

Covariance matrix spectrum of correlated charge insulators reveals hidden connections to coupled cluster, matrix product, and Rokhsar-Kivelson states

Izak Snyman *Mandelstam Institute for Theoretical Physics, School of Physics, University of the Witwatersrand, Johannesburg, South Africa*Serge Florens *Univ. Grenoble Alpes, CNRS, Grenoble INP, Institut Néel, 38000 Grenoble, France*

(Received 3 July 2025; revised 5 September 2025; accepted 8 September 2025; published 29 September 2025)

Charge ordering induced by strong short-range repulsion in itinerant fermion systems typically follows a two-site alternation pattern. However, the covariance matrix spectrum of the one-dimensional, half-filled, spinless t - V model reveals a post-Hartree-Fock picture at strong repulsion, with emergent four-site disruptions of the underlying staggered mean-field state. These disruptions are captured in a thermodynamically extensive manner by a compact four-fermion coupled cluster (doubles) state (CCS). Remarkably, all properties of this state may be computed analytically by combinatorial means, and also derived from an exactly solvable correlated hopping Hamiltonian. Furthermore, this coupled cluster state can be reexpressed as a low-rank matrix product state (MPS) with bond dimension exactly four. In addition, we unveil a hidden connection between this coupled cluster ansatz and a Rokhsar-Kivelson state (RKS), which is the ground state of a solvable parent quantum tetramer model. The broad picture that we uncover here thus provides deep connections between several core concepts of correlated fermions and quantum chemistry that have previously enjoyed limited synergy. In contrast to a recent perturbative treatment on top of Hartree-Fock theory, our approach asymptotically captures the correct correlations in the t - V model at small t/V , and remains a qualitatively accurate approximation even outside the perturbative regime. Our results make the case for further studies of the covariance matrix for correlated electron systems in which ground states have nontrivial unit-cell structure.

DOI: [10.1103/rs88-mdj9](https://doi.org/10.1103/rs88-mdj9)

I. INTRODUCTION

The covariance matrix $Q_{jj'} = \langle a_j^\dagger a_{j'} \rangle$ of an interacting many-body system of fermions $a_j^\dagger$ has a dual nature. As it constitutes a one-body density matrix, it allows one to evaluate any *single-particle* observable [1]. At the same time, its spectrum provides a basis-independent measure of *many-body* correlations [2]. Indeed, by counting the number of eigenvalues significantly different from zero or one, one can learn how far away the state under consideration is from a noninteracting state. The single-particle basis that diagonalizes the Q -matrix is called the natural orbital basis [3]. Of course, when the ground state of a many-fermion system has the same translation symmetry as the underlying lattice, the natural orbitals are momentum k Bloch waves [4], and the spectrum of Q is nothing but the Bloch momentum occupation numbers $\langle n_k \rangle$ that characterize the properties of the emergent Fermi liquid (or Luttinger liquid in one dimension [5]). However, in insulating systems like Mott or charge insulators, the ground state has lower translational symmetry than the underlying lattice because of charge or spin ordering, and an interesting situation arises. On the one hand, the symmetry-broken state is often understood in terms of Hartree-Fock mean-field theory [4], which yields an effective noninteracting system. Yet on the other hand, it is by virtue of correlation-inducing interactions that the insulating state is achieved. What can one learn from the covariance matrix of such a system, whose natural orbitals

are now not simply Bloch waves of the underlying lattice? Existing literature does not yet provide the answer. In this article, we start to fill in the gap. Along the way we uncover several emergent structures of correlated insulators that were deeply hidden from previous analytical or numerical understanding.

Quantum chemists have long understood the value of natural orbitals, when the state under consideration is the ground state of an interacting few-electron system [6]. More recently, it has stimulated the interest of condensed matter physicists. Studying the spectrum of the Q -matrix, it was realized that dynamical quantum impurity problems such as the ground state [7,8] and low-energy quasiparticle excitations [9] of the Kondo model, which are paradigmatic in quantum many-body theory, host only a handful of correlated natural orbitals, the complement being either completely filled or empty [10] for all practical purposes. When cast in the natural orbital basis, these problems thus become few-body problems [11], and this structure remains robust even under nonequilibrium dynamics [12]. The perspectives provided by calculating Q are less well explored when it comes to bulk-interacting systems, although the covariance matrix has been used as a diagnostics tool for many-body localization [13] or in quench protocols [14]. It was also pointed out recently that an entropy characterizing many-body complexity can be defined based on the covariance matrix spectrum [15]. Here, we complement such general perspectives by focusing on correlated insulators. We show that the spectrum of the covariance matrix of a

bulk interacting system unveils essential information enabling the construction of an accurate and tractable approximation for the many-body ground state that becomes asymptotically exact at strong interactions.

The framework for our study is the t - V model in one dimension [16]. It comprises spinless electrons with nearest-neighbor hopping t and repulsion V . It is arguably the simplest model to host a quantum phase transition from a Luttinger liquid to a correlated insulator at finite repulsive interactions. The t - V model is Bethe ansatz solvable [17], and also amenable to controlled numerical solutions via the density matrix renormalization group (DMRG) [18,19]. Our primary aim is to understand how features of the covariance matrix translate into the microscopic structure of electronic correlations, thereby revealing regimes of emergent simplicity, which remained obscure in the representations of the state employed in the Bethe ansatz or DMRG. While the reduced density matrix (and the more general spin-spin correlation functions) of the XXZ-model have been computed from integrability [20–23], fermionic counterparts in the t - V model defy analytical calculation because of the nonlocal nature of the mapping between spins and fermions.

Recently Gebhard *et al.* [24] implemented a self-consistent perturbative expansion for the t - V model on top of the Hartree-Fock approximation, that remarkably captures the quantum phase transition at finite repulsion, something that Hartree-Fock theory on its own fails at. However, this treatment yields inaccurate results for the covariance matrix spectrum at large repulsion $V \gg t$. The question of the asymptotically correct post-Hartree-Fock treatment in this regime is therefore still open. We show that the solution can essentially be read off from the perturbatively calculated correlation matrix spectrum. The key is to employ Wannierized natural orbitals to construct a state whose variational energy agrees with strong-coupling perturbation theory in next-to-next-to leading order, one order higher in $(t/V)^2$ than Hartree-Fock theory. In the Wannierized natural orbital basis, this correlated state consists of a superposition of terms in which the Hartree-Fock state is decorated by an arbitrary number of identical length four disruptions of the mean-field order (namely, the charge pattern is flipped along four consecutive sites). The amplitude of each of the exponentially many terms is exponentially small in the number of disruptions. As a result, the post-Hartree-Fock ground state of the t - V model is a particularly simple example of a class of states known in quantum chemistry as coupled cluster states [25]. Connecting the many-body states of correlated insulators, a general topic of condensed matter, to one of the core concepts of quantum chemistry is our first major result. It should be noted that coupled cluster methods have already been brought to bear on extended systems [26], including applications in condensed matter [27–29]. In these studies, coupled cluster states are invoked as part of a versatile but involved *ab initio* method, in conjunction with application-specific approximations, similar in spirit to diagrammatic Green’s function methods, to obtain approximate results that are, however, not variational. This stands in contrast to the simple and variational coupled cluster picture that we uncover in the t - V model at large V .

As a next decisive step, we explicitly construct a matrix product state (MPS) representation [30] of the coupled cluster

state. The bond dimension of the matrix product state is exactly four, and corresponds precisely to the size of the individual disruptions in the coupled cluster state. As far as we know, matrix product and coupled cluster states (CCS) have enjoyed only limited synergy [31,32] prior to our work, and the exact mapping between two instances of CCS and MPS constitutes our second major result.

A third important result is provided by an explicit connection between the coupled cluster state and the statistical mechanics of a one-dimensional lattice gas of tetramers (namely, rods of length four) [33]. Using a quantum-classical mapping, we derive a novel connection between the strong interaction regime of the t - V model and solvable quantum tetramer Hamiltonians of the Rokhsar-Kivelson form [34], initially introduced for dimers in the context of resonating valence bond states [35]. Our work clearly establishes several previously hidden connections between different nonperturbative methods for interacting fermions: coupled cluster states, matrix product states, and Rokhsar-Kivelson states.

Both the connection to coupled cluster and Rokhsar-Kivelson states allow us to derive parent Hamiltonians for which the coupled cluster state that we identified is the exact ground state. These two parent Hamiltonians are of interest in their own right, and have not been previously investigated as far as we know. (Note that a parent construction for the low-rank matrix product state is also possible, but is already well established [36,37], and will not be considered in detail.) The existence of exact parent Hamiltonians invites the study of the coupled cluster state at arbitrary correlation strengths, leading us to investigate new physical regimes that are inaccessible within the original t - V model. We obtain here an exact and compact expression for the covariance matrix elements of our coupled cluster state in the thermodynamic limit and at arbitrary correlation strength.

The rest of this article is structured as follows. In Sec. II, we introduce the one-dimensional t - V model and set notation. In Sec. III, we briefly consider the critical point for the case of attractive interactions. In this case, we calculate the covariance matrix exactly, and provide a closed-form expression Eq. (5) for its matrix elements at finite size, which we could not find in the existing literature. When we consider large repulsive interactions in subsequent sections, we will discover an interesting parallel with the critical state considered here. In Sec. IV, we examine the strong-coupling perturbation expansion of the covariance matrix of the t - V model, and notice a striking feature that determines the correct post-Hartree-Fock treatment in this regime. In Sec. V, we translate the insights we gained in the previous section into a coupled cluster ansatz. We show how the calculation of physical observables amounts to a combinatorial problem in classical statistical physics. We numerically demonstrate the accuracy of the coupled cluster state by comparing to Bethe ansatz and DMRG results for the ground-state energy and covariance matrix spectrum. In Sec. VI, we analytically construct an exact matrix product state representation for the coupled cluster state. In Sec. VII, we derive two parent Hamiltonians for which the coupled cluster state is the unique ground state. Viewing the coupled cluster state as the ground state of either Hamiltonian gives us license to consider arbitrary correlation strengths. In Sec. VIII we present an elegant derivation of an exact expression for

the nonzero matrix elements of the covariance matrix associated with the coupled cluster state at arbitrary correlation strength, before summarizing our conclusions and giving perspectives in Sec. IX. Various technical details can be found in Appendices.

II. MODEL

We consider here the t - V model of spinless fermions in one dimension [16]

$$H = -t \sum_{j=1}^{2N} (a_j^\dagger a_{j+1} + a_{j+1}^\dagger a_j) + V \sum_{j=1}^{2N} (a_j^\dagger a_j - 1/2)(a_{j+1}^\dagger a_{j+1} - 1/2). \quad (1)$$

The t - V model is fully equivalent to the solvable one-dimensional anisotropic Heisenberg (or XXZ) model [17] via the Jordan-Wigner transformation [5]. Without loss of generality, we can take $t > 0$. Periodic boundary conditions are implied, i.e., $a_j = a_{j+2N}$, unless noted otherwise (in particular for the construction of the MPS representation). The t - V Hamiltonian is particle-hole symmetric, i.e., $H = PHP$ where

$$P = \prod_{j=1}^N (a_{2j-1} - a_{2j-1}^\dagger)(a_{2j} + a_{2j}^\dagger) \quad (2)$$

is a Hermitian and unitary particle conjugation operator such that $Pa_jP = (-1)^j a_j^\dagger$. We will always work at half-filling. It is known that the Hamiltonian Eq. (1) undergoes quantum phase transitions at $V = \pm 2t$. For $V > 2t$ there are two macroscopically distinct gapped ground states, each of which represent a correlated insulator, with charge order, associated with broken particle-hole conjugation, as well as broken symmetry between the two sublattices of the original bipartite lattice. Because our main focus is this correlated insulator phase, we will work throughout with two-site unit cells $(2j-1, 2j)$ and use folded Bloch momenta that are rescaled in the range $k \in (-\pi, \pi]$, corresponding to the range $(-\pi/2, \pi/2]$ of the unfolded first Brillouin zone in the original lattice.

In the Luttinger liquid phase, the unbroken sublattice symmetry implies that the natural orbitals are simply the single-particle momentum eigenbasis. All physical information contained in the covariance matrix then resides in its spectrum, which corresponds to their occupation numbers $q_\pm(k)$. The subscript $\pm$ is necessary because of the folding of the Brillouin zone, i.e., q_- (q_+) refers to occupations of orbitals with momenta outside (inside) the interval $(-\pi/2, \pi/2]$ in the unfolded Brillouin zone. For $V = 0$, the ground state is a degenerate Fermi gas with $q_-(k) = 0$ and $q_+(k) = 1$ for all k . In the rest of the ungapped phase, namely for $|V| \leq 2t$, it is known [5] that $q_+(k)$ and $q_-(k)$ touch at $k = \pm\pi$, where particle-hole symmetry protects $q_\pm(k) = 1/2$, reflecting the fact that arbitrarily weak interactions destroy Fermi liquid behavior in one dimension. At $V = \pm 2t$, where lattice translation symmetry is spontaneously broken, gaps start to appear at $k = \pm\pi$. In Fig. 1, the behavior of the covariance matrix spectrum in the correlated insulator phase is shown. Since the ground state in the $V \rightarrow \infty$ limit is an uncorrelated

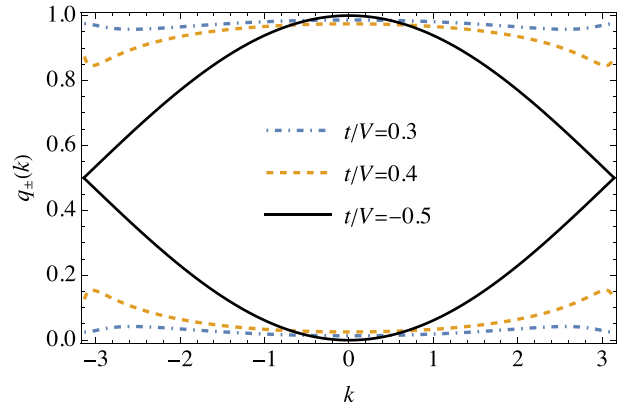


FIG. 1. The behavior of the covariance matrix spectrum in the rescaled and folded Brillouin zone. The dashed lines are numerical DMRG data for the gapped correlated insulator phase $V > 2t$, while the dashed curve is the analytical formula Eq. (8) for the critical point $V = -2t$ associated to a uniform gapless state.

single Slater determinant, the eigenvalues of the covariance matrix move closer to 0 or 1, as t/V decreases towards zero. The dashed lines are numerically exact DMRG data, obtained using the C++ ITensor library [38] for long chains with $2N = 2000$ sites, using up to 30 sweeps, a maximum bond dimension $D = 300$, and a truncation threshold 10^{-14} . The covariance matrix spectrum of the translationally invariant system was obtained from a long chain with open boundaries (which helps convergence of DMRG) by computing the Q -matrix $Q_{jj'} = \langle a_j^\dagger a_{j'} \rangle$ for sites j, j' near the center of the chain, in a region where translation invariance was carefully checked. In addition, the solid black curve in Fig. 1 indicates the analytically obtained gapless spectrum at $V = -2t$ (see Sec. III below). It shows the touching points at $k = \pm\pi$ that are generic for the whole gapless phase.

III. THE CRITICAL POINT AT $V = -2t$

Before studying the correlated insulator phase, we briefly focus on the gapless critical point at $V = -2t$, which is exactly solvable, and provides already some useful insights. In subsequent sections, we uncover interesting parallels and differences between this gapless ground state and the gapped insulating state found at large positive V .

At $V = -2t$, the critical ground state is the equal weight superposition of all arrangements of N particles among $2N$ sites, see [39] or Appendix A, and thus reads

$$|\text{GS}\rangle_{V=-2t} = \binom{2N}{N}^{-1/2} \sum_{\gamma \in C(2N, N)} a_{\gamma_1}^\dagger a_{\gamma_2}^\dagger \dots a_{\gamma_N}^\dagger |0\rangle. \quad (3)$$

Here, each $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_N)$ consists of N lattice indices chosen from the $2N$ sites of the chain, arranged such that $\gamma_1 < \gamma_2 < \dots < \gamma_N$, and the sum runs over all possible choices. For this state, we obtain analytically in Appendix B the matrix elements of the covariance matrix in position representation, $Q_{j,j} = 1/2$, while for $m > 0$,

$$Q_{j,j+m} = \frac{\sum_{n=\max(0, m-N)}^{\min(m-1, N-1)} (-1)^n \binom{m-1}{n} \binom{2N-m-1}{N-n-1}}{\binom{2N}{N}}. \quad (4)$$

This expression is easy to grasp. When $a_1^\dagger a_{1+m}$ is applied to a term in the sum Eq. (3), it yields a nonzero result if site 1 is empty and site $1+m$ is occupied. Each of these states is mapped into another state in the sum, up to a sign, and thus contributes $\pm \binom{2N}{N}^{-1}$ to the overlap with the ground state. The sign is determined by the number n of particles between sites 1 and $1+m$. We count the number of terms with site 1 empty and site $1+m$ occupied, that have n particles on sites 2 to m as follows. There are $\binom{m-1}{n}$ ways to arrange the n particles on sites 2 to m , and $\binom{2N-m-1}{N-n-1}$ ways of arranging the remaining $N-n-1$ particles among the remaining $2N-m-1$ sites. By translation invariance, $Q_{j,j+m} = Q_{1,1+m}$. The sum in Eq. (4) can be evaluated analytically. It is zero for m even. (This is guaranteed by particle-hole symmetry.) For $m = 2r + 1 \leq N$

$$Q_{j,j+2r+1} = (-1)^r \frac{2^{2(N-1)} \Gamma(r+1/2) \Gamma(N-r-1/2)}{(N-1)! \Gamma(1/2)^2 \binom{2N}{N}}. \quad (5)$$

(For $m > N$ we have $Q_{j,j+m} = Q_{j,2N+1-m}$ thanks to inversion symmetry.) It can be verified that $Q_{j,j+2r+1}$ decays like N^{-r} . See Appendix B for details on the derivation. To the best of our knowledge, the closed-form expression Eq. (5) does not appear in the literature, although it is known that the correlation length of the covariance matrix vanishes in the thermodynamic limit at $V = -2t$ [14]. The $m \ll N$ asymptotic behavior of Eq. (4), generalized to arbitrary filling, has also been derived in [40]. The extremely rapid decay of off-diagonal elements of the covariance matrix is a destructive interference effect between the contribution by states with an even or an odd number of fermions on sites $j+1, \dots, j+2r+1$. This fermionic sign effect was pointed out in [41]. In the thermodynamic limit

$$Q_{j,j'} = \frac{1}{4} (2\delta_{j,j'} + \delta_{j,j'+1} + \delta_{j+1,j'}). \quad (6)$$

Choosing to work in the folded Brillouin zone, we diagonalize Q by means of the unitary transform $Q' = F Q F^\dagger$ where

$$\begin{pmatrix} F_{2j-1,2l-1} & F_{2j-1,2l} \\ F_{2j,2l-1} & F_{2j,2l} \end{pmatrix} = \frac{e^{ik_l j}}{\sqrt{N}} \begin{pmatrix} -\frac{e^{-ik_l/2}}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{e^{ik_l/2}}{\sqrt{2}} \end{pmatrix},$$

$$\begin{pmatrix} Q'_{2l-1,2l'-1} & Q'_{2l-1,2l'} \\ Q'_{2l,2l'-1} & Q'_{2l,2l'} \end{pmatrix} = \delta_{l,l'} \begin{pmatrix} q_-(k_l) & 0 \\ 0 & q_+(k_l) \end{pmatrix}, \quad (7)$$

for $k_l = 2\pi(l - \frac{N-1}{2})/N$, $l = 1, 2, \dots, N$, and the spectrum of the covariance matrix is simply

$$q_{\pm}(k) = [1 \pm \cos(k/2)]/2. \quad (8)$$

Because of our choice of working in a folded Brillouin zone, the Q -matrix spectrum comprises two bands. These bands have 4π periodicity, owing to the full lattice symmetry of the ground state, and touch at $k = \pm\pi$, as expected in the gapless regime. Let us now contrast these results with the spectrum obtained for large electron-electron repulsion $V/t \gg 1$.

IV. LARGE V/t PERTURBATION THEORY

From here onwards, we focus on the correlated insulator phase $V > 2t$. Our starting point is to consider the spectrum of the covariance matrix at strong repulsion. In the small t/V

limit, Rayleigh-Schrödinger perturbation theory around the unperturbed ground state

$$|\Psi_0\rangle = a_2^\dagger a_4^\dagger \dots a_{2N}^\dagger |0\rangle, \quad (9)$$

yields the following result for the spectrum of the covariance matrix, to fourth order in t/V :

$$q(k)_{\pm} = \frac{1}{2} \pm \left[\frac{1}{2} - (4 - 2 \cos 2k) \left(\frac{t}{V} \right)^4 \right]. \quad (10)$$

See Appendix C for the details. Remarkably, the spectrum has k -periodicity π , even though the matrix $Q(k)$ only has 2π k -periodicity. What does this quartering of the period compared to Eq. (8) tell us about the structure of the gapped ground state?

The formal perturbative expansion of the ground-state wave function only answers this question in an indirect way. In the present context, this general statement can be elaborated as follows. When the perturbation $H_{\text{kin}} = -t \sum_{j=1}^{2N} (a_j^\dagger a_{j+1} + a_{j+1}^\dagger a_j)$ acts on the unperturbed state Eq. (9), it moves a single particle from the B -sublattice (even sites) to the A -sublattice. Thus the expansion of the wave-function to n th order contains at most n particles on the A sublattice. However, the true ground state at any nonzero t/V hosts on average a finite fraction of the total number of particles on the A -sublattice. In the thermodynamic limit, the formal perturbation expansion to finite order thus poorly approximates the true ground state. A symptom of this state of affairs is that the perturbatively calculated ground-state energy is not variational. In fact, if we denote the ground-state wave function and energy, expanded to n th order as respectively $|\Psi^{(n)}\rangle$ and $E_0^{(n)}$, then because of the nonextensive nature of ground-state corrections

$$\lim_{N \rightarrow \infty} \frac{1}{N} \frac{\langle \Psi^{(n)} | H | \Psi^{(n)} \rangle}{\langle \Psi^{(n)} | \Psi^{(n)} \rangle} = E_0^{(0)}, \quad (11)$$

that is, in the thermodynamic limit, finite-order perturbation theory is variationally no better than the unperturbed starting point. To obtain the answer $E_0^{(n)}$ correct to order n instead of $E_0^{(0)}$, the above expression must be expanded to order n before the limit $N \rightarrow \infty$ is taken. An order-by-order cancellation of certain terms between numerator and denominator then occurs. In the diagrammatic Green's function formalism, this corresponds to the cancellation of disconnected diagrams, leaving as the final result the sum of connected diagrams. This allows us to calculate expectation values of simple observables. However, the unsatisfactory fact remains that one cannot explicitly write a thermodynamically consistent perturbative wave function, and ask simple questions, such as how accurately can the perturbative ground state be approximated for a given MPS bond dimension?

More direct insight than that offered by perturbation theory of the ground-state wave function would be provided by a state whose variational energy equals $E_0^{(n)} + \mathcal{O}(t/V)^{n+1}$ for some n . Standard mean-field theory solves this problem to first nonvanishing order in the perturbation, namely at order $(t/V)^2$. Here we pursue the quest for a state that is accurate to the next nonvanishing order, namely at order $(t/V)^4$. This is the order necessary to shed light on the emergent π -periodicity of the covariance matrix spectrum, since the

correlation spectrum of the mean-field state is trivial (eigenvalues are either 0 or 1). Our main task is thus to construct a state describing a finite density of quantum fluctuations of the mean-field order parameter, with a variational energy that agrees with the perturbative result

$$E_0 = -\frac{NV}{2} \left[1 + 4 \left(\frac{t}{V} \right)^2 - 4 \left(\frac{t}{V} \right)^4 + \dots \right], \quad (12)$$

to order $(\frac{t}{V})^4$. [The expansion Eq. (12) can be verified from the exact Bethe ansatz expression Eq. (22) below.]

V. COUPLED CLUSTER DOUBLES ANSATZ

The Hartree-Fock mean-field theory of the correlated insulator at small t/V is reviewed in Appendix D. It yields a set of new fermion operators c_j , $j = 1, \dots, 2N$, with the following features. The Wannierized Hartree-Fock basis is invariant under translations by two lattice sites, under which action $c_j \rightarrow c_{j+2}$. The single-particle orbital associated with c_j is peaked at site j , and has a more pronounced amplitude on sites with the same parity as j , than on sites of opposite parity.

In this basis of Wannierized Hartree-Fock orbitals, the mean-field state reads

$$|\text{MF}\rangle = c_2^\dagger c_4^\dagger \dots c_{2N}^\dagger |0\rangle. \quad (13)$$

Of course, the same state could equivalently be obtained as the usual Slater determinant in which all Bloch momenta of the lower band of the mean field Hamiltonian are occupied, but the Wannierized version is more amenable to constructing a correlated state that is energetically favorable, since the repulsive term is quasilocal in space, as we discuss now.

Our ansatz introduces order parameter fluctuations on top of the state Eq. (13). These take the form of regions of opposite order (even sites empty, odd sites filled) with respect to the mean-field pattern [42,43]. Our key insight for reproducing the π -periodicity of the spectrum of the covariance matrix is that the fluctuations should be made of building blocks of size two unit cells of the symmetry broken system, i.e., four sites of the original lattice. We call four adjacent sites of which the two even orbitals c_{2j} and c_{2j+2} are empty, while the two odd orbitals c_{2j+1} and c_{2j+3} are filled, a disruption, depicted with a red ellipse in Fig. 2. Sites with the same occupation pattern as $|\text{MF}\rangle$ —odd sites empty, even ones filled—are said to be of regular order. Again, we point out that the $c_j^\dagger$ fermions on “sites” j correspond to Wannierized orbitals centered around the real space position j . The four-fermion operator

$$T = \sum_{j=1}^N (c_{2j} c_{2j+1}^\dagger c_{2j+2}^\dagger c_{2j+3} + c_{2j-1}^\dagger c_{2j} c_{2j+1}^\dagger c_{2j+2}), \quad (14)$$

creates one such a disruption at any of these possible locations $(l-3, l-2, l-1, l)$ of the lattice, in a translationally invariant manner. The unnormalized state

$$|m\rangle = \frac{1}{m!} T^m |\text{MF}\rangle, \quad (15)$$

is a superposition of all possible ways to make m disruptions, in which each unique term appears with weight equal to unity (see Fig. 2). Note that the product of any two terms in T that

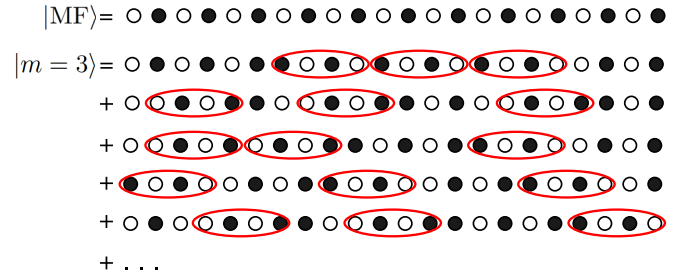


FIG. 2. Schematic representation of the mean-field state $|\text{MF}\rangle$ and the state $|m=3\rangle$ built from three disruptions of length four (red ellipses) on top of the mean-field state, for $N = 11$ fermions in a chain of $2N = 22$ sites. Filled/empty circles represent the center of mass of filled/empty Wannierized Hartree-Fock orbitals (and not localized sites of the original lattice). For $|m=3\rangle$, five of the 484 configurations in the equal weight superposition Eq. (15) are depicted here.

are associated with overlapping disruptions is zero. Disruptions therefore correspond to hard-core rods of size four. The norm of $|m\rangle$ is obtained by counting the number of nonzero terms in T^m . There are m disruptions and $2N - 4m$ sites of regular order to arrange into a string. This gives a combinatorial factor $\binom{2N-3m}{m}$. This way of counting includes a factor $2N - 3m$ corresponding to the number of cyclic permutations of the string, in which a disruption is never allowed to straddle the “seam” where site $2N$ is glued to site 1. We must, however, also include the arrangements where a disruption straddles the $2N-1$ seam, and therefore must multiply by $2N/(2N - 3m)$, i.e.,

$$\langle m|m\rangle = \frac{2N}{2N - 3m} \binom{2N - 3m}{m}. \quad (16)$$

If we assume that the fluctuations caused by multiple disruptions are independent of each other and are each associated with an amplitude α , then the resulting state is known as a coupled cluster doubles state built from the operator T [25]. It reads

$$|\alpha\rangle = \exp(\alpha T) |\text{MF}\rangle = \sum_{m=0}^{\lfloor N/2 \rfloor} \alpha^m |m\rangle. \quad (17)$$

(Note that $|\alpha\rangle$ is not normalized.) The exponential term $\exp(\alpha T)$ takes care of size extensivity by inducing a similarity transform $\exp(\alpha T) c_{2n}^\dagger \exp(-\alpha T)$ on each creation operator in Eq. (13). Each transformed operator has a finite amplitude to create particle-hole excitations, and as a result the typical number of total excitations is proportional to N [see Eqs. (41) and (42) below].

Usually, the extensive nature of coupled cluster states makes calculations for more than a handful of particles impossible, because of the exponential number of involved terms. We emphasize that this statement applies to calculations involving the full coupled cluster wave function, and not to the linearized treatment of coupled cluster equations, that has an algebraic complexity, but is not variational. Remarkably, the simple structure of the disruptions in the variational coupled cluster state that we consider avoids paying an exponential price. Indeed, although the states $|m\rangle$ are many-body

correlated states, for which Wick's theorem does not apply, operator matrix elements can nonetheless be calculated by combinatorial means.

We now sketch how analytical computations of the variational energy and the Q -matrix can be done with the disrupted states Eq. (15). When a single-particle additive operator $c_j^\dagger c_{j'}$ acts on a term in $|m\rangle$ such as one of those depicted in Fig. 2, it can only do one of the following things: (i) leave the term unchanged (if and only if $j = j'$), (ii) annihilate the term, (iii) move a sequence of head-to-toe disruptions by one site, thereby producing another term in $|m\rangle$ (up to a sign), or (iv) produce a state orthogonal to all $|m\rangle$. Since $c_j^\dagger c_{j'}$ cannot create or annihilate a full disruption, $\langle m | c_j^\dagger c_{j'} | m' \rangle = 0$ if $m \neq m'$. Expectation values of single-particle additive operators can be calculated by solving counting problems similar to the one we solved to compute the norm. For instance, $c_{2l}^\dagger c_{2l}$ counts terms that have a site of regular order at site $2l$ while $c_{2l+1}^\dagger c_{2l+1}$ counts terms with a disruption involving site $2l + 1$. This straightforwardly yields

$$\begin{aligned} \langle m | c_{2l}^\dagger c_{2l} | m \rangle &= \binom{2N - 3m - 1}{m}, \\ \langle m | c_{2l+1}^\dagger c_{2l+1} | m \rangle &= 4 \binom{2N - 3m - 1}{m - 1}. \end{aligned} \quad (18)$$

Process (iii) can occur when $c_{1+4l}^\dagger$ moves l head-to-toe disruptions initially occupying sites $2, \dots, 4l + 1$, one site to the left. An example is illustrated below.

$$\dots 0[0101] \dots \rightarrow \dots [1010]0 \dots$$

A single four-site disruption is depicted for simplicity, although we consider the general case of multiple head-to-toe disruptions. Here $[\dots]$ indicates the disruption that is being moved. Moving the l disruptions requires commuting $c_{1+4l}^\dagger$ past $2l - 1$ $c^\dagger$ operators, which is an odd number. If site 1 is initially empty, and sites $2, \dots, 4l + 1$ are occupied by l disruptions, then we have $2N - 4m - 1$ remaining sites of regular order and $m - l$ disruptions to arrange in a string. Hence,

$$\langle m | c_{1+4l}^\dagger | m \rangle = - \binom{2N - 3m - l - 1}{m - l}, \quad (19)$$

for $0 < l \leq m$. For $c_{2+4l}^\dagger$ essentially the same considerations apply. The only difference is that moving l disruptions, e.g.,

$$\dots [0101]1 \dots \rightarrow \dots 1[1010] \dots$$

now requires commuting $c_{2+4l}^\dagger$ past two $c^\dagger$ operators per disruption, and so no minus sign is generated. Including also processes that straddle the link between sites $2N$ and 1 we find

$$\begin{aligned} \langle m | c_{2+4l}^\dagger | m \rangle &= \langle m | c_{2N-4l+2}^\dagger | m \rangle \\ &= - \langle m | c_{1+4l}^\dagger | m \rangle = - \langle m | c_{1+4l}^\dagger | m \rangle. \end{aligned} \quad (20)$$

All nonzero $c_j^\dagger c_{j'}$ expectation values can be obtained from the above by invariance under translation by two lattice sites, or by Hermitian conjugation.

We note that every $4l'$ th diagonal (modulo $2N$) above or below the main diagonal of $Q_{jj'} = \langle \alpha | c_j^\dagger c_{j'} | \alpha \rangle / \langle \alpha | \alpha \rangle$ is non-

zero, and of order α^{2l} , with all other diagonals above or below the main diagonal equal to zero. This results in the $\cos(2k)$ component of the covariance matrix spectrum (coming from the $\pm 4l'$ th diagonals) being of order α^2 . The covariance matrix spectrum expanded to order α^2 , calculated from Eqs. (18) and (19), is given by $q_\pm(k) = 1/2 \pm [1/2 - \alpha^2(4 - 2 \cos 2k)]$. We thus see that we have to choose

$$\alpha = \left(\frac{t}{V} \right)^2, \quad (21)$$

to ensure agreement with the desired covariance matrix Eq. (10) for the ground state.

Apart from the processes already discussed, two-particle additive operators $c_{j_1}^\dagger c_{j_2} c_{j_3}^\dagger c_{j_4}$ produce further processes: (v) two strings of head-to-toe disruptions can each be moved independently, e.g.,

$$\begin{aligned} \dots 0[0 \ 1 \ 0 \ 1] \dots \ 0[0 \ 1 \ 0 \ 1] \dots \\ \rightarrow \dots [1 \ 0 \ 1 \ 0]0 \dots [1 \ 0 \ 1 \ 0]0 \dots; \end{aligned}$$

(vi) three strings of head-to-toe disruptions (of which some may be zero strings) can be fused into a single string containing one more disruption, e.g.,

$$\begin{aligned} \dots 0[0 \ 1 \ 0 \ 1]1[1 \ 0 \ 1 \ 0]0[0 \ 1 \ 0 \ 1]1 \dots \\ \rightarrow \dots [1 \ 0 \ 1 \ 0 \ 1 \ 0 \ 1 \ 0 \ 1 \ 0 \ 1 \ 0 \ 1 \ 0] \dots; \end{aligned}$$

(vii) and the reverse of process (vi) in which the number of disruptions decreases by one. We work out the matrix elements for two-particle additive operators, by solving the appropriate counting problems in Appendix E.

In this way we obtain a variational estimate $\langle H \rangle = \langle \alpha | H | \alpha \rangle / \langle \alpha | \alpha \rangle$ with respect to the state $|\alpha\rangle = (t/V)^2$ of the ground-state energy, that includes nontrivial correlation effects. See Appendix F for further details. To investigate the quality of the correlated ansatz, we consider the correlation energy $E_{\text{mf}} - \langle H \rangle$, where E_{mf} Eq. (D18) is the best estimate of the ground-state energy that can be obtained from an uncorrelated (mean-field) trial state. The exact ground-state energy is known via the Bethe ansatz [17],

$$\frac{E_0}{N} = t \left[\cosh^{-1} \lambda - 2 \sinh^{-1} \lambda \left(1 + 4 \sum_{n=1}^{\infty} \frac{1}{1 + e^{2n\lambda}} \right) \right], \quad (22)$$

where $\lambda = \cosh(V/2t)$. The main panel of Fig. 3 shows the correlation energy $(E_{\text{mf}} - \langle H \rangle) / (E_{\text{mf}} - E_0)$ relative to the exact value, for increasing t/V . At small t/V the ratio tends to unity, implying that in this regime, the correlated ansatz accounts for the full effect of correlations. The inset of Fig. 3 shows that the difference $\langle H \rangle - E_0$ grows like $(t/V)^6$ at small t/V . This indicates that the correlated ansatz gives a variational energy accurate to fourth order $\sim (t/V)^4$. In contrast, Hartree-Fock mean-field theory gives a variational energy that is only accurate to order $\sim (t/V)^2$ (see Appendix D). The correlated ansatz Eq. (17) thus achieves the goal of building a size-extensive post-Hartree-Fock wavefunction, that we had previously set in Sec. IV.

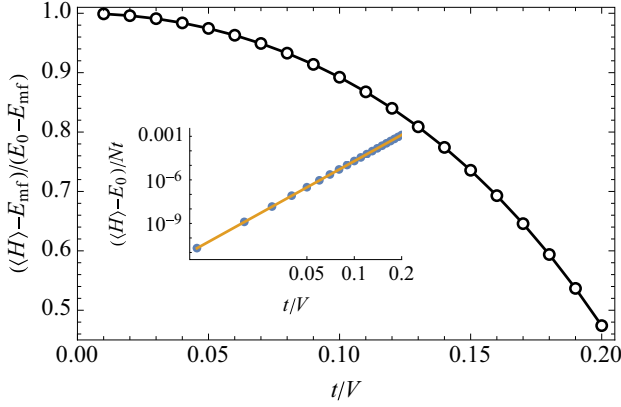


FIG. 3. Difference in ground-state energy between mean-field theory Eq. (D18) and the correlated ansatz Eq. (17) in units of the exact correlation energy of the t - V model. (Inset) The difference between the variational energy per particle of the correlated ansatz Eq. (17) and the true ground-state energy Eq. (22) in double-log scale. The solid line corresponds to $19.5(t/V)^6$, and indicates that the variational energy is correct to order $(t/V)^4$. Results are shown for a system of $N = 41$ particles occupying 82 sites.

In Fig. 4, we plot the lower band of the covariance matrix spectrum, calculated from the correlated ansatz Eq. (17) with the help of Eqs. (18) and (19), and compare to numerically exact results obtained with DMRG (to our knowledge, the correlation spectrum is not available from the Bethe ansatz), as well as to the perturbative result Eq. (10). This computation is done for $t/V = 0.1$, and for this not-so-small value of t/V , agreement of the ansatz and DMRG with perturbation theory is still good. Deviations between the ansatz and DMRG, because of higher order effects, are just becoming visible. From this comparison, we conclude that the regime where four-site disruptions of the mean-field order quantitatively accounts for the correlation effects is $t/V \lesssim 0.1$. It may be noted that our variational ansatz is very restrictive, containing no free parameters to be optimized (once α was set). In Appendix G, we

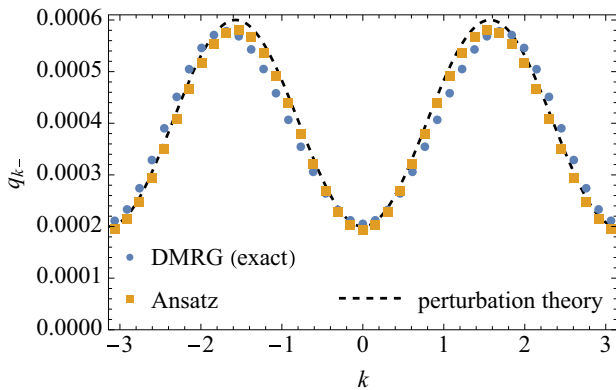


FIG. 4. The lower band q_{k-} of the correlation matrix spectrum of the t - V model, at $t/V = 0.1$, for a system of $N = 41$ particles occupying 82 sites. The variational results for the coupled cluster ansatz Eq. (17) are compared to numerically exact calculations obtained with the density matrix renormalization group (DMRG), and to the perturbative expression Eq. (10).

present numerical evidence that the coupled cluster doubles form assumed here is indeed optimal for a more general ansatz built out of the states $|m\rangle$.

Armed with a coupled cluster ansatz that is variationally accurate in this regime, we now turn to investigate the approximate matrix product state (MPS) structure of the ground state. It turns out that an exact MPS representation of the coupled cluster state $|\alpha\rangle$ can be constructed analytically, establishing a bridge from ideas about one-dimensional correlated fermions to concepts in quantum chemistry.

VI. EXACT MPS REPRESENTATION OF THE COUPLED CLUSTER STATE

As is often the case in the context of MPS, we find it convenient to consider open boundary conditions. We consider a system with L sites. Let D_j be the operator that creates a 4-site disruption ending on site j , i.e.,

$$D_j = \begin{cases} c_{j-3}c_{j-2}c_{j-1}c_j^\dagger & \text{if } j \text{ is odd,} \\ c_{j-3}^\dagger c_{j-2}c_{j-1}c_j & \text{if } j \text{ is even.} \end{cases} \quad (23)$$

We recall that the $c_j^\dagger$ fermions on “sites” j are associated to Wannierized orbitals centered around the real-space position j .

We define a disruption operator acting on sites 1 to l ,

$$T_l = \begin{cases} 0 & \text{if } l < 4, \\ \sum_{j=4}^l D_j & \text{if } l \geq 4. \end{cases} \quad (24)$$

We also find it convenient to introduce the following notation for states. Let $(c_l^\dagger)^1 \equiv c_l^\dagger$ and $(c_l^\dagger)^0 \equiv 1$. An arbitrary single Slater determinant in the Wannierized mean-field basis is then denoted

$$|n_1 n_2 \dots n_L\rangle \equiv (c_1^\dagger)^{n_1} (c_2^\dagger)^{n_2} \dots (c_L^\dagger)^{n_L} |0\rangle, \quad (25)$$

where $n_j \in \{0, 1\}$. Furthermore, let

$$r_l = \begin{cases} 0 & \text{if } l \text{ is odd,} \\ 1 & \text{if } l \text{ is even,} \end{cases} \quad d_l = \begin{cases} 1 & \text{if } l \text{ is odd,} \\ 0 & \text{if } l \text{ is even.} \end{cases} \quad (26)$$

(The notation r and d here denote “regular” and “disrupted” sites.) In this notation, the mean-field state reads

$$|\text{MF}\rangle = |r_1 r_2 \dots r_L\rangle, \quad (27)$$

while

$$D_j |\text{MF}\rangle = |r_1 \dots r_{j-4} d_{j-3} \dots d_j r_{j+1} \dots r_L\rangle. \quad (28)$$

Then the coupled cluster state that we considered previously is

$$|\alpha\rangle \equiv \exp(\alpha T_L) |r_1 r_2 \dots r_L\rangle. \quad (29)$$

At the heart of the MPS expansion is the task of relating states of the same form on systems that differ in size by one site. In view of this, we note that for $l \geq 4$,

$$T_l = T_{l-1} + D_l. \quad (30)$$

We further note that D_l kills all terms in T_{l-1} that involve sites $l-3, \dots, l-1$. Hence,

$$D_l (T_{l-1})^k = D_l (T_{l-4})^k. \quad (31)$$

From this together with $[D_j, D_{j'}] = 0$, and $D_j^2 = 0$ follows that

$$e^{\alpha T_l} = e^{\alpha T_{l-1}} + \alpha D_l e^{\alpha T_{l-4}}. \quad (32)$$

The next step is to define a set of states that have a desired form on sites left of site $l + 1$, and an arbitrary form on sites $l + 1, \dots, L$. We define

$$\begin{aligned} |1, n_{l+1} \dots n_L\rangle &= e^{\alpha T_l} |r_1 \dots r_l n_{l+1} \dots n_L\rangle, \\ |2, n_{l+1} \dots n_L\rangle &= e^{\alpha T_{l-1}} |r_1 \dots r_{l-1} d_l n_{l+1} \dots n_L\rangle, \\ |3, n_{l+1} \dots n_L\rangle &= e^{\alpha T_{l-2}} |r_1 \dots r_{l-2} d_{l-1} d_l n_{l+1} \dots n_L\rangle, \\ |4, n_{l+1} \dots n_L\rangle &= e^{\alpha T_{l-3}} |r_1 \dots r_{l-3} d_{l-2} d_{l-1} d_l n_{l+1} \dots n_L\rangle. \end{aligned} \quad (33)$$

State $|j, n_{l+1} \dots n_L\rangle$ is of the coupled cluster form on sites 1 to $l + 1 - j$, and has arbitrary occupations on sites $l + 1$ to L . For $j > 1$, it contains $j - 1$ disrupted sites situated from site $l + 2 - j$ to site l . For $l = 0$ we define $|1, n_1 \dots n_L\rangle \equiv |n_1 \dots n_L\rangle$. Employing Eq. (32), we find

$$\begin{aligned} |1, n_{l+1} \dots n_L\rangle &= e^{\alpha T_l} |r_1 \dots r_l n_{l+1} \dots n_L\rangle \\ &= |1, r_l n_{l+1} \dots n_L\rangle + \alpha |4, d_l n_{l+1} \dots n_L\rangle. \end{aligned} \quad (34)$$

Furthermore, we trivially have

$$|j, n_{l+1} \dots n_L\rangle = |j - 1, d_l n_{l+1} \dots n_L\rangle, \quad (35)$$

for $j > 1$. Thus we can write, with summation of repeated indices (including σ) implied,

$$|j, n_{l+1} \dots n_L\rangle = A(l)^{(\sigma)}_{jj'} |j', \sigma n_{l+1} \dots n_L\rangle, \quad (36)$$

where for $l \geq 4$, $A(l)^{(\sigma)}$, $\sigma \in \{0, 1\}$ is given by

$$A(l)^{(r_l)} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad (37)$$

and

$$A(l)^{(d_l)} = \begin{pmatrix} 0 & 0 & 0 & \alpha \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \quad (38)$$

For $l < 4$, the open boundary of the system at $l = 1$ is incorporated by taking $A(l)^{(\sigma)}$ as the upper left $(l + 1) \times l$ block of the above defined matrices. We now apply this recursively, starting from

$$|\alpha\rangle = A(L)^{(\sigma_L)}_{1j'} |j', \sigma_L\rangle. \quad (39)$$

Finally, we obtain the MPS expansion as

$$|\alpha\rangle = [A(L)^{(\sigma_L)} A(L - 1)^{(\sigma_{L-1})} \dots A(1)^{(\sigma_1)}]_{1,1} |\sigma_1 \sigma_2 \dots \sigma_L\rangle. \quad (40)$$

Thus, $|\alpha\rangle$ has bond dimension 4. Even though it applies to the ground state of the t - V model at large V , it is hard to imagine being able to derive it from a perturbative expansion.

We now turn to another insightful question that cannot easily be answered from perturbative approaches, namely how to find effective parent Hamiltonians that have asymptotically the same ground state as the t - V model at large V . We provide three complementary answers to this problem in the next section.

VII. EXACT PARENT HAMILTONIANS

We have seen that $|\alpha\rangle$ is a coupled cluster state; it admits a simple MPS representation, and further that the calculation of expectation values can be performed by solving combinatorial problems of rods in classical statistical physics. It turns out that each of these three properties of $|\alpha\rangle$ allows us to construct a distinct parent Hamiltonian for which it is the exact ground state. We review each construction in turn, which highlights novel connections between seemingly distinct topics in many-body physics.

A. Coupled cluster parent Hamiltonian

A first parent Hamiltonian containing only local terms, such that $|\alpha\rangle = \exp(\alpha T)|\text{MF}\rangle$ is its unique ground state, can be constructed by exploiting the coupled cluster property of $|\alpha\rangle$. We start the construction by noting that the mean-field state $|\text{MF}\rangle$ is the unique state that is annihilated by c_{2j-1} and $c_{2j}^\dagger$, for all $j = 1, \dots, N$. This means that $|\alpha\rangle = \exp(\alpha T)|\text{MF}\rangle$ is the unique state annihilated by

$$\begin{aligned} C_{2j-1} &\equiv e^{\alpha T} c_{2j-1} e^{-\alpha T} = c_{2j+1} + \alpha Q_{2j-1}, \\ C_{2j} &\equiv e^{\alpha T} c_{2j}^\dagger e^{-\alpha T} = c_{2j}^\dagger + \alpha Q_{2j}, \end{aligned} \quad (41)$$

where

$$\begin{aligned} Q_{2j-1} &= c_{2j-4} c_{2j-3}^\dagger c_{2j-2} - c_{2j-3}^\dagger c_{2j-2} c_{2j} \\ &\quad + c_{2j-2} c_{2j} c_{2j+1}^\dagger - c_{2j} c_{2j+1}^\dagger c_{2j+2}, \\ Q_{2j} &= c_{2j-3}^\dagger c_{2j-2} c_{2j-1}^\dagger - c_{2j-2} c_{2j-1}^\dagger c_{2j+1}^\dagger \\ &\quad + c_{2j-1}^\dagger c_{2j+1}^\dagger c_{2j+2} - c_{2j+1}^\dagger c_{2j+2} c_{2j+3}^\dagger. \end{aligned} \quad (42)$$

Note here that C_{2j-1} and C_{2j} are local operators in the Wannier basis associated with the c operators. Note further that C_{2j} and C_{2j-1} are obtained by a similarity transformation that is not unitary. Hence, $\{C_n, C_{n'}^\dagger\} \neq \delta_{n,n'}$, i.e., the transformed operators and their hermitian conjugates do not obey the canonical fermion creation-annihilation operator algebra. Nonetheless, the fact that $|\alpha\rangle$ is uniquely determined by demanding that $C_n |\alpha\rangle = 0$, $n = 1, \dots, 2N$, means that $|\alpha\rangle$ is the unique eigenstate of

$$H_{\text{CC}} \equiv \sum_{m=1}^{2N} C_m^\dagger C_m, \quad (43)$$

with eigenvalue zero. Since H_{CC} is semipositive definite, this means that $|\alpha\rangle$ is the unique ground state of H_{CC} . The fact that $|\alpha\rangle$ is simultaneously a zero-eigenstate of each local term $C_m^\dagger C_m$ puts H_p into a class of Hamiltonians that are called “frustration free”. Indeed, the strategy employed here, has previously been used in Ref. [44] in the context of spin chains, to construct frustration free correlated models and their ground-state manifolds from a simple frustration free starting point. After some algebra, H_p is finally cast into the form

$$\begin{aligned} H_{\text{CC}} &= \sum_{j=1}^N (c_{2j-1}^\dagger c_{2j-1} + c_{2j} c_{2j}^\dagger) - 4\alpha(T + T^\dagger) \\ &\quad + \alpha^2 \sum_{j=1}^N (Q_{2j-1}^\dagger Q_{2j-1} + Q_{2j}^\dagger Q_{2j}). \end{aligned} \quad (44)$$

Interestingly, the roles of single-particle and interaction terms are reversed compared to the original t - V Hamiltonian: The one-body term in Eq. (44) yields a flat dispersion relation, and all hopping terms are correlated. Flat dispersion and correlated hopping were previously encountered in an effective description of fermions on a Creutz ladder [45]. Here, these features emerge in another (simpler) one-dimensional setting. At small α , relevant for understanding correlations in the insulating phase at $V > 2t$ of the t - V model, there is a hierarchy, with the α^0 term, which defines the mean-field configuration, being dominant. For sufficiently small α , this guarantees that H_{CC} is gapped.

From H_{CC} , we see that three-body interactions are sufficient to stabilize the state $|\alpha\rangle$ for all α . Here, it is interesting to note a connection to η -pairing [46] generalized to spinless fermions [47,48]. The concept refers to towers of exact scar eigenstates generated by a pairing operator schematically of the form $\sum c_{j_1}^\dagger c_{j_2}^\dagger$. They were first noticed in the Hubbard model. In Ref. [47] they were found to describe many-body scars in arbitrary dimensions, in models including fine-tuned correlated hopping. In Ref. [48], such states were stabilized in a one-dimensional spinless fermion model by three-body interactions. Larger $\sum c_{j_1}^\dagger c_{j_2}^\dagger \dots c_{j_n}^\dagger$ multimers were also stabilized using n -body interactions. Our construction can be viewed as a particle-hole transformed variant of $n = 4$ tetramers that conserve particle number and that are stabilized as the ground state by three-body instead of four-body interactions.

B. MPS parent Hamiltonian

It is known that any MPS can be associated with a parent Hamiltonian involving local interactions [36,37]. However the states of the local Hilbert space do not necessarily correspond to sites of the underlying lattice, but may be blocks of sites. The starting point is to consider 2^n strings $A^{\sigma_1} \dots A^{\sigma_l}$ of the 4×4 matrices Eqs. (37) and (38) that comprise the MPS representation. In the general theory [37], the length n must be chosen such that the 2^n matrices that are generated by ranging over all possible choices of $\sigma_1, \dots, \sigma_n$ span the 16-dimensional space of 4×4 matrices. This means that $n \geq 4$. Brute force evaluations show that for the two matrices Eqs. (37) and (38), $n = 6$. The resulting parent Hamiltonian then contains terms describing interactions between adjacent six-site blocks. Since these terms have twelve-site footprints, the general MPS construction leads to a cumbersome parent Hamiltonian, which we will not report explicitly here.

C. Rokhsar-Kivelson parent Hamiltonian

We finally devise a mapping to generalized quantum dimer models, usually of relevance for insulating RVB spin liquids [49]. The precise connection can be built from quantum Hamiltonians that admit a stochastic matrix decomposition [34] and are dual to stochastic models in classical statistical physics. A preferred basis then exists for the quantum model, in which the expansion coefficients of the ground state are obtained by evaluating the partition function of the classical model. Even in cases where the ground state has a simple MPS structure, it may be very challenging to gain

insight into excited states [50]. It is therefore remarkable that for systems with stochastic matrix decomposition, equilibrium dynamic correlators of the quantum model are dual to the nonequilibrium stochastic Markovian dynamics of the classical model. Typically, the stochastic matrix decomposition requires a fine-tuning between the diagonal and off-diagonal matrix elements in the preferred basis. The fine-tuned Hamiltonian is then said to be at a Rokhsar-Kivelson point. The fact that calculations we performed for the state $|\alpha\rangle$ were equivalent to classical statistical physics calculations for a lattice gas of four-site rods (tetramers), which suggests that $|\alpha\rangle$ may be associated with a Rokhsar-Kivelson parent Hamiltonian. Here we confirm this intuition.

We have to work in the Hilbert space accounting for all possible ways that the mean-field order can be decorated with four-site disruptions. Formally, we define the index set $\mathcal{I}$ as the set of all sets of indices $\{j_1, \dots, j_m\}$ with $0 \leq 2m < N$, such that $1 \leq j_l \leq 2N$ for each index j_l , and furthermore $j_l \leq j_{l+1} - 4$. For J in $\mathcal{I}$, we define the state

$$|J\rangle = \prod_{j \in J} D_j |\text{MF}\rangle. \quad (45)$$

Periodic boundary conditions are handled by letting D_j in Eq. (23) wrap around past site 1 to sites $2N - 2, \dots, 2N$ for $j = 1, 2, 3$. The set of states $\mathcal{B} = \{|J\rangle | J \in \mathcal{I}\}$ forms a basis for the space $\mathcal{D}$ of all possible decorations of the mean-field order by four-site disruptions.

Next, we define local diagonal operators that will eventually be associated with potential energy terms in the parent Hamiltonian. The first projects out the component of a state in the $\mathcal{B}$ basis for which the five sites $j - 4, \dots, j$ are of regular order $\dots \circ \circ \circ \circ \circ \dots$,

$$|\circ \circ \circ \circ \circ\rangle \langle \circ \circ \circ \circ \circ|_j = \sum_{\{\sigma\}_j} |\{\sigma\}_j\rangle \langle \{\sigma\}_j|, \quad (46)$$

where

$$|\{\sigma\}_j\rangle = |\sigma_1 \dots \sigma_{j-5} r_{j-4} \dots r_j \sigma_{j+1} \dots \sigma_{2N}\rangle. \quad (47)$$

(We find it convenient here to denote sites of regular order with open circles and disrupted sites with dots. Note that this is different from the convention used in Fig. 2, where open/filled circles indicated the occupation of orbitals.) The second and third projects out components in the $\mathcal{B}$ basis in which a string of head-to-toe four-site disruption starts on site $j - 3$ or ends on site $j - 1$. Here and below “start” or “end” refers to where the string is bordered on respectively the left or right by a site of regular order $\dots \circ \bullet \bullet \bullet \bullet \dots$ or $\dots \bullet \bullet \bullet \bullet \circ \dots$.

$$\begin{aligned} |\circ \bullet \bullet \bullet \bullet\rangle \langle \circ \bullet \bullet \bullet \bullet|_j &= D_j |\circ \circ \circ \circ \circ\rangle \langle \circ \circ \circ \circ \circ|_j D_j^\dagger, \\ |\bullet \bullet \bullet \bullet \circ\rangle \langle \bullet \bullet \bullet \bullet \circ|_j &= D_{j-1} |\circ \circ \circ \circ \circ\rangle \langle \circ \circ \circ \circ \circ|_j D_{j-1}^\dagger. \end{aligned} \quad (48)$$

Finally, with the aid of Eq. (46) and D_j , we define four off-diagonal operators that will be associated with kinetic terms in the parent Hamiltonian. They create or destroy a four-site disruption starting at $j - 3$ when $j - 4$ is of regular order $\dots \circ \circ \circ \circ \circ \dots \leftrightarrow \dots \circ \bullet \bullet \bullet \bullet \dots$, or ending at $j - 1$ if j is

of regular order $\dots \circ \circ \circ \circ \circ \circ \dots \leftrightarrow \dots \bullet \bullet \bullet \bullet \circ \circ \dots$.

$$\begin{aligned} |\circ \bullet \bullet \bullet \bullet\rangle \langle \circ \circ \circ \circ \circ|_j &= D_j |\circ \circ \circ \circ \circ\rangle \langle \circ \circ \circ \circ \circ|, \\ |\circ \circ \circ \circ \circ\rangle \langle \circ \bullet \bullet \bullet \bullet|_j &= |\circ \circ \circ \circ \circ\rangle \langle \circ \circ \circ \circ \circ|_j D_j^\dagger, \\ |\bullet \bullet \bullet \bullet \circ\rangle \langle \circ \circ \circ \circ \circ|_j &= D_{j-1} |\circ \circ \circ \circ \circ\rangle \langle \circ \circ \circ \circ \circ|, \\ |\circ \circ \circ \circ \circ\rangle \langle \bullet \bullet \bullet \bullet \circ|_j &= |\circ \circ \circ \circ \circ\rangle \langle \circ \circ \circ \circ \circ|_j D_{j-1}^\dagger. \end{aligned} \quad (49)$$

Note that the operators Eq. (49) can be used to connect any state in $\mathcal{D}$ to any other state in $\mathcal{D}$: they allow one to add or remove a four-site disruption from either end of a string or head-to-toe disruptions. Combining two such operators allows us to move any four-site disruption whenever it is bordered by a site of regular order. We call this property “connectedness”.

In the space $\mathcal{D}$ of all possible decorations of the mean-field order by four-site disruptions, we define the Rokhsar Kivelson parent Hamiltonian

$$\begin{aligned} H_{\text{RK}} &= \sum_{j=1}^{2N} [2\alpha^2 |\circ \circ \circ \circ \circ\rangle \langle \circ \circ \circ \circ \circ|_j \\ &\quad + |\circ \bullet \bullet \bullet \bullet\rangle \langle \circ \bullet \bullet \bullet \bullet|_j + |\bullet \bullet \bullet \bullet \circ\rangle \langle \bullet \bullet \bullet \bullet \circ|_j \\ &\quad - \alpha (|\circ \bullet \bullet \bullet \bullet\rangle \langle \circ \circ \circ \circ \circ|_j + |\circ \circ \circ \circ \circ\rangle \langle \circ \bullet \bullet \bullet \bullet|_j \\ &\quad + |\bullet \bullet \bullet \bullet \circ\rangle \langle \circ \circ \circ \circ \circ|_j + |\circ \circ \circ \circ \circ\rangle \langle \bullet \bullet \bullet \bullet \circ|_j)] \\ &= \sum_{j=1}^{2N} (P_{1,j} + P_{2,j}), \end{aligned} \quad (50)$$

where

$$\begin{aligned} P_{1,j} &= (\alpha - D_j) |\circ \circ \circ \circ \circ\rangle \langle \circ \circ \circ \circ \circ|_j (\alpha - D_j^\dagger) \\ P_{2,j} &= (\alpha - D_{j-1}) |\circ \circ \circ \circ \circ\rangle \langle \circ \circ \circ \circ \circ|_j (\alpha - D_{j-1}^\dagger). \end{aligned} \quad (51)$$

This provides a phenomenological model for fluctuations of mean-field order generated by tetramers. There is a potential cost 1 associated with the ends $\circ \bullet \bullet \bullet \bullet$ and $\bullet \bullet \bullet \bullet \circ$ of strings of head-to-toe disruptions, but no energy cost associated with the bulk of the string. The kinetic terms of order α create new disruptions or grow/shrink the ends of existing ones (by four sites at a time). The fine-tuning characteristic of Rokhsar-Kivelson points amounts to a potential energy cost $2\alpha^2$ associated with not having disruptions in a region $\circ \circ \circ \circ \circ$ that could host them.

The connectedness property of the operators used to construct it, implies that H_{RK} is irreducible with respect to the basis $\mathcal{B}$: Every pair of elements in $\mathcal{B}$ has a nonzero matrix element for some power of H_{RK} . Note further that for an appropriate choice of Λ , all the matrix elements of $\Lambda I - H_{\text{RK}}$ are non-negative. The Perron-Frobenius theorem [51] then guarantees that H_{RK} has a nondegenerate ground state. H_{RK} is the sum of semipositive-definite terms. As a result, its ground-state energy is bounded from below by zero. It is straightforward to show that the nondegenerate ground state is $|\alpha\rangle = e^{\alpha T} |\text{MF}\rangle$ and that it has zero energy.

We start by writing $|\alpha\rangle = (1 + \alpha D_j) |\alpha, j\rangle$. Note that $|\alpha, j\rangle = e^{\alpha \sum_{j' \neq j} D_{j'}} |\text{MF}\rangle$ is formed without ever acting with D_j . As a result, none of the basis elements from $\mathcal{B}$ that make up $|\alpha, j\rangle$ contains a string of head-to-toe disruptions ending

on site j or starting on site $j - 3$. Then,

$$\begin{aligned} &|\circ \circ \circ \circ \circ\rangle \langle \circ \circ \circ \circ \circ|_j (\alpha - D_j^\dagger) |\alpha\rangle \\ &= \alpha |\circ \circ \circ \circ \circ\rangle \langle \circ \circ \circ \circ \circ|_j |\alpha, j\rangle \\ &\quad - \underbrace{\alpha |\circ \circ \circ \circ \circ\rangle \langle \circ \circ \circ \circ \circ|_j D_j^\dagger D_j}_{1} |\alpha, j\rangle \\ &\quad + \underbrace{\alpha^2 |\circ \circ \circ \circ \circ\rangle \langle \circ \circ \circ \circ \circ|_j D_j}_{2} |\alpha, j\rangle \\ &\quad - \underbrace{|\circ \circ \circ \circ \circ\rangle \langle \circ \bullet \bullet \bullet \bullet|_j}_{3} |\alpha, j\rangle. \end{aligned} \quad (52)$$

When acting to the left $D_j^\dagger D_j$ adds and then removes the same disruption from the regular sites $j - 3, \dots, j$ in the part marked (1), so that $|\circ \circ \circ \circ \circ\rangle \langle \circ \circ \circ \circ \circ|_j D_j^\dagger D_j = |\circ \circ \circ \circ \circ\rangle \langle \circ \circ \circ \circ \circ|_j$. The part marked (2) is zero because when acting to the left, D_j tries to remove a disruption from states that don't have any disruptions on sites $j - 3, \dots, j$. The part marked (3) is zero because $|\circ \circ \circ \circ \circ\rangle \langle \circ \bullet \bullet \bullet \bullet|_j$ looks for head-to-toe strings of disruptions that start on site $j - 3$ and $|\alpha, j\rangle$ has no such terms. Thus, $P_{1,j} |\alpha\rangle = 0$. When $|\alpha\rangle$ is written $|\alpha\rangle = (1 + \alpha D_j) |\alpha, j\rangle$, a similar argument applies for $P_{2,j+1} |\alpha\rangle$. The choice of j was arbitrary and hence $H_{\text{RK}} |\alpha\rangle = 0$.

It is immediately apparent from Eq. (50) that H_{RK} is of the stochastic matrix form {see Eqs. (15a) and (15b) in [34]}. The classical model that H_{RK} is dual to is a lattice gas of tetramers (four-site rods) in one dimension, with chemical potential $\beta\mu = 2 \ln \alpha$, but no kinetic energy. This is the physics relevant for linear adsorbates [33]. Here, the system is endowed with Markovian stochastic dynamics in the grand canonical ensemble involving random transitions $\dots \circ \circ \circ \circ \circ \circ \dots \leftrightarrow \dots \circ \bullet \bullet \bullet \bullet \circ \dots$ and $\dots \bullet \bullet \bullet \bullet \circ \dots \leftrightarrow \dots \circ \circ \circ \circ \circ \circ \dots$.

The parent Rokhsar-Kivelson Hamiltonian H_{RK} clearly differs from the parent coupled cluster Hamiltonian H_{CC} of Eq. (44). Local terms in H_{CC} have a footprint of up to seven sites, as opposed five sites for H_{RK} . Furthermore, there are terms at order α^2 in H_{CC} that move disruptions around, whereas in H_{RK} , moving a disruption is a second-order process first encountered in $(H_{\text{RK}})^2$. The translation between the different languages (coupled cluster state, matrix product state, Rokhsar-Kivelson state) describing the post-Hartree-Fock charge insulator is thus nontrivial.

VIII. NONPERTURBATIVE COUPLED CLUSTER STATE

Up to this point, we have viewed the coupled cluster state $|\alpha\rangle$ as an asymptotic description of the ground state of the t - V model for small t/V , which meant that we were only interested in the small α regime of the ansatz. However, as we argued in the previous section, the parent Hamiltonians that we identified there are of physical interest in their own right. When viewed as the ground state of such a parent Hamiltonian, the restriction that α should be small can be lifted. We set here the final goal to obtain the covariance matrix in the c_j basis at arbitrary α and in the thermodynamic limit. From

Eq. (19), we have

$$\langle \alpha | c_1^\dagger c_{1+4l} | \alpha \rangle = -\alpha^{2l} \sum_{j=0}^{\infty} \binom{2N-4l-1-3j}{j} \alpha^{2j}, \quad (53)$$

where we define $\binom{n}{j} = 0$ for $n < 0$ and we take $l > 0$. We can convert the sum into a contour integral,

$$\begin{aligned} \langle \alpha | c_1^\dagger c_{1+4l} | \alpha \rangle &= -\alpha^{2l} \sum_{n=0}^{\infty} \sum_{j=0}^n \binom{n}{j} \delta_{2N-4l-1-3j,n} \alpha^{2j} \\ &= -\alpha^{2l} \int_0^{2\pi} \frac{d\phi}{2\pi} e^{-(2N-4l-1)\phi} \\ &\quad \times \sum_{n=0}^{\infty} \sum_{j=0}^n \binom{n}{j} (e^{4i\phi} \alpha^2)^j (e^{i\phi})^{n-j} \end{aligned} \quad (54)$$

$$\begin{aligned} &= -\alpha^{2l} \int_0^{2\pi} \frac{d\phi}{2\pi} e^{-(2N-4l-1)\phi} \sum_{n=0}^{\infty} (e^{i\phi} + \alpha^2 e^{4i\phi})^n \\ &= -\alpha^{2l} \int_0^{2\pi} \frac{d\phi}{2\pi} \frac{e^{-(2N-4l-1)\phi}}{1 - e^{i\phi} - \alpha^2 e^{4i\phi}} \\ &= -\alpha^{2l} \oint_0^{2\pi} \frac{d\zeta}{2\pi i} \frac{1}{\zeta^{2N-4l} (1 - \zeta - \alpha^2 \zeta^4)}, \end{aligned} \quad (55)$$

where the integral traverses the unit circle in the complex plane in an anticlockwise sense. When applying the residue theorem to this contour integral, we encounter two types of poles. There is the high-order pole at $\zeta = 0$, which leads to a residue

$$\frac{1}{(2N-4l-1)!} \partial_{\zeta}^{2N-4l-1} \frac{1}{1 - \zeta - \alpha^2 \zeta^4} \Big|_{\zeta=0}. \quad (56)$$

It gives an $\mathcal{O}(N^0)$ contribution in the thermodynamic limit, owing to the factorial in the denominator. This contribution is overwhelmed by the residue at roots of $1 - \zeta - \alpha^2 \zeta^4$ inside the unit circle. In the thermodynamic limit and to exponential accuracy in N then,

$$\langle \alpha | c_1^\dagger c_{1+4l} | \alpha \rangle = -\frac{(\alpha^2 \zeta_0^4)^l}{\zeta_0^{2N}} \lim_{\zeta \rightarrow \zeta_0} \frac{\zeta - \zeta_0}{1 - \zeta - \alpha^2 \zeta^4}, \quad (57)$$

where ζ_0 is the real root of $1 - \zeta - \alpha^2 \zeta^4$ inside the unit circle that is closest to the origin. An explicit expression is provided in Appendix H. The root $\zeta_0(\alpha)$ is a monotonically decreasing function of α . In the small α regime, $\zeta_0 = 1 - \alpha^2 + \dots$ whereas at large α , $\zeta_0 \simeq 1/\sqrt{\alpha}$.

We can use the same method as above to obtain

$$\langle \alpha | \alpha \rangle = \frac{1 + 3\alpha^2 \zeta_0^4}{\zeta_0^{2N+1}} \lim_{\zeta \rightarrow \zeta_0} \frac{\zeta - \zeta_0}{1 - \zeta - \alpha^2 \zeta^4}. \quad (58)$$

This then gives in the thermodynamic limit

$$\frac{\langle \alpha | c_1^\dagger c_{1+4l} | \alpha \rangle}{\langle \alpha | \alpha \rangle} = -\frac{\alpha^{2l} \zeta_0^{4l+1}}{1 + 3\alpha^2 \zeta_0^4} = -\frac{(1 - \zeta_0)^l \zeta_0}{4(1 - \zeta_0) + \zeta_0}. \quad (59)$$

From Eq. (18) follows that $\langle \alpha | c_2^\dagger c_2 | \alpha \rangle / \langle \alpha | \alpha \rangle$ can be obtained from Eq. (59) by multiplying by -1 and setting $l = 0$. Since $\langle \alpha | c_1^\dagger c_1 | \alpha \rangle / \langle \alpha | \alpha \rangle = 1 - \langle \alpha | c_2^\dagger c_2 | \alpha \rangle / \langle \alpha | \alpha \rangle$, we then find for

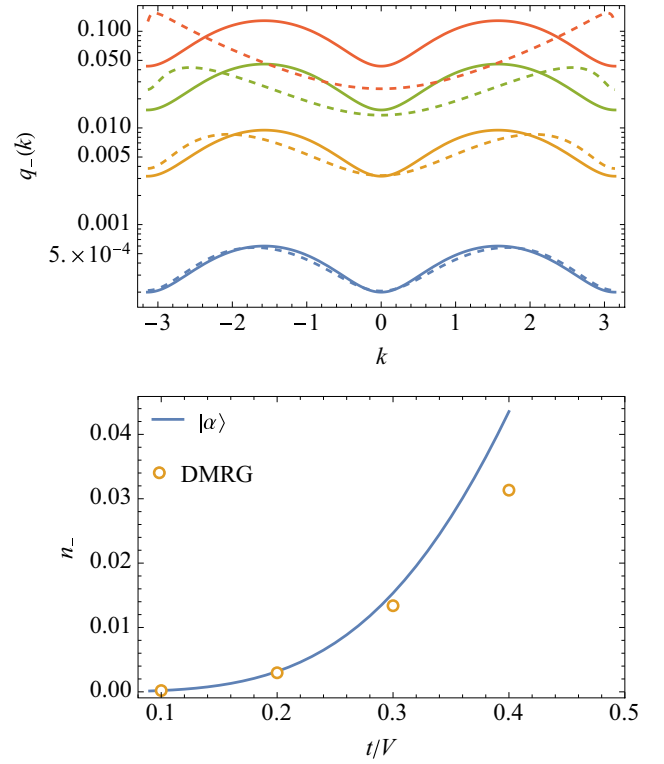


FIG. 5. (Top) Comparison of the numerically exact DMRG results (dashed curves) for the lower band of the covariance matrix spectrum of the t - V model, to (the Fourier transform of) the result Eq. (60) obtained for the nonperturbative coupled cluster state (solid curves). The value of $t/V = \sqrt{\alpha}$ for curves from bottom to top are $t/V = 0.1, 0.2, 0.3, 0.4$. (Bottom) The density n_- of quasiparticles associated with the upper band of the mean-field Hamiltonian vs t/V .

all l in the thermodynamic limit

$$\frac{\langle \alpha | c_1^\dagger c_{1+4l} | \alpha \rangle}{\langle \alpha | \alpha \rangle} = \delta_{l,0} - \frac{(1 - \zeta_0)^l \zeta_0}{4(1 - \zeta_0) + \zeta_0}. \quad (60)$$

The full covariance matrix $Q_{jj'} = \langle c_j^\dagger c_{j'} \rangle$ is deduced from the two-site translation symmetry of the charge ordered state. It is interesting to compare this exponential decay for the coupled cluster state to the decay found for the critical system at $V = -2t$. In both cases, an ordered state ($|\text{MF}\rangle$ in the former and the fermionic vacuum $|0\rangle$ in the latter case) is decorated with a finite density of randomly placed disruptions. For the coupled cluster state, the disruptions D_j contain four adjacent fermion operators, while for the critical state, the disruptions are single fermions. Randomness on its own causes an exponential decay of single-particle correlations with a finite-decay length Eq. (60). The additional destructive interference caused by the fermionic sign fluctuations when the disruptions are single fermions, cause exponential decay with a vanishing decay length $1/\ln N$ at large N , see Eq. (5).

While we have taken the view in this section that $|\alpha\rangle$ with $\alpha > 0.01$ is an interesting state in its own right, we conclude this section by pointing out that this state, although not quantitatively accurate, still captures reasonably well the extent of correlations in the t - V model at $t/V > 0.1$. In the top panel of Fig. 5 we compare the lower band of the covariance matrix

spectrum at values of t/V between 0.1 and 0.4, obtained from numerically exact DMRG calculations, to the result obtained by Fourier transforming Eq. (60) with $\alpha = (t/V)^2$. We see that the average position and bandwidth agree rather well between the two calculations, although the progressive change in periodicity is of course not captured by the coupled cluster state.

In the bottom panel of Fig. 5, we also compare results for the average density of the quasiparticles associated with the upper band of the mean-field Hamiltonian, i.e.,

$$n_- = \langle c_1^\dagger c_1 \rangle / 2. \quad (61)$$

(The factor 1/2 reflects the fact that the unit cell contains two chain sites.) We see good agreement that is not automatically guaranteed. For instance, the post-Hartree-Fock treatment in [24] cannot reproduce this agreement (see Figs. 12 and 13 in that work). While the actual disruption structure at $V \gtrsim 2t$ becomes more complicated than the simple four-site pattern of $|\alpha = (t/V)^2\rangle$, the latter state nonetheless still contains roughly the right amount of correlations, and constitutes a natural physical picture of correlated insulators.

IX. CONCLUSIONS AND PERSPECTIVES

In this work, we have uncovered a simple coupled cluster state, given by Eq. (17), that asymptotically describes the correlated charge insulating ground state of the t - V model in one dimension at large V/t . This emergent structure remains hidden in usual exact treatments of one-dimensional models, such as the Bethe ansatz or the density matrix renormalization group. Our result was obtained by bootstrapping from a perturbative calculation of the covariance matrix spectrum Eq. (10) to a coupled cluster state that variationally reproduces the ground-state energy to one order further than the Hartree-Fock mean-field theory of the correlated charge insulator. This enabled us to answer analytically questions that cannot be addressed by perturbative means. We showed that the leading order fluctuations beyond mean-field are the result of random decorations of the ordered state by four-site disruptions of opposite orientation to the underlying mean-field charge pattern (schematically illustrated in Fig. 2). We also showed that the four-site disruptions state leads to a matrix product state with bond dimension exactly four Eq. (40). While there is a general construction [36] that relates matrix product states to parent Hamiltonians with local terms only, that route does not seem to yield a parent Hamiltonian of much practical use in this case. We have instead derived two distinct parent Hamiltonians. One parent Hamiltonian, given by Eq. (50), constitutes a phenomenological model for quantum fluctuations induced by tetramers (length four rods), that connects the quantum system to the classical stochastic dynamics of linear adsorbates [33] when tuned to its Rokhsar-Kivelson point [34]. The other parent Hamiltonian, given by Eq. (44), is a correlated hopping Hamiltonian involving up to only three-body interactions. We have also obtained exact results for the covariance matrix Eq. (60) of the coupled cluster state in the thermodynamic limit, and showed that the level of correlation is similar to that of the t - V model even in a regime where there are quantitative differences between the coupled cluster and exact ground states.

Our results make the case for further studies of the covariance matrix of systems in which ground states have nontrivial unit-cell structure. We briefly speculate here regarding potential future developments. One interesting question is how to break free from the four-site periodic structure of the coupled cluster state based on four-site disruptions. This is not only required to improve agreement at larger values of t/V in the correlated insulator phase, but also to understand which are the new physical processes at play in this regime. Another promising avenue is to use the coupled cluster state structure that we uncovered as a starting point to build more intricate many-body states in one dimension. A direct and interesting extension of our work would be to consider one-dimensional models of Mott insulators with spinful fermions, as well as ladders, and to investigate the emergent structures in the presence of spin and charge degrees of freedom. In addition, since the exact gapless critical state was found to show some similarities to the gapped coupled cluster state, it is conceivable that further extension of our work could lead to an explicit construction of correlated wave functions for the Luttinger liquid phase. Studies of correlated phases of charge insulators beyond one dimension [52] could also be considered, using the mapping to generalized quantum multimer models. Another avenue to be explored relates to topologically nontrivial phases of interacting systems. For correlated systems with nontrivial unit-cell structure, the covariance matrix in k -space may play a similar role to a single-particle Hamiltonian, thus allowing one to borrow from the theory of topological band insulators [53]. In this context, there are possibly subtle but interesting questions about if and how the bulk-edge correspondence of the band insulator theory carries over to the theory of the covariance matrix of a correlated insulator.

ACKNOWLEDGMENTS

This work was supported by European Research Council (ERC) under Grant Agreement No. 101001310 466 (Super-Protected). I.S. thanks the Institut Néel for financial mobility support.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available upon publication because it is not technically feasible and/or the cost of preparing, depositing, and hosting the data would be prohibitive within the terms of this research project. The data are available from the authors upon reasonable request.

APPENDIX A: EXACT GROUND STATE AT $V = -2t$

In the Appendix of [39], it was proven that Eq. (3) is the ground state of the t - V model at $V = -2t$. The proof was performed in the fermionic language. Here, we sketch the equivalent proof in the spin-chain language, which is related to the fermion language by the Wigner-Jordan mapping to spin-1/2 operators

$$S_{j+} = S_{xj} + iS_{yj} = e^{i\pi \sum_{l=1}^{j-1} a_l^\dagger a_l} a_j^\dagger \quad (A1)$$

that commute if $j \neq j'$. In this representation the Hamiltonian at $V = -2t$ corresponds to the spin-isotropic, ferromagnetic Heisenberg spin chain,

$$H = -\frac{t}{2} \sum_{j=1}^{2N} \mathbf{S}_j \cdot \mathbf{S}_{j+1}. \quad (\text{A2})$$

The total spin and its z component are good quantum numbers. The fully polarized ground state is the state $|J = N, J_z = -N\rangle$, which corresponds to the empty state $|0\rangle$ in the fermionic language. The Hamiltonian commutes with $S_+ = \sum_{j=1}^{2N} S_{j+}$, and thus the ground state with spin projection $J_z = 0$ is given exactly by

$$|J = N, J_z = 0\rangle \propto (S_+)^N |J = N, J_z = -N\rangle. \quad (\text{A3})$$

In the fermion language, this leads to the half-filled state Eq. (3), as is easy to check using the mapping Eq. (A1).

APPENDIX B: COVARIANCE MATRIX AT $V = -2t$

Here, we perform the sum over combinatorial factors

$$S = \sum_{n=0}^{2r+1} (-1)^n \binom{2r}{n} \binom{2N-2r-2}{N-n-1} \quad (\text{B1})$$

that produces the result Eq. (5). It is straightforward to see that

$$\begin{aligned} & \binom{2r}{n} \binom{2N-2r-2}{N-n-1} \\ &= \frac{(2r)!(2N-2r-2)!}{[(N-1)!]^2} \binom{N-1}{n} \binom{N-1}{2r-n}, \end{aligned} \quad (\text{B2})$$

so that

$$S = \frac{(2r)!(2N-2r-2)!}{[(N-1)!]^2} \underbrace{\sum_{n=0}^{2r+1} (-1)^n \binom{N-1}{n} \binom{N-1}{2r-n}}_{(\text{B3})}. \quad (\text{B3})$$

The part marked with an underbrace is easily identified as the coefficient of the monomial x^{2r} in the polynomial $(1-x)^{N-1}(1+x)^{N-1} = (1-x^2)^{N-1}$. Thus, by the binomial theorem

$$\begin{aligned} S &= \frac{(2r)!(2N-2r-2)!}{[(N-1)!]^2} \binom{N-1}{r} \\ &= (-1)^r \frac{(2r)!(2N-2r-2)!}{r!(N-r-1)!(N-1)!} \\ &= (-1)^r \frac{2^{2N-2} \Gamma(r+1/2) \Gamma(N-r-1/2)}{\Gamma(1/2)^2 (N-1)!}, \end{aligned} \quad (\text{B4})$$

which decays like N^{-r} at large N , because of the r unbalanced factors of order N in the denominator of $(N-r-3/2)(N-r-5/2) \dots (1/2)/(N-1)(N-2) \dots (1)$.

APPENDIX C: COVARIANCE MATRIX AT ORDER $(t/V)^4$

In the main text we present the covariance matrix spectrum of the t - V model at small t/V , to order $(t/V)^4$. It is obtained from the following expansion of the covariance matrix, to

order $(t/V)^4$ in k -space:

$$\begin{aligned} Q(k) &= \frac{1}{2} \sigma_0 + \left[(1 + \cos k) \frac{t}{V} - (2 + 5 \cos k + 3 \cos 2k) \left(\frac{t}{V} \right)^3 \right] \sigma_x \\ &+ \left[\sin k \frac{t}{V} + (\sin k - 3 \sin 2k) \left(\frac{t}{V} \right)^3 \right] \sigma_y \\ &+ \left[-\frac{1}{2} + 2(1 + \cos k) \left(\frac{t}{V} \right)^2 \right. \\ &\quad \left. + (2 - 12 \cos k - 12 \cos 2k) \left(\frac{t}{V} \right)^4 \right] \sigma_z. \end{aligned} \quad (\text{C1})$$

As mentioned in the main text, the spectrum of $Q(k)$ has periodicity π whereas $Q(k)$ itself only has periodicity 2π .

APPENDIX D: MEAN-FIELD SOLUTION OF THE t - V MODEL

We seek a single-particle basis

$$b_\alpha^\dagger = \sum_{j=1}^{2N} f_{j\alpha} a_j^\dagger, \quad (\text{D1})$$

such that for a particular sub-set $\mathcal{F}$ of N of these orbitals, the state

$$\left(\prod_{\alpha \in \mathcal{F}} b_\alpha^\dagger \right) |0\rangle, \quad (\text{D2})$$

minimizes the expectation value of the Hamiltonian Eq. (1), within the manifold of single Slater determinants. Standard Hartree-Fock theory leads to a mean-field Hamiltonian

$$\varepsilon_\alpha f_{j\alpha} = \sum_{k=1}^{2N} h_{jk} f_{k\alpha}, \quad (\text{D3})$$

for the mean-field orbitals, where

$$\begin{aligned} h_{jk} &= -(t + \langle a_{j+1}^\dagger a_j \rangle) \delta_{k,j+1} - (t + \langle a_k^\dagger a_{k+1} \rangle) \delta_{j,k+1} \\ &+ V(\langle a_{j-1}^\dagger a_{j-1} \rangle + \langle a_{j+1}^\dagger a_{j+1} \rangle) \delta_{j,k}. \end{aligned} \quad (\text{D4})$$

Periodic boundary conditions here imply $2N+1 \equiv 1$ and $0 \equiv 2N$. We seek an inversion-symmetric solution that is invariant under translation by two lattice sites. We therefore take $\langle a_j^\dagger a_j \rangle = \langle a_{j+2}^\dagger a_{j+2} \rangle$, and $\langle a_{j+1}^\dagger a_j \rangle = \langle a_{j-1}^\dagger a_j \rangle$ for all j . We further assume $\langle a_{j+1}^\dagger a_j \rangle$ to be real.

This results in two bands with N orbitals each. For the lower band

$$f_{2j-1,\alpha,-} = \frac{x_\alpha}{\sqrt{N}} e^{ik_\alpha j}, \quad f_{2j,\alpha,-} = \frac{y_\alpha}{\sqrt{N}} e^{ik_\alpha j}, \quad (\text{D5})$$

with $k_\alpha = 2\pi\alpha/N$, and $\alpha = -\frac{N-1}{2}, \dots, \frac{N-1}{2}$, N being assumed odd. The coefficients x_α and y_α form the negative-energy eigenstate of the eigenvalue equation

$$-\varepsilon_\alpha \begin{pmatrix} x_\alpha \\ y_\alpha \end{pmatrix} = \begin{pmatrix} \Delta & -\tilde{t}(1 + e^{-ik_\alpha}) \\ -\tilde{t}(1 + e^{ik_\alpha}) & -\Delta \end{pmatrix} \begin{pmatrix} x_\alpha \\ y_\alpha \end{pmatrix}, \quad (\text{D6})$$

where the mean-field parameters

$$\begin{aligned}\Delta &= V(\langle a_{2j}^\dagger a_{2j} \rangle - \langle a_{2j-1}^\dagger a_{2j-1} \rangle), \\ \tilde{t} &= t + V \langle a_{j+1}^\dagger a_j \rangle,\end{aligned}\quad (\text{D7})$$

do not depend on j . From this we find

$$\varepsilon_\alpha = \sqrt{\Delta^2 + 4\tilde{t}^2 \cos^2 \frac{k_\alpha}{2}}, \quad (\text{D8})$$

$$x_\alpha = \frac{1}{\sqrt{2}} \sqrt{1 - \frac{z}{\sqrt{z^2 + \cos^2 \frac{k_\alpha}{2}}}} e^{-ik_\alpha/2}, \quad (\text{D9})$$

$$y_\alpha = \frac{1}{\sqrt{2}} \sqrt{1 + \frac{z}{\sqrt{z^2 + \cos^2 \frac{k_\alpha}{2}}}}. \quad (\text{D10})$$

where $z = \Delta/2\tilde{t}$. The upper band is described by orbitals

$$f_{2j-1,\alpha,+} = \frac{y_\alpha}{\sqrt{N}} e^{ik_\alpha j}, \quad f_{2j,\alpha,+} = -\frac{x_\alpha^*}{\sqrt{N}} e^{ik_\alpha j}. \quad (\text{D11})$$

To obtain self-consistency, we evaluate

$$\begin{aligned}\langle a_{2j}^\dagger a_{2j} \rangle - \langle a_{2j-1}^\dagger a_{2j-1} \rangle &= \frac{1}{N} \sum_{\alpha=-\frac{N-1}{2}}^{\frac{N-1}{2}} (y_\alpha^2 - |x_\alpha|^2) \\ &= \frac{1}{N} \sum_{\alpha=-\frac{N-1}{2}}^{\frac{N-1}{2}} \frac{\Delta}{\varepsilon_\alpha},\end{aligned}\quad (\text{D12})$$

$$\begin{aligned}\langle a_{2j}^\dagger a_{2j-1} \rangle &= \frac{1}{N} \sum_{\alpha=-\frac{N-1}{2}}^{\frac{N-1}{2}} x_\alpha y_\alpha \\ &= \frac{1}{N} \sum_{\alpha=-\frac{N-1}{2}}^{\frac{N-1}{2}} \frac{\tilde{t} \cos^2 \frac{k_\alpha}{2}}{\varepsilon_\alpha},\end{aligned}\quad (\text{D13})$$

which are independent of j . We also see that

$$\begin{aligned}\langle a_{2j+1}^\dagger a_{2j} \rangle &= \frac{1}{N} \sum_{\alpha=-\frac{N-1}{2}}^{\frac{N-1}{2}} x_\alpha^* e^{-ik_\alpha} y_\alpha \\ &= \frac{1}{N} \sum_{\alpha=-\frac{N-1}{2}}^{\frac{N-1}{2}} x_\alpha y_\alpha = \langle a_{2j}^\dagger a_{2j-1} \rangle,\end{aligned}\quad (\text{D14})$$

which confirms that it is consistent to take $\langle a_{j+1}^\dagger a_j \rangle$ independent of j and real. All the above leads to a transcendental equation

$$\frac{t}{V} = \frac{1}{2N} \sum_{\alpha=-\frac{N-1}{2}}^{\frac{N-1}{2}} \frac{\sin^2 \frac{k_\alpha}{2}}{\sqrt{z^2 + \cos^2 \frac{k_\alpha}{2}}}, \quad (\text{D15})$$

which determines z . Without loss of generality, we pick the $z > 0$ solution, which corresponds to a symmetry-broken state in which the even sites are more likely to be occupied than the odd ones.

The mean-field orbitals that we have found are Bloch states of the reduced symmetry mean-field Hamiltonian Eq. (D4).

Since we are studying an insulating state, a description in terms of localized orbitals is however more useful. To obtain such a description, we note that unitary transforms that do not mix orbitals from different bands, leave invariant the mean-field state that is constructed by completely filling a single band. We use this freedom to re-Wannierize the mean-field orbitals within a given band. The new Wannier orbitals are then given by

$$\begin{aligned}\xi(j) &= \frac{1}{N} \sum_{\alpha=-\frac{N-1}{2}}^{\frac{N-1}{2}} x_\alpha e^{i\phi_\alpha^-} e^{ik_\alpha j}, \\ \psi(j) &= \frac{1}{N} \sum_{\alpha=-\frac{N-1}{2}}^{\frac{N-1}{2}} y_\alpha e^{i\phi_\alpha^-} e^{ik_\alpha j}, \\ \tilde{\xi}(j) &= \frac{1}{N} \sum_{\alpha=-\frac{N-1}{2}}^{\frac{N-1}{2}} x_\alpha^* e^{i\phi_\alpha^+} e^{ik_\alpha j}, \\ \tilde{\psi}(j) &= \frac{1}{N} \sum_{\alpha=-\frac{N-1}{2}}^{\frac{N-1}{2}} y_\alpha e^{i\phi_\alpha^+} e^{ik_\alpha j},\end{aligned}\quad (\text{D16})$$

where the phases $\phi_\alpha^\pm$ have not been fixed yet, and may be adjusted to give optimal results in whatever scheme is employed on top of mean field theory. For our purposes, we used $\phi_\alpha^\pm = 0$, which gave a lower post-Hartree-Fock variational energy than other constant choices. We did not systematically vary over momentum-dependent choices.

We use these orbitals to construct new fermion creation operators $c_j^\dagger$, that are associated with Wannierized valence (even j) and conduction (odd j) band orbitals of the mean-field Hamiltonian. In the position basis they read

$$\begin{aligned}c_{2j}^\dagger &= \sum_{l=1}^N [\xi(l-j) a_{2l-1}^\dagger + \psi(l-j) a_{2l}^\dagger], \\ c_{2j-1}^\dagger &= \sum_{l=1}^N [\tilde{\psi}(l-j) a_{2l-1}^\dagger - \tilde{\xi}(l-j) a_{2l}^\dagger].\end{aligned}\quad (\text{D17})$$

It is useful to note that in terms of the parameters defined here, the expectation value of the Hamiltonian with respect to the mean-field state is simply given by

$$E_{\text{mf}} = -2NV \left(\frac{\tilde{t}^2 - t^2}{V^2} + \frac{\Delta^2}{4} \right). \quad (\text{D18})$$

In order to compare this result with perturbation theory, we wish to expand it in powers of t/V . As a first step, we expand Δ^2/V^2 in $2\tilde{t}/\Delta = 1/z$. From Eqs. (D7) and (D12) follows

$$\begin{aligned}\frac{\Delta^2}{V^2} &= \frac{1}{N} \left[\sum_{\alpha} \left(1 - \frac{1}{2z^2} \cos^2 \frac{k_\alpha}{2} + \frac{3}{8z^4} \cos^4 \frac{k_\alpha}{2} + \dots \right) \right]^2 \\ &= \left(1 - \frac{1}{4z^2} + \frac{9}{64z^4} + \dots \right)^2 \\ &= 1 - \frac{1}{2z^2} + \frac{11}{32z^4} + \dots\end{aligned}\quad (\text{D19})$$

We can find $1/z$ in terms of t/V with the help of Eq. (D15),

$$\begin{aligned} \frac{2t}{V} &= \frac{1}{Nz} \sum_{\alpha} \frac{\sin^2 \frac{k_{\alpha}}{2}}{\sqrt{1 + \frac{\cos^2 \frac{k_{\alpha}}{2}}{z^2}}} \\ &= \frac{1}{Nz} \sum_{\alpha} \left(\sin^2 \frac{k_{\alpha}}{2} - \frac{1}{2z^2} \sin^2 \frac{k_{\alpha}}{2} \cos^2 \frac{k_{\alpha}}{2} + \dots \right) \\ &= \frac{1}{2z} \left[1 - \frac{1}{8z^2} + \dots \right] \end{aligned} \quad (\text{D20})$$

Solving the above for $1/z$, we find

$$\frac{1}{z} = \frac{4t}{V} + \frac{8t^3}{V^3} + \dots \quad (\text{D21})$$

Substituting these results into Eq. (D18), we find

$$E_{\text{mf}} = -\frac{NV}{2} \left(1 + \frac{4t^2}{V^2} - \frac{8t^4}{V^4} + \dots \right). \quad (\text{D22})$$

Comparing to the expansion of the exact result Eq. (12), we see that the fourth order term of the Hartree-Fock solution is double what it should be. Correlations, included in the coupled cluster state, are responsible for the proper value of the fourth order term.

APPENDIX E: MATRIX ELEMENTS FOR TWO-BODY ADDITIVE OPERATORS

Products of four c operators, three of whose site indices have one parity (even/odd) and one whose site index is of opposite parity, necessarily change the number of sites of regular order by one. Such operators thus map $|m\rangle$ onto a state in the orthogonal complement to the space spanned by the $|m'\rangle$ $m' = 0, 1, \dots, [N/2]$ basis. We therefore only have to consider operator products in which all four site indices have the same parity, or where two have one parity while the other two have the opposite parity.

1. Operator products that increase m by one

Let us consider a two-body operator that increases m by one. It necessarily consists of two creation operators with odd site indices and two annihilation operators with even indices. They fuse three disruptions (some of which could be length zero), that are separated by two single sites of regular order into a single disruption, and grow the fused disruption by one site at each end. Due to invariance of the state $|m\rangle$ under translation by two lattice sites, we can restrict our attention to products in which one of the creation operators has site index 1. We use three labels l_1, l_2 and l_3 , corresponding to the number of atomic disruptions in each of the three disruptions that are fused, to label these operator products. There are four separate cases to consider, corresponding to how the fused disruption is positioned relative to the link between sites N and 1. Note that because N is odd, the initial Slater determinant on which the operator product acts always hosts an odd number of particles in the space between the ends of the eventual fused disruption.

(1) $c_1^\dagger c_{4l_1+2} c_{4(l_1+l_2)+3}^\dagger c_{4(l_1+l_2+l_3)+1}$: Suppose the fused disruption starts at site 1. The corresponding initial Slater determinant contains a site of regular order at site 1 (i.e.,

unoccupied), followed by a disruption on sites 2 to $4l_1 + 1$. A site of regular order at site $4l_1 + 2$ is occupied. Then follows a disruption from site $4l_1 + 3$ to site $4(l_1 + l_2) + 2$. A site of regular order at $4(l_1 + l_2) + 3$ is empty. Finally comes a disruption from site $4(l_1 + l_2) + 4$ to $4(l_1 + l_2 + l_3) + 3$ and an occupied site of regular order at site $4(l_1 + l_2 + l_3) + 3$.

$$\dots | 0[0 \ 1 \ 0 \ 1]1[1 \ 0 \ 1 \ 0]0[0 \ 1 \ 0 \ 1]1\dots$$

$$\rightarrow \dots | [1 \ 0 \ 1 \ 0 \ 1 \ 0 \ 1 \ 0 \ 1 \ 0 \ 1 \ 0 \ 1 \ 0] \dots$$

Here and below $|$ separates site $2N$ and site 1 and. Because of the initial presence of a particle on site $4l_1 + 2$ the operators $c_{4(l_1+l_2+l_3)+1}^\dagger$ and $c_{4(l_1+l_2)+3}^\dagger$ each has to be anticommutated past an odd number of fermion operators while c_{4l_1+2} must be anticommutated past an even number. This results in an overall sign of 1.

(2) $c_1^\dagger c_{2N-4(l_1+l_2+l_3)-2} c_{2N-4(l_2+l_3)-1}^\dagger c_{2N-4l_3}$: Suppose the fused disruption wraps around and ends at site 1. Initially sites 1 is empty while site $2N - 4(l_1 + l_2 + l_3) - 2$, just in front of the first unfused disruption is occupied. Site $2N - 4(l_2 + l_3) - 1$ between the first and second unfused disruption is empty.

$$|0\dots \ 1[1 \ 0 \ 1 \ 0]0[0 \ 1 \ 0 \ 1]1[1 \ 0 \ 1 \ 0]|$$

$$\rightarrow |1]\dots [0 \ 1 \ 0 \ 1 \ 0 \ 1 \ 0 \ 1 \ 0 \ 1 \ 0 \ 1 \ 0 \ 1]$$

The odd number of particles hosted in sites 2 to $2N - 4(l_1 + l_2 + l_3) - 2$ and the presence of a particle on site $2N - 4(l_1 + l_2 + l_3) - 2$ means that $2c_{2N-4(l_2+l_3)-1}$ and c_{2N-4l_3} must each be anticommutated past an even number of particles, while $c_{2N-4(l_1+l_2+l_3)-2}$ is commuted past an odd number of particles. The overall sign that results is -1 .

(3) $c_1^\dagger c_{4l_1+2} c_{2N-4(l_2+l_3)-1}^\dagger c_{2N-4l_3}$: One of the three fused disruptions fully wraps around past site 1 (so that site 1 is one of the fusion sites). In the initial Slater, site 1, is therefore empty. Site c_{4l_1+2} , just after the insertion that wrapped past 1, is initially filled. Site $2N - 4(l_2 + l_3) - 1$, just in front of the remaining two insertions is initially empty.

$$| \ 0[0 \ 1 \ 0 \ 1]1\dots \ 0[0 \ 1 \ 0 \ 1]1[1 \ 0 \ 1 \ 0]|$$

$$\rightarrow | \ 1 \ 0 \ 1 \ 0 \ 1 \ 0]\dots [1 \ 0 \ 1 \ 0 \ 1 \ 0 \ 1 \ 0 \ 1 \ 0 \ 1]$$

Anticommuting $c_{2N-4(l_2+l_3)-1}^\dagger$ and c_{2N-4l_3} past odd number of particles in the space between the ends of the eventual fused insertion, and the particle on site $4l_1 + 2$ gives a sign of 1. An even number of anticommutations puts c_{4l_1+2} in place. Thus, the overall sign is 1.

(4) $c_1^\dagger c_{4l_1+2} c_{4(l_1+l_2)+3}^\dagger c_{2N-4l_3}$: Two of the three fused disruptions fully wrap around past site 1. Site 1, in front of the two insertions that wrap past 1 is initially empty. Site $4l_1 + 2$ between the first and second wrapped insertion is initially filled. Site $4(l_1 + l_2) + 3$ just after the second wrapped insertion is initially empty.

$$| \ 0[0 \ 1 \ 0 \ 1]1[1 \ 0 \ 1 \ 0]0\dots \ 1[1 \ 0 \ 1 \ 0]|$$

$$\rightarrow | \ 1 \ 0 \ 1 \ 0 \ 1 \ 0 \ 1 \ 0 \ 1 \ 0 \ 1]\dots [0 \ 1 \ 0 \ 1 \ 0 \ 1] \quad (\text{E1})$$

Anticommuting c_{2N-4l_3} past the odd number of particles in the space between the ends of the eventual fused insertion, and past the particle on site $4l_1 + 2$, gives a sign of 1. Anticommuting $c_{4(l_1+l_2)+3}^\dagger$ past the particle on site $4l_1 + 2$, gives a sign

of -1 . Getting c_{4l_1+2} in place only requires anticommuting it past the particles in one of the initial unfused disruptions, and thus gives a sign 1. The overall sign is then -1 .

In each of the above cases, the number of Slater determinants in $|m\rangle$ that fit the bill can be counted as follow. The three initially unfused insertions, together with the sites in

$$\begin{aligned} \langle m+1 | c_1^\dagger c_{4l_1+2} c_{4(l_1+l_2)+3}^\dagger c_{4(l_1+l_2+l_3)+1} | m \rangle &= -\langle m+1 | c_1^\dagger c_{2N-4(l_1+l_2+l_3)-2} c_{2N-4(l_2+l_3)-1}^\dagger c_{2N-4l_3} | m \rangle \\ &= \langle m+1 | c_1^\dagger c_{4l_1+2} c_{2N-4(l_2+l_3)-1}^\dagger c_{2N-4l_3} | m \rangle = -\langle m+1 | c_1^\dagger c_{4l_1+2} c_{4(l_1+l_2)+3}^\dagger c_{2N-4l_3} | m \rangle = c_{l_1, l_2, l_3}^{1122}, \end{aligned} \quad (\text{E2})$$

where

$$c_{l_1, l_2, l_3}^{1122} = \binom{2N - (3m + l_1 + l_2 + l_3) - 4}{m - l_1 - l_2 - l_3}. \quad (\text{E3})$$

The restriction that the initial unfused disruptions do not overlap, and that there should be no more than m initial atomic disruptions, lead to the restrictions

$$\begin{aligned} l_1 &\in \left\{ 0, \min \left(\frac{N-1}{2} - 1, m \right) \right\}, \\ l_2 &\in \left\{ 0, \min \left(\frac{N-1}{2} - 1 - l_1, m - l_1 \right) \right\}, \\ l_3 &\in \left\{ 0, \min \left(\frac{N-1}{2} - 1 - l_1 - l_2, m - l_1 - l_2 \right) \right\}. \end{aligned} \quad (\text{E4})$$

Note that zero lengths are allowed for the unfused initial insertions. For instance, $l_1 = l_2 = l_3 = 0$ corresponds to a single atomic disruptions being created on the four sites from 1 to 4, if initially they are of regular order.

$$\dots | 0101 \dots \rightarrow \dots | [1010] \dots$$

Matrix elements of operators that reduce m by one are obtained by taking the complex conjugates of the operators that we calculated here.

2. A product in which all four site indices are distinct

Such an operator product changes each of the Slater determinants $|s\rangle = c_{s_1}^\dagger \dots c_{s_N}^\dagger |0\rangle$ that comprise $|m\rangle$, into another Slater determinant $|s'\rangle = c_{s'_1}^\dagger \dots c_{s'_N}^\dagger |0\rangle$, up to a sign. Calculating the associated expectation value amounts to counting the number of these new Slater determinants $|s'\rangle$ that are also part of $|m\rangle$, or $|m \pm 1\rangle$, and keeping track of signs.

In each case, two disruptions are moved. There are four subclasses, determined by the directions in which the disruptions are moved (left-left, right-right, left-right, and right-left). We use three integer labels j , l_1 , and l_2 to enumerate the different operator products that yield a nonzero expectation value.

(1) Four odd site indices: Without loss of generality, we can consider one creation operator to have site-index 1. All other cases give the same answer, because of invariance of the states under consideration by two lattice sites. These operator products preserve the number of sites of regular order, and thus only connect $|m\rangle$ to itself. The following mutually exclusive classes have nonzero $|m\rangle$ expectation values.

between, in front of, or behind them, leaves $m - l_1 - l_2 - l_3$ atomic disruptions and $2N - 4m - 4$ sites of regular order to place in the space between the ends of the eventual fused insertion. There are $\binom{2N - (3m + l_1 + l_2 + l_3) - 4}{m - l_1 - l_2 - l_3}$ ways to do this. Each of these terms in $|m\rangle$ is mapped onto a distinct term in $|m\rangle$, with prefactor ± 1 as determined above. Thus,

(a) $c_1^\dagger c_{4l_1+1} c_{2j+1}^\dagger c_{2j+1+4l_2}$: Suppose a Slater determinant contains a disruption starting on site 2 and extending to site $4l_1 + 1$ and another present at site $2j + 2$ and ending at site $2j + 4l_2 + 1$. The operator product moves the first disruption one site to the left. This is achieved by anticommuting c_{4l_1+1} past all but the last fermion in the first disruption, thus accumulating a sign of -1 . It also moves the second disruption one site to the left by anticommuting $c_{2j+1+4l_2}$ past all but the rightmost fermion in the second disruption, thus accumulating a sign -1 .

$$\begin{aligned} \dots | 0[0 \ 1 \ 0 \ 1] \dots \quad 0[0 \ 1 \ 0 \ 1] \dots \\ \rightarrow \dots | [1 \ 0 \ 1 \ 0] 0 \dots [1 \ 0 \ 1 \ 0] 0 \dots \end{aligned}$$

The total sign is thus 1.

(b) $c_1^\dagger c_{2j+1-4l_1} c_{2j+1}^\dagger c_{2N-4l_2+1}$: Suppose a Slater determinant contains a disruption present on site $2j + 1 - 4l_1$ and ending at site $2j$ and another present at site $2N + 4l_2 + 1$ and ending at site $2N$. The operator product moves the first disruption one site to the right. This is achieved by anticommuting $c_{2j+1}^\dagger$ past all the fermions in the first disruption, thus accumulating a sign of 1. It also moves the second disruption one site to the right, so that its end wraps around to site 1. It does so by anticommuting c_{2N-4l_2+1} past all the fermions that are not part of the right disruption. Since N is odd, and disruptions contain an even number of fermions, this produces a sign -1 .

$$\begin{aligned} | 0 \dots [1 \ 0 \ 1 \ 0] 0 \dots [1 \ 0 \ 1 \ 0] | \\ \rightarrow | 1] \dots 0[0 \ 1 \ 0 \ 1] \dots 0[0 \ 1 \ 0 \ 1] \end{aligned}$$

The total sign accumulated is then -1 .

(c) $c_1^\dagger c_{4l_1+1} c_{2j+1}^\dagger c_{2j+1+4l_2}$: Suppose a Slater determinant contains a disruption starting on site 2 and extending to site $4l_1 + 1$ and another present at site $2j + 1$ and ending at site $2j + 4l_2$. The operator product moves the first disruption one site to the left. This is achieved by anticommuting c_{4l_1+1} past all but the right-most fermion in the first disruption, thus accumulating a sign of -1 . It moves the second disruption one site to the right by anticommuting $c_{2j+1+4l_2}^\dagger$ past all but the fermions in the second disruption, thus accumulating a sign 1.

$$\begin{aligned} \dots | 0[0 \ 1 \ 0 \ 1] \dots [1 \ 0 \ 1 \ 0] 0 \dots \\ \rightarrow \dots | [1 \ 0 \ 1 \ 0] 0 \dots 0[0 \ 1 \ 0 \ 1] \dots \end{aligned}$$

The total sign is thus -1 .

(d) $c_1^\dagger c_{2j+1-4l_1}^\dagger c_{2j+1} c_{2N-4l_2+1}$: Suppose a Slater determinant contains a disruption starting on site $2j+2-4l_1$ and extending to site $2j+1$ and another present at site $2N+4l_2+1$ and ending at site $2N$. The operator product moves the first disruption one site to the left. This is achieved by anticommuting c_{2j+1} past all but the rightmost fermion in the first disruption, thus accumulating a sign of -1 . It also moves the second disruption one site to the right, so that it end wraps around to site 1. It does so by anticommuting c_{2N-4l_2+1} past all the fermions that are not part of the right disruption. This produces a sign -1 .

$$\begin{aligned} & | \ 0 \dots \ 0 [0 \ 1 \ 0 \ 1] \dots [1 \ 0 \ 1 \ 0] | \\ & \rightarrow | \ 1] \dots [1 \ 0 \ 1 \ 0] 0 \dots \ 0 [0 \ 1 \ 0 \ 1 | \end{aligned}$$

The total sign accumulated is then 1.

The number of Slater determinants in each subclass is the same and can be computed as follows: the two disruptions that are moved together contain $l_1 + l_2$ “atomic” (i.e., length= four sites) disruptions. Two sites of regular order (into which each disruption is moved) are also spoken for. This leaves $m - l_1 - l_2$ atomic disruptions and $2N - 2 - 4m$ sites of regular order. The number of Slater determinants in a subclass equals the number of ways we can arrange these atomic disruptions and sites of regular order in the space around the two moving disruptions. This space consists of two intervals, one of length $2j - 4l_1 - 1$ and one of length $2N - 2j - 1 - 4l_2$. Let q be the number of atomic insertions we assign to the latter interval. There are $\binom{2N-2j-4l_2-1-3q}{q}$ ways to do this. For each of these choices, there are $\binom{2j-4l_1-1-3(m-q-l_2)}{m-l_1-l_2-q}$ ways to arrange the remaining atomic disruptions within the interval of length $2j - 4l_1 - 1$. The total number of ways is obtained by summing over q .

(2) A creation operator and an annihilation operator with odd site indices, and a creation operator and an annihilation operator with even site indices: Without loss of generality, we can consider one creation operator to have site index 1. All other cases give the same answer, because of invariance of the states under consideration by two lattice sites. These operator products preserve the number of sites of regular order, and thus only connect $|m\rangle$ to itself. The counting of Slater determinants proceeds in as in the case of four odd site indices, with only minor modifications.

(3) Four even site indices: Without loss of generality, we can consider one creation operator to have site index 2. All other cases give the same answer, because of invariance of the states under consideration by two lattice sites. These operator products preserve the number of sites of regular order, and thus only connect $|m\rangle$ to itself. The counting of Slater determinants proceeds in as in the case of four odd site indices, with only minor modifications.

The above considerations then lead to the following results. For operator products with four odd site indices,

$$\begin{aligned} & \langle m | c_1^\dagger c_{4l_1+1} c_{2j+1}^\dagger c_{2j+1+4l_2} | m \rangle \\ & = -\langle m | c_1^\dagger c_{2j+1-4l_1} c_{2j+1}^\dagger c_{2N-4l_2+1} | m \rangle \\ & = -\langle m | c_1^\dagger c_{4l_1+1} c_{2j+1}^\dagger c_{2j+1+4l_2} | m \rangle \\ & = \langle m | c_1^\dagger c_{2j+1-4l_1} c_{2j+1}^\dagger c_{2N-4l_2+1} | m \rangle \\ & = c_{m,j,l_1,l_2}^{1111}, \end{aligned} \quad (E5)$$

where

$$c_{m,j,l_1,l_2}^{1111} = \sum_{q=\max\{0, \lceil m-l_2-\frac{j}{2}+\frac{1}{4} \rceil\}}^{\min\{m-l_1-l_2, \lfloor \frac{N-j}{2}-\frac{1}{4}-l_2 \rfloor\}} \binom{2N-2j-4l_2-1-3q}{q} \binom{2j-l_1-1-3(m-q-l_2)}{m-l_1-l_2-q}. \quad (E6)$$

The fact that the two disruptions are each at least four sites long, don’t overlap, and jointly have length $4m$ sites, restricts the labels to the respective ranges

$$j \in \{3, \dots, N-3\}, \quad l_1 \in \left\{1, \min\left(\left\lfloor \frac{j-1}{2} \right\rfloor, m-1\right)\right\}, \quad l_2 \in \left\{1, \min\left(\left\lfloor \frac{N-j-1}{2} \right\rfloor, m-l_1\right)\right\}. \quad (E7)$$

For an operator product with a creation operator and an annihilation operator with odd site indices, and a creation operator and an annihilation operator with even site indices

$$\begin{aligned} & \langle m | c_1^\dagger c_{4l_1+1} c_{2j+2}^\dagger c_{2j+2+4l_2} | m \rangle = -\langle m | c_1^\dagger c_{2j+2-4l_1}^\dagger c_{2j+2} c_{2N-4l_2+1} | m \rangle \\ & = -\langle m | c_1^\dagger c_{4l_1+1} c_{2j+2}^\dagger c_{2j+2+4l_2} | m \rangle \\ & = \langle m | c_1^\dagger c_{2j+2-4l_1} c_{2j+2}^\dagger c_{2N-4l_2+1} | m \rangle \\ & = c_{m,j,l_1,l_2}^{1212}, \end{aligned} \quad (E8)$$

where

$$c_{m,j,l_1,l_2}^{1212} = \sum_{q=\max\{0, \lceil m-l_2-\frac{j}{2} \rceil\}}^{\min\{m-l_1-l_2, \lfloor \frac{N-j-1}{2}-l_2 \rfloor\}} \binom{2N-2j-4l_2-2-3q}{q} \binom{2j-l_1-3(m-q-l_2)}{m-l_1-l_2-q}. \quad (E9)$$

Here,

$$\begin{aligned} j &\in \{2, \dots, N-3\}, \\ l_1 &\in \left\{1, \min\left(\left\lfloor \frac{j}{2} \right\rfloor, m-1\right)\right\}, \\ l_2 &\in \left\{1, \min\left(\left\lfloor \frac{N-j-1}{2} \right\rfloor, m-l_1\right)\right\}. \end{aligned} \quad (\text{E10})$$

For operator products with four even site indices,

$$\begin{aligned} &-\langle m | c_2^\dagger c_{2j-4l_1+2} c_{2j+2}^\dagger c_{2N-4l_2+2} | m \rangle \\ &= \langle m | c_2^\dagger c_{4l_1+2} c_{2j+2}^\dagger c_{2j+4l_2+2} | m \rangle \\ &= \langle m | c_2^\dagger c_{2j-4l_1+2} c_{2j+2}^\dagger c_{2j-4l_2+2} | m \rangle \\ &= -\langle m | c_2^\dagger c_{4l_1+2} c_{2j+2}^\dagger c_{2j+4l_2+2} | m \rangle \\ &= c_{m,j,l_1,l_2}^{2222}, \end{aligned} \quad (\text{E11})$$

where $c_{m,j,l_1,l_2}^{2222} = c_{m,j,l_1,l_2}^{1111}$, i.e., the combinatorial factor in the last line of the above equation is the same as for the case of four odd site indices.) The fact that the two disruptions are each at least four sites long, do not overlap, and jointly have length $4m$ sites, restricts the labels to the respective ranges

$$\begin{aligned} j &\in \{3, \dots, N-3\}, \\ l_1 &\in \left\{1, \min\left(\left\lfloor \frac{j-1}{2} \right\rfloor, m-1\right)\right\}, \\ l_2 &\in \left\{1, \min\left(\left\lfloor \frac{N-j-1}{2} \right\rfloor, m-l_1\right)\right\}. \end{aligned} \quad (\text{E12})$$

3. A product in which one of the annihilation operators have the same site index as one of the creation operators

In this case the two remaining operators must have site indices of the same parity, so that the only nonzero elements are between states with the same m . Owing to invariance under translation by two lattice sites, we can focus on the cases where the repeated site index is either 1 or 2, and choose to consider $\langle m | c_1 c_1^\dagger \dots | m \rangle$ and $\langle m | c_2^\dagger c_2 \dots | m \rangle$. These operator products check that site 1 or site 2 are of regular order, and further moves one disruption to the left or right by one site. The counting arguments are very similar to the case of four distinct site indices. We obtain

$$\begin{aligned} \langle m | c_1 c_1^\dagger c_{2j+1} c_{2j+4l+1}^\dagger | m \rangle &= -\langle m | c_1 c_1^\dagger c_{2j+1}^\dagger c_{2j+4l+1} | m \rangle \\ &= c_{m,j,0,l}^{1111}, \end{aligned} \quad (\text{E13})$$

with the restrictions on j and l being

$$\begin{aligned} j &\in \{1, \dots, N-3\}, \\ l &\in \left\{1, \min\left(\left\lfloor \frac{N-j-1}{2} \right\rfloor, m\right)\right\}. \end{aligned} \quad (\text{E14})$$

Furthermore,

$$\begin{aligned} &\langle m | c_1 c_1^\dagger c_{2j+2}^\dagger c_{2j+4l+2} | m \rangle \\ &= -\langle m | c_1 c_1^\dagger c_{2j+2} c_{2j+4l+2}^\dagger | m \rangle \\ &= c_{m,j,0,l}^{1212}, \end{aligned} \quad (\text{E15})$$

with

$$\begin{aligned} j &\in \{0, \dots, N-3\}, \\ l &\in \left\{1, \min\left(\left\lfloor \frac{N-j-1}{2} \right\rfloor, m\right)\right\}. \end{aligned} \quad (\text{E16})$$

Lastly,

$$\begin{aligned} &\langle m | c_2^\dagger c_{2j+2} c_{2j+4l+2}^\dagger | m \rangle \\ &= -\langle m | c_2^\dagger c_{2j+2} c_{2j+4l+2}^\dagger | m \rangle \\ &= c_{m,j,0,l}^{2222}, \end{aligned} \quad (\text{E17})$$

with

$$\begin{aligned} j &\in \{1, \dots, N-3\}, \\ l &\in \left\{1, \min\left(\left\lfloor \frac{N-j-1}{2} \right\rfloor, m\right)\right\}. \end{aligned} \quad (\text{E18})$$

4. A product with two pairs of creation and annihilation operators, where the members of each pair has the same site index

Invariance under translation by two lattice sites allows us to focus on three classes,

$$\langle m | c_1 c_1^\dagger c_{2j+1} c_{2j+1}^\dagger | m \rangle = c_{m,j,0,0}^{1111} \quad (\text{E19})$$

for $j \in \{1, \dots, N-1\}$.

$$\langle m | c_1 c_1^\dagger c_{2j+2}^\dagger c_{2j+2} | m \rangle = c_{m,j,0,0}^{1212} \quad (\text{E20})$$

for $j \in \{0, \dots, N-1\}$.

$$\langle m | c_2^\dagger c_{2j+2} c_{2j+2}^\dagger | m \rangle = c_{m,j,0,0}^{2222} \quad (\text{E21})$$

for $j \in \{1, \dots, N-1\}$.

APPENDIX F: HAMILTONIAN MATRIX ELEMENTS

All the ingredients necessary have now been collected and we can assemble them to express matrix elements of the Hamiltonian between $|m\rangle$ states in terms of the Wannier wave functions $\psi(l)$, $\xi(l)$, $\tilde{\psi}(l)$, and $\tilde{\xi}(l)$.

The kinetic energy does not connect states with different m ,

$$\begin{aligned} &\langle m | -t \sum_{j=1}^{2N} (a_j^\dagger a_{j+1} + a_{j+1}^\dagger a_j) | m \rangle \\ &= Nt \left[\frac{2N}{2N-3m} \binom{2N-3m}{m} X_0^{11} \right. \\ &\quad \left. - \sum_{l=0}^m (X_{2l}^{11} + X_{2l}^{22}) \binom{2N-3m-l-1}{m-l} \right] + \text{c.c.}, \end{aligned} \quad (\text{F1})$$

where

$$\begin{aligned} X_j^{11} &= \sum_{l=1}^N [\tilde{\psi}(l)^* \tilde{\xi}(l-j) + \tilde{\xi}(l)^* \tilde{\psi}(l+1-j)], \\ X_j^{22} &= \sum_{l=1}^N [\xi(l)^* \psi(l-j) + \psi(l)^* \xi(l+1-j)]. \end{aligned} \quad (\text{F2})$$

The interaction term is more involved. It connects states $|m\rangle$ to itself and to $|m \pm 1\rangle$,

$$\langle m | \sum_{j=1}^{2N} a_j^\dagger a_j a_{j+1}^\dagger a_{j+1} | m \rangle = N(T_m^{1111} + T_m^{1212} + T_m^{2222}), \quad (\text{F3})$$

and

$$\langle m+1 | \sum_{j=1}^{2N} a_j^\dagger a_j a_{j+1}^\dagger a_{j+1} | m \rangle = NT_m^{1122}. \quad (\text{F4})$$

Here,

$$T_m^{\sigma_1\sigma_2\sigma_3\sigma_4} = S_{0,m}^{\sigma_1\sigma_2\sigma_3\sigma_4} + S_{1,m}^{\sigma_1\sigma_2\sigma_3\sigma_4} + S_{1,m}^{\sigma_1\sigma_2\sigma_3\sigma_4}. \quad (\text{F5})$$

The S_0 contributions originate from four fermion operators that all have distinct Wannier orbital indices. For $\alpha = 1111, 1212, 2222$,

$$S_{0,m}^\alpha = \sum_{j,l_1,l_2}^\alpha c_{m,j,l_1,l_2}^\alpha [-W^\alpha(j+2l_2, 2l_1, j) + W^\alpha(j+2l_2, j, 2l_1) + W^\alpha(j, j+2l_2, 2l_1) - W^\alpha(j, 2l_1, j+2l_2) - W^\alpha(j, -2l_2, j-2l_1) + W^\alpha(j, j-2l_1, -2l_2) + W^\alpha(j-2l_1, j, -2l_2) - W^\alpha(j-2l_1, -2l_2, j)]. \quad (\text{F6})$$

For $\alpha = 1122$,

$$S_{0,m}^{1122} = \sum_{l_1,l_2,l_3}^{1122} c_{l_1,l_2,l_3}^{1122} [-W^{1122}(2(l_1+l_2)+1, 2(l_1+l_2+l_3)+1, 2l_1) + W^{1122}(2(l_1+l_2)+1, 2l_1, 2(l_1+l_2+l_3)+1) + W^{1122}(-2(l_2+l_3)-1, -2l_3-1, -2(l_1+l_2+l_3)-2) - W^{1122}(-2(l_2+l_3)-1, -2(l_1+l_2+l_3)-2, -2l_3-1) - W^{1122}(-2(l_2+l_3)-1, -2l_3-1, 2l_1) + W^{1122}(-2(l_2+l_3)-1, 2l_1, -2l_3-1) + W^{1122}(2(l_1+l_2)+1, -2l_3-1, 2l_1) - W^{1122}(2(l_1+l_2)+1, 2l_1, -2l_3-1)]. \quad (\text{F7})$$

The summations are

$$\begin{aligned} \sum^{1111} &= \sum_{j=3}^{N-3} \sum_{l_1=1}^{\min\{\lfloor \frac{j-1}{2} \rfloor, m-1\}} \sum_{l_2=1}^{\min\{\lfloor \frac{N-j-1}{2} \rfloor, m-l_1\}} \\ \sum^{1212} &= \sum_{j=2}^{N-3} \sum_{l_1=1}^{\min\{\lfloor \frac{j}{2} \rfloor, m-1\}} \sum_{l_2=1}^{\min\{\lfloor \frac{N-j-1}{2} \rfloor, m-l_1\}} \\ \sum^{2222} &= \sum_{j=3}^{N-1} \sum_{l_1=1}^{\min\{\lfloor \frac{j-1}{2} \rfloor, m-1\}} \sum_{l_2=1}^{\min\{\lfloor \frac{N-j-1}{2} \rfloor, m-l_1\}} \end{aligned} \quad (\text{F8})$$

and

$$\sum^{1122} = \sum_{l_1=0}^{\min\{\frac{N-1}{2}-1, m\}} \sum_{l_2=0}^{\min\{\frac{N-1}{2}-1-l_1, m-l_1\}} \sum_{l_3=0}^{\min\{\frac{N-1}{2}-1-l_1-l_2, m-l_1-l_2\}}. \quad (\text{F9})$$

If the upper bound of a sum is lower than the lower bound, then the result is zero. The W^α 's read

$$\begin{aligned} W^{1111}(j_1, j_2, j_3) &= \sum_{l=1}^N [\tilde{\psi}(l)^* \tilde{\xi}(l-j_1)^* \tilde{\xi}(l-j_2) \tilde{\psi}(l-j_3) + \tilde{\xi}(l)^* \tilde{\psi}(l+1-j_1)^* \tilde{\psi}(l+1-j_2) \tilde{\xi}(l-j_3)], \\ W^{1212}(j_1, j_2, j_3) &= \sum_{l=1}^N [\tilde{\psi}(l+1)^* \psi(l-j_1)^* + \tilde{\xi}(l)^* \xi(l+1-j_1)^*] [\tilde{\psi}(l+1-j_2) \psi(l-j_3) + \tilde{\xi}(l-j_2) \xi(l+1-j_3)] \\ &\quad + [\tilde{\psi}(l)^* \psi(l-j_1)^* + \tilde{\xi}(l)^* \xi(l-j_1)^*] [\tilde{\psi}(l-j_2) \psi(l-j_3) + \tilde{\xi}(l-j_2) \xi(l-j_3)], \\ W^{2222}(j_1, j_2, j_3) &= \sum_{l=1}^N [\psi(l)^* \xi(l+1-j_1)^* \xi(l+1-j_2) \psi(l-j_3) + \xi(l)^* \psi(l-j_1)^* \psi(l-j_2) \xi(l-j_3)], \\ W^{1122}(j_1, j_2, j_3) &= \sum_{l=1}^N [\tilde{\psi}(l)^* \tilde{\xi}(l-j_1)^* \psi(l-j_2) \xi(l-j_3) + \tilde{\xi}(l)^* \tilde{\psi}(l+1-j_1)^* \xi(l+1-j_2) \psi(l-j_3)], \end{aligned} \quad (\text{F10})$$

The S_1 terms contain two Fermion c -operators with the same index. The 1122 contribution is zero,

$$\begin{aligned} S_{1,m}^{1111} &= \sum_{j=1}^{N-3} \sum_{l=1}^{\min\{\lfloor \frac{N-1}{2} - \frac{j}{2} \rfloor, m\}} [V^{1111}(j, j+2l) + V^{1111}(j+2l, j)] \left[c_{m,j,0,l}^{1111} - \binom{2N-3m-1-l}{m-l} \right], \\ S_{1,m}^{2222} &= \sum_{j=1}^{N-3} \sum_{l=1}^{\min\{\lfloor \frac{N-1}{2} - \frac{j}{2} \rfloor, m\}} [V^{2222}(j, j+2l) + V^{2222}(j+2l, j)] c_{m,j,0,l}^{1111}, \end{aligned} \quad (\text{F11})$$

where

$$V^\alpha(j_1, j_2) = W^\alpha(j_1, j_2, 0) - W^\alpha(j_1, 0, j_2) + W^\alpha(-j_1, -j_1, j_2 - j_1) - W^\alpha(-j_1, j_2 - j_1, -j_1). \quad (\text{F12})$$

$$\begin{aligned} S_{1,m}^{1212} &= \sum_{j=0}^{N-3} \sum_{l=1}^{\min\{\lfloor \frac{N-1}{2} - \frac{j}{2} \rfloor, m\}} \left\{ [W^{1212}(j, 0, j+2l) + W^{1212}(j+2l, 0, j)] \binom{2N-3m-1-l}{m-l} \right. \\ &\quad - [W^{1212}(j, 0, j+2l) + W^{1212}(j+2l, 0, j)] \\ &\quad \left. + W^{1212}(-j-1, 2l, -j-1) + W^{1212}(-j-2l-1, -2l, -j-2l-1) \right] c_{m,j,0,l}^{1212} \Big\}. \end{aligned} \quad (\text{F13})$$

The S_2 terms contain two pairs of Fermion c -operators with the same index. Again the 1122 contribution is zero,

$$\begin{aligned} S_{2,m}^{1111} &= \sum_{j=1}^{N-1} [W^{1111}(j, j, 0) - W^{1111}(j, 0, j)] \left[\frac{2N}{2N-3m} \binom{2N-3m}{m} - 2 \binom{2N-1-3m}{m} + c_{m,j,0,0}^{1111} \right], \\ S_{2,m}^{1212} &= \sum_{j=0}^{N-1} W^{1212}(j, 0, j) \left[\binom{2N-1-3m}{m} - c_{m,j,0,0}^{1212} \right], \\ S_{2,m}^{2222} &= \sum_{j=1}^{N-1} [W^{2222}(j, j, 0) - W^{2222}(j, 0, j)] c_{m,j,0,0}^{1111}. \end{aligned} \quad (\text{F14})$$

APPENDIX G: A MORE GENERAL VARIATIONAL CALCULATION

In the main text, we showed that the exponential coupled cluster state $|\alpha\rangle = \exp(\alpha T)|\text{MF}\rangle$ with $\alpha = (t/V)^2$ yields a variational energy that is accurate to order $(t/V)^4$. In this Appendix we investigate a more general state built from states $|m\rangle = (T)^m|\text{MF}\rangle/m!$. We namely consider the ansatz

$$|\Psi\rangle = \sum_{m=0}^{\lfloor N/2 \rfloor} \Psi_m |m\rangle. \quad (\text{G1})$$

The simplest variational scheme is to treat the Wannierized single-particle basis obtained in mean-field theory as cast in stone, and to vary only the amplitudes Ψ_m in Eq. (17). However, one may hope to obtain better results by also adjusting the single-particle basis as well. We have implemented the simplest version of this optimization by varying the parameter z , that defines the $c_j^\dagger$ orbitals through Eqs. (D9), (D10), and (D16), along with the Ψ_m , rather than fixing it by solving the mean-field Eq. (D15).

In Fig. 6, we again plot the ratio between the correlation energy as estimated by the more general ansatz Eq. (G1), to the true correlation energy, as well as the difference in energy per particle between the more general ansatz and the exact result. This should be compared to Fig. 3 in the main text. We see that

the error in energy per particle goes like $10(t/V)^6$ at small t/V , as compared to $19.5(t/V)^6$ for the simpler exponential

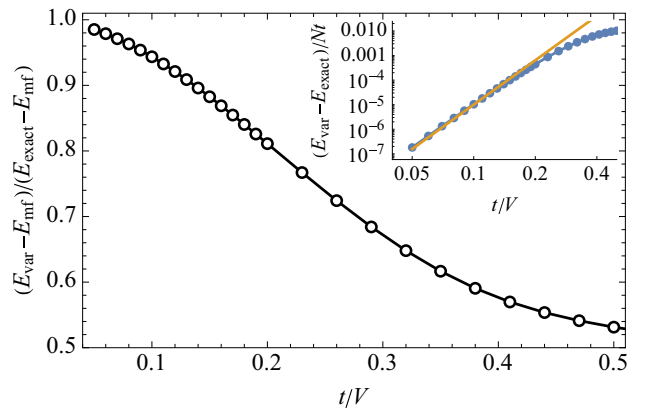


FIG. 6. The difference in ground-state energy between mean-field theory Eq. (D18) and the correlated ansatz Eq. (G1) in units of the exact correlation energy. (Inset) The difference between the variational energy per particle of the correlated ansatz Eq. (G1) and the true ground-state energy Eq. (22) in double log scale. The solid line corresponds to $10(t/V)^6$. Results are shown for a system of $N = 41$ particles occupying 82 sites.

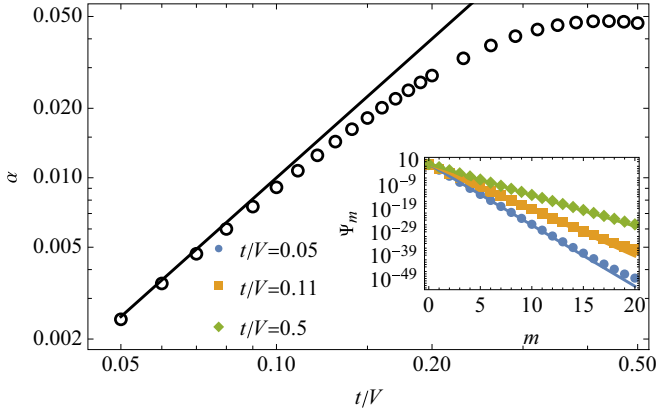


FIG. 7. (Inset) The optimal variational amplitudes Ψ_m Eq. (G1) versus number of disruptions m , plotted in logarithmic scale on the vertical axis. Results for three values of t/V are shown. Solid lines represent fits to the coupled cluster formula $\Psi_m = N\alpha^m$. The extracted cluster strength α versus t/V plotted in double log scale. The solid line corresponds to $\alpha = (t/V)^2$. Results are shown for a system of $N = 41$ particles occupying 82 sites.

ansatz of the main text, giving a small improvement. At larger t/V , the more general ansatz more qualitatively capture the physics, as it can account for about 50% of the correlation energy even close to the phase transition at $t/V = 0.5$.

In the inset of Fig. 7, we plot the state amplitudes Ψ_m of the optimal correlated states for various t/V , including ones outside the regime where the correlated ansatz is quantitatively accurate. In all cases, we see that the amplitudes are very well approximated by a function that is exponential in m . Thus, the improved accuracy of more general ansatz Eq. (G1) does not come from abandoning the coupled cluster doubles form. Rather it is caused by a renormalization of α away from

$(t/V)^2$ when this parameters becomes large, and to the modification of the mean-field parameter z . At small t/V where the correlated ansatz is intended to be quantitatively accurate, we find that varying z makes little difference. For instance, at $t/V = 0.1$ the optimal z is 2.40 while the mean-field solution is 2.45. At larger t/V , there is a more significant renormalization of z (around 20% close to the critical point), which leads to improved correlation energy estimates. It should however be kept in mind that in this regime, the correlated ansatz provides at best a qualitative description anyway, because the periodicity of the covariance matrix spectrum is not correctly obtained. In the main panel of Fig. 7, we plot the α values versus t/V that we extracted from our improved variational calculation. In the regime $t \lesssim 0.1V$, where the ansatz is quantitatively accurate, we still find that $\alpha \simeq (t/V)^2$, but at larger t/V it is suppressed below this quadratic growth.

APPENDIX H: EXPLICIT EXPRESSION FOR $\zeta_0(\alpha)$

In Sec. VII, we expressed the covariance matrix associated with the state $|\alpha\rangle$ in terms of the real root ζ_0 of $1 - \zeta - \alpha^2 \zeta^4$ that is closest to the origin. Using Ferrari's method to solve this depressed quartic equation we obtain

$$\zeta_0(\alpha) = \frac{1}{2\alpha^{2/3}} \left[\sqrt{\frac{2}{u(\alpha)} - u(\alpha)^2} - u(\alpha) \right], \quad (\text{H1})$$

where $u(\alpha)$ takes the value

$$\sqrt{\left(\frac{\sqrt{1 + \frac{256}{27}\alpha^2} + 1}{2} \right)^{1/3} - \left(\frac{\sqrt{1 + \frac{256}{27}\alpha^2} - 1}{2} \right)^{1/3}}. \quad (\text{H2})$$

- [1] G. F. Giuliani and G. Vignale, *Quantum Theory of the Electron Liquid* (Cambridge University Press, Cambridge, 2005).
- [2] P.-O. Löwdin, Quantum theory of many-particle systems. I. Physical interpretations by means of density matrices, natural spin-orbitals, and convergence problems in the method of configurational interaction, *Phys. Rev.* **97**, 1474 (1955).
- [3] E. R. Davidson, Properties and uses of natural orbitals, *Rev. Mod. Phys.* **44**, 451 (1972).
- [4] H. Bruus and K. Flensberg, *Many-Body Quantum Theory in Condensed Matter Physics* (Oxford University Press, Oxford, 2015).
- [5] T. Giamarchi, *Quantum Physics in One Dimension* (Oxford University Press, Oxford, 2003).
- [6] J. Olsen, The CASSCF method: A perspective and commentary, *Int. J. Quantum Chem.* **111**, 3267 (2011).
- [7] R.-Q. He and Z.-Y. Lu, Quantum renormalization groups based on natural orbitals, *Phys. Rev. B* **89**, 085108 (2014).
- [8] M. Debertolis, S. Florens, and I. Snyman, Few-body nature of Kondo correlated ground states, *Phys. Rev. B* **103**, 235166 (2021).
- [9] I. Snyman, Structure of quasiparticles in a local Fermi liquid, *Phys. Rev. B* **108**, 205120 (2023).
- [10] S. Bravyi and D. Gosset, Complexity of quantum impurity problems, *Commun. Math. Phys.* **356**, 451 (2017).
- [11] M. Debertolis, I. Snyman, and S. Florens, Simulating realistic screening clouds around quantum impurities: Role of spatial anisotropy and disorder, *Phys. Rev. B* **106**, 125115 (2022).
- [12] Y. Núñez-Fernández, M. Debertolis, and S. Florens, Resolving space-time structures of quantum impurities with a numerically-exact algorithm using few-body revealing, [arXiv:2503.13706](https://arxiv.org/abs/2503.13706).
- [13] S. Bera, H. Schomerus, F. Heidrich-Meisner, and J. H. Bardarson, Many-body localization characterized from a one-particle perspective, *Phys. Rev. Lett.* **115**, 046603 (2015).
- [14] M. Thamm, H. Radhakrishnan, H. Bargathi, B. Rosenow, and A. Del Maestro, One-particle entanglement for one-dimensional spinless fermions after an interaction quantum quench, *Phys. Rev. B* **106**, 165116 (2022).
- [15] T. I. Vanhala and T. Ojanen, Complexity of fermionic states, *Phys. Rev. Res.* **6**, 023178 (2024).
- [16] J. Des Cloizeaux, A soluble Fermi-gas model. Validity of transformations of the Bogoliubov type, *J. Math. Phys.* **7**, 2136 (1966).

- [17] C. N. Yang and C. P. Yang, One-dimensional chain of anisotropic spin-spin interactions. II. Properties of the ground-state energy per lattice site for an infinite system, *Phys. Rev.* **150**, 327 (1966).
- [18] S. R. White, Density matrix formulation for quantum renormalization groups, *Phys. Rev. Lett.* **69**, 2863 (1992).
- [19] C. Karrasch and J. E. Moore, Luttinger liquid physics from the infinite-system density matrix renormalization group, *Phys. Rev. B* **86**, 155156 (2012).
- [20] M. Jimbo, K. Miki, T. Miwa, and A. Nakayashiki, Correlation functions of the XXZ model for $\Delta < -1$, *Phys. Lett. A* **168**, 256 (1992).
- [21] F. Göhmann, A. Klümper, and A. Seel, Integral representation of the density matrix of the XXZ chain at finite temperatures, *J. Phys. A: Math. Gen.* **38**, 1833 (2005).
- [22] N. Kitanine, J. Maillet, and V. Terras, Correlation functions of the XXZ Heisenberg spin-1/2 chain in a magnetic field, *Nucl. Phys. B* **567**, 554 (2000).
- [23] C. Babenko, F. Göhmann, K. K. Kozłowski, J. Sirker, and J. Suzuki, Exact real-time longitudinal correlation functions of the massive XXZ chain, *Phys. Rev. Lett.* **126**, 210602 (2021).
- [24] F. Gebhard, K. Bauerbach, and O. Legeza, Accurate localization of Kosterlitz-Thouless-type quantum phase transitions for one-dimensional spinless fermions, *Phys. Rev. B* **106**, 205133 (2022).
- [25] R. J. Bartlett and M. Musiał, Coupled-cluster theory in quantum chemistry, *Rev. Mod. Phys.* **79**, 291 (2007).
- [26] S. Hirata, R. Podeszwa, M. Tobita, and R. J. Bartlett, Coupled-cluster singles and doubles for extended systems, *J. Chem. Phys.* **120**, 2581 (2004).
- [27] R. F. Bishop and K. H. Lührmann, Electron correlations: I. Ground-state results in the high-density regime, *Phys. Rev. B* **17**, 3757 (1978).
- [28] J. J. Shepherd, A. Grüneis, G. H. Booth, G. Kresse, and A. Alavi, Convergence of many-body wave-function expansions using a plane-wave basis: From homogeneous electron gas to solid state systems, *Phys. Rev. B* **86**, 035111 (2012).
- [29] M. Georgiou, I. Rousochatzakis, D. J. J. Farnell, J. Richter, and R. F. Bishop, Spin- S Kitaev-Heisenberg model on the honeycomb lattice: A high-order treatment via the many-body coupled cluster method, *Phys. Rev. Res.* **6**, 033168 (2024).
- [30] J. I. Cirac, D. Pérez-García, N. Schuch, and F. Verstraete, Matrix product states and projected entangled pair states: Concepts, symmetries, theorems, *Rev. Mod. Phys.* **93**, 045003 (2021).
- [31] S. Sharma and A. Alavi, Multireference linearized coupled cluster theory for strongly correlated systems using matrix product states, *J. Chem. Phys.* **143**, 102815 (2015).
- [32] L. Veis, A. Antalík, J. Brabec, F. Neese, Ö. Legeza, and J. Pittner, Coupled cluster method with single and double excitations tailored by matrix product state wave functions, *J. Phys. Chem. Lett.* **7**, 4072 (2016).
- [33] A. J. Ramirez-Pastor, T. P. Eggarter, V. D. Pereyra, and J. L. Riccardo, Statistical thermodynamics and transport of linear adsorbates, *Phys. Rev. B* **59**, 11027 (1999).
- [34] C. Castelnovo, C. Chamon, C. Mudry, and P. Pujol, From quantum mechanics to classical statistical physics: Generalized Rokhsar–Kivelson Hamiltonians and the “Stochastic Matrix Form” decomposition, *Ann. Phys.* **318**, 316 (2005).
- [35] D. S. Rokhsar and S. A. Kivelson, Superconductivity and the quantum hard-core dimer gas, *Phys. Rev. Lett.* **61**, 2376 (1988).
- [36] M. Fannes, B. Nachtergaele, and R. F. Werner, Finitely correlated states on quantum spin chains, *Commun. Math. Phys.* **144**, 443 (1992).
- [37] C. Fernández-González, N. Schuch, M. M. Wolf, J. I. Cirac, and D. Pérez-García, Frustration free gapless Hamiltonians for matrix product states, *Commun. Math. Phys.* **333**, 299 (2015).
- [38] M. Fishman, S. R. White, and E. M. Stoudenmire, The ITensor software library for tensor network calculations, *SciPost Phys. Codebases* **4** (2022).
- [39] H. Barghathi, E. Casiano-Diaz, and A. Del Maestro, Operationally accessible entanglement of one-dimensional spinless fermions, *Phys. Rev. A* **100**, 022324 (2019).
- [40] H. Yoshida and H. Katsura, Exact eigenstates of extended $SU(n)$ hubbard models: Generalization of η -pairing states with n -particle off-diagonal long-range order, *Phys. Rev. B* **105**, 024520 (2022).
- [41] M. Kohno, M. Arikawa, J. Sato, and K. Sakai, Spectral properties of interacting one-dimensional spinless fermions, *J. Phys. Soc. Jpn.* **79**, 043707 (2010).
- [42] S. Fratini, B. Valenzuela, and D. Baeriswyl, Incipient quantum melting of the one-dimensional Wigner lattice, *Synth. Met.* **141**, 193 (2004), michael J. Rice Memorial Festschrift.
- [43] M. Mayr and P. Horsch, Domain-wall excitations and optical conductivity in one-dimensional Wigner lattices, *Phys. Rev. B* **73**, 195103 (2006).
- [44] J. Wouters, H. Katsura, and D. Schuricht, Interrelations among frustration-free models via Witten’s conjugation, *SciPost Phys. Core* **4**, 027 (2021).
- [45] I. Mahyaeh, T. Köhler, A. M. Black-Schaffer, and A. Kantian, Superconducting pairing from repulsive interactions of fermions in a flat-band system, *Phys. Rev. B* **106**, 125155 (2022).
- [46] C. N. Yang, η pairing and off-diagonal long-range order in a Hubbard model, *Phys. Rev. Lett.* **63**, 2144 (1989).
- [47] K. Tamura and H. Katsura, Quantum many-body scars of spinless fermions with density-assisted hopping in higher dimensions, *Phys. Rev. B* **106**, 144306 (2022).
- [48] L. Gotta, L. Mazza, P. Simon, and G. Roux, Exact many-body scars based on pairs or multimers in a chain of spinless fermions, *Phys. Rev. B* **106**, 235147 (2022).
- [49] R. Moessner and S. L. Sondhi, Resonating valence bond phase in the triangular lattice quantum dimer model, *Phys. Rev. Lett.* **86**, 1881 (2001).
- [50] S. Moudgalya, S. Rachel, B. A. Bernevig, and N. Regnault, Exact excited states of nonintegrable models, *Phys. Rev. B* **98**, 235155 (2018).
- [51] C. Meyer, *Matrix Analysis and Applied Linear Algebra* (Society for Applied and Industrial Mathematics, Philadelphia, PA, 2000).
- [52] A. Menczer, K. Kapás, M. A. Werner, and O. Legeza, Two-dimensional quantum lattice models via mode optimized hybrid CPU-GPU density matrix renormalization group method, *Phys. Rev. B* **109**, 195148 (2024).
- [53] J. D. Hannukainen, M. F. Martínez, J. H. Bardarson, and T. K. Kivornig, Interacting local topological markers: A one-particle density matrix approach for characterizing the topology of interacting and disordered states, *Phys. Rev. Res.* **6**, L032045 (2024).