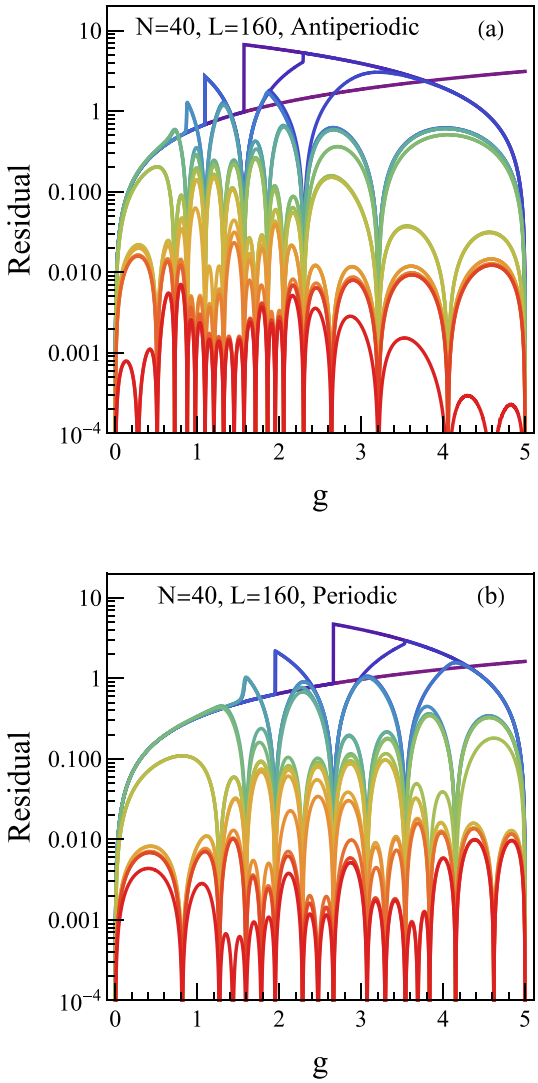


FIG. 8. Evolution of the residual (5), in units of $t_1$, during the offline emulation phase of the RGK Hamiltonian (15) for a chain of $L = 160$ at quarter filling with antiperiodic (a) and periodic boundary conditions (b). The sampling interval for the reduced strength $g$ (20) is $0 \leqslant g \leqslant 5$, in units of $t_1$.

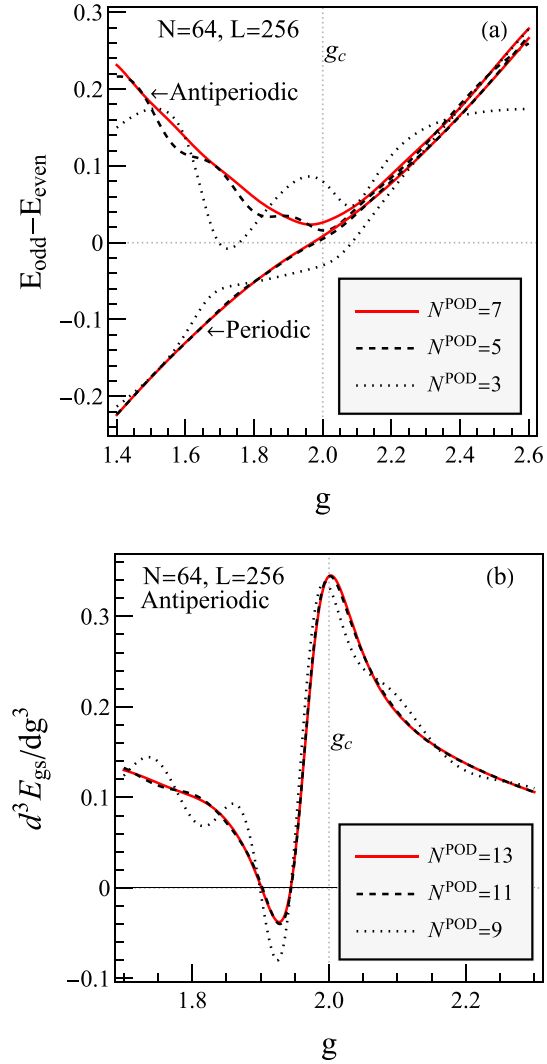


FIG. 9. Even-odd energy differences (19), (a), and ground-state energy third derivative, (b), in units of $t_1$, for various reduced bases of dimensionality $N^{POD}$ obtained upon POD compression for a RGK chain of $L = 256$ sites at quarter filling. The thermodynamic limit critical strength $g_c = 2$ is evidenced. The sampling interval for the reduced strength $g$ (20) is $1.4 \leqslant g \leqslant 2.6$, in units of $t_1$.

corresponding to the third-order topological phase transition (in the thermodynamic limit) is $g_c = 2$, the topologically nontrivial state being entered for $g < g_c$.

We present in Fig. 8 a typical evolution of the residual profile during the sampling of the interval $0 \leqslant g \leqslant 5$ using the greedy algorithm described in Sec. II, with an emulation accuracy threshold of 0.01. We observe how the ground-state evolution with increasing $g$ is strongly dependent on the boundary conditions: in the antiperiodic case the range of greatest variation (with the highest density of sampling points) is observed to be $1 < g < 2$ preceding the phase transition, while in the periodic case there is a pronounced variation on both sides of the phase transition point, more or less within $1 < g < 4$.

In the following, we restrict our attention to the interval $1.4 < g < 2.6$ and show in Fig. 9 the signatures of (the precursor to) the topological phase transition in a system of $N = 64$ particles on $L = 256$ sites. The RBM emulator is able to account well for both the gross features of the phase transition (top panel) and for the finer details such as correctly capturing the order of the phase transition (lower panel, note that in the case of a finite system the discontinuity in the third-order derivative is naturally smoothed out).

To reach the 0.01 emulation accuracy threshold (as given by the residual's maximum) for each case with $N = 63, 64, 65$ and periodic/antiperiodic boundary conditions necessary to obtain Fig. 9, no more than 13 sampling points are required in the interval $1.4 < g < 2.6$ during the offline step. In Fig. 9 we also confirm that the gross features of the phase transition are still present when performing a moderate POD compression (of the final reduced basis obtained in the offline step), while the more sensitive energy third derivative requires an almost lossless description to be faithfully reproduced.

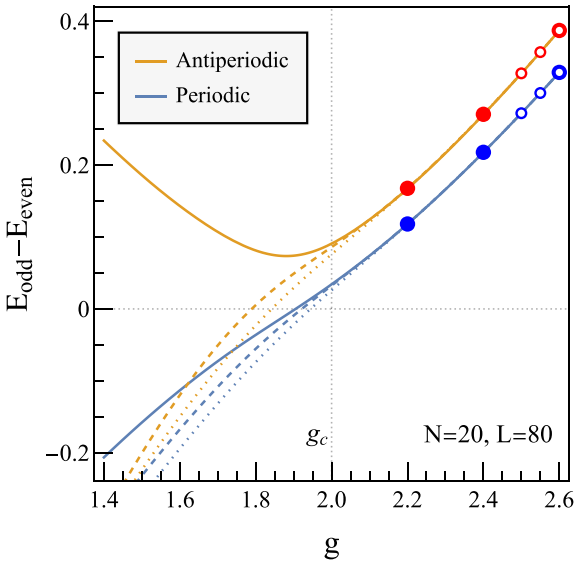


FIG. 10. Extrapolated (dotted curve for open circle sampling points and dashed curve for filled dot sampling points) even-odd energy difference (19), in units of $t_1$, for a RGK chain on $L = 80$ sites at quarter filling. The thermodynamic limit critical strength $g_c = 2$ is evidenced. The sampling interval for the reduced strength $g$ (20) is $1.4 \leqslant g \leqslant 2.6$, in units of $t_1$.

In concluding this section, we remark that an emulator trained only within the topologically trivial phase cannot extrapolate well into the topologically nontrivial phase even for a relatively small system, as displayed in Fig. 10. This is contrary to one of the earliest applications of Eigenvector Continuation in Ref. [7] where the extrapolation from the dilute gas regime of a 3D Hubbard model was possible well into the clusterized regime (for a small four-particle system). In our specific topological transition, the failure of the extrapolation may be diagnosed by the large concentration of sampling points in Fig. 8, especially for the case of antiperiodic boundary conditions, indicating a rapid and significant change in the ground-state structure in the phase transition region.

We leave a more detailed study of the RGK chain, within the DMRG + RBM approach, away from exact solvability and in various $t_2/t_1$ regimes to a future work, and investigate in the next section another variation on the Richardson pairing model particularly relevant to the state-of-the-art technological applications of condensed matter phenomena.

## V. EMULATING SUPER-SEMI HYBRID DEVICES

### A. Mesoscopic superconductor-semiconductor hybrid devices

As a second generalization of the Richardson Hamiltonian, we consider in this section its coupling to an impurity level: this introduces pair-breaking processes and breaks integrability, but nevertheless the resulting model may still be accurately solved using DMRG [32]. The physical situation corresponds to a quantum dot (QD) coupled to a small superconducting island (SIs) into a strongly correlated hybrid device where the three-way competition of pairing correlations, Coulomb blockade and Kondo screening results in a richness of emergent phenomena (see, e.g., Refs. [32,54] and references therein). In particular, for a small mesoscopic superconducting island with a considerable charging energy and strong even-odd occupancy effects, the subgap states have properties quite unlike those of the standard Yu-Shiba-Rusinov (YSR) states [35] formed inside a macroscopic superconductor's spectral gap by binding a Bogoliubov quasiparticle at the impurity [55–58].

While the theoretical results are in excellent qualitative agreement with experiment regarding the many-body properties of such devices [34–36], their efficient modeling is limited by the large number of independent parameters, i.e., SI and QD charging energies and gate-defined occupancies, QD-SI tunnel and capacitive coupling strengths, QD and SI effective $g$-factors or spin-orbit coupling strengths. With the only studies currently available having been carried out for the simplest QD-SI and SI-QD-SI configurations, the efficient modeling of more realistic systems involving multiple interconnected SIs and QDs could be enabled by the RBM-based emulation.

As a proof-of-concept application of the RBM to accelerate the theoretical modeling of this class of hybrid devices, we consider in this section the QD-SI system of Refs. [32,35]. The total Hamiltonian is obtained by combining the Anderson model for the interacting quantum dot and the Richardson model for the small superconducting island into

$$H = H_{\rm QD} + H_{\rm SI} + H_{\rm hyb} + H_{\rm cc}, \tag{21}$$

where the QD, SI, hybridization, and capacitive coupling terms are

$$\begin{aligned} H_{\rm QD} &= \epsilon_{\rm QD}\, N_{\rm QD} + U\, N_{{\rm QD},\uparrow}\, N_{{\rm QD},\downarrow} \\ &= (U/2)(N_{\rm QD} - \nu)^2 + \text{const}, \\ H_{\rm SI} &= H_{\rm Richardson} + E_c (N_{\rm SI} - n_0)^2, \\ H_{\rm hyb} &= \frac{t}{\sqrt{L}} \sum_{i=1}^{L} \sum_{\sigma=\uparrow,\downarrow} (c_{i\sigma}^\dagger d_\sigma + \text{H.c.}), \\ H_{\rm cc} &= U_{\rm cc} (N_{\rm QD} - \nu)(N_{\rm SI} - n_0). \end{aligned} \tag{22}$$

For the quantum dot, $N_{{\rm QD},\sigma} = d_\sigma^\dagger d_\sigma$ is the number operator for each spin projection $\sigma = \uparrow, \downarrow$, and $N_{\rm QD} = N_{{\rm QD},\uparrow} + N_{{\rm QD},\downarrow}$. The gate-defined optimal QD occupation $\nu$ is related to its single particle energy $\epsilon_{\rm QD}$ and electron-electron repulsion strength $U$ by $\nu = 1/2 - \epsilon_{\rm QD}/U$. The SI pairing correlations are modeled by the Richardson Hamiltonian of Eq. (7) which involves here $L = 50$ doubly degenerate levels spaced by $d = 2D/L$, where $D$ is the half bandwidth (in this section we follow Ref. [32] and take $D = 1$ as the unit of energy). We tune the pairing strength to $G = 0.4d$ which gives rise to the superconducting gap $\Delta = 0.166$ and ensures that an appropriate number of levels participate in the pairing interaction, thus minimizing the finite-size effects [33]. The SI Coulomb repulsion effects are taken into account by the second term in $H_{\rm SI}$ written in terms of its charging energy $E_c$ and its gate-defined optimal occupation $n_0$. We fix the QD-SI tunneling amplitude $t$ by choosing an intermediately strong value for the hybridization strength $\Gamma = \pi \rho t^2 = 0.4\Delta$, where $\rho = 1/2D$ is the SI normal state density of states.

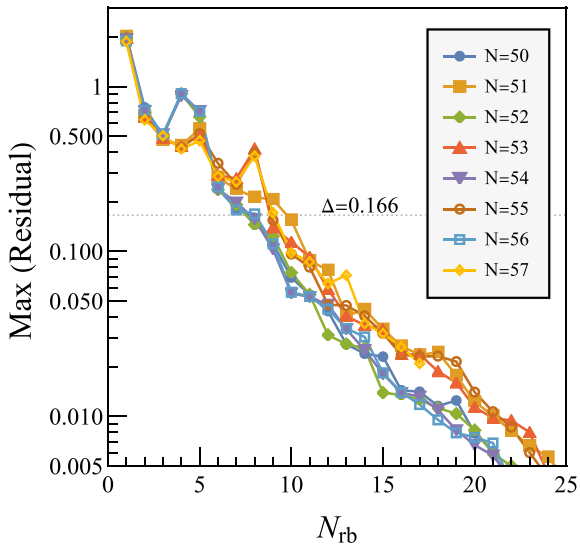


FIG. 11. Evolution of the residual's maximum, in units of $D$, versus the reduced-basis dimension $N_{rb}$ during the offline emulation phase of the QD-SI Hamiltonian (21) for various particle numbers $N$. See main text below Eqs. (22) and (24) for information on the used parameters and sampling regions.

### B. QD-SI emulation: RBM results

In this application, we take the control parameters for the RBM emulation to be $U$, $\nu$, $E_c$, $n_0$, and $U_{cc}$. By using the particle number conservation $[H, N] = 0$ with $N = N_{QD} + N_{SI}$, we may restrict our attention to each particle number sector individually and absorb the effects of the SI's charging energy into $H_{QD}$ by a redefinition the QD parameters. We thus work with the effective Hamiltonian

$$H(\xi_1, \xi_2) = H_0 + \xi_1 N_{QD} + \xi_2 N_{QD}^2, \tag{23}$$

where $H_0 = H_{Richardson} + H_{hyb}$ and

$$\begin{aligned} \xi_1 &= (U_{cc} - U)\,\nu - (2E_c + U_{cc})(N - n_0), \\ \xi_2 &= U/2 + E_c - U_{cc}, \end{aligned} \tag{24}$$

reducing the number of independent control parameters to two. In the following, we restrict our analysis to the parameter domain $(\xi_1, \xi_2) \in [-5.8, 11.6] \times [0, 2.5]$ which is sufficient to cover the physical domain $(n_0, \nu) \in [49.5, 51.5] \times [-0.5, 2.5]$ up to moderately large values of $U$, $E_c$, and $U_{cc}$ for all relevant particle numbers $N \in \{50, 51, ..., 57\}$.

We show in Fig. 11 the residual's maximum exponential decrease during the greedy learning algorithm which confirms the efficient RBM representation of the QD-SI parameter space. The learning rates are almost identical for the range of system sizes considered, which is consistent with the reduced finite size effects. Additionally, we observe that the emulation for odd-parity states requires a slightly larger number of sample points than for even-parity states to reach the same accuracy (as quantified by the residual's maximum).

Remarkably, a very small basis dimensionality (corresponding to a modest emulation accuracy of the order of the superconducting gap) is sufficient for reaching convergence in emulating the QD-SI charge stability diagram, as seen in Figs. 12 and 13. The evolution of the ground-state parity transition lines (by "ground state," we mean the lowest energy state over all particle number sectors) is shown in Fig. 12 for varying SI charging energy $E_c$, and in Fig. 13 for varying capacitive coupling strength $U_{cc}$. In the former case, our emulated results agree with those presented in Fig. 1 of Ref. [32], regarding the evolution from the YSR regime (with $2e$ periodicity along the $n_0$ axis) to the Coulomb blockaded regime (with $1e$ periodicity). In the latter case, the phase boundaries acquire an angle due to the capacitive coupling influencing the occupancy in both parts of the system, in agreement with the findings in Refs. [35,59].

The occupancies constitute another sensitive metric for the convergence of the reduced-basis emulation. The $\nu$ dependence of the emulated average quantum dot occupancy (for even $n_0$) is shown in Fig. 14 to be in agreement with Fig. 5(a) of Ref. [32]. The RBM description of the occupancy achieves convergence for a reduced-basis dimensionality similar to that of the energy emulation. The RBM curves accurately capture the rapid variation of the lowest-energy (subgap) excited states around half filling, which is directly related to an equally rapid variation of the excitation spectrum [32], as also shown in Fig. 14. The latter agrees qualitatively with Fig. 4(d) of Ref. [32], with the only quantitative difference being that our excitation energies deviate slightly from the values $\pm(1 + E_c/\Delta)$ at small/large values of $\nu$, due to finite size effects.

## VI. CONCLUSIONS

This work provides the first indications on the RBM capability to capture efficiently the ground-state correlations

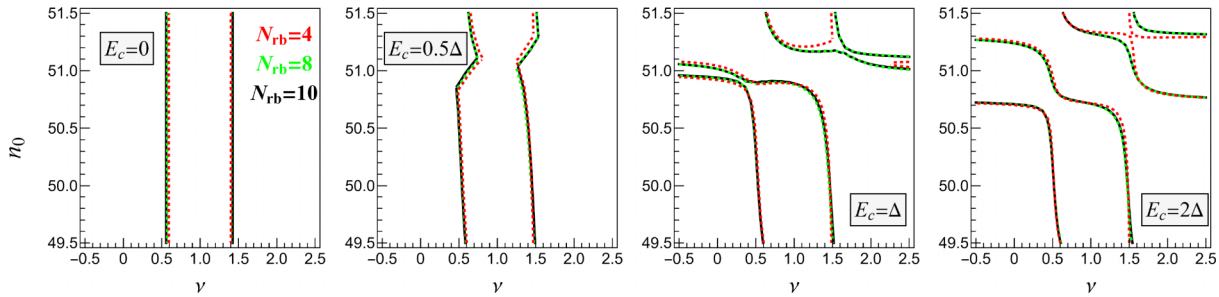


FIG. 12. QD-SI ground-state parity transition curves as a function of gate voltages applied to the QD ($\nu$) and to the SI ($n_0$) for varying SI charging energy $E_c$ and for different reduced-basis dimensionalities $N_{rb}$. We consider fixed $U = 4\Delta$, $U_{cc} = 0$. See main text below Eqs. (22) and (24) for information on other parameters and sampling regions.

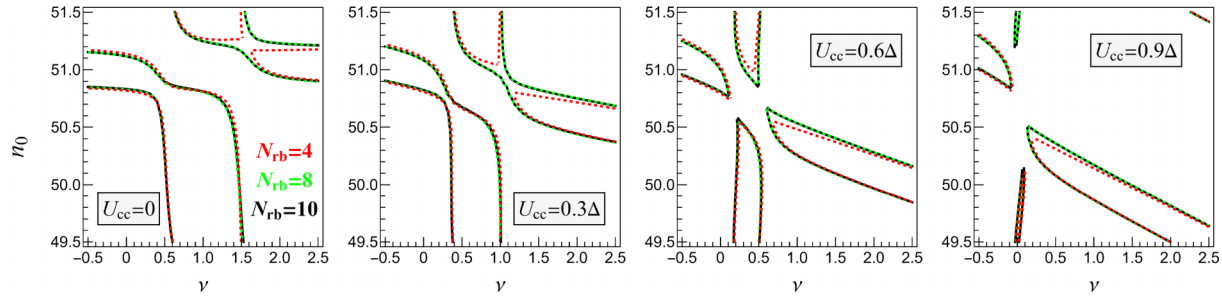


FIG. 13. QD-SI ground-state parity transition curves as a function of gate voltages applied to the QD ($\nu$) and to the SI ($n_0$) for varying QD-SI capacitive coupling strength $U_{cc}$ and for different reduced-basis dimensionalities $N_{rb}$. We consider fixed $U = 4\Delta$, $E_c = 1.3\Delta$. See main text below Eqs. (22) and (24) for information on other parameters and sampling regions.

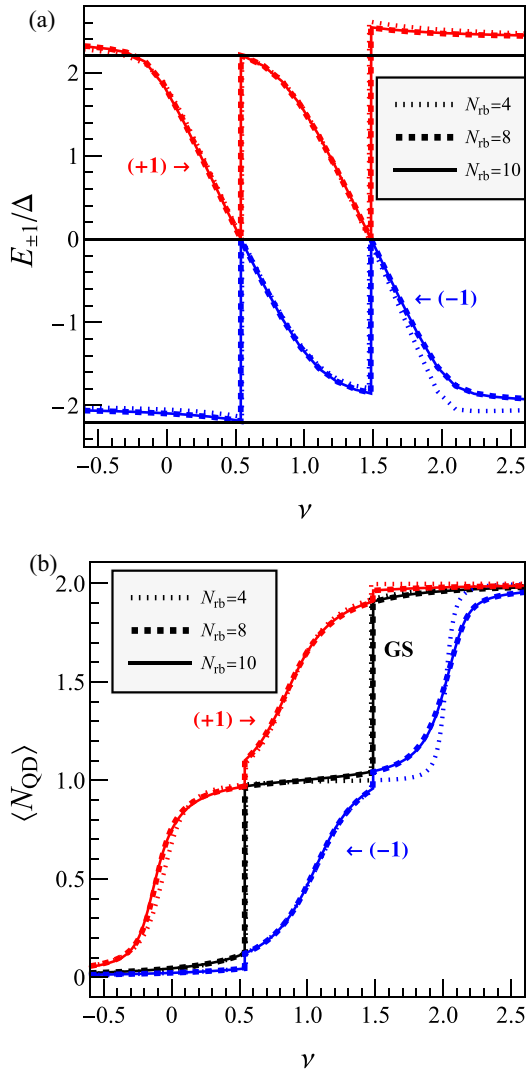


FIG. 14. Excitation energies $E_{\pm1}$ (a) and quantum dot average occupations $\langle N_{QD}\rangle$ (b) as a function of the gate voltage applied to the QD ($\nu$) for the ground state (GS) and the lowest-energy (subgap) excited particle-like and hole-like states ($\pm1$), for even $n_0 = 52$ and for different reduced-basis dimensionalities $N_{rb}$. The solid horizontal black lines in the top panel are situated at 0 and $\pm(1 + E_c/\Delta)$. We consider fixed $U = 4\Delta$, $E_c = 1.2\Delta$. See main text below Eqs. (22) and (24) for information on the used parameters and sampling regions.

induced by the pairing force across entire parameter spaces of various interacting many-body systems.

Very small bases were confirmed to accurately describe the weak-to-strong pairing crossover in the Richardson model (in Sec. III), the third-order topological phase transition of the interacting Richardson-Kitaev chain (in Sec. IV, and the charge stability diagram of a hybrid quantum dot–superconductor device (in Sec. V). In all cases the efficient sampling of the relevant parameter space was confirmed by the exponential decay of the residual (5) during the construction of the reduced basis.

On the one hand, the residual is a direct measure for the emulation error; its magnitude is given by a combination of the error in the energy and the error in the ground-state eigenvector itself. Thus, it is typically larger than each of the two individual errors, which in practice ensures that the actual energy error will be much smaller than the chosen residual threshold for the reduced-basis construction, e.g., compare Figs. 1 and 6(b) and see also Ref. [38].

On the other hand, the distribution of sampling points generated by the greedy residual optimization strategy may provide qualitative insight into the evolution of the ground-state structure across the parameter domain. A higher density of sampling points implies a rapid change in the eigenstate structure, specific to crossover phenomena, see, e.g., the discussion around Fig. 1, or quantum phase transitions, see, e.g., the discussion around Fig. 8 and also Ref. [30].

More generally, beyond investigating phase diagrams as outlined in Ref. [30], the RBM could actually enable the otherwise expensive but desirable microscopic models to better interface with experimental design and data analysis. For example, the precise location of possible qu*d*it operational sweet-spots (regions with reduced noise sensitivity; see also Ref. [33]) for more realistic systems involving multiple interconnected SIs and QDs could be precisely determined by the fast and accurate RBM scans of their enlarged parameter spaces, impractical only with direct DMRG solvers. For such high-dimensional parameter space problems, several approaches have been devised for addressing the issue of optimal training set sampling. We mention here the Multi-Stage Greedy algorithm of Ref. [60] and the adaptively enriching greedy algorithm of Ref. [61], and refer the reader to the Introduction of Ref. [62] for a thorough review of existing methods.

Conversely, fitting the model parameters to the experimental data, as in Refs. [34,35] for this class of systems, could

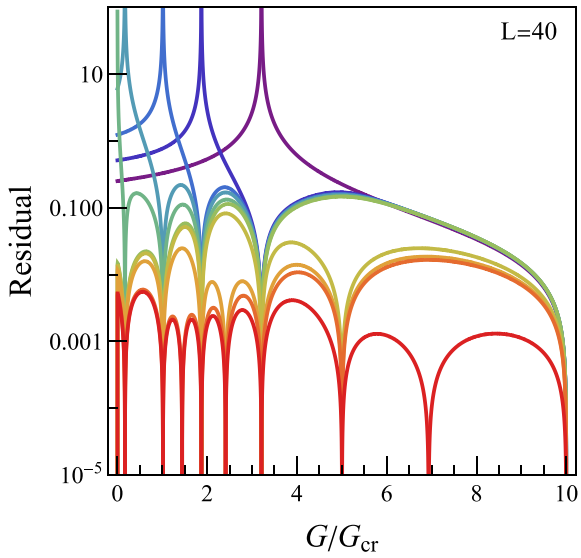


FIG. 15. Evolution of the residual (A1), in units of the level spacing $\epsilon$, during the offline emulation phase of the pairing Hamiltonian (7) for a system of $N_p = 20$ pairs distributed over $L = 40$ levels. The sampling interval for the pairing strength $G$ is $0 \leqslant G \leqslant 10G_{\rm cr}$, with the corresponding $G_{\rm cr} \simeq 0.22\epsilon$ for $L = 2N_p = 40$. The color spectrum (violet, blue, green, orange, red) correlates with the increasing iteration number of the greedy self-learning algorithm.

become effortless once a cheap RBM effective representation is found. Perhaps a more challenging novel RBM application would consist of building emulators able to accelerate transport simulations (e.g., those of Ref. [63]) upon adapting the RBM treatment of time-dependent differential equations [64]. This would be a valuable tool in mitigating the rise in computational complexity when extending the transport simulation to finite temperature, where more demanding density matrix or thermofield descriptions are necessary (see Ref. [63] and references therein).

## ACKNOWLEDGMENTS

We thank Jens Paaske, Max Geier, Michele Burrello, Matteo Rizzi, Andrea Maiani, Ben Joecker, Luka Pavešić, Rok Žitko, Jorge Dukelsky, and Alexandru Nemnes for useful discussions and suggestions. V.V.B. acknowledges many fruitful research interactions at the Condensed Matter Theory Group and QDev (Niels Bohr Institute, Copenhagen University). This work was supported by the grant of the Romanian Ministry of Education and Research, CNCS-UEFISCDI, Project No. PN-III-P4-ID-PCE-2020-1142, within PNCDI III.

## APPENDIX

In this Appendix, we consider a dynamic renormalization of the residual designed to assess the emulation accuracy on the relevant energy scale for each region of the parameter space.

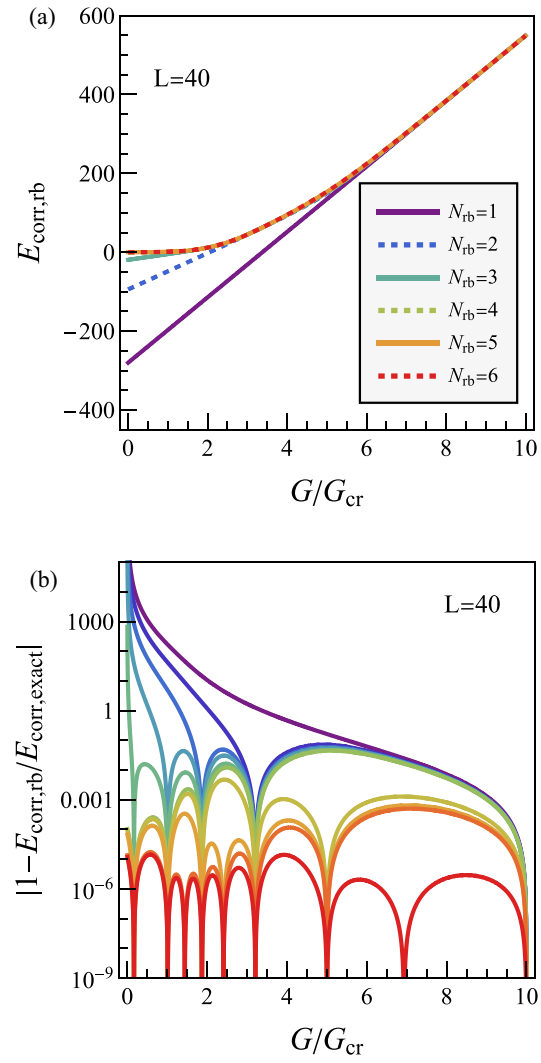


FIG. 16. Evolution of the reduced-basis correlation energy $E_{\rm corr}$ (13) (a), in units of the level spacing $\epsilon$, and the corresponding error relative to its exact value (b) during the offline emulation phase. The sampling interval for the pairing strength $G$ is $0 \leqslant G \leqslant 10G_{\rm cr}$, with the corresponding $G_{\rm cr} \simeq 0.22\epsilon$ for the chosen system size $L = 2N_p = 40$.

Due to the residual defined in Eq. (5) scaling as an energy, the greedy algorithm based on it may preferentially sample the strong-coupling regime [large $|f_p(\boldsymbol{\xi})|$ in Eq. (1)] to reach the imposed global energy-based threshold for the emulation accuracy. This over-sampling may be observed for instance in Figs. 6 and 7, where the errors of the correlation energy and of the canonical gap (relative to their exact values) are seen to decay exponentially into the strong-pairing (large-$G$) regime of the Richardson pairing Hamiltonian (7), for a $0.01\epsilon$ global emulation accuracy threshold.

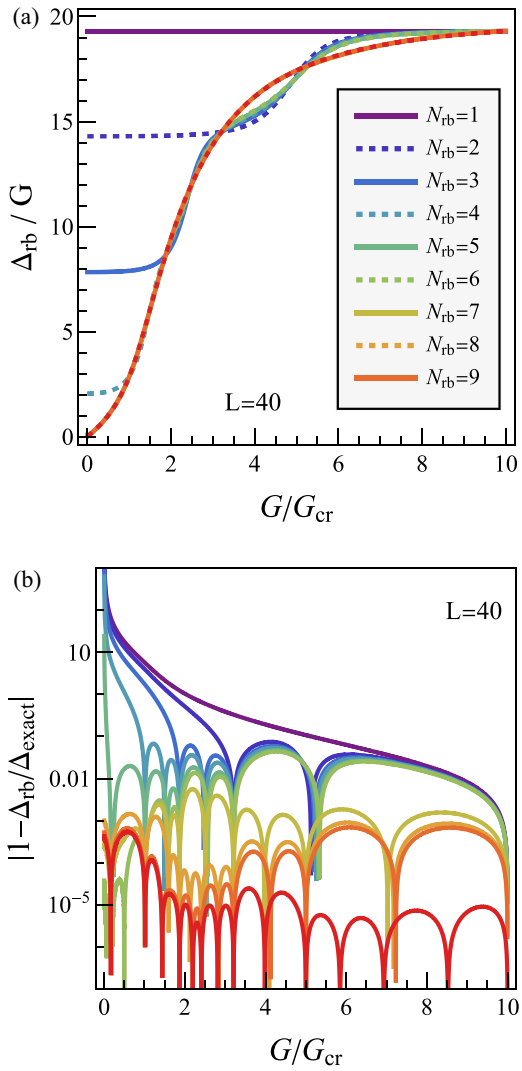


FIG. 17. Evolution of the reduced-basis canonical gap $\Delta$ (14) (a), and the corresponding error relative to its exact value, (b), during the offline emulation phase. The sampling interval for the pairing strength $G$ is $0 \leqslant G \leqslant 10G_{cr}$, with the corresponding $G_{cr} \simeq 0.22\epsilon$ for the chosen system size $L = 2N_p = 40$.

We could allow instead for the emulation accuracy to be measured locally against the relevant energy scale for each region of the parameter space, thus imposing a global threshold for the relative emulation accuracy. This could avoid the previously mentioned over-sampling issue, thus potentially providing an even more efficient reduced-basis construction methodology.

To illustrate these concepts with a specific example, we consider once again the Richardson pairing Hamiltonian of Eq. (7) with the pairing strength $G$ as the control parameter, $\xi = G$. As the renormalized version of the residual, we work in this Appendix with the definition

$$\mathrm{Res}(G) \equiv \frac{\|\mathcal{H}(G)|\psi^{(\mathrm{rb})}\rangle - E^{(\mathrm{rb})}|\psi^{(\mathrm{rb})}\rangle\|}{|E^{(G=0)} - E^{(\mathrm{rb})}|}, \tag{A1}$$

with all reduced-basis (rb) quantities having an implicit $G$-dependence. As a natural local energy scale, we are thus considering the gain in energy obtained when turning on the pairing interaction, $E^{(G=0)} - E(G)$. This still needs to be evaluated in the reduced-basis approximation for the error estimation procedure to remain efficient.

The evolution of the renormalized residual profile during the greedy sampling algorithm is shown in Fig. 15, for the particular choice of $G = 10G_{cr}$ as a first sampling point (the $G = 0$ Hartree-Fock state leads to an inconclusive constant renormalized residual at the first iteration). As seen in Fig. 15, our chosen 0.01 relative emulation accuracy threshold is achieved with only 10 sampling points. During the first few iterations (until the $G = 0$ HF state is included in the basis), new sampling points are selected by the vanishing of the denominator in Eq. (A1). This signals where the reduced-basis energy approximation becomes worse than its uncorrelated value, thus prompting for a new exact evaluation.

Correspondingly, in Fig. 16 the correlation energy of Eq. (13), $E_{\mathrm{corr}}(G) \equiv E_{\mathrm{HF}}(G) - E(G)$, is shown to become negative during the first iterations, with its zeros closely related to the residual's peaks in Fig. 15. Due to the vanishing of the exact correlation energy at $G = 0$, the corresponding relative error diverges here until the Hartree-Fock state is included in the basis (which is needed to ensure a sufficiently accurate description of the weak pairing regime).

Remarkably, for both the correlation energy and for the canonical gap of Eq. (14), the local maxima in the final relative errors exhibit only only small variations across the entire $G$-interval, as seen in Figs. 16 and 17 [compare with Figs. 6 and 7 obtained using the un-normalized residual of Eq. (5)]. This confirms that the definition (A1) is able to address the previously discussed strong-coupling over-sampling issue and thus to provide a sensible emulation accuracy measure for more efficient sampling strategies.

We finally note that the choice of Eq. (A1) for the residual normalization is not unique. We obtained a qualitatively similar residual profile evolution with that of Fig. 15 (same sampling points, with smoothed residual peaks during the first iterations) by adapting the prescription of Ref. [38] and using the normalization factor $\langle\psi^{(\mathrm{rb})}|[E^{(G=0)} - \mathcal{H}]^2|\psi^{(\mathrm{rb})}\rangle^{1/2}$.

[1] J. von Delft and D. C. Ralph, Spectroscopy of discrete energy levels in ultrasmall metallic grains, Phys. Rep. **345**, 61 (2001).
[2] J. Dukelsky, S. Pittel, and G. Sierra, Colloquium: Exactly solvable Richardson-Gaudin models for many-body quantum systems, Rev. Mod. Phys. **76**, 643 (2004).
[3] R. A. Broglia and V. Zelevinsky, *Fifty Years of Nuclear BCS* (World Scientific, Singapore, 2012).
[4] 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).
[5] J. J. Thiagarajan, B. Venkatesh, R. Anirudh, P.-T. Bremer, J. Gaffney, G. Anderson, and B. Spears, Designing accurate emulators for scientific processes using calibration-driven deep models, Nat. Commun. **11**, 5622 (2020).

[6] M. F. Kasim, D. Watson-Parris, L. Deaconu, S. Oliver, P. Hatfield, D. H. Froula, G. Gregori, M. Jarvis, S. Khatiwala, J. Korenaga, J. Topp-Mugglestone, E. Viezzer, and S. M. Vinko, Building high accuracy emulators for scientific simulations with deep neural architecture search, Mach. Learn.: Sci. Technol. **3**, 015013 (2022).

[7] D. Frame, R. He, I. Ipsen, D. Lee, D. Lee, and E. Rrapaj, Eigenvector Continuation with Subspace Learning, Phys. Rev. Lett. **121**, 032501 (2018).

[8] R. J. Furnstahl, A. J. Garcia, P. J. Millican, and X. Zhang, Efficient emulators for scattering using eigenvector continuation, Phys. Lett. B **809**, 135719 (2020).

[9] J. A. Melendez, C. Drischler, A. J. Garcia, R. J. Furnstahl, and X. Zhang, Fast and accurate emulation of two-body scattering observables without wave functions, Phys. Lett. B **821**, 136608 (2021).

[10] C. Drischler, M. Quinonez, P. G. Giuliani, A. E. Lovell, and F. M. Nunes, Toward emulating nuclear reactions using eigenvector continuation, Phys. Lett. B **823**, 136777 (2021).

[11] D. Bai and Z. Ren, Generalizing the calculable *R*-matrix theory and eigenvector continuation to the incoming-wave boundary condition, Phys. Rev. C **103**, 014612 (2021).

[12] D. Bai, New extensions of eigenvector continuation *R*-matrix theory based on analyticity in momentum and angular momentum, Phys. Rev. C **106**, 024611 (2022).

[13] X. Zhang and R. J. Furnstahl, Fast emulation of quantum three-body scattering, Phys. Rev. C **105**, 064004 (2022).

[14] N. Yapa and S. König, Volume extrapolation via eigenvector continuation, Phys. Rev. C **106**, 014309 (2022).

[15] S. Yoshida and N. Shimizu, Constructing approximate shell-model wave functions by eigenvector continuation, Prog. Theor. Exp. Phys. **2022**, 053D02 (2022).

[16] P. Demol, T. Duguet, A. Ekstrom, M. Frosini, K. Hebeler, S. Konig, D. Lee, A. Schwenk, V. Somà, and A. Tichai, Improved many-body expansions from eigenvector continuation, Phys. Rev. C **101**, 041302(R) (2020).

[17] P. Demol, M. Frosini, A. Tichai, V. Somà, and T. Duguet, Bogoliubov many-body perturbation theory under constraint, Ann. Phys. **424**, 168358 (2021).

[18] M. Companys Franzke, A. Tichai, K. Hebeler, and A. Schwenk, Excited states from eigenvector continuation: The anharmonic oscillator, Phys. Lett. B **830**, 137101 (2022).

[19] A. Ekström, and G. Hagen, Global Sensitivity Analysis of Bulk Properties of an Atomic Nucleus, Phys. Rev. Lett. **123**, 252501 (2019).

[20] S. König, A. Ekström, K. Hebeler, D. Lee, and A. Schwenk, Eigenvector continuation as an efficient and accurate emulator for uncertainty quantification, Phys. Lett. B **810**, 135814 (2020).

[21] T. Djärv, A. Ekström, C. Forssén, and H. T. Johansson, Bayesian predictions for $A = 6$ nuclei using eigenvector continuation emulators, Phys. Rev. C **105**, 014005 (2022).

[22] A. L. Anderson, G. L. O'Donnell, and J. Piekarewicz, Applications of reduced-basis methods to the nuclear single-particle spectrum, Phys. Rev. C **106**, L031302 (2022).

[23] E. Bonilla, P. Giuliani, K. Godbey, and D. Lee, Training and projecting: A reduced basis method emulator for many-body physics, Phys. Rev. C **106**, 054322 (2022).

[24] J. A. Melendez, C. Drischler, R. J. Furnstahl, A. J. Garcia, and X. Zhang, Model reduction methods for nuclear emulators, J. Phys. G: Nucl. Part. Phys. **49**, 102001 (2022).

[25] B. O. Almroth, P. Stern, and F. A. Brogan, Automatic choice of global shape functions in structural analysis, AIAA J. **16**, 525 (1978).

[26] A. Quarteroni, A. Manzoni, and F. Negri, *Reduced Basis Methods for Partial Differential Equations* (Springer International Publishing, Cham, Switzerland).

[27] J. S. Hesthaven, G. Rozza, and B. Stamm, *Certified Reduced Basis Methods for Parametrized Partial Differential Equations* (Springer International Publishing, Cham, Switzerland).

[28] *Reduced Order Methods for Modeling and Computational Reduction* (Springer International Publishing, Cham, Switzerland).

[29] S. L. Brunton and J. N. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control* (Cambridge University Press, Cambridge, UK, 2019).

[30] M. F. Herbst, B. Stamm, S. Wessel, and M. Rizzi, Surrogate models for quantum spin systems based on reduced-order modeling, Phys. Rev. E **105**, 045303 (2022).

[31] J. Dukelsky and G. Sierra, Density Matrix Renormalization Group Study of Ultrasmall Superconducting Grains, Phys. Rev. Lett. **83**, 172 (1999).

[32] L. Pavešić, D. Bauernfeind, and R. Žitko, Subgap states in superconducting islands, Phys. Rev. B **104**, L241409 (2021).

[33] L. Pavešić and R. Žitko, Qubit based on spin-singlet Yu-Shiba-Rusinov states, Phys. Rev. B **105**, 075129 (2022).

[34] J. C. E. Saldaña, A. Vekris, L. Pavešič, R. Žitko, K. Grove-Rasmussen, and J. Nygård, Two Bogoliubov quasiparticles entangled by a spin, arXiv:2203.00104 [cond-mat.mes-hall].

[35] J. C. Estrada Saldaña, A. Vekris, L. Pavešić, P. Krogstrup, R. Žitko, K. Grove-Rasmussen, and J. Nygård, Excitations in a superconducting Coulombic energy gap, Nat. Commun. **13**, 2243 (2022).

[36] F. K. Malinowski, R. K. Rupesh, L. Pavešić, Z. Guba, D. de Jong, L. Han, C. G. Prosko, M. Chan, Y. Liu, P. Krogstrup, A. Pályi, R. Žitko, and J. V. Koski, Quantum capacitance of a superconducting subgap state in an electrostatically floating dot-island, arXiv:2210.01519 [cond-mat.mes-hall].

[37] G. Ortiz, J. Dukelsky, E. Cobanera, C. Esebbag, and C. Beenakker, Many-Body Characterization of Particle-Conserving Topological Superfluids, Phys. Rev. Lett. **113**, 267002 (2014).

[38] A. Sarkar and D. Lee, Self-learning emulators and eigenvector continuation, Phys. Rev. Res. **4**, 023214 (2022).

[39] M. Barrault, Y. Maday, N. C. Nguyen, and A. T. Patera, An "empirical interpolation" method: Application to efficient reduced-basis discretization of partial differential equations, C. R. Math. **339**, 667 (2004).

[40] M. A. Grepl, Y. Maday, N. C. Nguyen, and A. T. Patera, Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations, ESAIM: M2AN **41**, 575 (2007).

[41] S. R. White, Density Matrix Formulation for Quantum Renormalization Groups, Phys. Rev. Lett. **69**, 2863 (1992).

[42] U. Schollwöck, The density-matrix renormalization group in the age of matrix product states, Ann. Phys. **326**, 96 (2011).

[43] M. Fishman, S. R. White, and E. M. Stoudenmire, The ITensor software library for tensor network calculations, SciPost Phys. Codebases, 4 (2022).

[44] M. Fishman, S. R. White, and E. M. Stoudenmire, Codebase release 0.3 for ITensor, SciPost Phys. Codebases, 4 (2022).

[45] V. V. Baran and J. Dukelsky, Variational theory combining number-projected BCS and coupled-cluster doubles, Phys. Rev. C **103**, 054317 (2021).

[46] P. Ring and P. Schuck, *The Nuclear Many-Body Problem* (Springer, Berlin, 1980).

[47] R. Žitko and L. Pavešić, Yu-Shiba-Rusinov states, BCS-BEC crossover, and exact solution in the flat-band limit, Phys. Rev. B **106**, 024513 (2022).

[48] R. W. Richardson, Numerical study of the 8-32-particle eigenstates of the pairing Hamiltonian, Phys. Rev. **141**, 949 (1966).

[49] Y. Lin and A. J. Leggett, Some questions concerning Majorana fermions in 2D ( $p + ip$) Fermi superfluids, Quant. Front. **1**, 4 (2022).

[50] G. Ortiz, R. Somma, J. Dukelsky, and S. Rombouts, Exactly solvable models derived from a generalized Gaudin algebra, Nucl. Phys. B **707**, 421 (2005).

[51] A. Y. Kitaev, Unpaired Majorana fermions in quantum, Phys.-Usp. **44**, 131 (2001).

[52] A. Akhmerov, Topology in condensed matter, https://ocw.tudelft.nl/course-readings/bulk-topological-invariant-bulk-edge-correspondence/.

[53] K. Flensberg, F. von Oppen, and A. Stern, Engineered platforms for topological superconductivity and Majorana zero modes, Nat. Rev. Mater. **6**, 944 (2021).

[54] T.-F. Fang, A.-M. Guo, and Q.-F. Sun, Strongly correlated electrons in superconducting islands with fluctuating Cooper pairs, Phys. Rev. B **106**, 075117 (2022).

[55] L. Yu, Bound state in superconductors with paramagnetic impurities, Acta Phys. Sin. **21**, 75 (1965) .

[56] H. Shiba, Classical spins in superconductors, Prog. Theor. Phys. **40**, 435 (1968).

[57] A. I. Rusinov, Superconductivity near a paramagnetic impurity, Zh. Eksp. Teor. Fiz. **9**, 146 (1969) [JETP Lett. **9**, 85 (1969)].

[58] A. Sakurai, Comments on superconductors with magnetic impurities, Prog. Theor. Phys. **44**, 1472 (1970).

[59] J. C. Estrada Saldaña, A. Vekris, R. Žitko, G. Steffensen, P. Krogstrup, J. Paaske, K. Grove-Rasmussen, and J. Nygård, Two-impurity Yu-Shiba-Rusinov states in coupled quantum dots, Phys. Rev. B **102**, 195143 (2020).

[60] S. Sen, Reduced-basis approximation and *a posteriori* error estimation for many-parameter heat conduction problems, Numer. Heat Transfer, Part B **54**, 369 (2008).

[61] J. S. Hesthaven, B. Stamm, and S. Zhang, Efficient greedy algorithms for high-dimensional parameter spaces with applications to empirical interpolation and reduced-basis methods, ESAIM: M2AN **48**, 259 (2014).

[62] S. Chellappa, L. Feng, and P. Benner, A training set subsampling strategy for the reduced-basis method, J. Sci. Comput. **89**, 63 (2021).

[63] C.-M. Chung, M. M. Wauters, and M. Burrello, Matrix product state simulations of quantum quenches and transport in Coulomb blockaded superconducting devices, Phys. Rev. B **106**, 094308 (2022).

[64] J. S. Hesthaven, C. Pagliantini, and G. Rozza, Reduced-basis methods for time-dependent problems, Acta Numer. **31**, 265 (2022).