

Efficient simulation of moiré materials using the density matrix renormalization group

Tomohiro Soejima (副島智大) ^{1,*}, Daniel E. Parker ^{1,*}, Nick Bultinck,^{1,2} Johannes Hauschild ¹ and Michael P. Zaletel^{1,3}¹Department of Physics, University of California, Berkeley, California 94720, USA²Department of Physics, Ghent University, 9000 Ghent, Belgium³Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

(Received 21 September 2020; revised 24 October 2020; accepted 26 October 2020; published 11 November 2020)

We present an infinite density-matrix renormalization group (DMRG) study of an interacting continuum model of twisted bilayer graphene (tBLG) near the magic angle. Because of the long-range Coulomb interaction and the large number of orbital degrees of freedom, tBLG is difficult to study with standard DMRG techniques—even constructing and storing the Hamiltonian already poses a major challenge. To overcome these difficulties, we use a recently developed compression procedure to obtain a matrix product operator representation of the interacting tBLG Hamiltonian which we show is both efficient and accurate even when including the spin, valley, and orbital degrees of freedom. To benchmark our approach, we focus mainly on the spinless, single-valley version of the problem where, at half filling, we find that the ground state is a nematic semimetal. Remarkably, we find that the ground state is essentially a k -space Slater determinant, so that Hartree-Fock and DMRG give virtually identical results for this problem. Our results show that the effects of long-range interactions in magic angle graphene can be efficiently simulated with DMRG and open up a new route for numerically studying strong correlation physics in spinful, two-valley tBLG, and other moiré materials in future work.

DOI: [10.1103/PhysRevB.102.205111](https://doi.org/10.1103/PhysRevB.102.205111)

I. INTRODUCTION

Magic angle twisted bilayer graphene (tBLG) hosts a diverse array of correlated insulating and superconducting phases [1–23]. This rich system has inspired intensive theoretical efforts to understand the origin and mechanism(s) behind these phases, and a large number of theories have already been proposed. One way to assess these proposals—especially when they are not associated with clear experimental signatures—is numerical calculation. To that end, this work presents a proof-of-concept density matrix renormalization group (DMRG) [24] study of a microscopically realistic, strongly interacting model of tBLG [25] (Fig. 1).

A. Challenges of tBLG numerics

Let us review what makes the tBLG problem so numerically challenging and identify a viable path around the obstacles. The first obstacle is the separation in scales between the graphene lattice constant a and the moiré length scale L_M ; at the magic angle $L_M/a \sim 1/\theta_M \sim 50$, so the moiré unit cell contains over 10 000 carbon atoms, and consequently the superlattice band structure contains $N_B \sim 10\,000$ bands. Fortunately, various treatments of the band structure [26–29] (including the Bitzritzer-MacDonald (BM) continuum model [26] used here) reveal that the flat bands of interest are separated from the tower of “remote” bands by gaps of order 20 meV–25 meV [see Fig. 2(c)]. Since these gaps are larger than the Coulomb scale $E_C = \frac{e^2}{4\pi\epsilon_0\epsilon_r L_M} \sim 10$ meV–20 meV (using a relative permittivity $\epsilon_r = 12$ –6), it is a reasonable starting point to project the Coulomb interaction $V(r)$ into the

flat bands.¹ Each spin and valley of the graphene has *two* flat bands, for a total of eight, winning us a reduction from $N_B = 10\,000 \searrow 8$. We refer to this as the “interacting Bitzritzer-MacDonald (IBM) model,” although our method works just as well for improved continuum models of tBLG which take into account effects like lattice relaxation. The touching of these two bands is locally protected by a crucial $C_2\mathcal{T}$ symmetry (a 180-degree rotation combined with time reversal), which distinguishes tBLG from other moiré materials. The Coulomb scale E_C is much larger than the bandwidth $t \sim 5$ meV, so *a priori* unbiased, strongly-interacting numerical approaches such as exact diagonalization [32], determinantal quantum Monte Carlo, or DMRG are required.

Most strongly-interacting approaches proceed from a real space lattice model, so a natural next step is to construct a lattice model via 2D Wannier localization of the continuum Bloch bands. In real space, the density of states of the flat bands is predominantly located on the AA-stacking regions of the moiré unit cell, which form a triangular lattice [see Fig. 2(d)]. So one might hope that the physics is then well described by an eight-component triangular lattice Hubbard model. However, there is a topological obstruction which complicates this approach: The flat bands possess “fragile topology” which makes their Wannier localization very subtle [29,33–38]. In particular, the presence of $C_2\mathcal{T}$, valley conservation $U_v(1)$, and translation make it impossible to Wannier localize the flat bands in a manner where $U_v(1)$ and $C_2\mathcal{T}$ both

¹Hartree-Fock studies which include the remote bands do find that they have a quantitative effect (for example, on the magnitudes of the symmetry-broken gaps), but there are some discrepancies regarding their qualitative importance [30,31].

*These two authors contributed equally.

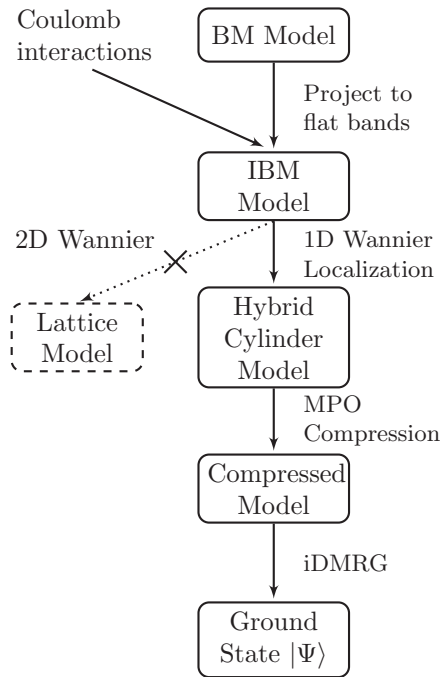


FIG. 1. Flowchart of our approach to DMRG for tBLG. First we start from a continuum BM model and add Coulomb interactions, projected to the flat bands to reduce the number of degrees of freedom down to a manageable level. Second, we perform hybrid Wannier localization, which maps the model to a cylinder in mixed- xk space, thereby avoiding a topological obstruction and allowing all symmetries to act locally. Third, we use a compression procedure to represent the long-range interactions with a reasonable bond dimension to make DMRG numerically tractable. This allows us to perform DMRG with all $N_B = 8$ components at moderate cylinder radius ($L_y = 6$).

act in a strictly local fashion. This is somewhat analogous to the obstruction to finding a Wannier basis for a 2D topological insulator under the requirement that $\mathcal{T}$ acts as a permutation of the orbitals [39].

Two resolutions to the Wannier obstruction issue have been proposed in the literature. The conceptually simplest is to include some number of remote bands, at minimum $N_B = 8 \rightarrow 20$, which removes the topological obstruction and allows for a local symmetry action [34]. But from a DMRG standpoint, a model with 20 orbitals per unit cell, all strongly interacting, appears to be numerically intractable. The other approach is to simply ignore the symmetry considerations and Wannier localize in a basis which hybridizes different valleys or $C_2\mathcal{T}$ sectors. In this approach, for example, valley number conservation $U_v(1)$ becomes slightly nonlocal, and the associated charge takes the form $Q_V = \sum_{i,j,m,n} Q_{mn}^{ij} \hat{c}_{m,i}^\dagger \hat{c}_{n,j}$ where the sum runs over all sites i, j and internal degrees of freedom m, n . The matrix elements Q_{mn}^{ij} fall off with distance $|r_i - r_j|$ [33]. Intriguingly, the Wannier orbitals then take the shape of three-lobed “fidget spinners” connecting three nearby AA regions [33,40,41]. In this basis, the Coulomb interaction is not dominated by a $U\hat{n}^2$ Hubbard interaction but instead contains a profusion of all allowed $V_{mnop}^{ijkl} \hat{c}_{m,i}^\dagger \hat{c}_{n,j}^\dagger \hat{c}_{o,k} \hat{c}_{p,\ell}$ terms which decay exponentially over a few moiré sites [40–42].

Numerically, however, the interactions must be cut off at some finite range, which will spuriously break either the $U_v(1)$ or $C_2\mathcal{T}$ symmetry due to the nonlocal form they take. This runs the risk of biasing the results by explicitly breaking a symmetry which should be preserved and would require careful extrapolation of the tails to ensure the correct results. While not necessarily unworkable (in particular, see Ref. [43]), in our estimation this approach makes numerical results delicate to interpret.

Fortunately, DMRG is a 1D algorithm, which allows us to avoid the construction of 2D Wannier orbitals altogether. When DMRG is applied to the cylinder geometry, the model must be in a localized basis along the length of the cylinder, to ensure favorable entanglement properties, but it does not need to be in a localized basis around its circumference. Therefore, we can consider “hybrid” real-space/momentum-space Wannier states which are maximally localized along the length of the cylinder x but T_y eigenstates around its circumference y (see Fig. 3). There is no topological obstruction to the construction of hybrid Wannier states, making them an attractive basis for the flat bands of magic angle graphene, as was also recognized by the authors of Refs. [25,44–46]. Geometrically, this defines a model on a cylinder, with real space in the x direction and k space around the circumference [47]. The hybrid approach allows the $U_v(1)$, $C_2\mathcal{T}$, and translation symmetries to all act locally without adding extraneous degrees of freedom. This is exactly the approach used by Kang and Vafeek in their recent DMRG study of tBLG [25], and it is the approach we take as well.

The hybrid approach is not without challenge, however, because upon mapping the orbitals to a 1D fermion chain for input into the DMRG, the effective Hamiltonian is quite long ranged. The localization width of the Wannier orbitals is comparable to the moiré scale, so all the sites in a single column of the cylinder are strongly overlapping, generating a panoply of couplings $V_{mnop}^{ijkl} \hat{c}_{m,i}^\dagger \hat{c}_{n,j}^\dagger \hat{c}_{o,k} \hat{c}_{p,\ell}$. Though these decay exponentially with distance, a cylinder with circumference $L_y = 6$ with both spin and valley has on the order of 850 000 non-negligible (i.e., above 10^{-2} meV) matrix elements per unit cell.

A similar problem is encountered in the context of cylinder DMRG for the fractional quantum Hall effect [48] or finite-DMRG simulations for quantum chemistry problems [49]. There, as here, it is essential to use tensor network methods to “compress” the V^{ijkl} as a matrix product operator (MPO). To do so, we leverage a recent algorithm for black-box compression of Hamiltonian MPOs with various optimality properties [50]. We find that for a circumference $L_y = 6$ cylinder, the spinless/single-valley problem ($N_B = 2$) requires an MPO bond dimension of $D \sim 100$ for physical observable to obtain a relative precision of 10^{-2} , while in the spinful/valleyful $N_B = 8$ case, we estimate the required bond dimension to be $D \sim 1000$. While large, these values are tractable, especially when exploiting the charge, spin, valley, and k_y quantum numbers.

B. Overview of DMRG results

After presenting details of the interacting tBLG Hamiltonian and its MPO compression, we apply our approach in

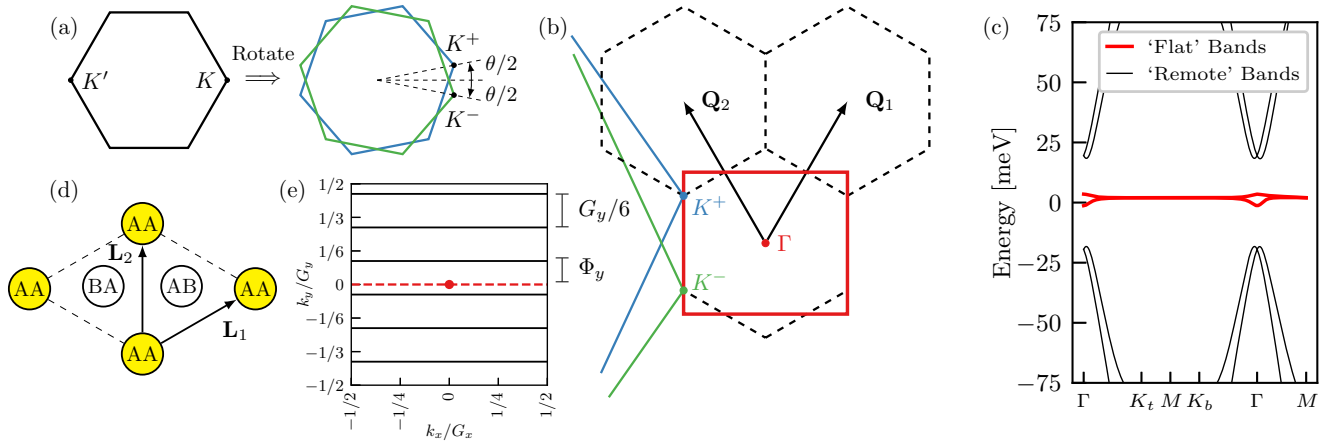


FIG. 2. (a) The BM model for bilayer graphene is constructed from two regular graphene Brillouin zones, rotated by $\pm\theta/2$. (b) Zoomed view of (a) showing the mini (or moiré) Brillouin zone. We choose a square mBZ (thick red lines) for numerical convenience. (c) The band structure of the BM model over the mBZ showing the flat bands (red lines). The interacting BM model is defined by adding Coulomb interactions to the BM model and projecting to the flat bands. (d) Schematic of the real space moiré unit cell with Bravais lattice vectors $\mathbf{L}_{1,2}$. (e) The IBM model on a cylinder has N_y discrete momentum cuts at k_y values given by Eq. (3), offset by Φ_y .

detail to a “toy” $N_B = 2$ problem in which we keep only valley K and spin $\uparrow$; more physical models are reserved for future work. When filling one of the two bands, this scenario is conceptually similar to fillings $\nu = -3, 3$ of tBLG under the assumption that these fillings are spin and valley polarized. However, we caution the reader that our results are not a quantitative prediction for these fillings because the toy model differs from $|\nu| = 3$ of tBLG by a $4\times$ difference in the magnitude of the Hartree potential generated relative to neutrality. We’ve made this choice so that we can quantitatively compare with Refs. [25] prior results; the “physical” $|\nu| = 3$ result, which differs in some interesting respects, will be presented in a future work.

Following Ref. [25], we fix $\theta = 1.05^\circ$ and vary the ratio of the AA and AB interlayer tunneling hopping strengths w_0/w_1 from 0 to 0.9. While physically $w_0/w_1 \sim 0.8$ [27,28,41], the resulting phase diagram is conceptually interesting because the dominant effect of w_0/w_1 is to redistribute the Berry curvature of the flatbands, rather than changing their bandwidth,

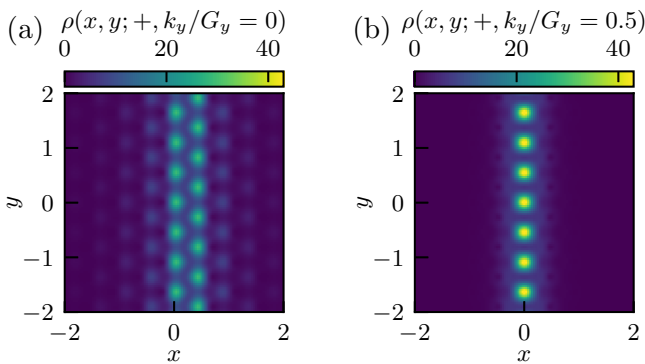


FIG. 3. Real space charge density of Wannier orbitals. The orbitals are maximally localized in the x direction and periodic along y , with charge densities concentrated in AA stacked regions. The Wannier center and the character changes with k_y .

revealing that the former is crucial to the physics. As a précis of our findings,

(1) In agreement with Ref. [25], we find that below a critical value $w_0/w_1 \lesssim 0.8$, the ground state spontaneously breaks the $C_2\mathcal{T}$ symmetry, forming a quantum anomalous Hall state (Chern insulator), with $C = \pm 1$.

(2) In agreement with Ref. [25], above $w_0/w_1 \gtrsim 0.8$, the $C_2\mathcal{T}$ is restored. In this region the DMRG results of Ref. [25] did not reliably converge, but their mean-field calculations suggested either a “nematic $C_2\mathcal{T}$ symmetric semimetal” first proposed in Ref. [51] or a gapped $C_2\mathcal{T}$ -symmetry stripe [25]. Our DMRG numerics reliably converge to a state in excellent agreement with the nematic $C_2\mathcal{T}$ semimetal, with two band touchings near the Γ point.

(3) We analyze the k -space electron correlation function $P_{mn}(\mathbf{k}) = \langle c_{n,\mathbf{k}}^\dagger c_{m,\mathbf{k}} \rangle$ of the DMRG ground state, which can be directly compared with Hartree-Fock calculations. We find that both phases are extremely well captured by a single k -space Slater determinant (to within $\approx 1\%$), strongly supporting the validity of recent Hartree-Fock studies [7,10,25,30,31,45,46,51–53].

(4) Finally, we compare the energy of the DMRG ground state with various competing variational ansatz such as the $C_2\mathcal{T}$ -stripe ansatz proposed in Ref. [25]. In agreement with their result, we find that the nematic semimetal and the $C_2\mathcal{T}$ stripe compete at the order of 0.1 meV per unit cell.

The remainder of this work is organized as follows. Section II introduces our model: an interacting Bistritzer-MacDonald model, equipped with long range Coulomb interactions and projected to the flat bands. Section III discusses how the model may be expressed as a matrix product operator and both why and how it must be compressed to perform DMRG. Section IV provides the results of DMRG calculations and shows that Hartree-Fock accurately captures the ground state physics in this model. Section V discusses the nature of the nematic $C_2\mathcal{T}$ semimetal. We conclude in Sec. VI. Extensive appendices describe all details needed to reproduce our results. Appendix A details the IBM model. Appendix B

deals with the Wannier localization and the gauge choice we make. Appendix C constructs the precompression MPO: an infinite MPO with arbitrary long range four-body interactions. Appendix D provides the algorithm for MPO compression, as well as rigorous error bounds. Finally, Appendix E explains the extensive numerical cross checks we performed to ensure the accuracy of our results.

II. THE IBM MODEL

This section describes the interacting generalization of the Bistritzer-MacDonald (BM) model we use in this work. We first briefly recall the BM model and the geometry of the mini-Brillouin zone (mBZ), then discuss how interactions are added. We then show how the model can be placed on a cylindrical geometry and conclude with the symmetries of the model.

A. Continuum model

Our starting point is the single-particle Bistritzer-MacDonald (BM) model [26], composed of two layers of graphene, with relative twist angle θ , coupled together by a spatially-varying moiré potential. The potential is governed by two parameters, w_0 and w_1 , which specify the AA/AB interlayer tunneling, respectively. DFT calculations which account for lattice relaxation find that $w_1 = 109$ meV and $w_0/w_1 \approx 0.8$ [27,28,41], but here we will treat w_0/w_1 as an axis of the phase diagram. We maximize the ratio of band gap to band width for the flat bands by setting $\theta_{BM} \sim 1.05^\circ$. Figure 2 details our choice of conventions. In particular, we work with a rectangular mBZ grid for numerical convenience.

We now define an interacting Bistritzer-MacDonald (IBM) model where double-gate screened Coulomb interactions are added to the single-particle model. As the interactions are much larger than the spectral width of the flat bands but smaller than the gap to nearby bands, we expect interactions to act quite nonperturbatively inside the flat bands and perturbatively between separated bands. We therefore project the interactions to the two flat bands, akin to models of the fractional quantum Hall effect [54]. Our presentation will focus on a single spin and valley, but their inclusion is conceptually identical: We promote $2 \rightarrow 8$. Consider a vector of fermions $\mathbf{f}_k = (f_{1,k}, f_{2,k})$ running over the two nearly flat bands. The Hamiltonian is then given by

$$\hat{H} = \sum_{k \in \text{mBZ}} \mathbf{f}_k^\dagger h(\mathbf{k}) \mathbf{f}_k + \frac{1}{2A} \sum_q V_q : \rho_q \rho_{-q} : . \quad (1)$$

The single-particle term $h(\mathbf{k})$ contains not only the flat band energies of the BM model but also band renormalization terms coming from the interaction with the filled remote bands and a subtraction to avoid double counting of Coulomb interaction effects. We refer to Appendix A for more details.

The second term in Eq. (1) corresponds to the dual gate-screened Coulomb interaction, with A being the sample area and $V_q = e^2 \tanh(|\mathbf{q}|d)/(2\epsilon_r \epsilon_0 |\mathbf{q}|)$. The screened Coulomb potential depends on two parameters ϵ_r and d , which, respectively, are the relative permittivity and the distance between the twisted bilayer graphene device and the metallic gates. While the effective dielectric constant of the typical substrate,

hBN, is $\epsilon_r \approx 4.4$, here we use $\epsilon_r = 12$ in order to phenomenologically account for screening from the remote bands of the tBLG A. This sets the typical interaction energy scale to be several meV. For the gate distance we choose $d = 10$ nm to facilitate comparison with Ref. [25]. The Fourier components of the flat-band projected charge density operator are given by

$$\rho_q = \sum_{k \in \text{mBZ}} \mathbf{f}_k^\dagger \Lambda_q(\mathbf{k}) \mathbf{f}_{k+q}, \quad (2)$$

where the 2×2 form factor matrices $[\Lambda_q(\mathbf{k})]_{ab} = \langle \psi_{a,\mathbf{k}} | e^{-i\mathbf{q} \cdot \mathbf{r}} | \psi_{b,\mathbf{k}+\mathbf{q}} \rangle$ are defined in terms of overlaps between the Bloch states of the BM model.

The model enjoys several global symmetries: time reversal followed by in-plane rotation $C_2\mathcal{T}$, out-of-plane C_{2x} rotation, and C_3 rotation. We will describe their action on the basis states explicitly below. In summary, the spinless, single-valley IBM model we have described is a strongly interacting many-body problem defined in momentum space over the mini-Brillouin zone.

B. Cylinder model

Our goal is to perform quasi-2D DMRG on Eq. (1). To this end, we work in an infinite cylinder geometry of circumference N_y with a mixed real and momentum space representation of the model. In the momentum space, this corresponds to having N_y momentum cuts through the mBZ at

$$k_y/G_y = \frac{n + \Phi_y/(2\pi)}{N_y} \pmod{1}, \quad 0 \leq n \leq N_y - 1, \quad (3)$$

where Φ_y is the amount of flux threaded through the cylinder, which offsets the y momentum as shown in Fig. 2. We will Fourier transform each of these momentum cuts in the x direction, such that our basis states are hybrid Wannier orbitals, periodic in the y direction and localized in the x direction.² We will sometimes call this mixed xk representation.

The choice of real-space basis in the x direction is not unique, but we choose the basis of maximally localized Wannier orbitals. Using the maximally-localized orbitals ensures that the interactions are as short ranged as possible and hence minimizes the range of the interaction terms in the Hamiltonian and the entanglement of the ground state. Due to the relation between maximal Wannier localization and the Bloch Berry connection $\mathbf{A}_k$, this basis will also make manifest their topology. We perform the change of basis:

$$\begin{aligned} \hat{c}_{\pm, k_x, k_y}^\dagger &:= U_{\pm, b}(\mathbf{k}) \hat{f}_{b, k_x, k_y}^\dagger, \\ \hat{c}_{\pm, n, k_y}^\dagger &:= \int \frac{dk_x}{\sqrt{G_x}} e^{ik \cdot \mathbf{R}_n} \hat{c}_{\pm, k_x, k_y}^\dagger, \end{aligned} \quad (4)$$

where $\hat{c}_{\pm, n, k_y}^\dagger$ is the creation operator for the Wannier orbital for unit cell n in the x direction, $\mathbf{R}_n = n\mathbf{L}_1$ with $\mathbf{L}_1$ the Bravais lattice vector, and $U(\mathbf{k})$ is a 2×2 change-of-basis matrix for the internal (band index) degrees of freedom. The nontrivial

²A further advantage of this mixed representation over “snaking” around a real space cylinder is that k_y becomes a good quantum number, reducing the resource cost for a given radius [47,55].

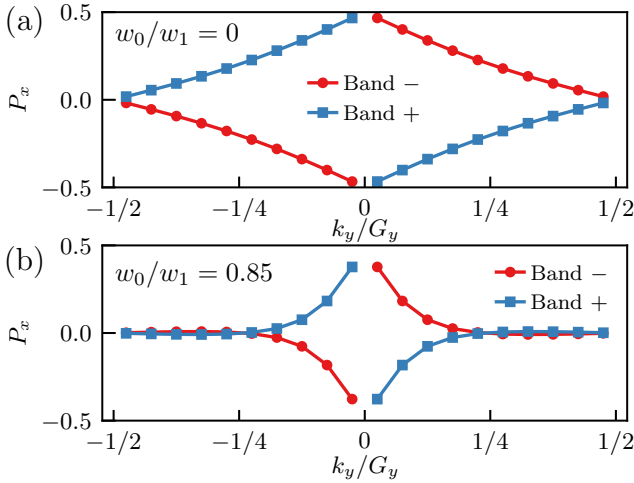


FIG. 4. Polarization P_x as a function of k_y and w_0/w_1 . At each value of w_0/w_1 , there are two bands with Chern number ± 1 , which wraps once as k_y/G_y increases by 1. This requires a single discontinuity in P_x , which we have chosen to place at $k_y = 0$. One can see that P_x is more linear for $w_0/w_1 = 0$, reflecting flatter Berry curvature.

topology of the tBLG flat bands [29,33–38] is made explicit in the hybrid Wannier basis by the fact that the states with subscripts $\pm$ are constructed from bands with Chern numbers ± 1 . We will explain this in more detail below. We choose the internal rotations $U(\mathbf{k})$ so that the Wannier orbitals are maximally localized (i.e., their spread in the x direction is minimized). Since the problem is effectively 1D for each k_y cut, we can employ a well-known algorithm [56] to deterministically calculate the unique $U(\mathbf{k})$ [up to $(k_y, \pm)$ dependent phases]. Figure 3 shows examples of the Wannier orbitals. One can see they are localized in the x direction but extended and periodic in y . The charge density is also not uniform in the y direction but is concentrated in certain regions corresponding to the AA region [26]. For later notational convenience, we also define $|w(\pm, n, k_y)\rangle = \hat{c}_{\pm, n, k_y}^\dagger |0\rangle$. We emphasize that the $\pm$ basis is not the energy eigenbasis of the single-particle Hamiltonian. In the $\pm$ basis with the gauge convention described later, the $k.p$ expansion of the two band Hamiltonian around $K^\pm$ points takes the form $h(K^\pm + \mathbf{q}) \propto \mp(q_x \cdot \sigma_x + q_y \cdot \sigma_y)$.

A key physical property of Wannier orbitals is their polarization $P_x(\pm, k_y)$, which can be derived via modern theory of polarization [57–60]. They can be thought of as the center of Wannier orbital inside the zeroth unit cell:

$$P_x(\pm, k_y) = \frac{\langle w(\pm, 0, k_y) | \hat{x} | w(\pm, 0, k_y) \rangle}{L_1 \cdot \mathbf{e}_1}, \quad (5)$$

where we normalize the polarization by the x extent of the unit cell. The polarization is related to the Berry phase along each momentum cut via the Wilson loop $e^{2\pi i P_x(k_y)} = e^{i \int_{A_x(k_x, k_y)} dk_x / G_x}$ and is only defined modulo 1 [60]. Therefore P_x returns to itself as k_y sweeps across the (mini) BZ. Furthermore, the Chern number of a band is conveniently expressed in terms of the total winding of the polarization, $C = \int dk_y \frac{dP_x}{dk_y}$. In Fig. 4, we plot the polarization versus k_y momentum at various different w_0 . We make two observations: First, we see that the polarization of the plus (minus) band winds from 0 to 1 (0 to -1) as momentum goes from

-0.5 to 0.5 . We may therefore identify these bands as having Chern numbers ± 1 —hence our index convention. On the other hand, the profile of the polarization changes as w_0/w_1 increases from 0 to 0.85. At $w_0/w_1 = 0$, the slope is constant, and the Wannier orbitals are almost equally spaced in the x direction, reminiscent of the lowest Landau level of a 2D electron gas in a magnetic field. At $w_0/w_1 = 0.85$, however, P_x is constant for most k_y values and suddenly changes around the Γ point.

There is a subtle issue relating our convention for polarization to our choice of gauge for single-particle wave functions in the mBZ. Since the polarization increases by ± 1 as k_y increases by 2π , we must choose a k_y where the polarization wraps around. We pick the convention that the wrapping $P_x \rightarrow P_x \pm 1$ occurs at $k_y = 0$, as shown in Fig. 4. In terms of the Wannier orbitals, this means that their centers of charge move continuously with k_y , except at $k_y = 0$ where they “exit” the unit cell and “enter” the neighboring unit cell. In terms of the momentum space creation operator $\hat{c}_{\pm, \mathbf{k}}^\dagger$ this corresponds to a choice of gauge that is smooth in the upper and lower halves of the Brillouin zone but *discontinuous* across $k_y = 0$. This discontinuity will appear in several figures below.

Finally, let us give the explicit action of global symmetries on our basis states. We first note the C_3 symmetry of the continuum model is weakly broken by the cylindrical geometry and is no longer an explicit symmetry of the model. We also note that for flux values $\Phi_y \neq 0 \pmod{\pi}$, the C_{2x} symmetry is not present.

Similar to Ref. [25], we partially fix the gauge of the flat band Bloch states such that the symmetries act in a simple way on the hybrid Wannier orbitals:

$$\begin{aligned} T_{L_1} |w(\pm, n, k_y)\rangle &= |w(\pm, n+1, k_y)\rangle \\ T_{L_2} |w(\pm, n, k_y)\rangle &= e^{i2\pi k_y} |w(\pm, n, k_y)\rangle \\ C_2 \mathcal{T} |w(\pm, n, k_y)\rangle &= |w(\mp, -n, k_y)\rangle \\ C_{2x} |w(\pm, n, k_y)\rangle &= \mp i e^{-i2\pi k_y n} |w(\mp, n, -k_y)\rangle, \end{aligned} \quad (6)$$

where the last equation holds only at C_{2x} symmetric flux values. The first two definitions are the consequence of Eq. (4), while the latter two come from demanding the following actions in momentum space:

$$\begin{aligned} C_2 \mathcal{T} \hat{c}_{\pm, \mathbf{k}}^\dagger (C_2 \mathcal{T})^{-1} &= \sigma^x K \hat{c}_{\pm, \mathbf{k}}^\dagger, \\ C_{2x} \hat{c}_{\pm, k_x, k_y}^\dagger (C_{2x})^{-1} &= \sigma^y \hat{c}_{\pm, k_x, -k_y}^\dagger, \end{aligned} \quad (7)$$

where σ^x acts on $\pm$ indices, and K is the complex conjugation operator. This, together with a continuity criterion such that the Wannier functions are a smooth function of k_y , fixes the phase ambiguity up to an overall minus sign (Appendix B).³ Now that we have described the interacting Bistritzer-MacDonald model in detail, we proceed to discuss how we will solve for its ground state using DMRG.

³In the absence of C_{2x} symmetry, we use a heuristic such that the gauge is continuous as a function of Φ_y .

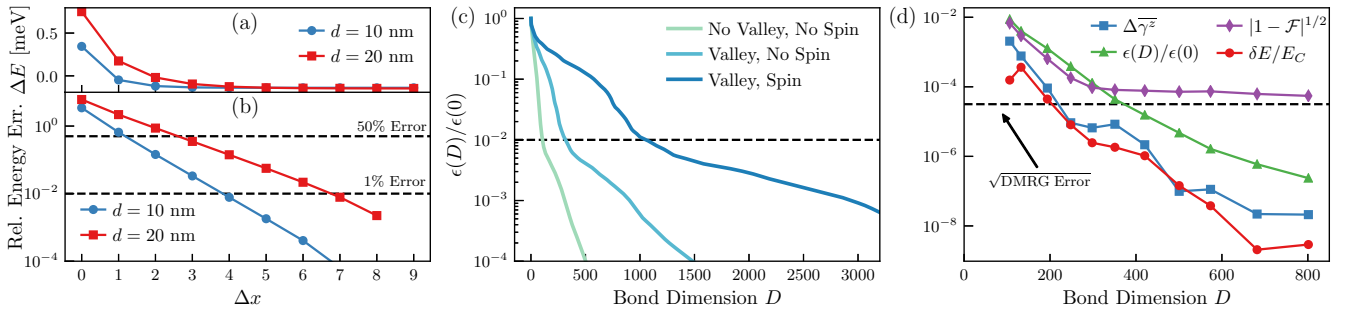


FIG. 5. (a) Energy difference between the QAH and the SM ansatz at different gate distances at $w_0/w_1 \sim 0.8$. (b) The relative error in energy difference between the QAH and SM ansatz. The black dashed lines indicate 50% error and 1% error. (c) Precision of the compressed MPO as a function of bond dimension for the IBM model with and without spin and valley degrees of freedom. The Hamiltonian is the IBM model at the chiral limit $w_0 = 0$ with parameters given by Table I. For (b), the cutoff range is reduced to $\Delta x = 3$. (d) Relative precision of MPO compression as a function of the postcompression bond dimension D . The precision is controlled by $\epsilon(D) = (\sum_{a=D+1} s_a^2)^{1/2}$, as described in Eq. (D28). Here $\Delta E(D) = E(D = 1000) - E(D)$ is the energy error in the ground state, $\mathcal{F}$ is the fidelity per unit cell between the ground state at D and the ground state at $D = 1000$, and $\Delta \bar{\gamma}^z$ is the error in the polarization versus $D = 1000$, described in Sec. IV. One can see that the precision improves roughly in proportion to ϵ , except for $|1 - \mathcal{F}|^{1/2}$, which is limited by the precision of DMRG (black dashed line).

III. MPO COMPRESSION AND DMRG

In this section we consider the practical details of performing infinite DMRG on the IBM model defined in the last section and the necessity of MPO compression. To perform infinite DMRG, we must express the Hamiltonian [Eq. (1)] as an infinite 1D matrix product operator (MPO) whose size D is called the bond dimension⁴ [61]. To map from 2D to a 1D chain, we order the Wannier orbitals $|w(\pm, n, k_y)\rangle$ by the positions $P_x(\pm, k_y) + n$ of their Wannier centers. Translation along L_1 simply increments $|w(\pm, n, k_y)\rangle \rightarrow |w(\pm, n + 1, k_y)\rangle$, so the 1D chain is periodic with a unit cell of size $N_B N_y$ sites ($N_B = 8$ with spin and valley). Once the MPO is obtained, we can in principle find its ground state with DMRG.

However, the long-range nature of the Coulomb interaction complicates matters. Although the screened Coulomb interaction decays exponentially in real space, truncating it at short range can lead to physically incorrect results. To demonstrate

this, Fig. 5(a) examines the energy of two ground state wave function ansätze “QAH” and “SM_y” as a function of the truncation distance of the interaction Δx ⁵ (these physical states are defined and used in Sec. V below). In particular, we examine the energy difference $\Delta E = E_{\text{QAH}} - E_{\text{SM}}$ in panel (a) and the relative energy difference in panel (b). The true energy difference between these states is $|\Delta E(\Delta x \rightarrow \infty)| \approx 0.1$ meV, yet the energy difference achieves 0.1 meV precision only at $\Delta x \geq 3$. Going from $\Delta x = 1 \rightarrow 2$, for example, their energies change by almost 0.1 meV.

More importantly, the relative error in energy difference $(\Delta E(\Delta x) - \Delta E(\infty))/\Delta E(\infty)$ reaches 1% only at the cutoff $\Delta x \geq 4$. This means that in order to resolve closely competing ground state candidates—which we will encounter in practice in Sec. II B below—we require a relatively large cutoff Δx .

Furthermore, the required cutoff is highly dependent on the model parameters. For example, if we increase the gate distance to 20 nm, then the screening distance is increased, and the relative energy gap does not achieve 50% precision until $\Delta x \approx 3$ [Fig. 5(b)].⁶ Together, these results suggest that premature truncation may lead to physically incorrect results, and we are forced to retain relatively long-range interactions in the Hamiltonian.

After mapping to a 1D chain, this means we must keep track of interactions up to range $R = N_B N_y \Delta x$ orbitals. An exact representation has an optimal bond dimension which scales as $D = O(R^2)$ (Appendix C). However, this still produces an MPO of size $D \sim 1 \times 10^4$ for $\Delta x \leq 4$ without spin and valley, and if we were to add in spin and valley it would be $D \sim 1 \times 10^5$. As the computational complexity of DMRG increases as $O(D^2)$, and D is usually a few hundred at most, the Hamiltonian for BLG is far too large for DMRG to be

⁴The MPO bond dimension is always denoted by D , and χ is reserved for the MPS bond dimension.

TABLE I. Parameters of the IBM model, DMRG calculation, and relevant energy scales. See main text for the definition of each entry.

Parameter	Value(s)
θ_{BM}	$\sim 1.05^\circ$
w_1	~ 109 meV
w_0/w_1	$[0, 1]$
Gate distance	10 nm
Relative permittivity	12
N_y	6
Φ_y	$\pi, \pi/10$
χ	≤ 1024
Δx	10
ϵ_{MPO}	$< 10^{-2}$ meV
Kinetic energy scale (t)	< 1 meV
Interaction energy scale (V)	< 10 meV

⁵We define Δx as the distance between the first and last field operators along the cylinder.

⁶Note, however, that larger *precompression* MPO bond dimension does not necessarily mean larger *postcompression* MPO bond dimension. The relationship between Hamiltonian parameters and the postcompression bond dimension is a subject of future work.

practical. The DMRG results of Ref. [25] considered a single spin and valley with interactions truncated at $\Delta x = 2$, resulting in an MPO of $D \sim 2000$ at $L_y = 6$. But increasing Δx , or adding spin and valley, makes the problem impractical.

On a finite system, the MPO can be viewed as a two-sided MPS and compressed by SVD truncation (this approach is implemented in the AutoMPO feature of the iTensor library [62]). However, in the infinite limit we wish to take here, this naive SVD truncation is unstable and actually destroys the locality of the Hamiltonian. To avoid this, Ref. [50] developed a modification of SVD compression which guarantees that the compressed Hamiltonian remains Hermitian and local in the thermodynamic limit.

As in finite SVD compression, an intermediate step of the algorithm produces a singular value spectrum s_a , and the bond dimension can be reduced by discarding the lowest values of the spectrum. For an appropriate notion of distance this truncation is optimal, and when applied to a single cut, the discarded weight $\epsilon(D) = \sqrt{\sum_{i=D+1}^{D'} s_a^2}$ upper bounds the error in $\hat{H}$ with respect to the Frobenius norm. In Appendix D we present efficient algorithms for finite-length unit cells and derive error bounds for various quantities. When exploiting quantum numbers, the algorithm is capable of compressing MPOs with bond dimensions 5×10^4 or larger on a cluster node.

With the bond dimension thus reduced to a reasonable value, we may perform DMRG. We use the standard TenPy library [63], written by one of us, taking full advantage of symmetries. Careful checks guaranteeing the accuracy and precision of our code, benchmarks, and other numerical details are given in Appendix E.

Figure 5 showcases the precision of our DMRG results. We performed DMRG at the chiral limit $w_0 = 0$ and computed the relative error in the ground state energy, ground state fidelity, and expectation values as a function of postcompression bond dimension D , relative to $D = 1000$. The relative precision $\epsilon(D)/\epsilon(0)$ improves quickly with D , dipping below 10^{-6} by $D = 800$. In accordance with the error bound on H , the ground state energy, wave function, and expectation values converge quickly as $\epsilon \rightarrow 0$.

As a proof of principle, we also performed MPO compression for the IBM model with spin and valley at $L_y = 6$ and $w_0 = 0$. Due to constraints on the size of the uncompressed MPO we can handle, we chose a cutoff range of $\Delta x = 3$, which resulted in a $D \sim 35000$ uncompressed MPO. The singular value spectrum of the MPO is shown in Fig. 5. If we define $\mathcal{F}$ as the fidelity per unit cell⁷ between the ground state of the compressed MPO with bond dimension D and the ground state of the MPO with $D = 1000$, then we see from Fig. 5 that in the spinless, single-valley calculation, $\epsilon(D)/\epsilon(0)$ and $|1 - \mathcal{F}|^{1/2}$ have roughly the same order of magnitude. Using this fact as a guide, we can estimate the bond dimension, where $|1 - \mathcal{F}|^{1/2} \sim 10^{-2}$ by looking at the value of $\epsilon(D)/\epsilon(0)$. This gives us bond dimensions $D = \{106, 317, 1057\}$ for spinless/single-valley, spinless/valley,

spin/valley MPO. While still relatively large, such bond dimensions are tractable with a standard workstation or cluster node when exploiting quantum numbers.

Of course the IBM model itself is only an approximation to the physical system, neglecting effects such as lattice relaxation, phonons, and twist angle disorder which, though small, are expected to enter at the 1 meV level. This provides a limit on the amount of precision which is physically useful. To be safe, we choose $\Delta x = 10$,⁸ $\epsilon = 10^{-2}$ meV, which results in postcompression bond dimensions of $D \approx 600$ –1000, depending on the value of w_0/w_1 . In conclusion, we have used MPO compression to reduce the Hamiltonian to a computational tractable size, incurring a precision error on the order of 10^{-2} meV—three orders of magnitude below the relevant energy scale of the problem. We now discuss the results of DMRG and the implications for the ground state physics of bilayer graphene.

IV. GROUND STATE PHYSICS AT HALF FILLING

In this section we report the results of our DMRG calculations and discuss the ground state physics of the (spinless, single valley) IBM model at half filling. We will show there is a clear transition from a quantum anomalous Hall state at small w_0/w_1 to a nematic semimetallic state at large w_0/w_1 . Furthermore, we will show that these ground states are almost exactly described by the k -space Slater determinants predicted by Hartree Fock.

A. Single particle projector and order parameter

We start by defining several crucial observables and order parameters. Because we find the DMRG ground state is translation invariant, all one-body expectation values can be obtained from the correlation matrix

$$P(\mathbf{k}) := \begin{pmatrix} \langle c_{+,k_x,k_y}^\dagger c_{+,k_x,k_y} \rangle & \langle c_{-,k_x,k_y}^\dagger c_{+,k_x,k_y} \rangle \\ \langle c_{+,k_x,k_y}^\dagger c_{-,k_x,k_y} \rangle & \langle c_{-,k_x,k_y}^\dagger c_{-,k_x,k_y} \rangle \end{pmatrix}. \quad (8)$$

This matrix is a projector when the expectation values are taken with respect to a Slater determinant, and it is the central variational object for k -space Hartree-Fock calculations. For DMRG in mixed- xk space, we calculate $P(\mathbf{k})$ by Fourier transforming two-point correlation functions.⁹

The one-body observables are spanned by the expectation values of Pauli matrices σ in the $\pm$ band space,

$$\gamma^z(\mathbf{k}) := \text{tr}[P(\mathbf{k})\sigma^z], \quad (9)$$

and similarly for $\gamma^x, \gamma^y, \gamma^+ = \gamma^x + i\gamma^y$. We denote mBZ averages by

$$\overline{\gamma^\alpha} := \frac{1}{A_{\text{mBZ}}} \int_{\text{mBZ}} d^2\mathbf{k} \gamma^\alpha(\mathbf{k}), \quad (10)$$

⁸This gives us an uncompressed bond dimension of order 5×10^4 , close to what would be necessary for spinful/valleyful calculation.

⁹Explicitly, $P(\mathbf{k})$ is defined for $\mathbf{k}$ on a $108 \times L_y$ grid of $\mathbf{k}$ points in the mBZ by computing expectations $\langle c_{\pm,0,k_y}^\dagger c_{\pm,n,k_y} \rangle$ with respect to the DMRG ground state on the mixed- xk space cylinder for $-53 \leq n \leq 54$ and performing a discrete Fourier transform with respect to L_y .

⁷We define the fidelity per unit cell in the thermodynamic limit as $\mathcal{F} = \lim_{N \rightarrow \infty} |\langle \psi_{N,D} | \psi_{N,D=1000} \rangle|^{2/N}$, where N is the number of unit cells.

where A_{mBZ} is the area of the mBZ. We will focus particularly on γ^z —which is an order parameter for $C_2\mathcal{T}$ and C_{2x} , as follows from Eq. (6) which implies that $\gamma^z(\mathbf{k}) = 0$ for a $C_2\mathcal{T}$ symmetric state, and $\gamma^z(C_{2x}\mathbf{k}) = -\gamma^z(\mathbf{k})$ for a C_{2x} symmetric state.

In the case where the state is indeed a momentum-diagonal Slater determinant, $P(\mathbf{k})$ acquires several special properties. In particular, if a momentum mode $\mathbf{k}$ is occupied by one electron, $P(\mathbf{k})$ takes values on the unit sphere and can be parametrized in spherical coordinates as

$$P(\mathbf{k}) = \frac{1}{2}(\sigma^0 + \cos \theta_k \sigma^z + \sin \theta_k \cos \varphi_k \sigma^x + \sin \theta_k \sin \varphi_k \sigma^y) \quad (11)$$

which implies $|\gamma^+|^2 + |\gamma^z|^2 = 1$. If the projector respects $C_2\mathcal{T}$ and C_{2x} symmetries, then, respectively, $\theta_k = \pi/2$ and $\varphi_k = -\varphi_{C_{2x}\mathbf{k}} + \pi$ at all $\mathbf{k}$. Finally, since $P(\mathbf{k})$ is a projector for momentum-diagonal Slater determinants, it satisfies $S_{\text{vN}}(\mathbf{k}) := -\text{Tr}[P(\mathbf{k}) \log P(\mathbf{k})] = 0$. In general, then, $S_{\text{vN}}(\mathbf{k}) \geq 0$ measures the deviation of a state from a translationally-invariant Slater determinant.

B. iDMRG details and parameter choices

Infinite DMRG (iDMRG) calculations were performed using the open source TeNPy package [63]. The numerical parameters and physical energy scales of the problem are summarized in Table I. In particular, we take $N_y = 6$ and $\Phi_y = \pi$ as the “default” values. The MPO bond dimension was compressed down to 600–1000, such that the expected error is of order 10^{-2} meV, as described in Sec. II B. To ensure that iDMRG was converged, we varied the MPS bond dimension χ between 200 and 1024. We found that DMRG converged well even at very low bond dimensions, except near the transition. We also allowed ground states with broken translation invariance with a doubled unit cell, but we found a fully translation invariant ground state for all parameters we tested.

C. Ground state transition and the QAH phase

We performed iDMRG at 44 values of w_0/w_1 in the range $[0,1]$. Figure 6(b) shows that the order parameter γ^z is nonzero for $w_0/w_1 \leq 0.8$ and vanishes for larger values of w_0/w_1 , signaling a transition from a $C_2\mathcal{T}$ and C_{2x} broken phase to a $C_2\mathcal{T}$ and C_{2x} symmetric phase. For low w_0/w_1 , not only is $C_2\mathcal{T}$ broken, but the state is almost perfectly polarized, with $\gamma^z \approx 1$. This implies that the state has a large overlap with the product state in which all “+” orbitals are occupied:

$$|\text{QAH}\rangle \approx \prod_{x,k_y} \hat{c}_{+,x,k_y}^\dagger |0\rangle. \quad (12)$$

Since the $\pm$ bands carry Chern number $C = \pm 1$, this state is a quantum anomalous Hall (QAH) insulator [31,44,64]. This approximation is quite good: The QAH state is well described by a product state plus small corrections, $|\langle \Psi_0 | \Psi_{\text{QAH}} \rangle| \approx 0.846$ per unit cell at $w_0 = 0$. Consequently, the QAH state has low entanglement entropy (Fig. 6) and DMRG converges at quite moderate bond dimensions.

Above $w_0/w_1 \approx 0.8$, $\gamma^z = 0$ and the state instead develops a large expectation value for $\gamma^+ = \gamma^x + i\gamma^y$. Section V below is devoted to the large w_0/w_1 phase, and we will see that it

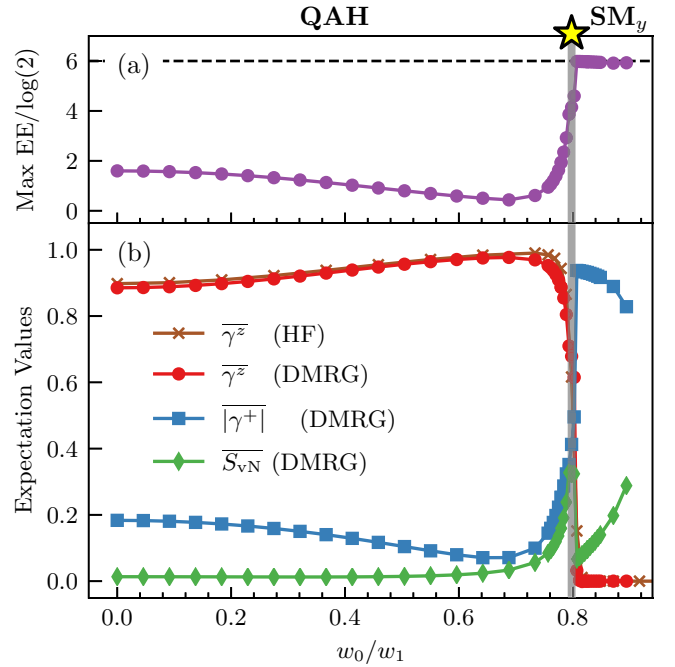


FIG. 6. The phase diagram of the IBM at half filling as a function of w_0/w_1 . There is a transition from a quantum anomalous Hall (QAH) phase to a semimetallic (SM) phase at $w_0/w_1 = 0.798 \approx 0.8$, represented by a yellow star. (a) Entanglement entropy of the DMRG ground state with hybrid Wannier orbitals ordered according to their polarization, maximized over all 12 entanglement cuts dividing the system into left and right halves. (b) Expectation values of various observables (defined in the text) in the DMRG or Hartree-Fock (HF) ground states. The polarization in Chern band space γ^z is an order parameter for the transition. Its drop across the transition is accompanied by a commensurate increase in $|\gamma^+|$, such that the DMRG ground state remains close to a Slater determinant. DMRG is performed at bond dimension $\chi = 1024$, $\epsilon_{\text{MPO}} = 10^{-2}$ meV, and is convergent away from the transition. (Gray shading indicates where DMRG is not well converged.)

is a nematic semimetal [51], which we refer to as “SM_y,” in reference to the ordering in the x/y plane. First, however, we analyze a surprising structure in the ground state correlations.

D. The remarkable accuracy of Hartree-Fock

The ground states of the strongly interacting IBM model are—quite surprisingly—very well described by k -space Slater determinants. For all values of w_0/w_1 away from the transition, the difference between the ground state and a Slater determinant as quantified by $S_{\text{vN}}(\mathbf{k})$ is small. In particular, Fig. 6 shows that $\bar{S}_{\text{vN}}$ is low in the QAH phase, increases or diverges near the transition, and is relatively small but growing in the SM_y phase. In the QAH phase this behavior is expected due to the large overlap with the simple Chern band polarized Slater determinant Eq. (12).

To provide further evidence that the ground state is essentially a Slater determinant, we compare DMRG results with Hartree-Fock (HF) calculations. Hartree-Fock determines an optimal Slater determinant approximation to the ground state of a many-body problem through a self consistent equation. Computationally, HF scales only polynomially in the number

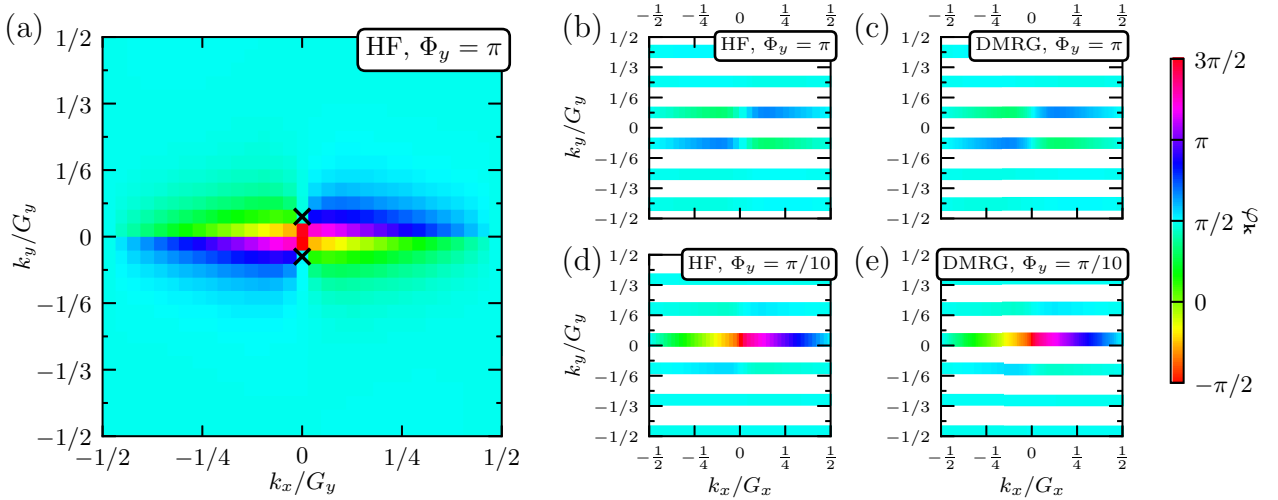


FIG. 7. Comparison of HF and DMRG calculations of $\phi_k = \arg[\gamma^+(\mathbf{k})] \in [-\pi, \pi]$ over the mBZ at $w_0/w_1 = 0.85$. (a) HF on a 30×29 grid with $\Phi_y = \pi$. The crosses represent the approximate location of Dirac points. (b) HF on a 30×6 grid at $\Phi_y = \pi$. (c) DMRG correlation function on a 108×6 grid at $\Phi_y = \pi$. (d) HF on a 30×6 grid at $\Phi_y = \pi/10$. (e) DMRG correlation function on a 108×6 grid at $\Phi_y = \pi/10$. The horizontal bands in (b)–(d) are centered on the k_y cuts used, given by Eq. (3). One can see that the DMRG and HF calculations are virtually identical. DMRG are performed using $\chi = 1024$. The discontinuity at $k_y = 0^+$ is a gauge choice, described in Sec. II B.

of k_y cuts (rather than exponentially for DMRG), so it provides a much cheaper alternative—when it is applicable. When the ground state of the IBM model is close to a Slater determinant, the HF ground state should be quite accurate and would have high overlap with the true ground state. We performed HF calculations on a $N_x \times N_y$ grid in the mBZ; numerical details of our HF calculations have been reported elsewhere [30].

We find that HF and DMRG results are nearly identical. The $C_2\mathcal{T}$ order parameter γ^z differs by around 2% (Fig. 6). Figure 7 shows a side-by-side comparison of the DMRG and HF predictions for $\phi_k = \arg[\gamma^+(\mathbf{k})]$ in the SM_y phase, where it completely specifies the Slater determinant because $\theta \equiv \pi/2$ is fixed by $C_2\mathcal{T}$ symmetry [see Eq. (11)]. Panel (a) shows a high-resolution HF calculation with $N_x = 30$, $N_y = 29$, which shows that $\phi_k \approx \pi/2$ over most of the mBZ but winds through 2π for k_y cuts that go near the Γ point. Panels (b)–(d) demonstrate that the same pattern appears with only $N_y = 6$ discrete momentum cuts. Both HF and DMRG produce a C_{2x} symmetric ϕ_k and the results obtained from both methods are almost indistinguishable. Other observables are similarly accurate in HF. We may therefore use HF to study large system sizes or observables that are not easily accessible in DMRG. For instance, Koopman’s theorem implies that the energies of single-fermion excited states are given by the self-consistent Hartree-Fock spectrum. We found the Hartree-Fock spectrum at the chiral limit has a gap of order 20 meV, showing the QAH state is gapped. We conclude that DMRG and HF agree to a remarkable degree and may be used almost interchangeably in this regime.

V. THE NEMATIC SEMIMETAL

We now show that the large- w_0/w_1 phase is a nematic semimetal, first described in Ref. [51], with energetics governed by the Berry curvature of the flat bands. This is an altogether different state than the Dirac semimetal which

appears in the noninteracting BM model. Our analysis is based on combination of DMRG (at $L_y = 6$) and HF (at $L_y \sim 30$), which agree wherever they can be compared. After establishing the nature of the nematic semimetal, we make contact with recent ideas in the literature [25,44,51]. Namely, we explain how the Ginzburg-Landau-like functional for the interband coherence φ proposed in Ref. [51] provides an intuitive description of the nematic state and the transition, and also confirm that the stripe state proposed in Ref. [25] is extremely competitive, with an energy only 0.2 meV/electron above the DMRG ground state.

A. The large w_0/w_1 phase is nematic

The large w_0/w_1 phase is a nematic state which breaks C_3 but preserves $C_2\mathcal{T}$. $C_2\mathcal{T}$ requires $\gamma^z = 0$ but allows for finite $\gamma^{x/y}$, so within the spherical coordinate description of Eq. (11) the state is characterized by $\theta = \pi/2$ and an azimuthal angle ϕ_k . This is clearly visible in Fig. 7(a): Throughout *most* of the mBZ, including at the mini- $K^\pm$ points, the state is $\gamma^y \approx 1$ ($\phi \sim \pi/2$). A state with finite $\gamma^x + i\gamma^y$ at the $K^\pm$ points breaks C_3 symmetry (nematicity) because C_3 acts as $C_3 = e^{i\pi\sigma^z/3}$ there. This is in contrast to the BM ground state, which has Dirac nodes at $K^\pm$: The BM Dirac structure $h(K^\pm + \mathbf{q}) \propto q_x\sigma_x + q_y\sigma_y$ instead causes $\gamma^x + i\gamma^y$ to wind by $+2\pi$.

Consequently we denote the large w_0/w_1 state “ SM_y .” Presumably the other C_3 -rotated versions are not found because the cylinder geometry weakly breaks the C_3 symmetry for finite N_y .

While the SM_y phase is close to a Slater determinant, it is not a small perturbation to the noninteracting ground state. To quantify this, Fig. 8(c) shows the trace distance between the BM ground state projector and the SM_y projector over the mBZ. Around the $K^\pm$ points, the trace distance rotates between complete agreement and orthogonality, consistent with the winding of γ^+ in the BM state versus the fixed $\gamma^y \approx 1$

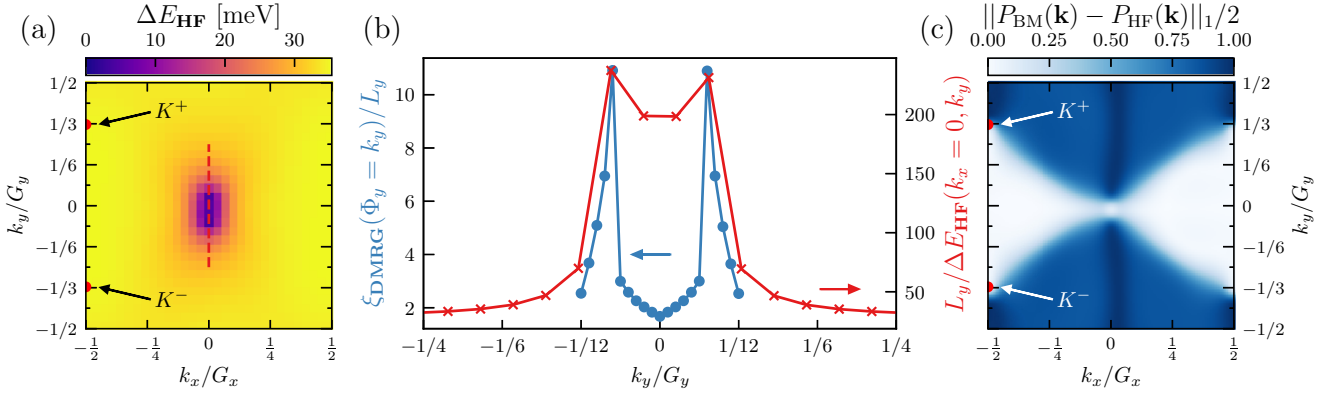


FIG. 8. Result of HF calculation at $N_x \times N_y = 30 \times 29$. (a) Energy gap between the top band and the bottom band. We see there are two Dirac points near the gamma point. (b) Correlation length obtained from thin cylinder DMRG at different flux values and the inverse HF gap size as a function of k_y . We see the correlation length becomes large near the Dirac points found in HF calculation. (c) The trace distance between the density matrices for the BM ground state and Hartree-Fock ground state. One can see that they are unrelated over much of the mBZ.

in SM_y . The trace distance also provides us with a gauge invariant way to identify the nematicity of the phase: The BM and SM_y projectors achieve near complete agreement along the y axis.

B. $C_2\mathcal{T}$ protected Dirac points

Near the Γ point, however, $\varphi_{\mathbf{k}}$ deviates from $\varphi = \pi/2$. We now show that this is because the SM_y phase features two Dirac points in the vicinity of Γ , which cause $\varphi_{\mathbf{k}}$ to wind there. This behavior is in fact enforced by topological properties [33,44]. For a generic two band problem in the presence of $C_2\mathcal{T}$, any Wilson loop is quantized: $W(C) := i \oint_C \mathbf{A} \cdot d\mathbf{k} = n\pi$ with $n \in \mathbb{Z}$ [65].¹⁰ In particular, $W(C) = (2n+1)\pi$ if and only if it encloses a Dirac cone.

This is the well-known topological protection of Dirac cones. In the case of the single valley BM model, Dirac cones at the mini $K^\pm$ points have the *same* chirality $W(C) = +\pi$. Therefore, not only are the Dirac cones locally protected, but even if they move away from $K^\pm$ they cannot meet and annihilate, enforcing the existence of either a pair of Dirac points or quadratic band touching.¹¹

The semimetallic nature of SM_y is borne out in both HF and DMRG numerics. The spectrum of the self-consistent HF Hamiltonian, shown in Fig. 8(a), has a large ($\Delta \approx 30$ meV) gap across most of the mBZ, except near the Γ point. The structure of Δ near the Γ point is consistent with two Dirac points at $\mathbf{k}_\pm := (k_x, k_y) \approx (0, \pm 0.05G_y)$ [Fig. 8(b)]. In contrast, the BM model is gapless at the $K^\pm$ points but gapped near Γ .

DMRG numerics can also detect these nodes, but special care is required. This is because the allowed momentum cuts,

Eq. (3), generically avoid $\mathbf{k}_\pm$. To confirm their existence, we continuously adjust the flux Φ_y through the cylinder [see Fig. 2(e)] and monitor the behavior of the DMRG ground state as the allowed momenta pass through the putative Dirac points. Figure 8(b) shows that the DMRG correlation length appears to diverge right as the allowed momenta pass through the location of the Dirac points $\mathbf{k}^\pm$ found in HF, consistent with the gap closing. We conclude that the large w_0/w_1 phase is a nematic semimetal: SM_y preserves $C_2\mathcal{T}$, breaks C_3 , and has two Dirac nodes on the y axis near Γ .

C. Ginzburg-Landau-like description of the SM_y phase

There is a very appealing Hartree-Fock picture for why the Coulomb interactions reconstruct the single particle Dirac semimetal into the nematic SM_y semimetal. When $C_2\mathcal{T}$ is preserved, the state is specified entirely by the phase of the interband coherence $\varphi_{\mathbf{k}}$, so the HF energy is a functional $E_{\text{HF}}[\varphi_{\mathbf{k}}]$. Reference [51] analytically computed this functional for the IBM model, Eq. (1), and found that the dominant contribution takes the form

$$E_{\text{HF}}[\varphi_{\mathbf{k}}] = E_{\text{HF}}^{\text{QAH}} + \frac{1}{2} \int g_{\mathbf{k}} (\nabla_{\mathbf{k}} \varphi_{\mathbf{k}} - 2\mathbf{a}_{\mathbf{k}})^2 d^2\mathbf{k} + \dots$$

$$g_{\mathbf{k}} = \frac{1}{A} \sum_{\mathbf{q}} V_{\mathbf{q}} \mathbf{q}^2 |\Lambda_{\mathbf{q}}(\mathbf{k})|^2. \quad (13)$$

Here $g_{\mathbf{k}}$ is an E_C -scale function independent of $\varphi_{\mathbf{k}}$, and $E_{\text{HF}}^{\text{QAH}}$ is the Coulomb energy of the QAH state. Finally, $\mathbf{a}_{\mathbf{k}}$ is a $U(1)$ vector potential which encodes the band geometry: Due to the $C_2\mathcal{T} = \sigma^x K$ symmetry, the $SU(2)$ Berry connection of the Bloch states is constrained to take the diagonal form $\mathbf{A}(\mathbf{k}) = \sigma^z \mathbf{a}_{\mathbf{k}}$, reducing it to a $U(1)$ connection.¹²

We see that the energy is similar to the Ginzburg-Landau functional for a superconductor in a magnetic field $\mathbf{F}_{\mathbf{k}} = d\mathbf{a}_{\mathbf{k}}$.

¹⁰The Berry phase is computed for the filled band only.

¹¹Note however that this “global” protection implicitly assumes translation symmetry: If the unit cell doubles, the bands fold and the band count doubles. Beyond two bands, $W(C)$ is only defined modulo π , so Dirac points can meet and annihilate. This mechanism underlies the $C_2\mathcal{T}$ stripe phase [25].

¹²Note that $\mathbf{a}_{\mathbf{k}}$ here does not have a quantized Wilson loop, unlike $\mathbf{A}_{\mathbf{k}}$ discussed in Sec. VB. This is because $\mathbf{a}_{\mathbf{k}}$ is defined in terms of Chern bands, which are not invariant under $C_2\mathcal{T}$.

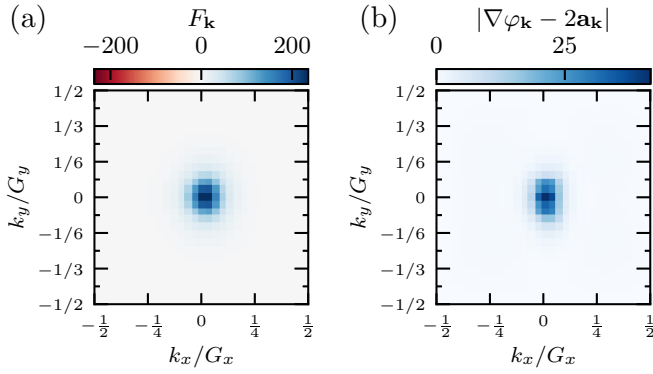


FIG. 9. (a) The Berry curvature $F_{\mathbf{k}}$ for $w_0/w_1 = 0.85$ shows concentration at the Γ point. (b) The k -space “supercurrent” $|\nabla\varphi_{\mathbf{k}} - 2\mathbf{a}_{\mathbf{k}}|$ is concentrated in the same region as $F_{\mathbf{k}}$. Both (a) and (b) are calculated in the units $G_x = G_y = 1$, $A_{\text{BZ}} = 1$.

This isn’t a coincidence: $\nabla_{\mathbf{k}}\varphi_{\mathbf{k}}$ can only appear via a gauge-covariant derivative because $\varphi_{\mathbf{k}}$, $\mathbf{a}_{\mathbf{k}}$ transform as a gauge pair under a $C_2\mathcal{T}$ -preserving phase redefinition of the Bloch states, $\hat{c}_{\pm,\mathbf{k}} \rightarrow e^{\pm i\varphi_{\mathbf{k}}} \hat{c}_{\pm,\mathbf{k}}$. However, there is no exact $U(1)$ symmetry, so the small “...” terms we neglect (for example the dispersion $\hbar$) do couple directly to φ .

The superconducting analogy can be made more concrete by applying a particle-hole transformation to only the $C = -1$ band, so that the coherence φ between $C = \pm 1$ bands maps to “superconducting pairing” between two $C = 1$ bands [44]. The Berry curvature $F_{\mathbf{k}}$ then appears with the same form as a magnetic field, albeit in k space, similar to how Berry curvature manifests as a “ k -space magnetic field” in the semiclassical equations of motion for Bloch electrons [66].

If we treat the mBZ as the unit cell, the Chern number $\frac{1}{2\pi} \int F_{\mathbf{k}} d^2\mathbf{k} = 1$ implies that there is one flux quantum per unit cell. Just like the vortex lattice of a superconductor in a magnetic field $F_{\mathbf{k}}$, this forces φ to have two vortices per unit cell. Each vortex ($+2\pi$ winding) is equivalent to a Dirac point, so this recovers the topological protection of the Dirac points discussed earlier. In the BM ground state, the two vortices are pinned to the $K^{\pm}$ points, while in the SM_y state they lie near Γ [Fig. 7(a)].

The vortices lead to an energy penalty relative to the QAH state, explaining Eq. (13). However, in our case the Berry curvature $F_{\mathbf{k}}$ is not uniform: Instead, $F_{\mathbf{k}}$ is concentrated near the Γ point [Fig. 9(a)]. By analogy to a superconductor in a nonuniform field, the lowest energy configuration of Eq. (13) will place the vortices in the region of concentrated $F_{\mathbf{k}}$, explaining their shift from $K^{\pm} \rightarrow \Gamma$. In Fig. 9(b), we confirm that $\nabla_{\mathbf{k}}\varphi_{\mathbf{k}} - 2\mathbf{a}_{\mathbf{k}} \approx 0$ in the region where $F_{\mathbf{k}}$ is small but is finite near Γ where $F_{\mathbf{k}}$ is concentrated. Accordingly, most of the energy penalty comes from near the Γ point. Increasing w_0/w_1 makes $F_{\mathbf{k}}$ increasingly concentrated, reducing the Coulomb penalty of SM_y relative to QAH.

The final ingredient driving the finite w_0/w_1 transition is the small terms like the dispersion $\hbar$ hidden in “...,” which slightly prefer the SM_y phase.¹³ As w_0/w_1 increases, the

TABLE II. Probability that some states are occupied by 0, 1, or 2 electrons in a given unit cell at $w_0/w_1 = 0.85$, $\Phi_y = \pi$, and $\chi = 1024$. N_{k_y} is the number of electrons with momentum k_y in the unit cell.

k_y	$-\frac{5}{12}$	$-\frac{1}{4}$	$-\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{5}{12}$
$p(N_{k_y} = 0)$	0.021	0.021	0.049	0.049	0.021	0.021
$p(N_{k_y} = 1)$	0.958	0.959	0.902	0.902	0.959	0.958
$p(N_{k_y} = 2)$	0.021	0.021	0.049	0.049	0.021	0.021

Coulomb penalty for the SM_y phase decreases due to the Berry curvature concentration, and these subleading terms win out. Consequently, while increasing w_0/w_1 does slightly increase the bandwidth, its primary effect is actually via the redistribution of Berry curvature, which enters at the Coulomb scale ($g_{\mathbf{k}} \sim E_C$).

D. “Thin cylinder” DMRG analysis

The DMRG ground state at $N_y = 6$ can be approximated by a particularly simple “thin cylinder” ansatz, provided none of the momentum cuts cross the region of large Berry curvature near Γ . We ensure this by taking $\Phi_y = \pi$ and focus on $w_0/w_1 \approx 0.85$. In this case, the DMRG ground state has a relatively small particle number fluctuation in the unit cell. This is because away from Γ , $P(\mathbf{k})$ varies slowly with k_x , so in the xk space the correlations are local (intra-unit cell).

Table II shows the probability for each momentum mode in the unit cell to be occupied by 0, 1, or 2 electrons. All modes have $p(N_{k_y} = 1)$ larger than 0.9 and many of them larger than 0.95. This suggests there is a simple “thin-cylinder” ansatz with no particle number fluctuation per momentum mode: Following Ref. [25], we define

$$|\Psi_{\text{TC}}\rangle = \prod_{x,k} \left(\cos \frac{\theta(x, k_y)}{2} \hat{c}_{+,x,k_y}^\dagger + \sin \frac{\theta(x, k_y)}{2} e^{i\varphi(x, k_y)} \hat{c}_{-,x,k_y}^\dagger \right) |0\rangle. \quad (14)$$

It is easy to check that if θ and φ are independent of x , these θ and φ correspond to those in Eq. (11). By inspecting Fig. 7(b), we find $\theta = \pi/2$ and $\varphi = \pi/2$ is a good approximation for the SM_y state¹⁴ so long as the momentum cuts are not close to $k_y = 0$. The fidelity per unit cell of the resulting ansatz $|\Psi_{\text{SM}}\rangle$ and the DMRG ground state is $\mathcal{F} \sim 0.98$ —remarkably accurate for such a simple ansatz. This ansatz also captures well the behavior of entanglement entropy in Fig. 6: Each momentum mode contributes a “Bell pair” to entanglement, making the maximum entanglement $6 \log 2$ with our choice of orbital ordering.

We can further motivate the thin-cylinder ansatz from the polarization $P_x(\pm, k_y)$ (Fig. 4). At large w_0/w_1 , P_x is close to 0 for most k_y values, and orbitals in one unit cell are well separated from those in neighboring unit cells. The interaction therefore strongly couples modes within the same unit cell, resulting in vanishing number fluctuation per unit cell.

¹³For example, in the SM_y phase, $\varphi_{\mathbf{k}}$ can perturbatively deform to follow the dispersion $\hbar$, particularly in the vicinity of Γ .

¹⁴Taking our different gauge convention into account, this is in agreement with Ref. [25].

TABLE III. Energy per electron of various trial states, evaluated with respect to the IBM Hamiltonian at $w_0/w_1 = 0.85$. Here “Dirac” refers to ground state of the single-particle BM Hamiltonian, and the parameters for SM and stripe ansätze are described in the text. We see that the stripe is very close to the DMRG ground state, and the Dirac state is well separated from the rest, reflecting the dominant importance of the Coulomb interactions. Note that the energies are negative because we have subtracted off the $q = 0$ part of the Coulomb interaction.

State	Energy (meV)
DMRG ground state (SM _y)	−28.24
QAH ansatz [Eq. (14)]	−28.04
SM _y ansatz [Eq. (14)]	−27.92
$C_2\mathcal{T}$ - stripe ansatz [Eq. (14)]	−28.08
Dirac (BM ground state)	−20.62

We also see why the thin-cylinder ansatz breaks down near the Γ point: $P(\mathbf{k})$ changes rapidly there [Fig. 7(e)], leading to inter-unit cell correlations. We stress that the ansatz is thus a crude approximation to the true HF ground state, since it fails to capture the complex Berry curvature contribution to the energy that is dominant near the Γ point. As $N_y \rightarrow \infty$, the momentum cuts unavoidably approach Γ , and the thin-cylinder ansatz will break down.

Finally, we comment on the energy competition between different candidate ground states. The simple form of the ansatz enabled Kang and Vafeek [25] to put forward another candidate for the ground state, which breaks translation symmetry in the L_1 direction in favor of a period-2 stripe state with screw symmetry $C_{2x}T_{L_1}$. To crude approximation, this state corresponds to $\theta(n, k_y) = \pi/2$ and $\varphi(n, k_y) = (-1)^n\pi/2 + \pi/2$ in the parametrization of Eq. (14). In order to test this ansatz, we computed the energy of the SM ansatz and the stripe ansatz and compared it to the DMRG ground state energy (Table III) and the QAH ansatz ($\theta = 0$). We also computed the energy of the ground state of the BM model with respect to the IBM model to establish the relevant Coulomb energy scale. We find that the energy of the QAH, SM_y, and the stripe ansatz are within 1/3 meV of the DMRG energy. While the stripe is not the true ground state for the parameter values studied here,¹⁵ this confirms the assertion in Ref. [25] that the stripe is a viable candidate in the wider phase diagram.

VI. CONCLUSIONS

In this work we have introduced a method for studying tBLG (and any other moiré material) using DMRG. Our method (Fig. 1) starts with the BM model, adds interactions, and compresses the resulting MPO down to a reasonable size so that the DMRG is computationally tractable. We carefully verified the correctness of our approach and showed that it is sufficiently precise to capture the ground state physics of the IBM model. To benchmark our approach we focused mostly

on the spinless, single-valley case. However, we showed in Sec. III that our method can be extended to the spinful, two-valley case with only moderately greater computational resources. Therefore we have identified a method to study the ground state physics of tBLG using DMRG.

Even though the spinless, single-valley model is not strictly physical, our results have several important conceptual implications for the study of tBLG. Remarkably, we found that the DMRG ground state is well approximated as a k -space Slater determinant for all w_0/w_1 . However, we stress that the ground state nevertheless has *no* relation to the ground state of the single-particle BM model: It mixes states very far from the Fermi surface of the BM Hamiltonian [Fig. 8(c)]. At least near the magic angle this suggests that weak-coupling approaches to tBLG, which rely on various details of the BM Fermi surface, will miss the essential physics. Instead, the energetics are dominated by the exchange physics of the Coulomb interaction. As a result, the effective band structure (as would be computed from the self-consistent Hartree-Fock Hamiltonian) is entirely different than that of the BM model, with a width set by E_C [Fig. 8(a)]. This is true even in the large w_0/w_1 nematic semimetal phase SM_y, which may have some relation to the $\nu = 0$ semimetallic resistance peak found in experiment.

Furthermore, it is subtle to describe the observed ground states within a 2D Wannier-localized “Mott insulating” picture. For small w_0/w_1 we have a QAH phase: The filled states have net Chern number, so the projector onto the filled states cannot be 2D Wannier localized. This is not to say that it is impossible to find these states within a numerical approach which starts from 2D Wannier orbitals (which is just a change of basis), but rather that in such a basis the order would manifest as a set of coherences $\langle c_i^\dagger c_j \rangle$ between sites rather than an onsite order parameter. Presumably this would complicate any mean-field approach which depends on a site-local self energy.

Taken together, this supports the point of view that tBLG is more closely related to quantum Hall ferromagnetism, where symmetry breaking is driven by the combination of band topology and Coulomb exchange, than it is to the Mott insulating physics of the Hubbard model. But of course in contrast to quantum Hall systems, tBLG comes with time-reversal symmetry, making it amenable to superconductivity. Future work will explore the physics of tBLG upon restoring the spin and valley degrees of freedom.

ACKNOWLEDGMENTS

We thank Xiangyu Cao and Eslam Khalaf for guiding our understanding of MPO compression and the nematic semimetal, as well as discussions with Shubhayu Chatterjee, Frank Pollmann, Senthil Todadri, and Ashvin Vishwanath. We also thank Jian Kang and Oskar Vafeek for carefully explaining their DMRG results to us. D.E.P. acknowledges support from the NSF Graduate Research Fellowship Program Grant No. NSF DGE 1752814. M.P.Z. was supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division of the US Department of Energy under Contract No. DE-AC02-05-CH11231 (van der Waals heterostructures program, KCWF16). J.H. was

¹⁵We verify this by initializing the DMRG using the doubled unit cell stripe ansatz and find the stripe reverts to the SM_y phase at convergence.

funded by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division under Contract No. DE-AC02-05-CH11231 through the Scientific Discovery through Advanced Computing (SciDAC) program (KC23DAC Topological and Correlated Matter via Tensor Networks and Quantum Monte Carlo). This research used the Savio computational cluster resource provided by the Berkeley Research Computing program at the University of California, Berkeley (supported by the UC Berkeley Chancellor, Vice Chancellor for Research, and Chief Information Officer).

APPENDIX A: INTERACTING BISTRITZER-MACDONALD MODEL

In this Appendix, we review the interacting BM model projected into the flat bands. We use the conventions and definition of $\mathbf{h}(\mathbf{k})$ from Supp. Mat. I of Ref. [30].

Let us first consider the Coulomb interaction

$$\begin{aligned}\hat{H}_C &= \frac{1}{2} \int d\mathbf{r} \int d\mathbf{r}' V(\mathbf{r} - \mathbf{r}') \psi_\alpha^\dagger(\mathbf{r}) \psi_\beta^\dagger(\mathbf{r}') \psi_\beta(\mathbf{r}') \psi_\alpha(\mathbf{r}) \\ &= \frac{1}{2A} \sum_{\mathbf{k}, \mathbf{k}', q} V_q \psi_{\alpha, \mathbf{k}+q}^\dagger \psi_{\beta, \mathbf{k}'-q}^\dagger \psi_{\beta, \mathbf{k}'} \psi_{\alpha, \mathbf{k}},\end{aligned}\quad (\text{A1})$$

where A is the sample area, $V_q = \int d\mathbf{r} V(\mathbf{r}) e^{iq\cdot\mathbf{r}}$, and α, β are combined layer-sublattice indices. Summation over repeated indices is implicit. The Fourier components of the Fermi operators are defined to satisfy the canonical anticommutation relations:

$$\{\psi_{\alpha, \mathbf{k}}^\dagger, \psi_{\beta, \mathbf{k}'}\} = \delta_{\alpha, \beta} \delta_{\mathbf{k}, \mathbf{k}'}.\quad (\text{A2})$$

Next, we relabel the sums over the momenta $\mathbf{k}$ and $\mathbf{k}'$ as

$$\sum_{\mathbf{k}} \rightarrow \sum_{\mathbf{k} \in \text{mBZ}} \sum_{\tau} \sum_{\mathbf{G}},\quad (\text{A3})$$

where $\tau = \pm$ is a valley label, and $\mathbf{G}$ are the moiré reciprocal lattice vectors. We can now approximate the Coulomb interaction as

$$\begin{aligned}\hat{H}_C &= \frac{1}{2A} \sum_{\tau, \tau'} \sum_{\mathbf{q}} \sum_{\mathbf{k}, \mathbf{k}' \in \text{mBZ}} \sum_{\mathbf{G}, \mathbf{G}'} V_q \psi_{\alpha, \tau, \mathbf{G}(\mathbf{k} + \mathbf{q})}^\dagger \psi_{\beta, \tau', \mathbf{G}'}^\dagger \\ &\quad \times (\mathbf{k}' - \mathbf{q}) \psi_{\beta, \tau', \mathbf{G}'}(\mathbf{k}') \psi_{\alpha, \tau, \mathbf{G}(\mathbf{k})},\end{aligned}\quad (\text{A4})$$

where $\psi_{\alpha, \tau, \mathbf{G}(\mathbf{k})}^\dagger = \psi_{\alpha, \mathbf{k} + \tau \mathbf{K}_\Gamma + \mathbf{G}}^\dagger$ and $\mathbf{K}_\Gamma$ denotes the Γ point of the mBZ centered at the K points of the graphene layers. Note that by definition, $\psi_{\alpha, \tau, \mathbf{G}(\mathbf{k} + \mathbf{G}')}^\dagger = \psi_{\alpha, \tau, \mathbf{G} + \mathbf{G}'}^\dagger(\mathbf{k})$. Equation (A4) is only an approximation to the complete Coulomb interaction, as intervalley scattering terms have been neglected. This can be justified because of the long-range nature of the interaction, which suppresses intervalley scattering by a factor of order $V_{2\mathbf{K}_\Gamma}/V_0$.

Next, we perform a unitary transformation to the BM band basis and define

$$f_{m, \tau, \mathbf{k}}^\dagger = \sum_{\alpha, \mathbf{G}} u_{m, \tau, \alpha, \mathbf{G}}(\mathbf{k}) \psi_{\alpha, \tau, \mathbf{G}}^\dagger(\mathbf{k}),\quad (\text{A5})$$

where m labels the bands of the single-valley BM model, and $u_{m, \tau, \alpha, \mathbf{G}}(\mathbf{k})$ are the periodic part of the Bloch states of the BM Hamiltonian. Note that $f_{m, \tau, \mathbf{k} + \mathbf{G}'}^\dagger = f_{m, \tau, \mathbf{k}}^\dagger$ because the BM

Bloch states satisfy $u_{m, \tau, \alpha, \mathbf{G}}(\mathbf{k} + \mathbf{G}') = u_{m, \tau, \alpha, \mathbf{G} + \mathbf{G}'}(\mathbf{k})$. With this definition, Eq. (A4) takes the following form in the BM band basis:

$$\begin{aligned}\hat{H}_C &= \frac{1}{2A} \sum_{\tau, \tau'} \sum_{\mathbf{q}} \sum_{\mathbf{k}, \mathbf{k}' \in \text{mBZ}} V_q [\Lambda_q^\tau(\mathbf{k})]_{mn} [\Lambda_{-q}^{\tau'}(\mathbf{k}')]_{m'n'} \\ &\quad \times f_{m, \tau, \mathbf{k} + \mathbf{q}}^\dagger f_{m', \tau', \mathbf{k}' - \mathbf{q}}^\dagger f_{n', \tau', \mathbf{k}'} f_{n, \tau, \mathbf{k}},\end{aligned}\quad (\text{A6})$$

where the sums over band indices are implicit, and the form factors are given by

$$[\Lambda_q^\tau(\mathbf{k})]_{mn} = \sum_{\alpha, \mathbf{G}} u_{m, \tau, \alpha, \mathbf{G}}^*(\mathbf{k} + \mathbf{q}) u_{n, \tau, \alpha, \mathbf{G}}(\mathbf{k}).\quad (\text{A7})$$

In this work, we consider the single-valley model, which means that we fix all valley labels, i.e., $\tau = +$ everywhere. The single-valley Coulomb interaction is then given by

$$\hat{H}_{C,sv} = \frac{1}{2A} \sum_{\mathbf{q}, \mathbf{k}, \mathbf{k}'} V_q : [f_{\mathbf{k} + \mathbf{q}}^\dagger \Lambda_q(\mathbf{k}) f_{\mathbf{k}}] [f_{\mathbf{k}' - \mathbf{q}}^\dagger \Lambda_{-q}(\mathbf{k}') f_{\mathbf{k}'}] : \quad (\text{A8})$$

where $\Lambda_q(\mathbf{k}) = \Lambda_q^+(\mathbf{k})$ and $f_{\mathbf{k}}^\dagger = f_{+, \mathbf{k}}^\dagger$ is a vector of creation operators running over the BM bands.

As a final step, we now project $\hat{H}_C$ into the subspace where all remote valence bands are occupied, and all remote conduction bands are empty. To do that, we first define the following Hartree Hamiltonian functional:

$$\hat{H}_h[P(\mathbf{k})] = \frac{V_0}{A} \sum_{\mathbf{G}} \left[\sum_{\mathbf{k}'} \text{tr}(P(\mathbf{k}') \Lambda_{\mathbf{G}}(\mathbf{k}')) \right] \sum_{\mathbf{k}} f_{\mathbf{k}}^\dagger \Lambda_{-\mathbf{G}}(\mathbf{k}) f_{\mathbf{k}},\quad (\text{A9})$$

where the fermion operators are restricted to the flat bands and which depends on a general Slater determinant correlation matrix $\langle f_{m, \mathbf{k}}^\dagger f_{n, \mathbf{k}'} \rangle = \sum_{\mathbf{G}} \delta_{\mathbf{k} + \mathbf{G}, \mathbf{k}'} [P(\mathbf{k})]_{nm}$. We also similarly define a Fock Hamiltonian functional:

$$\hat{H}_f[P(\mathbf{k})] = -\frac{1}{A} \sum_{\mathbf{q}, \mathbf{k}} V_q f_{\mathbf{k}}^\dagger \Lambda_q(\mathbf{k} - \mathbf{q}) P(\mathbf{k} - \mathbf{q}) \Lambda_{-q}(\mathbf{k}) f_{\mathbf{k}},\quad (\text{A10})$$

where again the fermion operators are restricted to the flat bands. With these definitions, one can write the flat-band projected Coulomb interaction as

$$\hat{H}_{C,sv|FB} = \tilde{H}_{C,sv} + \hat{H}_h[P_r(\mathbf{k})] + \hat{H}_f[P_r(\mathbf{k})],\quad (\text{A11})$$

where $\tilde{H}_{C,sv}$ is obtained from $\hat{H}_{C,sv}$ by simply restricting all band indices to the flat bands, and $P_r(\mathbf{k})$ is the correlation matrix of the Slater determinant where only the remote valence bands are filled.

Having obtained the flat-band projected single-valley Coulomb interaction, we now have to be careful not to double count certain interaction effects. In particular, the value of the hopping parameter in the tight-binding model of monolayer graphene is chosen to best reproduce the experimentally observed Dirac velocity. Importantly, this Dirac velocity is already renormalized by the Coulomb interaction. So if we want to explicitly add back the complete Coulomb interaction, we must make sure not to forget to subtract off the renormalization of the dispersion. In practice, this means that we have to subtract off the following Hartree-Fock Hamiltonian:

$$\hat{H}_{\text{sub}} = \hat{H}_h[P_0(\mathbf{k})] + \hat{H}_f[P_0(\mathbf{k})],\quad (\text{A12})$$

where $P_0(\mathbf{k})$ is the correlation matrix of the charge-neutrality Slater determinant of two *decoupled* graphene layers [31] restricted single spin and valley, expressed in the BM band basis. The complete projected single-valley BM model thus takes the form

$$\hat{H} = \hat{H}_{\text{BM},sv} + \hat{H}_{C,sv|\text{FB}} - \hat{H}_{\text{sub}}. \quad (\text{A13})$$

Because the interlayer tunneling is only a small perturbation compared to the intralayer hopping, it does not significantly change the remote bands. It thus holds to a very good approximation that

$$[P_r(\mathbf{k})]_{mn} = [P_0(\mathbf{k})]_{mn} \text{ for } m, n \in \text{remote bands}. \quad (\text{A14})$$

Now combining all the single-particle terms in Eq. (A13), one obtains the matrix $h(\mathbf{k})$ defined in Eq. (1). The remaining interaction Hamiltonian of Eq. (A13) is then exactly the second term of Eq. (1).

Finally, let us also comment on the value of the dielectric constant ϵ_r that we use for the projected model. The dielectric constant of the unprojected model is determined by the insulating hexagonal boron-nitride (hBN) substrate. In particular, it is given by the geometric mean of the dielectric constants of hBN parallel and orthogonal to the atomic plane: $\epsilon_{\text{hBN}} = \sqrt{\epsilon_{\perp}\epsilon_{\parallel}} \approx 4.4$. When we restrict to the flat bands we have to take into account that the filled remote bands which have been projected out act as another insulating background, whose polarization will also contribute to the dielectric constant. We phenomenologically incorporate this effect by screening the Coulomb potential with the static RPA polarization of eight Dirac cones at neutrality. In doing so, we ignore the small interlayer tunneling terms (whose effect on the remote bands is expected to be small), and we overcount a small contribution coming from the states in the flat bands themselves. The polarization bubble of Dirac fermions in monolayer graphene was calculated in Ref. [67]. Using the results from that paper, we obtain a dielectric constant $\epsilon_r = \epsilon_{\text{hBN}} + N_D * 0.73 \approx 10.3$, where $N_D = 8$ is the number of Dirac cones. This gives an order-of-magnitude estimate for the dielectric constant of the projected model. Given the many approximations used in deriving this estimate, it is safest to assume that the true dielectric constant lies somewhere in the interval $\epsilon_r \sim 6\text{--}12$.

We also note that the effective gate distance d appearing in $V_{\mathbf{q}} = \frac{e^2}{4\pi\epsilon_q} \tanh(qd)$ is modified by the anisotropy of the substrate dielectric constant. Specifically, if d_g is the physical distance to the gate, then $d = \sqrt{\frac{\epsilon_{\perp}}{\epsilon_{\parallel}}} d_g$, due to the bending of field lines in an anisotropic substrate.

APPENDIX B: WANNIER LOCALIZATION, GAUGE FIXING, AND SYMMETRIZATION

In this section, we list several invariants enforced by the maximal localization of Wannier orbitals and gauge fixing. We also comment on the difference between our gauge choice and the gauge choice in Ref. [25], and how we can map states from one gauge to another.

We denote the periodic part of momentum space orbitals $c_{\pm,\mathbf{k}}^\dagger$ by $|u_{\pm}(k_x, k_y)\rangle$. We define 2×2 overlap matrix as

$$O_{\alpha\beta}(\mathbf{k}, \mathbf{k}') = \langle u_{\alpha}(\mathbf{k}) | u_{\beta}(\mathbf{k}') \rangle. \quad (\text{B1})$$

For each overlap matrix O , we define unitary overlap matrix $\tilde{O}$ by the unitary part of the polar decomposition of O . Wannier localization [56] guarantees that the unitary overlap matrix in the k_x direction is always given by

$$\tilde{O}_{\alpha\beta}(\mathbf{k}, \mathbf{k} + \Delta k_x) = \delta_{\alpha\beta} e^{iP_x(\alpha, k_y)}, \quad (\text{B2})$$

where Δk_x is the unit of discretization in the x direction. Physically, this corresponds to choosing constant $A_x = P_x/G_x$.

Wannier localization fixes relative phases within each k_y mode. To fix relative phases between different k_y modes, we demand a continuity criterion in the k_y direction. One natural choice is to demand the unitary overlap matrix in the y direction be the identity matrix:

$$\tilde{O}_{\alpha\beta}(\mathbf{k}, \mathbf{k} + \Delta k_y) = \delta_{\alpha\beta}, \quad (\text{B3})$$

where the x component of $\mathbf{k}$ is some fixed k_{x0} . The condition notably does *not* wraparound the mBZ: The mode near $k_y/G_y = 0.5$ and the mode near $k_y/G_y = -0.5$ are not subject to the continuity condition with each other.

This leaves us with a global $U(1) \times U(1)$ phase ambiguity. Fortunately, the Wannier localization and continuity conditions are compatible with the symmetry requirements

$$\begin{aligned} C_2 \mathcal{T} |u_{\pm}(\mathbf{k})\rangle &= |u_{\mp}(\mathbf{k})\rangle, \\ C_{2x} |u_{\pm}(\mathbf{k})\rangle &= \mp i |u_{\mp}(-\mathbf{k})\rangle, \end{aligned} \quad (\text{B4})$$

which fixes the gauge up to an overall minus sign.

We note that the gauge fixing condition is different from Ref. [25] on two accounts: We use a different continuity criterion, and we fix C_{2x} to act as σ_y rather than σ_x . The latter is easy to account for by a $e^{i\pi\sigma_x/4}$ rotation, which maps σ_y to σ_x , but there is no guarantee that the gauge is the same even after such rotation.

Luckily, in the case of $L_y = 6$, $\Phi_y = \pi$, $w_0 = 0.185$, $e^{i\pi\sigma_x/4}$ rotation maps our ground state ansatz in Sec. VD to their ground state ansatz, suggesting our gauge choice is the same up to the rotation. In particular, this means we can go from their parametrization to our parametrization simply by changing $\varphi \rightarrow \varphi + \pi/2$. We make use of this fact when we write down the stripe ansatz in our gauge.

APPENDIX C: AN UNCOMPRESSED MPO FOR BLG

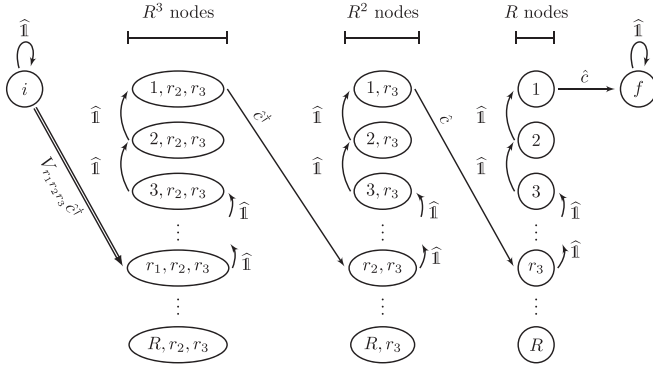
This Appendix details the construction of the uncompressed infinite MPO for bilayer graphene. More concretely, we show how to construct an iMPO which encodes arbitrary four-fermion interactions up to range R .¹⁶ Schematically, what is the iMPO for the Hamiltonian

$$\hat{H} = \sum_{ij} t_{ij} c_i^\dagger c_j + \sum V_{ijkl} c_i^\dagger c_j^\dagger c_k c_l, \quad (\text{C1})$$

for arbitrary interactions t and V ?

We proceed in several stages. First we provide a straightforward construction of such an MPO of size $D = O(R^3)$. However, this gives an iMPO which is too large to even

¹⁶For the construction of *finite* MPO for interacting fermions, as is relevant to quantum chemistry applications, we refer the reader to Ref. [49].

FIG. 10. A straightforward but rather inefficient MPO for $\hat{H}_{\text{simple}}$.

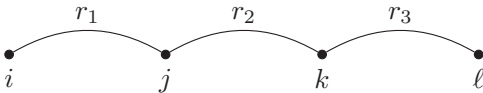
begin compressing ($D \approx 8\,000\,000$ for our standard Hamiltonian parameters). Second, therefore, we provide a more efficient MPO which represents the same operator at size $D = O(R^2)$, which will give $D \approx 100\,000$ —small enough to be compressed.

1. A straightforward MPO construction

This section will give a relatively straightforward way to construct an MPO for a long-range four-fermion Hamiltonian. For simplicity, we restrict ourselves to a 1D spin chain of spinless fermions and no internal degrees of freedom and construct a iMPO representation for a class of “toy” four-body Hamiltonians

$$\hat{H}_{\text{simple}} = \sum_{i < j < k < \ell} V_{ijkl} c_i^\dagger c_j^\dagger c_k c_\ell \quad (\text{C2})$$

with $|i - j|, |j - k|, |k - \ell| \leq R$.¹⁷ In particular, we have imposed the artificial properties that H contains only terms of the form $c^\dagger c^\dagger c c$ (which makes this non-Hermitian) is completely translationally invariant and has $i < j < k < \ell$ with strict inequalities. These restrictions will be lifted below.



Let us work in the finite state machine picture for the iMPO. For convenience, define $r_1 = j - i$, $r_2 = k - j$, $r_3 = \ell - k$. As $0 < r_1, r_2, r_3 \leq R$, our operator can have up to R^3 terms. For each one, we must first place a $c^\dagger$, then place $r_1 - 1$ identity operators $\hat{\mathbb{I}}$, then place another $c^\dagger$, and so on. To encode these into the finite state machine, we make a unique path for each term from the initial to final nodes, as shown in Fig. 10. For each term in the Hamiltonian, there is an edge $V_{r_1 r_2 r_3}$ from the initial node (i) to node (r_1, r_2, r_3) . After that, there is a unique path from node (r_1, r_2, r_3) to node (f). The nodes are labeled by the distances to nonidentity nodes that have yet to be placed. In each column, the path simply “counts down” in the first index, until it reaches 1. At that site

a nonidentity onsite operator is placed, and the path goes on to the next column. As a matrix, the operation of counting down is encoded by a “skipping matrix,” with identities on the first superdiagonal:

$$\hat{\mathbb{S}} = \begin{pmatrix} 0 & \hat{\mathbb{I}} & & & \\ & 0 & \hat{\mathbb{I}} & & \\ & & \ddots & \ddots & \\ & & & 0 & \hat{\mathbb{I}} \\ & & & & 0 \end{pmatrix}.$$

We therefore adopt a concise diagrammatic notation where *rectangular* nodes represent nodes that count down, as shown in Fig. 11. In this notation, $\hat{\mathbb{S}}_i$ acts on a node as¹⁸

$$\hat{\mathbb{S}}_i(\hat{O}[r_i \dots]) = (\hat{O}[r_i - 1 \dots]). \quad (\text{C3})$$

We will use this notation below. We also note that some transitions are deterministic, in the sense that the next node is fully specified by the current node [e.g., $(2, r_2, r_3) \rightarrow (1, r_2, r_3)$]. On the other hand, the transition from the initial node is highly branching, giving rise to the R^3 different terms in the Hamiltonian. We show such highly branching transitions in the figure by double arrows, while deterministic transitions are shown by single solid arrows.

Overall, this construction requires $C_1(R) := R^3 + R^2 + R + 2$ nodes. This is already somewhat efficient, as multiple terms will partially reuse the same paths. For instance, the paths through the state machine for V_{5,r_2,r_3} and V_{6,r_2,r_3} will be the same after the first few edges. To add other types of terms, such as $cc^\dagger cc^\dagger$ or $ccc^\dagger c^\dagger$, one must duplicate all these nodes, giving $4C_1(R)$ nodes. For bilayer graphene, we use a unit cell of size 12 and interaction cutoff of 10 unit cells, giving $D = 6\,970\,088$ just for the four-body terms. This makes the $O(D^3)$ compression algorithm impractical, so we must seek a more efficient way to encode the MPO.

2. An efficient MPO construction

We now describe a more efficient way to encode the uncompressed MPO and give explicit state machines for two-body, three-body, and four-body interaction terms. The key idea is to place the coefficient $V_{r_1 r_2 r_3}$ in the *middle* of the path rather than at the beginning.

Figure 11 (bottom) shows the more efficient construction of the MPO for the same operator as before, Eq. (C2). Let us unpack how it works. For each term $V_{r_1 r_2 r_3} c^\dagger c^\dagger c c$, we start at node (i), jump to the node ($\hat{c}^\dagger[1]$) by placing $c^\dagger$, then “count up” to r_1 , at which point we place the onsite operator $V_{r_1 r_2 r_3} c^\dagger$. The state machine then “counts down” for distance r_2 , places a c operator, counts down for distance r_3 , places a second c , and reaches the final node.

To distinguish “count up” skipping matrix and “count down” skipping matrix, we introduced “transposed skipping matrix” $\hat{\mathbb{S}}_i^T$ such that

$$\hat{\mathbb{S}}_i^T(\hat{O}[r_i \dots]) = (\hat{O}[r_i + 1 \dots]). \quad (\text{C4})$$

¹⁷This is slightly different from how we implemented Δx cutoff for BLG, where we demanded $\ell - i < R$. This will only change the complexity by a constant factor, so we stick to the simpler cutoff.

¹⁸This is slight abuse of notation, since it doesn’t convey that $\hat{\mathbb{S}}_i$ places an identity operator upon the transition.

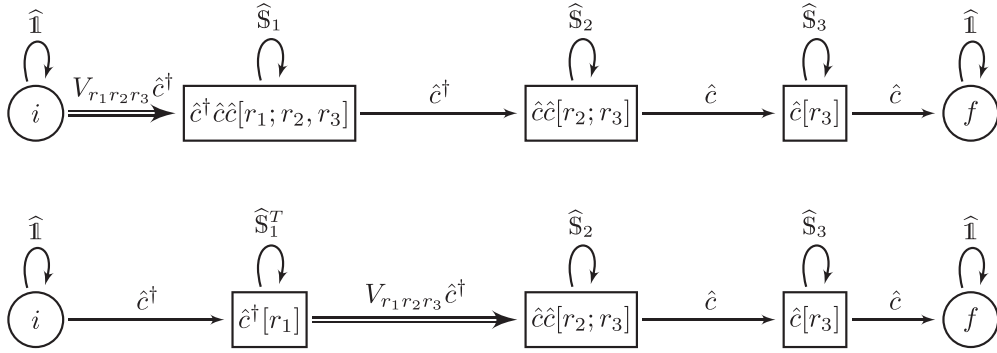


FIG. 11. (Top) Definition of the compact state machine notation. This is the same state machine as in Fig. 10 but where columns have been replaced by rectangular nodes. Rectangular nodes label the onsite operators that are yet to be placed, as well as the distances to them. The first rectangle contains R^3 nodes, the second contains R^2 , and the third contains R . The self-loop $\hat{S}$ means one should place an identity $\hat{1}$ and decrement the first index of the node. (Bottom) Another MPO which represents the same Hamiltonian with only $D = O(R^2)$ nodes instead of $O(R^3)$. Rectangular nodes to the right of the branching arrow are labeled similarly to the top, while the rectangular node to the left of the branching arrow is labeled by which operator it has already picked up and the distance to it.

The advantage over the previous method is that, instead of having a unique path for each term in the Hamiltonian, many of the paths are partially shared. For example the terms $V_{r_1 5 r_3}$ and $V_{r_1 7 r_3}$ will share the same first r_1 and last r_3 steps in their path through the state machine. This means that instead of requiring columns of sizes R^3 , R^2 , and R , we instead only need R , R^2 , and R , which is vastly more efficient. Another way to see this is to observe when the highly-branching transitions occur. In this more efficient algorithm, branching occurs between $\hat{c}^\dagger$ to $\hat{c}^\dagger \hat{c}^\dagger$, thereby reducing the number of paths to keep track of to R^2 . Of course, one could continue to further optimize the layout of the state machine, but we do not need a generic or completely optimal solution to this problem, only one that will render the BLG Hamiltonian small enough to fit in memory to be compressed.

Now that we have described the technique for writing down a sufficient efficient MPO construction, we will relax the artificial assumptions. Let us first relax the assumption that only $\hat{c}^\dagger \hat{c}^\dagger c c$ terms appear. In general, we will have a Hamiltonian of the form

$$\begin{aligned}
 H &= H_{\text{hop}} + H_{\text{int}} = H_{\text{hop}} + H_2 + H_3 + H_4 \\
 H_{\text{hop}} &= \sum_{i < j} \tilde{V}_{ij}^{\hat{c}c} \hat{c}_i^\dagger \hat{c}_j + \sum_{i < j} \tilde{V}_{ij}^{cc} \hat{c}_i \hat{c}_j^\dagger \\
 H_2 &= \sum_{i < j} \tilde{V}_{ij}^{nn} n_i n_j \\
 H_3 &= \sum_{i < j < k} \tilde{V}_{ijk}^{\hat{c}nc} \hat{c}_i^\dagger n_j \hat{c}_k + \tilde{V}_{ijk}^{\hat{c}nc} n_i \hat{c}_j^\dagger \hat{c}_k + \tilde{V}_{ijk}^{\hat{c}cn} \hat{c}_i \hat{c}_j^\dagger n_k \\
 &\quad + \tilde{V}_{ijk}^{\hat{c}nc} n_i \hat{c}_j \hat{c}_k^\dagger + \tilde{V}_{ijk}^{\hat{c}nc} \hat{c}_i n_j \hat{c}_k^\dagger + \tilde{V}_{ijk}^{\hat{c}cn} \hat{c}_i \hat{c}_j^\dagger n_k \\
 H_4 &= \sum_{i < j < k < \ell} \tilde{V}_{ijkl}^{\hat{c}ccc} \hat{c}_i^\dagger \hat{c}_j^\dagger \hat{c}_k \hat{c}_\ell + \tilde{V}_{ijkl}^{\hat{c}ccc} \hat{c}_i^\dagger \hat{c}_j \hat{c}_k^\dagger \hat{c}_\ell \\
 &\quad + \tilde{V}_{ijkl}^{\hat{c}ccc} \hat{c}_i^\dagger \hat{c}_j \hat{c}_k \hat{c}_\ell^\dagger + \tilde{V}_{ijkl}^{\hat{c}ccc} \hat{c}_i \hat{c}_j \hat{c}_k^\dagger \hat{c}_\ell^\dagger \\
 &\quad + \tilde{V}_{ijkl}^{\hat{c}ccc} \hat{c}_i \hat{c}_j^\dagger \hat{c}_k \hat{c}_\ell^\dagger + \tilde{V}_{ijkl}^{\hat{c}ccc} \hat{c}_i \hat{c}_j^\dagger \hat{c}_k^\dagger \hat{c}_\ell, \quad (C5)
 \end{aligned}$$

with all two-body, three-body, and four-body interactions involving $\hat{c}^\dagger$, $\hat{c}$, and n operators, and $\hat{c}^\dagger$ is represented by $\hat{c}$ for

concision. We can make this representation more amenable to finite state machine description by mapping $\tilde{V}_{ijkl}$ to $V_{r_0 r_1 r_2 r_3}$ such that $V_{i(j-i)(k-j)(\ell-k)} = \tilde{V}_{ijkl}$ (see Fig. 12). By combining all of these, one can generate the full Hamiltonian Eq. (C5) with $D(R) = 4R^2 + 6R + 2$ nodes. In practice, then, the spinless, single-valley BLG Hamiltonian on a cylinder with six momentum cuts (i.e., a unit cell of 12 sites) and interactions of range 10 unit cells is size $D \approx 58\,000$ before compression.

APPENDIX D: COMPRESSION OF INFINITE MPOS WITH GENERAL UNIT CELLS

This Appendix will present an algorithm for compressing infinite matrix product operators (iMPOs) with nontrivial unit cells, a generalization of Ref. [50]. The main application for these algorithms to compress the Hamiltonian for bilayer graphene down to a sufficiently small bond dimension that DMRG may be performed easily, while retaining sufficient precision to determine its physical properties. However, as the technique may be of independent interest, we will keep our discussion sufficiently general that the results apply to any iMPO that shows up in 2D DMRG.

Let us briefly review the context in which this algorithm is useful. Suppose that you have a Hamiltonian $\hat{H}$ for a 2D system and wish to find the ground state with DMRG [24,68]. A standard technique is to use a ‘thin cylinder’ geometry of circumference N_y sites and length $N_x \rightarrow \infty$. One then chooses a linear (1D) ordering for the sites on the cylinder by wrapping around in a helical pattern. This effectively reduces the problem to a 1D chain but at a cost: Interactions at distance r in 2D can be as far as $N_y \times r$ in 1D. Furthermore, the resource cost grows hugely, as the matrix product state (MPS) bond dimension needed to accurately capture a 2D area law state grows as $\chi \sim e^{N_y}$ [68]. In practice, therefore, 2D DMRG is often limited to around $N_y = 6 - 12$, even with bond dimensions of $\chi \sim 10\,000$ or more. For sufficiently long-range interactions in 2D, however, the bottleneck is not the MPS bond dimension but rather the MPO bond dimension needed to encode the Hamiltonian. For example, long-range four-body interactions of range R result in iMPOs of bond dimension $D \sim R^2$ (see

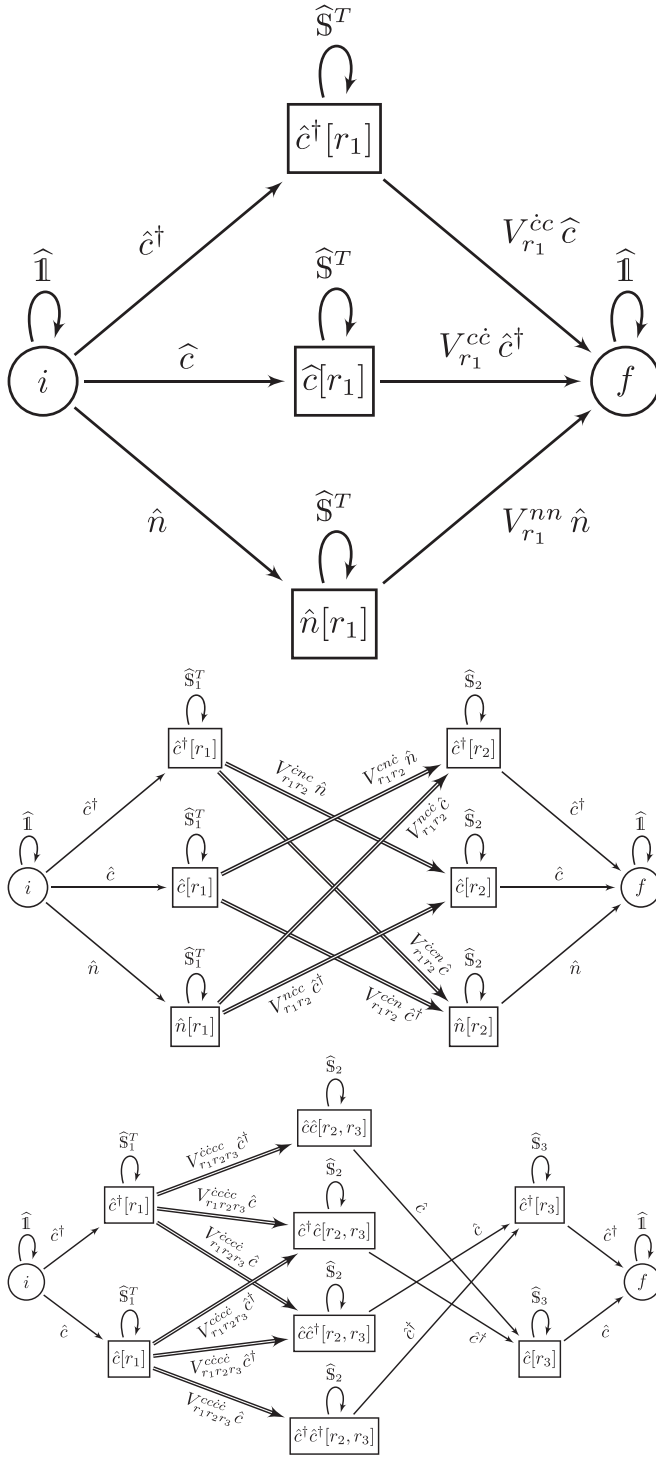


FIG. 12. Finite state machines for constructing the two-body, three-body, and four-body interactions of Eq. (C5). The notation is the same as in Fig. 11: Boxes stand for collections of many nodes with $1 \leq r_i \leq R$. For concision, the operator $c^\dagger$ is represented by $\hat{c}$ in superscripts. Note that some of the nodes can be reused among two, three, and four-body paths. We have suppressed the r_0 index for clarity on the initial node and for the V coefficients; this shows the case of unit cell of size $N = 1$.

Appendix C), and hence DMRG scales as $D^2 \sim R^4$, which becomes quickly impractical. The algorithms given below

allow one to proceed by finding the best approximation for the iMPO of bond dimension $D' < D$. For many physical Hamiltonians, this compression incurs only a minor penalty (say 10^{-4}) in the precision of the eventual ground state. In the case of the single-valley IBM model we have used in this paper, the bond dimension may be reduced by a factor of 10^3 , vastly improving the speed of DMRG.

The rest of this Appendix is organized as follows. We first give an overview of iMPO compression in the case without unit cells to set notation. We then discuss how an iMPO may be “topologically ordered,” how this vastly speeds up compression, and present a practical compression algorithm. Afterwards we examine the properties of compressed Hamiltonians: Under reasonable assumptions (1) Hermiticity is retained, (2) the fidelity of the compressed ground state versus the true ground state is high, and (3) their ground state energy and expectations values are accurate.

1. Lightning review of iMPO compression

We briefly review the notion of iMPO compression from Ref. [50]. It is well known that the optimal way to compress a 1D matrix product *state* is to perform a Schmidt decomposition and drop the smallest singular values (see Ref. [69] or [63] for a review). For iMPOs, we employ the same basic technique with a few modifications to preserve the locality of the operator.

We first present the algorithm in terms of (nonmatrix product) local operators. Consider a local operator $\hat{H}$ on an infinite 1D chain. We can split $\hat{H}$ into left and right halves at some bond:

$$\hat{H} = \hat{H}_L \hat{1}_R + \hat{1}_L \hat{H}_R + \sum_{a,b=1}^D \hat{h}_L^a \mathbf{M}_{ab} \hat{h}_R^b, \quad (\text{D1})$$

where a subscript L or R means the operator is supported entirely on the left or right half, respectively, and the matrix $\mathbf{M}$ keeps track of the terms which straddle the cut. This decomposition is not unique; we have the freedom to apply arbitrary unitaries on the left and right. We can take advantage of this freedom to put $\hat{H}$ into almost-Schmidt form

$$\hat{H} = \hat{H}_L \hat{1}_R + \hat{1}_L \hat{H}_R + \sum_{a=1}^D \hat{\mathcal{O}}_L^a s_a \hat{\mathcal{O}}_R^a, \quad (\text{D2})$$

where both $\{\hat{\mathcal{O}}_L^a\}$ and $\{\hat{\mathcal{O}}_R^a\}$ are orthonormal collections under the scaled Frobenius norm

$$(A, B) := \text{Tr}[\hat{A}^\dagger \hat{B}] / \text{Tr}[I], \quad (\text{D3})$$

and where $s_1 \geq s_2 \geq \dots \geq s_D \geq 0$ are referred to as the singular values.¹⁹ We further require that the operators $\hat{\mathcal{O}}_{L,R}^a$ are identity free, i.e., $(\hat{1}_{L,R}, \hat{\mathcal{O}}_{L,R}^a) = 0 \forall a$. The compressed operator is then simply the truncation of the sum to the largest

¹⁹We note that the singular values resulting from an almost-Schmidt decomposition and a true Schmidt decomposition are slightly different. Their relation is described in Sec. 8 of Ref. [50].

$D' < D$ singular values

$$\hat{H}' := \hat{H}_L \hat{1}_R + \hat{1}_L \hat{H}_R + \sum_{a=1}^{D'} \hat{O}_L^a s_a \hat{O}_R^a. \quad (\text{D4})$$

One can show that $\hat{H}'$ is the best approximation to $\hat{H}$ with only D' ‘terms’ straddling the cut [50]. Furthermore, as we shall see below, the accuracy of the approximation is controlled by the truncated singular value weight

$$\epsilon^2(D') := \sum_{a=D'+1}^D s_a^2. \quad (\text{D5})$$

To compute almost-Schmidt forms and compress operators, we work in the framework of matrix-product operators. We now recall their definitions to set notation and a few essential properties. Suppose we have a space of onsite operators with an orthonormal basis $\{\hat{1}, \hat{O}_2, \dots, \hat{O}_d\}$ where $\langle \hat{O}_\alpha, \hat{O}_\beta \rangle = \delta_{\alpha\beta}$ is an inner product.²⁰ Any translation-invariant operator, with unit cell of size N , can be written as²¹

$$\hat{H} = \dots [\hat{W}^{(1)} \hat{W}^{(2)} \dots \hat{W}^{(N)}] [\hat{W}^{(1)} \dots \hat{W}^{(N)}] \dots, \quad (\text{D6})$$

where each $\hat{W}^{(n)} = \sum_{\alpha=1}^d [W^{(n)}]^\alpha \hat{O}_\alpha$ is an operator-valued matrix of size $D^{(n)} \times D^{(n+1)}$. We require each $\hat{W}^{(n)}$ to have blocks of size $(1, D^{(n)} - 2, 1) \times (1, D^{(n+1)} - 2, 1)$

$$\hat{W} = \begin{pmatrix} \hat{1} & \hat{c} & \hat{d} \\ & \hat{A} & \hat{b} \\ & & \hat{1} \end{pmatrix}. \quad (\text{D7})$$

This ensures that each operator is a sum of terms that are the identity far enough to the left or right—a physical and mathematical necessity for a local operator.

An MPO is said to be in left-canonical form if all but the last column of each $\hat{W}^{(n)}$ are mutually orthonormal.²²

$$\sum_{a=1}^{D-1} \langle \hat{W}_{ab}^{(n)}, \hat{W}_{ac}^{(n)} \rangle = \delta_{bc}, \quad \forall n \in \mathbb{Z}/N\mathbb{Z}, 1 \leq b, c \leq D-1. \quad (\text{D8})$$

Similarly, an MPO is right-canonical if all the rows except the first are mutually orthonormal.

The representation (D6) is not unique, which is a manifestation of gauge freedom. Two MPOs $\{\hat{W}^{(n)}\}$ and $\{\hat{W}^{(n')}\}$ are gauge equivalent if there exist gauge matrices $\{G_n\}$ such that

$$G_{n-1} \hat{W}^{(n')} = \hat{W}^{(n)} G_n, \quad n \in \mathbb{Z}/N\mathbb{Z}. \quad (\text{D9})$$

So long as $\hat{H}$ is sufficiently local, one can show [50] there exists a gauge where the $\hat{W}^{(n)}$ ’s are left canonical (and another gauge for right canonical). Now that we have set definitions, the next section describes the compression algorithm for unit cell MPOs.

²⁰For example, a spin- $\frac{1}{2}$ chain has an onsite basis of the Pauli matrices $\{\hat{1}, \hat{X}, \hat{Y}, \hat{Z}\}$.

²¹This is notation for multiplication along the virtual indices: $[\hat{W}^{(1)} \hat{W}^{(2)}]_{ac} := \sum_{b=1}^{D^{(2)}} [W^{(1)}]_{ab}^\alpha [W^{(2)}]_{bc}^\beta \hat{O}_\alpha \otimes \hat{O}_\beta$.

²²More explicitly, if $\sum_{a=1}^d \sum_{b=1}^{D-1} [W^{(n)}]_{ab}^{\alpha*} [W^{(n)}]_{ac}^\alpha = \delta_{bc}$ for each $\hat{W}^{(n)}$.

2. The unit cell compression algorithm

The rough idea of the compression algorithm for unit cell MPOs $\{\hat{W}^{(n)}\}$ is as follows.

- (1) Compute the right-canonical form $\{\hat{W}_R^{(n)}\}$.
- (2) Find the gauge transform $\{G_n\}$ needed to transform to left-canonical form $\{\hat{W}_L^{(n)}\}$.
- (3) Take the SVD decomposition of the gauge transformation matrix: $G_n = U_n S_n V_n^\dagger$ and absorb the unitaries into the $\hat{W}$ ’s. This realizes the almost-Schmidt decomposition of Eq. (D2).

- (4) Truncate the number of singular values in S_n from $D^{(n)}$ to $D^{(n')}$ and correspondingly reduce the bond dimensions of the $\hat{W}$ ’s, producing the compressed Hamiltonian.

The rest of this section is devoted to showing the correctness of this procedure and filling in the details. It turns out that the most subtle part by far is canonicalization—the algorithm for putting an MPO into left/right canonical form. We therefore delay the discussion of canonicalization to Appendix D 3 below and for now simply assume it can be done.

We specialize to the case of $N = 2$ sites in the unit cell for concision, as larger unit cells are a direct generalization. Suppose that

$$R_{n-1} \hat{W}_R^n = \hat{W}^{(n)} R_n \quad (\text{D10})$$

is a gauge transformation so that the $\hat{W}_R^{(n)}$ ’s are right canonical. Then

$$\hat{H} = \dots \hat{W}^{(1)} \hat{W}^{(2)} \hat{W}^{(1)} \hat{W}^{(2)} \dots \quad (\text{D11})$$

$$= \dots \hat{W}_R^{(1)} \hat{W}_R^{(2)} \hat{W}_R^{(1)} \hat{W}_R^{(2)} \dots \quad (\text{D12})$$

by introducing R_2 at ∞ and sweeping to the right. We can then impose a further gauge transformation to make the first row of each $\hat{W}_R^{(n)}$ simultaneously identity free. This is done by

$$R'_n := \begin{pmatrix} 1 & \mathbf{t}_n & 0 \\ 0 & I & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (\text{D13})$$

where the $1 \times (D^{(n)} - 2)$ -dimensional vectors $\mathbf{t}_n$ are chosen such that $0 = \mathbf{c}_0^{(n)} + \mathbf{t}_n - \mathbf{t}_{n+1} A_0^{(n)}$ where $A_0^{(n)}$ and $\mathbf{c}_0^{(n)}$ are the $\hat{1}$ components of $\hat{A}^{(n)}$ and $\mathbf{c}^{(n)}$, respectively.

$$\begin{pmatrix} \mathbf{c}_0^{(1)} & \dots & \mathbf{c}_0^{(N)} \end{pmatrix} = (\mathbf{t}_1 \quad \dots \quad \mathbf{t}_N) \begin{pmatrix} I & & \dots & -A_0^{(n)} \\ -A_0^{(1)} & I & & \\ & \ddots & \ddots & \\ & & -A_0^{(n-1)} & I \end{pmatrix}. \quad (\text{D14})$$

In practice, one should solve this by imposing the identity free condition column by column. For each new column, this requires solving a linear equation of size N . Using the same technique from Eq. (71) of Ref. [50], the total operation can be performed in $O(ND^2)$ operations.

Imposing this gauge we may assume $\hat{W}_R^{(n)}$ has no identity components in its first row. This implies that the first column of $\hat{W}_R^{(n)}$ is already orthogonal to all the other columns, such that the gauge transformation to the left-canonical form can

Algorithm 1 Unit Cell iMPO Compression.

Require: $\{\widehat{W}^{(n)}\}$ is a first-order (see Ref. [50]) unit cell iMPO.

1: **procedure** UNITCELLCOMPRESS($\widehat{W}^{(n)}, \eta$) $\triangleright$ Cutoff η

2: $\widehat{W}_R^{(n)} \leftarrow \text{RIGHTCAN}[\widehat{W}^{(n)}]$

3: $\widehat{W}_R^{(n)} \leftarrow R_{n-1}' \widehat{W}_R^{(n)} R_{n-1}'^{-1}$ so that $\widehat{c}_0^{(n)} = 0$ $\triangleright$ Use t_n from Eq. (D14).

4: $\widehat{W}_L^{(n)}, C_n \leftarrow \text{LEFTCAN}[\widehat{W}_R^{(n)}]$

5: $(U_n, S_n, V_n^\dagger) \leftarrow \text{SVD}[C_n]$

6: $\widehat{Q}^{(n)}, \widehat{P}^{(n)} \leftarrow U_{n-1}^\dagger \widehat{W}_L^{(n)} U_n, V_{n-1}^\dagger \widehat{W}_R V$

7: $D^{(n)} \leftarrow \max_a \{1 \leq a \leq D^{(n)} : S_{aa}^{(n)} > \eta\}$ $\triangleright$

Defines the projector $\mathbb{P}_n$.

8: $\widehat{Q}^{(n)} \leftarrow \mathbb{P}_{n-1}^\dagger \widehat{Q}^{(n)} \mathbb{P}_n$

9: $S_n \leftarrow \mathbb{P}_n^\dagger S^{(n)} \mathbb{P}_n$

10: $\widehat{P}^{(n)} \leftarrow \mathbb{P}_{n-1}^\dagger \widehat{P}^{(n)} \mathbb{P}_n$

11: **return** $\widehat{Q}^{(n)}$

be written as

$$C_{n-1} \widehat{W}_R^{(n)} = \widehat{W}_L^{(n)} C_n, \quad (\text{D15})$$

with block-diagonal gauge transformation matrices $C_n = \text{diag}(1 \ C_n \ 1)$. Putting in a C_2 matrix at $-\infty$ and sweeping it to the center, we arrive at a mixed canonical form

$$\widehat{H} = \dots \widehat{W}_L^{(1)} \widehat{W}_L^{(2)} C_2 \widehat{W}_R^{(1)} \widehat{W}_R^{(2)} \dots \quad (\text{D16})$$

$$= \dots \widehat{W}_L^{(2)} \widehat{W}_L^{(1)} C_1 \widehat{W}_R^{(2)} \widehat{W}_R^{(1)} \dots \quad (\text{D17})$$

As C_n are block diagonal, we can compute their SVD's

$$C_n = U_n S_n V_n^\dagger \quad (\text{D18})$$

which will also be block diagonal. Define

$$\widehat{Q}^{(n)} := U_{n-1}^\dagger \widehat{W}_L^{(n)} U_n \quad (\text{D19})$$

$$\widehat{P}^{(n)} := V_{n-1}^\dagger \widehat{W}_R^{(n)} V_n \quad (\text{D20})$$

for $n \in \mathbb{Z}/N\mathbb{Z}$. Then, since $U_n U_n^\dagger = I = V_n V_n^\dagger$, we have

$$\widehat{H} = \dots \widehat{Q}^{(1)} \widehat{Q}^{(2)} S_2 \widehat{P}^{(1)} \widehat{P}^{(2)} \dots \quad (\text{D21})$$

$$= \dots \widehat{Q}^{(2)} \widehat{Q}^{(1)} S_1 \widehat{P}^{(2)} \widehat{P}^{(1)} \dots \quad (\text{D22})$$

which is analogous to the center-canonical form for MPS. The center bond can be swept back and forth via the gauge relation

$$S_{n-1} \widehat{P}^{(n)} = \widehat{Q}^{(n)} S_n, \quad n \in \mathbb{Z}/N\mathbb{Z}. \quad (\text{D23})$$

To see how compression works, we adopt the technique of assuming that the operator can be represented *exactly* by an MPO of lower bond dimension, i.e., that a number of the singular values vanish exactly. Finding the lower bond dimension MPO uses the same algorithm as compression when the small singular values are truncated, so this shows the correctness of the algorithm.

Thus we assume, temporarily, that only $D^{(n) \prime}$ of the $D^{(n)}$ singular values of S_n are nonzero. Hence there are projection operators $\mathbb{P}_n$ from bond dimension $D^{(n)}$ to bond dimension $D^{(n) \prime}$ with $\mathbb{P}_n \mathbb{P}_n^\dagger$ a projector and $\mathbb{P}_n^\dagger \mathbb{P}_n = I_{1+\chi^{(n) \prime}+1}$ and

$$S_n = S_n \mathbb{P}_n \mathbb{P}_n^\dagger = \mathbb{P}_n S_n' \mathbb{P}_n^\dagger = \mathbb{P}_n S_n' S_N, \quad (\text{D24})$$

where S_n' is the projected diagonal matrix of nonzero singular values.

We can then introduce pairs of projectors on each bond:

$$\begin{aligned} \widehat{H} &= \dots \widehat{Q}^{(1)} \widehat{Q}^{(2)} S_2 \mathbb{P}_2 \mathbb{P}_2^\dagger \widehat{P}^{(1)} \widehat{P}^{(2)} \dots \\ &= \dots \widehat{Q}^{(1)} S_1 \widehat{P}^{(2)} \mathbb{P}_2 \mathbb{P}_2^\dagger \widehat{P}^{(1)} \widehat{P}^{(2)} \dots \\ &= \dots \widehat{Q}^{(1)} S_1 \mathbb{P}_1 \mathbb{P}_1^\dagger \widehat{P}^{(2)} \mathbb{P}_2 \mathbb{P}_2^\dagger \widehat{P}^{(1)} \widehat{P}^{(2)} \dots \\ &= \dots \mathbb{P}_2^\dagger \widehat{Q}^{(1)} \mathbb{P}_1 \mathbb{P}_1^\dagger \widehat{P}^{(2)} S_2' \mathbb{P}_2^\dagger \widehat{P}^{(1)} \mathbb{P}_1 \mathbb{P}_1^\dagger \widehat{P}^{(2)} \mathbb{P}_2 \dots \\ &= \dots \mathbb{P}_1^\dagger \widehat{Q}^{(2)} \mathbb{P}_2 \mathbb{P}_2^\dagger \widehat{P}^{(1)} \mathbb{P}_1 S_1' \mathbb{P}_1^\dagger \widehat{P}^{(2)} \mathbb{P}_2 \mathbb{P}_2^\dagger \widehat{P}^{(1)} \mathbb{P}_1 \dots \quad (\text{D25}) \end{aligned}$$

It is now clear how to define a new representation for $\widehat{H}$ with a reduced bond dimension:

$$\widehat{P}^{(n) \prime} := \mathbb{P}_{n-1}^\dagger \widehat{P}^{(n)} \mathbb{P}_n \quad (\text{D26})$$

$$\widehat{Q}^{(n) \prime} := \mathbb{P}_{n-1}^\dagger \widehat{Q}^{(n)} \mathbb{P}_n \quad (\text{D27})$$

whereupon

$$\widehat{H} = \dots \widehat{Q}^{(1) \prime} \widehat{Q}^{(2) \prime} S_2' \widehat{P}^{(1) \prime} \widehat{P}^{(2) \prime} \dots \quad (\text{D28})$$

is a representation of $\widehat{H}$ with lower bond dimension. If we now relax the requirement that the truncated singular values were exactly zero, the strict equality of the new representation becomes approximate.

3. Canonicalization & topological sorting for unit cell MPOs

In this section we provide the “missing link” needed to complete the compression procedure: a canonicalization algorithm. Any unit cell MPO (UCMPO) can be put into left or right canonical form using QR iteration [50] with cost $O(ND^3)$. As many as 40 iterations can be necessary to reach high precision, making this quite slow in practice. However, the MPOs for Hamiltonians in DMRG have a special property, a “topological ordering,” which enables canonicalization to be performed with cost $O(ND^3)$ but *without iteration*. For large MPOs such as the one for BLG with $D \sim 100\,000$, this is a crucial speed-up. We first define a “topological ordering,” then provide the canonicalization algorithm and a proof of its correctness and runtime. We conclude the section with a few remarks on practical implementation details.

An MPO can be thought of as a finite state machine (FSM) for placing onsite operators in a certain order [70]. For MPOs with N tensors in a unit cell, the FSM gains an additional structure: The FSM has N parts, with the nodes of part n corresponding to the bond between $\widehat{W}^{(n-1)}$ and $\widehat{W}^{(n)}$ and edges between parts $n-1$ and n corresponding to tensor elements $\widehat{W}_{ab}^{(n)}$. See Fig. 13 for an example.

When one writes down an (non-unit-cell) MPO $\widehat{W}$ for a Hamiltonian “by hand,” then the MPO generally has a special structure: $\widehat{W}$ is upper triangular as a matrix. In Ref. [50], the upper triangular structure was shown to permit a fast canonicalization algorithm. However, this does not immediately generalize to a unit cell MPO, for a simple reason: If a unit cell MPO $\{\widehat{W}^{(n)}\}$ has bond dimensions $D^{(1)}, D^{(2)} \dots D^{(n)}$, not all equal, then the matrices are rectangular and cannot all be upper triangular. To find a good generalization of triangularity, we must look to the finite state machine.

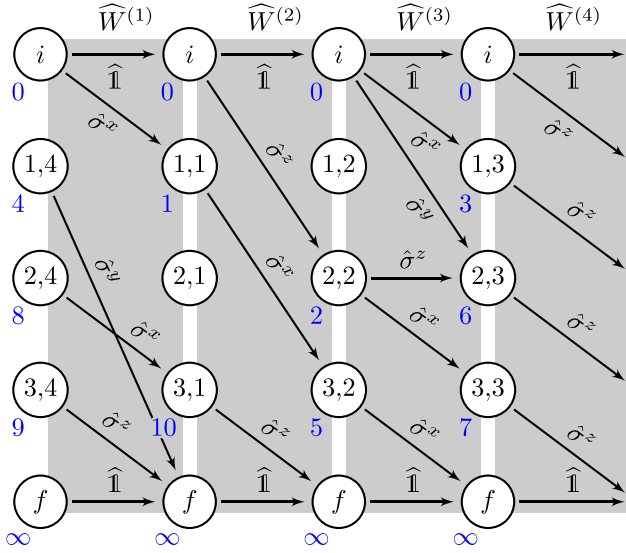


FIG. 13. An example of the finite state machine for an MPO with unit cell size 4. One onsite operator is placed for each arrow, and the arrows wrap around from right to left. Each gray box represents the data stored in one tensor. The UCMPO has bond dimension $(D_4, D_1, D_2, D_3) = (5, 4, 4, 5)$ and is loop free. The blue numbers are a (nonunique) topological ordering for the nodes.

A UCMPO $\{\hat{W}^{(n)}\}_{n=1}^N$ is said to be loop free if its finite state machine contains no loops *after* the initial and final nodes are removed from the graph. For $N = 1$, then an upper-triangular MPO is always loop free, and a loop-free MPO is always upper-triangular (up to permutation). We stress that both the upper triangular and loop free conditions are gauge dependent. Furthermore, for any N , if each $\hat{W}^{(n)}$ is square and upper-triangular, then the UCMPO is loop free. The converse is almost true as well; any loop free UCMPO is an upper-triangular MPO “in disguise.” To see this, we need a definition, which will be at the heart of this section.

Definition 1. A topological ordering for a UCMPO $\{\hat{W}^{(n)}\}_{n=1}^N$ is an ordering of the nodes of the FSM (excluding the initial and final nodes)

$$O = \{(a_1, n_1) < (a_2, n_2) < \dots < (a_D, n_D)\} \quad (\text{D29})$$

such that

$$\hat{W}_{ab}^{(n)} = 0 \text{ whenever } (a, n) \geq (b, n+1), \quad (\text{D30})$$

where $n \in \mathbb{Z}/N\mathbb{Z}$ indexes the bonds (explicitly, bond $n+1$ connects $\hat{W}^{(n)}$ to $\hat{W}^{(n+1)}$), $a \in \mathbb{N}$ indexes the node within the bond, and $D = \sum_n D^{(n)}$ is the total number of nodes.

If an MPO is loop free, then its finite state machine (excluding the initial and final nodes) is a directed acyclic graph and thus contains at least one topological ordering. This is easily computed by Kahn’s algorithm (a standard result in graph theory) with cost linear in the number of nodes plus edges in the FSM. With this, we can show that loop free UCMPOs are upper triangular ones “in disguise” and then use this ordering as the basis for an efficient canonicalization algorithm.

Lemma 1. Suppose $\{\hat{W}^{(n)}\}_{n=1}^N$ is a loop free MPO. Then, by inserting rows and columns of zeros and permuting the rows

Algorithm 2 Unit Cell iMPO (Left) Canonicalization.

Require: $\{\hat{W}^{(n)}\}_{n=1}^N$ is a loop free UCMPO.

- 1: **procedure** UNITCELLLEFTCANONICAL $\hat{W}^{(n)}, \eta$
- 2: $O = \text{KAHN'SALGORITHM}[\text{FSM}[\{\hat{W}^{(n)}\}]]$
- 3: **for** $(b, n+1) \in O$ **do**
- 4: $P \leftarrow \{a : O \ni (a, n+1) < (b, n+1)\}$
- 5: $r_a \leftarrow \sum_c \langle \hat{W}_{ca}^{(n)}, \hat{W}_{cb}^{(n)} \rangle, \quad \forall a \in P$
- 6: $R \leftarrow I_{D_n}, R_{ab} \leftarrow r_a, \quad \forall a \in P$
- 7: $\hat{W}^{(n)} \leftarrow \hat{W}^{(n)} R, \hat{W}^{(n+1)} \leftarrow R^{-1} \hat{W}^{(n+1)}$
- 8: $R \leftarrow I_{D_n}, R_{bb} \leftarrow (\sum_c \langle \hat{W}_{cb}^{(n)}, \hat{W}_{cb}^{(n)} \rangle)^{-1/2}$
- 9: $\hat{W}^{(n)} \leftarrow \hat{W}^{(n)} R, \hat{W}^{(n+1)} \leftarrow R^{-1} \hat{W}^{(n+1)}$
- 10: **return** $\{\hat{W}^{(n)}\}$

and columns of the matrices (which is a gauge transform), $\{\hat{W}^{(n)}\}$ can be made upper triangular.

Proof. Suppose the bond dimension of $\hat{W}^{(n)}$ is $D^{(n)}$ on the left and $D^{(n+1)}$ on the right, with $D = \sum_n D^{(n)}$. Let O be a topological ordering for $\{\hat{W}^{(n)}\}$ of the form (D29). Define a gauge matrix $\mathbb{P}_n$ of dimension $D^{(n)} \times D$ with matrix elements

$$[\mathbb{P}_n]_{b,i} = \begin{cases} 1 & \text{if } (b, n+1) = O_i \\ 0 & \text{otherwise.} \end{cases} \quad (\text{D31})$$

This “blows up” $\hat{W}^{(n)}$ on the right to bond dimension $D > D^{(n+1)}$ by inserting zeros and puts the indices into topological order. One can check that $\mathbb{P}_n^\dagger \mathbb{P}_n = I_{D^{(n+1)}}$, so we may define $\hat{W}^{(n')} := \mathbb{P}_{n-1}^\dagger \hat{W}^{(n)} \mathbb{P}_n$ of size $D \times D$ (which obeys $\mathbb{P}_{n-1} \hat{W}^{(n')} = \hat{W}^{(n)} \mathbb{P}_n$, making it a gauge transformation).

The new MPO $\hat{W}^{(n')}$ is upper triangular. To see this, take $i \geq j$. Then either $\hat{W}_{ij}^{(n')} = W_{ab}^{(n)}$ for $O_i = (a, n)$ and $O_j = (b, n+1)$, or $\hat{W}_{ij}^{(n')} = 0$. But $O_i = (a, n) \geq (b, n+1) = O_j$, so $W_{ab}^{(n)} = 0$ regardless. Therefore $\hat{W}^{(n')}$ is upper triangular. ■

We present the algorithm for (left) canonicalization of loop free UCMPOs in Algorithm 2. We now prove its correctness, then analyze its cost.

Proposition 1. Suppose $\{\hat{W}^{(n)}\}_{n=1}^N$ is loop free. The output of Algorithm 2 is a left canonical UCMPO.

Proof. The main idea is to iterate over the columns of $\{\hat{W}^{(n)}\}$ in topological order, orthogonalizing each column against all the previous ones as in Gram-Schmidt.

Let O be a topological order for the nodes as in (D29). As each $\hat{W}^{(n)}$ has the form (D7), the first column of each is already orthonormal. We proceed by induction. Suppose that we have orthogonalized columns up to $(d, n+1) \in O$. Then for $(a, m), (b, m) < (d, n+1)$,

$$\sum_c \langle \hat{W}_{ca}^{(m)}, \hat{W}_{cb}^{(m)} \rangle = \delta_{ab}. \quad (\text{D32})$$

Let the *predecessor* nodes be $P := \{(a, n+1) : (a, n+1) < (d, n+1)\}$ and for each (a, n) , define the inner products with all previous columns as

$$r_a := \sum_c \langle \hat{W}_{ca}^{(n)}, \hat{W}_{cd}^{(n)} \rangle, \quad (\text{D33a})$$

$$R := I_{D^{(n+1)}} - \sum_{a \in P} r_a e_{ad}, \quad (\text{D33b})$$

where $\mathbf{e}_{ad}$ is the elementary matrix where entry ad is 1 and the rest are zero: $(\mathbf{e}_{ad})_{ij} = \delta_{ai}\delta_{dj}$. Here R is only nonidentity in column d , and it performs elementary column operations when acting to the left and elementary row operations acting to the right. In particular, we have chosen it to perform one Gram-Schmidt step, orthogonalizing column d against previous columns of $\widehat{W}^{(n)}$. It is easy to invert R , $R^{-1} = I_{D^{(n)}} + \sum_{a \in P} r_a \mathbf{e}_{ad}$, so we can cast this Gram-Schmidt step as a gauge transform with a single nonidentity gauge matrix:

$$\widehat{W}^{(n)'} := \widehat{W}^{(n)} R, \quad (\text{D34a})$$

$$\widehat{W}^{(n+1)'} := R^{-1} \widehat{W}^{(n+1)}. \quad (\text{D34b})$$

We then have two things to show: (i) that this gauge transform really does orthogonalize column d of $\widehat{W}^{(n)}$ against previous columns and (ii) that the gauge transform does not ruin the orthogonality condition of Eq. (D32). Both are easy computations.

For (i), the effect of R acting on $\widehat{W}^{(n)}$ on the right is to add column c to column d with coefficient r_c :

$$\widehat{W}^{(n)} R = \widehat{W}^{(n)} - \sum_{c \in P} r_c \widehat{W}^{(n)} \mathbf{e}_{cd}.$$

The matrix $\widehat{W}^{(n)} \mathbf{e}_{cd}$ is the matrix with only column d nonzero and whose values are those from column c of $\widehat{W}^{(n)}$. Let $\widehat{W}_{:,a}$ denote the a th column vector of $\widehat{W}$, as usual. Then

$$\langle \widehat{W}_{:,a}^{(n)}, [\widehat{W}^{(n)} \mathbf{e}_{cd}]_{:,d} \rangle = \langle \widehat{W}_{:,a}^{(n)}, \widehat{W}_{:,c}^{(n)} \rangle = \delta_{ac}$$

by (D32). Therefore, for any $(a, n) \prec (d, n)$,

$$\begin{aligned} \langle \widehat{W}_{:,a}^{(n)'}, \widehat{W}_{:,d}^{(n)'} \rangle &= \langle \widehat{W}_{:,a}^{(n)}, \widehat{W}_{:,d}^{(n)} \rangle - \sum_{c \in P} r_c \langle \widehat{W}_{:,a}^{(n)}, [\widehat{W}^{(n)} \mathbf{e}_{cd}]_{:,d} \rangle \\ &= r_a - \sum_{c \in P} r_c \delta_{ca} = 0. \end{aligned}$$

Therefore column d of $\widehat{W}^{(n)}$ is orthogonal to each previous column.

For (ii), the effect of R^{-1} acting to the left on $\widehat{W}^{(n+1)}$ is to add row d to row c with coefficient r_c :

$$\widehat{W}^{(n+1)'} = \widehat{W}^{(n+1)} + \sum_{c \in P} r_c \sum_e (\widehat{W}_{de}^{(n+1)}) \mathbf{e}_{ce} =: \widehat{W} + \delta \widehat{W}.$$

Take $(a, n+1), (b, n+1) \prec (d, n)$. Then

$$[\delta \widehat{W}]_{:,a} = \sum_{c \in P} r_c (\widehat{W}_{da}^{(n+1)}) \mathbf{e}_{ca} = 0$$

since $\widehat{W}_{da}^{(n+1)} = 0$ as $(d, n) \succeq (a, n+1)$ by (D30), and similarly $[\delta \widehat{W}]_{:,b} = 0$. Therefore,

$$\begin{aligned} \langle \widehat{W}_{:,a}^{(n+1)'}, \widehat{W}_{:,b}^{(n+1)'} \rangle &= \langle \widehat{W}_{:,a} + \delta \widehat{W}_{:,a}, \widehat{W}_{:,b} + \delta \widehat{W}_{:,b} \rangle \\ &= \langle \widehat{W}_{:,a}, \widehat{W}_{:,b} \rangle + 0 = \delta_{ab}, \end{aligned}$$

so the induction hypothesis (D32) holds for $\{\widehat{W}^{(n)'}\}$.

As the gauge transform R adds previous columns to column d of $\widehat{W}^{(n)}$ and adds row d of $\widehat{W}^{(n+1)}$ to previous rows, the transformed UCMPO is also loop free. Thus after this gauge transform, column d of $\widehat{W}^{(n)'}$ is orthogonal to all previous columns and all of the structure of the UCMPO is preserved.

A similar, simpler gauge transform

$$R = I_{D^{(n+1)}} + \left(\sum_c \langle \widehat{W}_{cd}^{(n)'}, \widehat{W}_{cd}^{(n)'} \rangle \right)^{-1/2} \mathbf{e}_{dd}$$

can then be used to normalize column d of $\widehat{W}^{(n)'}$. Repeating the previous arguments, one can show that this similarly does not disrupt the orthogonality of $\widehat{W}^{(n+1)'}$ or the loop free condition. Therefore we have made one more column orthonormal to the previous ones, completing the proof. ■

Algorithm 2 is quite efficient, with cost that scales as $O(\sum_n [D^{(n)}]^3)$. This is somewhat surprising, as it seems we are doing a total of $D = \sum_n D_n$ gauge transformations, each of which is a matrix multiplication. However, the R matrices are particularly simple: They only differ from the identity in a single column. The transformations $\widehat{W}^{(n)'} = \widehat{W}^{(n)} R$ and $\widehat{W}^{(n+1)'} = R \widehat{W}^{(n+1)}$ to orthogonalize a column may be performed with rank-1 matrix updates whose cost is only $O([D^{(n)}]^2)$. Similarly, the gauge transform to normalize a column, which simply scales a row or column, costs only $O(D^{(n)})$. As we must iterate over every column of every tensor, the total cost is then $O(\sum_n [D^{(n)}]^3)$. However, each iteration requires only elementary matrix operations, for which highly optimized libraries are available, and a low constant factor on the algorithm. One can also employ these algorithms with charge-conserving MPOs, which vastly decreases the runtime in practice.

4. Properties of compressed Hamiltonians

We now show that compressed Hamiltonians are accurate approximations to the original Hamiltonian. This will give us guarantees that the (ground state) physics we are interested in is unchanged by compression. In fact, just as with matrix product states, the error is controlled by the weight of the truncated singular values. We demonstrate three properties of the compressed Hamiltonian H' when we truncate a single bond:

$$\widehat{H}_{\text{BLG}} \rightarrow \widehat{H}'(\epsilon) \quad (\text{D35})$$

where $\widehat{H}'$ satisfies the following:

- (1) $\widehat{H}'$ is Hermitian.
- (2) The ground state energy is accurate: $\delta E \leq 4^2 \epsilon$.
- (3) Observables are accurate: $\Delta \langle \widehat{O} \rangle \leq \frac{4^3}{\Delta E} \|\widehat{O}\| \epsilon$.
- (4) The ground state wave function is accurate: $|1 - \mathcal{F}| \leq \frac{4^4}{\Delta E^2} \epsilon^2$.

We reiterate that these are *local* bounds, corresponding to truncating a single bond. It is reasonable to expect that these results can be generalized to *global* bounds which apply when all bonds are truncated simultaneously, just as they can for matrix product states. However, such generalizations are often highly technical and therefore beyond the scope of this work. As a practical matter, Fig. 5 demonstrates that the global errors in the ground state energy, fidelity, and expectation values are small and decrease as $\epsilon(D) \rightarrow 0$.

a. Compressed Hamiltonians are Hermitian

All Hamiltonians in quantum mechanics are Hermitian. We now show that Hermiticity is preserved by dropping singular

values of a Hamiltonian. There is just one caveat: If the spectrum contains a set of degenerate singular values, then one must drop either all of them or none of them:

Proposition 2. Let $\hat{H}$ be a Hermitian operator with the following almost-Schmidt form:

$$\hat{H} = \hat{H}_L \hat{1}_R + \hat{1}_L \hat{H}_R + \sum_{a=1}^{N_s} \sum_{i=1}^{D_a} \hat{O}_L^{a,i} s_a \hat{O}_R^{a,i}, \quad (\text{D36})$$

where a labels degenerate singular values, N_s is the number of distinct singular values, and D_a is the degeneracy of the a th Schmidt value. Then, for any subset $\mathcal{A} \subset \{1, 2, \dots, N_s\}$, the compressed operator

$$\hat{H}' = \hat{H}_L \hat{1}_R + \hat{1}_L \hat{H}_R + \sum_{a \in \mathcal{A}} \sum_{i=1}^{D_a} \hat{O}_L^{a,i} s_a \hat{O}_R^{a,i} \quad (\text{D37})$$

is Hermitian.

For concision, we sketch the proof. Due to the orthonormality condition, $\hat{H}_L \hat{1}_R$, $\hat{1}_L \hat{H}_R$, and $\sum_{a=1}^{N_s} \sum_i^{D_a} \hat{O}_L^{a,i} s_a \hat{O}_R^{a,i}$ must be independently Hermitian. This implies $\sum_{a=1}^{N_s} \sum_i^{D_a} \hat{O}_L^{a,i} s_a \hat{O}_R^{a,i} = \sum_{a=1}^{N_s} \sum_i^{D_a} (\hat{O}_L^{a,i})^\dagger s_a (\hat{O}_R^{a,i})^\dagger$. Note that both the LHS and the RHS of this equation can be regarded as a singular value decomposition in operator space. It follows from the uniqueness of singular value decomposition that $\sum_i^{D_a} \hat{O}_L^{a,i} s_a \hat{O}_R^{a,i} = \sum_i^{D_a} (\hat{O}_L^{a,i})^\dagger s_a (\hat{O}_R^{a,i})^\dagger$ for any a . The proposition follows. In practice, this means that one should always drop singular values by imposing a minimum value to retain rather than a maximum number.

We note that this result readily generalizes to a case when we truncate all bonds at the same time. To see this, we note that the proof above shows the action of Hermitian conjugation commutes with the singular value matrix. Then, just like in the case of symmetric MPS, singular values can be dropped without ruining Hermiticity.

b. Compressed Hamiltonians are accurate

The accuracy of a compressed Hamiltonian is controlled by the weight of the truncated singular values in almost-Schmidt form, Eq. (D5). Conceptually, one should think of truncation as introducing a small perturbation to the Hamiltonian. If the truncated weight is small, then the perturbation is small, and its effects to the ground state energy, the fidelity, and other observables are also small.

To quantify these effects, we employ the sup norm, an operator norm well suited for ground state properties. If $\hat{H}$ is an operator, then its sup norm $\|\hat{H}\|$ is given by

$$\|\hat{H}\|^2 := \sup_{|\psi\rangle} \frac{\langle \psi | \hat{H} \hat{H} | \psi \rangle}{\langle \psi | \psi \rangle}. \quad (\text{D38})$$

As the sup norm is extensive, we work with the sup norm per unit cell, so that it is finite. We first quote a result from Ref. [50]: The change in the ground state energy is small under truncation.

Proposition 3 (Proposition 5 of Ref. [50]). Suppose $\hat{H}$ is a k -body Hamiltonian with onsite dimension d .²³ If $\hat{H}$ is com-

pressed from bond dimension D to $\hat{H}'$ with D' with truncated weight

$$\epsilon^2 := \sum_{a=D+1}^{D'} s_a^2, \quad (\text{D39})$$

then the change in the ground state energy is bounded by

$$\delta E \leq \|\hat{H} - \hat{H}'\| \leq d^{\frac{k}{2}} \epsilon. \quad (\text{D40})$$

In practice, the singular values for a Hamiltonian fall off quite quickly—often exponentially, or as a power law at worse. So retaining only a small number of singular values can produce a highly accurate approximation for the ground state energy.

It is natural to assume that if the ground state energy is accurate, then the other ground state properties—such as expectation values of observables and even the entire ground state wave function—are accurate as well. Unfortunately, there is a rare but severe failure of this assumption. Near a first-order phase transition, a tiny perturbation to a Hamiltonian can push the system across the phase transition, changing the properties of the ground state in a discontinuous manner (except for the energy). However, as long as the competing states have large energy difference away from the transition, this will only cause infinitesimal shift of critical parameters. We can therefore understand the generic case by simply assuming we are far from a phase transition and the ground state changes continuously.

To do this, we work in first order perturbation theory. Suppose $\hat{H}$ is a k -body Hamiltonian with a unique ground state with gap ΔE . Suppose we write $\hat{H} = \hat{H}' + \delta \hat{H}$ with truncated weight ϵ^2 as in (D39) and consider an observable of interest $\hat{O}$. Then we can write the new ground state as

$$|E_0(\delta')\rangle = |E_0\rangle + |\delta E_0\rangle + O(\epsilon^2),$$

$$|\delta E_0\rangle = \sum_{\lambda \neq 0} \frac{\langle E_\lambda | \delta \hat{H} | E_0 \rangle}{E_\lambda - E_0} |E_\lambda\rangle.$$

Then

$$\begin{aligned} \Delta O &:= |\langle E_0(\delta') | \hat{O} | E_0(\delta') \rangle - \langle E_0 | \hat{O} | E_0 \rangle| \\ &= 2|\text{Re} \langle E_0 | \hat{O} | \delta E_0 \rangle| + O(\epsilon^2), \end{aligned}$$

so

$$\begin{aligned} |\langle E_0 | \hat{O} | \delta E_0 \rangle| &\leq \left| \sum_{\lambda \neq 0} \frac{\langle E_0 | \delta \hat{H} | E_\lambda \rangle \langle E_\lambda | \hat{O} | E_0 \rangle}{E_\lambda - E_0} \right| \\ &\leq \frac{1}{\Delta E} \left| \sum_{\lambda \neq 0} \langle E_0 | \hat{O} | E_\lambda \rangle \langle E_\lambda | \delta \hat{H} | E_0 \rangle \right| \\ &\leq \frac{1}{\Delta E} |\langle E_0 | \hat{O} \delta \hat{H} | E_0 \rangle - \langle E_0 | \hat{O} | E_0 \rangle \langle E_0 | \delta \hat{H} | E_0 \rangle| \\ &\leq \frac{2}{\Delta E} \|\hat{O}\| \cdot \|\delta \hat{H}\|, \end{aligned}$$

where we have used $\sum_{\lambda \neq 0} |E_\lambda\rangle \langle E_\lambda| = I - |E_0\rangle \langle E_0|$ and submultiplicativity of the norm. Using (D40), the change in the

²³E.g., $d = 4$ for spin- $\frac{1}{2}$'s or spinless fermions.

TABLE IV. Energy of $|\psi_{\text{kin}}\rangle$ per momentum per band at $N_y = 2$, $w_0/w_1 = 0.825$, $d = 30$ nm. MPO compression was performed with singular value truncation cutoff at 10^{-3} meV. The energy difference is calculated against the k space result.

	k space	xk space	MPO
E_{kin} (meV)	-69.099	-69.095	-69.095
E_{int} (meV)	32.263	32.257	32.257
ΔE_{kin} (meV)		4.2×10^{-3}	4.2×10^{-3}
ΔE_{int} (meV)		7.0×10^{-3}	6.4×10^{-3}

expectation value is bounded by

$$\Delta O \leq \frac{4d^{\frac{k}{2}}}{\Delta E} \|\hat{O}\| \epsilon. \quad (\text{D41})$$

We may therefore conclude that the error in expectation values should be small, provided that the uncompressed Hamiltonian is sufficiently far from a first-order phase transition. The condition of a gapped ground state may be relaxed, in which case the error will be controlled by the matrix elements of $\hat{O}$ between the ground state and low-lying excited states.

c. Compressed Hamiltonians have high fidelity

We have now seen that the ground state energy and expectation values of observables are accurately captured by the approximate, compressed Hamiltonian. In fact, the entire ground state wave function $|E'_0\rangle$ of $\hat{H}'$ is very close to the original ground state wave function $|E_0\rangle$ of $\hat{H}$. This allows us to use structural properties of $|\psi'\rangle$, such as its correlation length as a function of MPS bond dimension, as an accurate stand-in for the true ones and use them to, e.g., diagnose the scaling properties of phase transitions.

To see this, we again work in perturbation theory, this time to second order. Let $\hat{H} = \hat{H}' + \delta\hat{H}$ and take the same assumptions as above. Then we write

$$|E_0(\delta)'\rangle = |E_0\rangle + |\delta E_0\rangle + |\delta^2 E_0\rangle + O(\epsilon^3),$$

so

$$\langle E_0 | E_0(\delta)' \rangle = 1 + 0 - \frac{1}{2} \sum_{\lambda \neq 0} \frac{\langle E_0 | \delta \hat{H} | E_\lambda \rangle \langle E_\lambda | \delta \hat{H} | E_0 \rangle}{(E_\lambda - E_0)^2}.$$

By the same argument as above the error in the ground state fidelity is bounded as

$$|1 - \langle \psi' | \psi \rangle| \leq \frac{d^k}{\Delta E^2} \epsilon^2. \quad (\text{D42})$$

In conclusion, we have now seen that the compressed MPOs should accurately reproduce the true ground state physics and provided error bounds on the precision. This justifies our use of compressed Hamiltonians to study twisted bilayer graphene.

APPENDIX E: NUMERICAL CROSS CHECKS

In addition to the analytic error bounds from the previous section, we also performed extensive numerical checks to verify that our computations were correct in practice as well as in principle. As mentioned above, we used the standard TeNPy

library [63], written by one of us, for all DMRG calculations. The MPO compression code was carefully verified by unit testing, benchmarking, and a variety of cross-checks. The two primary cross-checks, which we now describe, verify the accuracy of the compression algorithm and the accuracy of the transformations between the various representations of the Hamiltonian.

1. Gauge transform verification

It is crucial that the compression algorithm is not only precise, as we have shown in previous sections, but also accurate. That is, the output of the implementation of the compression algorithm is indeed the compressed MPO described analytically. To verify this, we use the fact that Algorithm 1 is a gauge transformation, up until the truncation step. This gauge transformation obeys gauge relations given in Sec. D 3, and reproduced here for convenience:

$$\begin{aligned} R_{n-1} \hat{W}_R^n &= \hat{W}^{(n)} R_n, \\ C_{n-1} \hat{W}_R^{(n)} &= \hat{W}_L^{(n)} C_n, \\ S_{n-1} \hat{P}^{(n)} &= \hat{Q}^{(n)} S_n. \end{aligned} \quad (\text{E1})$$

Due to the large number of small matrix elements, many indexing errors and other accuracy problems only manifest as small errors in the gauge relations. We therefore verified the gauge relations to precision 10^{-13} , nearly the floating point limit. Together with checks for canonicity of $\hat{W}_{R,L}$, this constitutes a sufficient check for the correctness of the algorithmic implementation.

2. Cross checks for tBLG Hamiltonian

In order to perform HF and DMRG calculations, one needs to represent the Hamiltonian in a variety of ways, as shown in Fig. 1, and it is imperative to make sure there is no error when we transform one representation to another. This section reviews a list of numerical checks we performed to guarantee correctness.

We first start from the momentum space representation, which is suitable for HF calculations. This interaction is specified by V_q , together with the form factors Λ_q . This representation can be Wannier localized (i.e., Fourier transformed) to obtain the mixed- xk space representation for iDMRG. Finally we use the mixed- xk space representation to construct an MPO, and compress it down to a smaller bond dimension. The first transformation is a unitary transformation, and is in principle exact up to numerical precision, whereas the precision of the compression is limited by MPO singular value cutoff.

In order to check if these transformations are accurate, we calculate a physical observable using each representation. The kinetic part H_{kin} of the BM Hamiltonian $h(k)$ is gapped for a wide range of parameters for small N_y , and the ground state $|\psi_{\text{kin}}\rangle$ is easy to calculate in each representation. Therefore, we can easily evaluate the following energies in momentum space, mixed- xk space, and DMRG.

$$\begin{aligned} E_{\text{kin}} &= \langle \psi_{\text{kin}} | H_{\text{kin}} | \psi_{\text{kin}} \rangle \\ E_{\text{int}} &= \langle \psi_{\text{kin}} | H_{\text{int}} | \psi_{\text{kin}} \rangle, \end{aligned} \quad (\text{E2})$$

where H_{kin} and H_{int} are the kinetic and interaction part of the Hamiltonian, respectively. The comparison for $N_y = 2$ is shown in Table IV. We see that the energy error is very small. We further check that these errors decrease as the accuracy

of each calculation is increased (e.g., by increasing the cutoff range for MPO creation). We may conclude that the transformations between Hamiltonian representations are sufficiently accurate.

-
- [1] Y. Cao, V. Fatemi, A. Demir, S. Fang, S. L. Tomarken, J. Y. Luo, J. D. Sanchez-Yamagishi, K. Watanabe, T. Taniguchi, E. Kaxiras, R. C. Ashoori, and P. Jarillo-Herrero, Correlated insulator behavior at half-filling in magic-angle graphene superlattices, *Nature (London)* **556**, 80 (2018).
 - [2] Y. Cao, V. Fatemi, S. Fang, K. Watanabe, T. Taniguchi, E. Kaxiras, and P. Jarillo-Herrero, Unconventional superconductivity in magic-angle graphene superlattices, *Nature (London)* **556**, 43 (2018).
 - [3] M. Yankowitz, S. Chen, H. Polshyn, Y. Zhang, K. Watanabe, T. Taniguchi, D. Graf, A. F. Young, and C. R. Dean, Tuning superconductivity in twisted bilayer graphene, *Science* **363**, 1059 (2019).
 - [4] A. Kerelsky, L. J. McGilly, D. M. Kennes, L. Xian, M. Yankowitz, S. Chen, K. Watanabe, T. Taniguchi, J. Hone, C. Dean, A. Rubio, and A. N. Pasupathy, Maximized electron interactions at the magic angle in twisted bilayer graphene, *Nature (London)* **572**, 95 (2019).
 - [5] Y. Jiang, X. Lai, K. Watanabe, T. Taniguchi, K. Haule, J. Mao, and E. Y. Andrei, Charge order and broken rotational symmetry in magic-angle twisted bilayer graphene, *Nature (London)* **573**, 91 (2019).
 - [6] X. Lu, P. Stepanov, W. Yang, M. Xie, M. A. Aamir, I. Das, C. Urgell, K. Watanabe, T. Taniguchi, G. Zhang *et al.*, Superconductors, orbital magnets and correlated states in magic-angle bilayer graphene, *Nature (London)* **574**, 653 (2019).
 - [7] Y. Xie, B. Lian, B. Jäck, X. Liu, C.-L. Chiu, K. Watanabe, T. Taniguchi, B. A. Bernevig, and A. Yazdani, Spectroscopic signatures of many-body correlations in magic-angle twisted bilayer graphene, *Nature (London)* **572**, 101 (2019).
 - [8] P. Stepanov, I. Das, X. Lu, A. Fahimniya, K. Watanabe, T. Taniguchi, F. H. L. Koppens, J. Lischner, L. Levitov, and D. K. Efetov, Untying the insulating and superconducting orders in magic-angle graphene, *Nature* **583**, 375 (2020).
 - [9] Y. Saito, J. Ge, K. Watanabe, T. Taniguchi, and A. F. Young, Independent superconductors and correlated insulators in twisted bilayer graphene, *Nat. Phys.* **16**, 926 (2020).
 - [10] Y. Choi, J. Kemmer, Y. Peng, A. Thomson, H. Arora, R. Polski, Y. Zhang, H. Ren, J. Alicea, G. Refael, F. von Oppen, K. Watanabe, T. Taniguchi, and S. Nadj-Perge, Electronic correlations in twisted bilayer graphene near the magic angle, *Nat. Phys.* **15**, 1174 (2019).
 - [11] H. Yoo, R. Engelke, S. Carr, S. Fang, K. Zhang, P. Cazeaux, S. H. Sung, R. Hovden, A. W. Tsen, T. Taniguchi, K. Watanabe, G.-C. Yi, M. Kim, M. Lusk, E. B. Tadmor, E. Kaxiras, and P. Kim, Atomic and electronic reconstruction at the van der Waals interface in twisted bilayer graphene, *Nat. Mater.* **18**, 448 (2019).
 - [12] A. L. Sharpe, E. J. Fox, A. W. Barnard, J. Finney, K. Watanabe, T. Taniguchi, M. A. Kastner, and D. Goldhaber-Gordon, Emergent ferromagnetism near three-quarters filling in twisted bilayer graphene, *Science* **365**, 605 (2019).
 - [13] M. Serlin, C. L. Tschirhart, H. Polshyn, Y. Zhang, J. Zhu, K. Watanabe, T. Taniguchi, L. Balents, and A. F. Young, Intrinsic quantized anomalous hall effect in a moiré heterostructure, *Science* **367**, 900 (2020).
 - [14] S. L. Tomarken, Y. Cao, A. Demir, K. Watanabe, T. Taniguchi, P. Jarillo-Herrero, and R. C. Ashoori, Electronic Compressibility Of Magic-Angle Graphene Superlattices, *Phys. Rev. Lett.* **123**, 046601 (2019).
 - [15] D. Wong, K. P. Nuckolls, M. Oh, B. Lian, Y. Xie, S. Jeon, K. Watanabe, T. Taniguchi, B. Andrei Bernevig, and A. Yazdani, Cascade of electronic transitions in magic-angle twisted bilayer graphene, *Nature (London)* **582**, 198 (2020).
 - [16] U. Zondiner, A. Rozen, D. Rodan-Legrain, Y. Cao, R. Queiroz, T. Taniguchi, K. Watanabe, Y. Oreg, F. von Oppen, A. Stern, E. Berg, P. Jarillo-Herrero, and S. Ilani, Cascade of phase transitions and Dirac revivals in magic-angle graphene, *Nature (London)* **582**, 203 (2020).
 - [17] H. S. Arora, R. Polski, Y. Zhang, A. Thomson, Y. Choi, H. Kim, Z. Lin, I. Z. Wilson, X. Xu, J.-H. Chu, K. Watanabe, T. Taniguchi, J. Alicea, and S. Nadj-Perge, Superconductivity in metallic twisted bilayer graphene stabilized by WSe₂, *Nature (London)* **583**, 379 (2020).
 - [18] K. P. Nuckolls, M. Oh, D. Wong, B. Lian, K. Watanabe, T. Taniguchi, B. A. Bernevig, and A. Yazdani, Strongly correlated Chern insulators in magic-angle twisted bilayer graphene, *arXiv:2007.03810*.
 - [19] S. Wu, Z. Zhang, K. Watanabe, T. Taniguchi, and E. Y. Andrei, Chern insulators and topological flat-bands in magic-angle twisted bilayer graphene, *arXiv:2007.03735*.
 - [20] C. L. Tschirhart, M. Serlin, H. Polshyn, A. Shragai, Z. Xia, J. Zhu, Y. Zhang, K. Watanabe, T. Taniguchi, M. E. Huber, and A. F. Young, Imaging orbital ferromagnetism in a moiré Chern insulator, *arXiv:2006.08053*.
 - [21] X. Lu, B. Lian, G. Chaudhary, B. A. Piot, G. Romagnoli, K. Watanabe, T. Taniguchi, M. Poggio, A. H. MacDonald, B. A. Bernevig, and D. K. Efetov, Fingerprints of fragile topology in the Hofstadter spectrum of twisted bilayer graphene close to the second magic angle, *arXiv:2006.13963*.
 - [22] X. Liu, Z. Wang, K. Watanabe, T. Taniguchi, O. Vafek, and J. I. A. Li, Tuning electron correlation in magic-angle twisted bilayer graphene using Coulomb screening, *arXiv:2003.11072*.
 - [23] Y. Cao, D. Rodan-Legrain, J. M. Park, F. Noah Yuan, K. Watanabe, T. Taniguchi, R. M. Fernandes, L. Fu, and P. Jarillo-Herrero, Nematicity and competing orders in superconducting magic-angle graphene, *arXiv:2004.04148*.
 - [24] S. R. White, Density Matrix Formulation For Quantum Renormalization Groups, *Phys. Rev. Lett.* **69**, 2863 (1992).
 - [25] J. Kang and O. Vafek, Non-abelian Dirac node braiding and near-degeneracy of correlated phases at odd integer filling in magic-angle twisted bilayer graphene, *Phys. Rev. B* **102**, 035161 (2020).

- [26] R. Bistritzer and A. H. MacDonald, Moiré bands in twisted double-layer graphene, *Proc. Natl. Acad. Sci.* **108**, 12233 (2011).
- [27] N. N. T. Nam and M. Koshino, Lattice relaxation and energy band modulation in twisted bilayer graphene, *Phys. Rev. B* **96**, 075311 (2017).
- [28] S. Carr, S. Fang, Z. Zhu, and E. Kaxiras, Exact continuum model for low-energy electronic states of twisted bilayer graphene, *Phys. Rev. Research* **1**, 013001 (2019).
- [29] Z. Song, Z. Wang, W. Shi, G. Li, C. Fang, and B. A. Bernevig, All Magic Angles In Twisted Bilayer Graphene Are Topological, *Phys. Rev. Lett.* **123**, 036401 (2019).
- [30] N. Bultinck, E. Khalaf, S. Liu, S. Chatterjee, A. Vishwanath, and M. P. Zaletel, Ground State And Hidden Symmetry Of Magic-Angle Graphene At Even Integer Filling, *Phys. Rev. X* **10**, 031034 (2020).
- [31] M. Xie and A. H. MacDonald, Nature Of The Correlated Insulator States In Twisted Bilayer Graphene, *Phys. Rev. Lett.* **124**, 097601 (2020).
- [32] C. Repellin, Z. Dong, Y.-H. Zhang, and T. Senthil, Ferromagnetism In Narrow Bands Of Moiré Superlattices, *Phys. Rev. Lett.* **124**, 187601 (2020).
- [33] H. C. Po, L. Zou, A. Vishwanath, and T. Senthil, Origin Of Mott Insulating Behavior And Superconductivity In Twisted Bilayer Graphene, *Phys. Rev. X* **8**, 031089 (2018).
- [34] H. C. Po, L. Zou, T. Senthil, and A. Vishwanath, Faithful tight-binding models and fragile topology of magic-angle bilayer graphene, *Phys. Rev. B* **99**, 195455 (2019).
- [35] L. Zou, H. C. Po, A. Vishwanath, and T. Senthil, Band structure of twisted bilayer graphene: Emergent symmetries, commensurate approximants, and wannier obstructions, *Phys. Rev. B* **98**, 085435 (2018).
- [36] K. Hejazi, C. Liu, H. Shapourian, X. Chen, and L. Balents, Multiple topological transitions in twisted bilayer graphene near the first magic angle, *Phys. Rev. B* **99**, 035111 (2019).
- [37] J. Liu, J. Liu, and X. Dai, Pseudo landau level representation of twisted bilayer graphene: Band topology and implications on the correlated insulating phase, *Phys. Rev. B* **99**, 155415 (2019).
- [38] J. Ahn, S. Park, and B.-J. Yang, Failure Of Nielsen-Ninomiya Theorem And Fragile Topology In Two-Dimensional Systems With Space-Time Inversion Symmetry: Application To Twisted Bilayer Graphene At Magic Angle, *Phys. Rev. X* **9**, 021013 (2019).
- [39] A. A. Soluyanov and D. Vanderbilt, Wannier representation of $\mathbb{Z}_2$ topological insulators, *Phys. Rev. B* **83**, 035108 (2011).
- [40] J. Kang and O. Vafek, Symmetry, Maximally Localized Wannier States, And A Low-Energy Model For Twisted Bilayer Graphene Narrow Bands, *Phys. Rev. X* **8**, 031088 (2018).
- [41] M. Koshino, N. F. Q. Yuan, T. Koretsune, M. Ochi, K. Kuroki, and L. Fu, Maximally Localized Wannier Orbitals And The Extended Hubbard Model For Twisted Bilayer Graphene, *Phys. Rev. X* **8**, 031087 (2018).
- [42] J. Kang and O. Vafek, Strong Coupling Phases Of Partially Filled Twisted Bilayer Graphene Narrow Bands, *Phys. Rev. Lett.* **122**, 246401 (2019).
- [43] X. Wang and O. Vafek, Diagnosis of explicit symmetry breaking in the tight-binding constructions for symmetry-protected topological systems, *Phys. Rev. B* **102**, 075142 (2020).
- [44] N. Bultinck, S. Chatterjee, and M. P. Zaletel, A Mechanism For Anomalous Hall Ferromagnetism In Twisted Bilayer Graphene, *Phys. Rev. Lett.* **124**, 166601 (2020).
- [45] K. Hejazi, X. Chen, and L. Balents, Hybrid Wannier Chern bands in magic angle twisted bilayer graphene and the quantized anomalous Hall effect, *arXiv:2007.00134*.
- [46] Y. H. Kwan, G. Wagner, N. Chakraborty, S. H. Simon, and S. A. Parameswaran, Orbital Chern insulator domain walls and chiral modes in twisted bilayer graphene, *arXiv:2007.07903*.
- [47] J. Motruk, M. P. Zaletel, R. S. K. Mong, and F. Pollmann, Density matrix renormalization group on a cylinder in mixed real and momentum space, *Phys. Rev. B* **93**, 155139 (2016).
- [48] M. P. Zaletel, R. S. K. Mong, F. Pollmann, and E. H. Rezayi, Infinite density matrix renormalization group for multicomponent quantum hall systems, *Phys. Rev. B* **91**, 045115 (2015).
- [49] G. K.-L. Chan, A. Keselman, N. Nakatani, Z. Li, and S. R. White, Matrix product operators, matrix product states, and ab initio density matrix renormalization group algorithms, *J. Chem. Phys.* **145**, 014102 (2016).
- [50] D. E. Parker, X. Cao, and M. P. Zaletel, Local matrix product operators: Canonical form, compression, and control theory, *Phys. Rev. B* **102**, 035147 (2020).
- [51] S. Liu, E. Khalaf, J. Y. Lee, and A. Vishwanath, Nematic topological semimetal and insulator in magic angle bilayer graphene at charge neutrality, *arXiv:1905.07409*.
- [52] T. Cea and F. Guinea, Band structure and insulating states driven by Coulomb interaction in twisted bilayer graphene, *Phys. Rev. B* **102**, 045107 (2020).
- [53] Y. H. Kwan, Y. Hu, S. H. Simon, and S. A. Parameswaran, Exciton band topology in spontaneous quantum anomalous Hall insulators: Applications to twisted bilayer graphene, *arXiv:2003.11560*.
- [54] X.-L. Qi, Generic Wave-Function Description Of Fractional Quantum Anomalous Hall States And Fractional Topological Insulators, *Phys. Rev. Lett.* **107**, 126803 (2011).
- [55] G. Ehlers, S. R. White, and R. M. Noack, Hybrid-space density matrix renormalization group study of the doped two-dimensional hubbard model, *Phys. Rev. B* **95**, 125125 (2017).
- [56] N. Marzari and D. Vanderbilt, Maximally localized generalized Wannier functions for composite energy bands, *Phys. Rev. B* **56**, 12847 (1997).
- [57] R. D. King-Smith and D. Vanderbilt, Theory of polarization of crystalline solids, *Phys. Rev. B* **47**, 1651 (1993).
- [58] R. Resta, Macroscopic electric polarization as a geometric quantum phase, *Europhys. Lett. (EPL)* **22**, 133 (1993).
- [59] D. Vanderbilt and R. D. King-Smith, Electric polarization as a bulk quantity and its relation to surface charge, *Phys. Rev. B* **48**, 4442 (1993).
- [60] N. Marzari, A. A. Mostofi, J. R. Yates, I. Souza, and D. Vanderbilt, Maximally localized Wannier functions: Theory and applications, *Rev. Mod. Phys.* **84**, 1419 (2012).
- [61] B. Pirvu, V. Murg, J. I. Cirac, and F. Verstraete, Matrix product operator representations, *New J. Phys.* **12**, 025012 (2010).
- [62] M. Fishman, S. R. White, and E. M. Stoudenmire, The ITensor Software Library for Tensor Network Calculations, *arXiv:2007.14822*.
- [63] J. Hauschild and F. Pollmann, Efficient numerical simulations with tensor networks: Tensor network python (TeNPy), *SciPost Phys. Lect. Notes* **5** (2018).

- [64] Y.-H. Zhang, D. Mao, and T. Senthil, Twisted bilayer graphene aligned with hexagonal boron nitride: Anomalous hall effect and a lattice model, [Phys. Rev. Research **1**, 033126 \(2019\)](#).
- [65] Y. Kim, B. J. Wieder, C. L. Kane, and A. M. Rappe, Dirac Line Nodes In Inversion-Symmetric Crystals, [Phys. Rev. Lett. **115**, 036806 \(2015\)](#).
- [66] G. Sundaram and Q. Niu, Wave-packet dynamics in slowly perturbed crystals: Gradient corrections and berry-phase effects, [Phys. Rev. B **59**, 14915 \(1999\)](#).
- [67] E. H. Hwang and S. D. Sarma, Dielectric function, screening, and plasmons in two-dimensional graphene, [Phys. Rev. B **75**, 205418 \(2007\)](#).
- [68] S. Liang and H. Pang, Approximate diagonalization using the density matrix renormalization-group method: A two-dimensional-systems perspective, [Phys. Rev. B **49**, 9214 \(1994\)](#).
- [69] U. Schollwöck, The density-matrix renormalization group in the age of matrix product states, [Ann. Phys. **326**, 96 \(2011\)](#).
- [70] G. M. Crosswhite and D. Bacon, Finite automata for caching in matrix product algorithms, [Phys. Rev. A **78**, 012356 \(2008\)](#).