Robustness of Majorana edge modes and topological order: Exact results for the symmetric interacting Kitaev chain with disorder

Max McGinley, Johannes Knolle, and Andreas Nunnenkamp
Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, United Kingdom
(Received 30 June 2017; revised manuscript received 8 November 2017; published 27 December 2017)

We investigate the robustness of Majorana edge modes under disorder and interactions. We exploit a recently found mapping of the interacting Kitaev chain in the symmetric region ($\mu = 0, t = \Delta$) to free fermions. Extending the exact solution to the disordered case allows us to calculate analytically the topological phase boundary for all interaction and disorder strengths, which has been thought to be only accessible numerically. We discover a regime in which moderate disorder in the interaction matrix elements enhances topological order well into the strongly interacting regime $U > t$. We also derive the explicit form of the many-body Majorana edge wave function, revealing how it is dressed by many-particle fluctuations from interactions. The qualitative features of our analytical results are valid beyond the fine-tuned integrable point, as expected from the robustness of topological order and as corroborated here by an exact diagonalization study of small systems.

DOI: 10.1103/PhysRevB.96.241113

Majorana edge modes in condensed matter physics have recently received a great deal of attention [1] primarily due to their applications in topological quantum computation [2]. In a seminal paper [3], Kitaev introduced the minimal model of a one-dimensional (1D) $p$-wave superconducting wire, now known as the Kitaev chain. One of its remarkable properties is the presence of zero-energy states localized at the two ends of the chain. Paired together, these Majorana edge modes can form a qubit which is largely protected from decoherence due to its nonlocal nature.

Compelling experimental evidence of their existence has been reported in semiconducting nanowires in proximity to $s$-wave superconductors [4–7] and in ferromagnetic atomic chains [8–10]. However, there remains a possibility that the zero-bias conductance peak measured in these experiments is due to disorder rather than to Majorana modes [11–13], and so it is important to include disorder in theoretical investigations. Additionally, the nature of these experimental platforms inevitably leads to the presence of interactions between the low-energy degrees of freedom [14–16].

The majority of analytical studies of the Kitaev chain have focused on the clean, noninteracting case [1]. Beyond this, for clean, interacting chains, only few exact results are known [17–20], and numerical/perturbative studies have shown that Majorana edge modes can be stable up to moderate interaction strengths [14,15,21–23]. Similarly, a number of works on noninteracting, disordered/quasiperiodic chains find a relatively broad parameter region of stability [24–26]. The combined effect of interactions and disorder in Kitaev chains has recently been studied numerically [27,28], as well as through a weak-disorder renormalization group (RG) approach [29]. However, an analytic treatment of both strong interactions and strong disorder has been thought to be impossible.

In this Rapid Communication, we investigate analytically the combined effects of disorder and interactions on topological order. Focusing on the example of a Kitaev chain supplemented with disorder and density-density interactions, we exploit a recently found exact solution of Miao et al. [30] to classify its topological order and to construct the Majorana edge modes explicitly. The solution is valid in the symmetric region, which is particle-hole symmetric $\mu = 0$ and has hopping amplitude and superconducting pairing equal on each site $t_i = \Delta_D$. Exact results are obtainable for any configuration of hopping/pairing amplitudes and interaction strengths, allowing us to access the strongly interacting and disordered regions of the phase diagram.

By considering the normalizability of the topological edge modes, we are able to obtain an analytic condition for the ordered phase for arbitrary disorder distributions. Focusing on the specific case of uniform disorder, we calculate the phase diagram as a function of the mean and width of the disorder (Fig. 1). We find that moderate disorder enhances the topological phase well into the strongly disordered and interacting regime $U \gtrsim t$. This suppression of the interaction-driven
with a hopping term \( t_j \), an on-site chemical potential \( \mu_j \), a p-wave superconductor pairing \( \Delta_j \), and a nearest-neighbor interaction \( U_j \). Kitaev's original model had \( U_j = 0 \), so we refer to \( U_j \neq 0 \) as an interacting Kitaev chain.

The topological properties of our system are best studied using a basis of Majorana operators, defined as \( \gamma^A_j = c_j + c_j^\dagger \) and \( \gamma^B_j = -i(c_j - c_j^\dagger) \) which satisfy the anticommutation relation \( \{ \gamma^A_j, \gamma^B_k \} = 2\delta_{jk}\delta^{AB} \). The clean noninteracting model has two gapped phases which differ by the existence of a pair of Majorana modes localized at either edge. Such a mode can be described by a Hermitian operator \( Q \) which commutes with the Hamiltonian \( [H, Q] = 0 \) (up to corrections that decay exponentially with the system size). As an example, if \( \mu_j = U_j = t_j - \Delta_j = 0 \), we get \( Q = \gamma^A_j \) or \( \gamma^B_j \), each of which are absent from the Hamiltonian and so commute exactly. Within this phase, the action of either \( Q \) operator on any eigenstate produces a different eigenstate of the same energy, and so the energy spectrum is doubly degenerate. Importantly, the two states in each pair have opposite fermion number parity, i.e., are eigenstates of the operator

\[
Z^\pm = (-1)^{\sum c_i c_i^\dagger}
\]

(2)

(which itself commutes with \( H \)), with eigenvalues \pm 1, corresponding to even and odd numbers of fermions, respectively. This nonlocal observable acts as a topological order parameter distinguishing between even and odd sectors. If degenerate states have opposite \( Z^\pm \), then \( Q \) must anticommute with \( Z^\pm \). In this case, the phase is topologically ordered, with twofold degenerate ground states in opposite topological sectors.

Topological order is robust against local perturbations, and so when we include interactions in our system, we expect the Majorana edge modes to persist up to some critical interaction strength. While the microscopic nature of the Majorana modes will inevitably be different from the noninteracting case, they should still satisfy the same requirements of being localized at either edge, commuting with the Hamiltonian, and anticommuting with \( Z^\pm \) [33]. However, unlike the noninteracting case, if \( U_j \neq 0 \), then \( Q \) cannot be written as a linear combination of \( \gamma \) operators because the Majorana modes will be dressed by higher-order multiple-particle contributions. In general, the Majorana mode generalizes to a many-body Majorana operator with an expansion [34]

\[
Q = \sum_{c=A,B} \alpha^L_c \gamma^L_j + \sum_{c,c',c''=A,B} \alpha^L_{c,c',c''} \gamma^L_{j,c} \gamma^L_{j,c'} \gamma^L_{j,c''} + \cdots,
\]

(3)

featuring terms with an odd number of Majorana operators. If all the coefficients \( \alpha \) are real, then this describes a Hermitian operator that anticommutes with \( Z^\pm \). If \( \alpha \rightarrow 0 \) as \( j \rightarrow \infty \) for the left mode, and similarly for the right mode).

Constructing explicit expressions for the \( \alpha \) is generally only possible numerically due to the complexity of the many-body problem. However, here for the symmetric chain we derive closed expressions for the coefficients of the many-body Majorana operator (3) for an interacting disordered system.
This allows us to classify the phases as topological if the many-body Majorana operator exists.

**Exact solution.** To achieve this, we make use of an exact solution due to Miao et al. [30]. Using two successive Jordan-Wigner transformations and a spin rotation, they showed that for the clean case in the symmetric region $\mu = 0$ and $t = \Delta$, the nonlocal transformation

\[
\lambda^A_j = \begin{cases} 
\frac{1}{\sqrt{2}} \left( i \gamma^A_j | j \rangle \langle j+1 | + \gamma^B_j | j \rangle \langle j+1 | \right), & \text{odd}, \\
\frac{1}{\sqrt{2}} \left( i \gamma^A_j | j \rangle \langle j+1 | - \gamma^B_j | j \rangle \langle j+1 | \right), & \text{even},
\end{cases}
\]

\[
\lambda^B_j = \begin{cases} 
\frac{1}{\sqrt{2}} \left( i \gamma^B_j | j \rangle \langle j+1 | + \gamma^A_j | j \rangle \langle j+1 | \right), & \text{odd}, \\
\frac{1}{\sqrt{2}} \left( i \gamma^B_j | j \rangle \langle j+1 | - \gamma^A_j | j \rangle \langle j+1 | \right), & \text{even},
\end{cases}
\]

(4)

preserves the Majorana anticommutation relations, so we have $\{\lambda^A_j, \lambda^B_k\} = 2 \delta_{j,k} \delta^{\alpha \beta}$. This allows us to express the Hamiltonian (1) in terms of $\lambda$-fermion bilinears. We note that this also holds for disordered $t_j$ and $U_j$, yielding the Hamiltonian

\[
H = \sum_{j=1}^{N-1} \left[ -it_j \lambda^A_{j+1} \lambda^B_j + iU_j \lambda^A_j \lambda^B_{j+1} \right].
\]

(5)

We depict the Hamiltonian (5) visually in the inset of Fig. 1, using lines to represent fermion bilinears. It is evident that one half of the Majorana operators decouples from the other half, so we can consider two subsystems separately, which we label with roman numerals I and II. We make this explicit by redefining $\phi^A_{1,j} = \lambda^A_{j-1,j}$, $\phi^B_{1,j} = \lambda^B_{j-1,j}$, $\phi^A_{II,j} = \lambda^A_{j,j+1}$, $\phi^B_{II,j} = \lambda^B_{j,j+1}$, so the Hamiltonian is the sum of two uncoupled chains,

\[
H = \sum_{j=1}^{N/2} \left[ -it_{2,j} \phi^A_{I,j+1} \phi^B_{I,j} + iU_{2,j} \phi^A_{I,j} \phi^B_{I,j+1} \right]
\]

\[
+ \sum_{j=1}^{N/2} \left[ it_{2,j} \phi^A_{II,j+1} \phi^B_{II,j} - it_{2,j-1} \phi^A_{II,j} \phi^B_{II,j+1} \right].
\]

(6)

Each of the two subsystems is equivalent to a noninteracting Kitaev chain of length $N/2$ with $t_j = \Delta_j$, one of which has the parameters $\mu_I = 2U_{2,j-1}$ and $t_I = t_{2,j}$, and the other of which has the parameters $\mu_j = -2U_{2,j}$ and $t_I = -t_{2,j-1}$. Mutatis mutandis, from our knowledge of noninteracting Kitaev chains, we can identify quantum phase transitions in the clean case at $U = \pm t \Delta$ at which one subsystem becomes topological and the other becomes trivial in this transformed basis. We note that this decoupling is analogous to the equivalence between the $XY$ model and two independent transverse-field Ising models [35], as each can be related to our system via Jordan-Wigner transformations.

**Topological phase boundary.** Having reduced the original interacting Hamiltonian (1) to a quadratic one (6), we know that any zero-energy boundary mode in the new basis is a linear combination of single fermion operators

\[
Q^A_I = \sum_j \alpha^I_j \phi^A_{I,j}, \quad Q^A_{II} = \sum_j \alpha^I_j \phi^A_{II,j},
\]

with similar expressions for $Q^B_I$ and $Q^B_{II}$. Because the system after the nonlinear transformation is equivalent to a conventional noninteracting Kitaev chain, we can use the standard expression for a noninteracting boundary mode with $t = \Delta$ [33], with the appropriate reassignments of $\mu$ and $t$, giving us $\alpha^A_j \propto (\pm U_j/t_j)\delta_{1,j}$ and $\alpha^A_{II,j} \propto (\pm U_j/t_j)\delta_{1,j}$. Within each phase, only one of the subsystems has a normalizable mode $Q^2 < \infty$.

We can transform these operators back into the original basis using Eq. (4). In the $|U| < t$ phase, subsystem I possesses Majorana modes and the many-body Majorana operator is

\[
Q^A_I = \alpha^I_1 \phi^A_{I,1} + \alpha^I_2 \phi^A_{I,2} (\phi^A_{I,1} \phi^B_{I,2}) + \ldots.
\]

This expression is a generalization of a noninteracting Majorana mode and a special case of Eq. (3) for which the coefficients $\alpha$ can be given explicitly. It is an edge mode in the sense that terms featuring the operators $\gamma^A_j \beta^B_j$ decay exponentially with $j$, and again it is a Majorana operator since it is Hermitian and anticommutes with the fermion parity operator $Z^2$. Additionally, the mode is adiabatically connected to a noninteracting Majorana mode—for $U \to 0$ all multiparticle terms vanish, leaving us with a single $\gamma$ operator.

On the other hand, when $|U| > t$, the edge mode changes to $Q^B_{II}$, which in the original basis is

\[
Q^B_{II} = \alpha^B_1 (\phi^A_{I,1} \phi^B_{I,2}) + \alpha^B_2 (\phi^A_{I,2} \phi^B_{I,1}) (\phi^A_{I,1} \phi^B_{I,2}) + \ldots.
\]

It has a form similar to that of the $Q^A_I$ mode (8), but with the crucial difference that it commutes with $Z^2$ and does not have the form of Eq. (3). Therefore, Eq. (9) cannot represent a topological edge mode, and cannot be adiabatically connected to any other Majorana mode. Acting on states with $Q^B_{II}$ does indeed generate different states of the same energy, but this is an accidental degeneracy of the symmetric chain (specifically $\mu = 0$), so arbitrarily small perturbations from the fine-tuned point will destroy the degeneracy and the edge mode, as shown below by ED. We thus classify this phase as topologically trivial.

**Disordered phase diagram.** We now consider the case where the parameters $U_j$ and $t_j$ are sampled from probability distributions $P(U)$ and $P(t)$. To calculate the wave-function coefficients $\alpha^I_1$ and $\alpha^A_{II,j}$, we impose the condition $[H, Q^2] = 0$, as done for the noninteracting disordered case [36], yielding $\alpha^{I+1}_j = -(U_{2,j-1}/t_j)\alpha^I_j$. For such a mode to exist we will need to be able to normalize it, i.e., $\langle Q^A_I \rangle^2 = 1$, and thus the sums of the squares of the coefficients $\alpha^I_j$ need to be bounded. As before, the condition for the topological phase is that $Q^A_I$ exists, and so $\alpha^I_1$ must decay sufficiently fast to the right $j \to \infty$. Specifically,

\[
\lim_{N \to \infty} \sum_{j=1}^{N} \left| \sum_{k=1}^{j} U_{2,k-1}/t_{2,k} \right|^2 < \infty.
\]

Repeating the argument for $B$-flavor modes, we find that the above is also the condition for $Q^B_{II}$ to be localized on the right. The $j$th term $S_j$ in the sum above can be written as $S_j = \exp(\sum_{k=1}^{j} 2 \ln |U_{2,k-1}/t_{2,k}|)$ which tends to $e^{2(\ln U - \ln t)}$ as $j \to \infty$, as the sum is self-averaging. Clearly, if

\[
\int dU P(U) \ln |U| < \int dt P(t) \ln |t|.
\]
then $S_j$ decays exponentially with $j$ and the sum converges. Equation (10) is the condition for the topological phase in the disordered system, which represents one of our main results. Within this phase, the Majorana modes have the interacting form (8) with coefficients as calculated above. The argument above also gives us the characteristic decay length of the Majorana mode as $\xi = (\ln t - \ln |U|)^{-1}$, which for uniform or Gaussian distributions diverges as $\sim(|U| - \langle U_{\text{cn}}\rangle)^{-1}$ at the phase transition.

Having derived the condition for the topological phase for arbitrary disorder distributions (10), let us consider the specific example of constant $t_j = t$ and a uniform distribution for $U_j \in [−\sqrt{3}\sigma_U, \sqrt{3}\sigma_U]$, in units for which $t = 1$. We can construct the topological phase diagram as a function of $U$ and $\sigma_U$ by solving Eq. (10) for a uniform $P(U)$, giving an analytic expression for the topological-trivial phase boundary

$$
\left(\frac{U}{\sqrt{3}\sigma_U} + \frac{1}{2}\right) \ln |U + \sqrt{3}\sigma_U| - \left(\frac{U}{\sqrt{3}\sigma_U} - \frac{1}{2}\right) \ln |U - \sqrt{3}\sigma_U| = 1, \quad (11)
$$

where we fixed $U > 0$ as the phase diagram is invariant under $U_j \rightarrow -U_j$. The critical $U$ grows quadratically for weak disorder and reaches a maximum at $(\sigma_U \approx 1.0451, U \approx 1.5089)$. We note that this is a significantly enhanced maximal interaction strength $U$ for which the phase is topological. For $\sigma_U > 2t/\sqrt{3} \approx 1.5694$ the system is trivial regardless of the mean interaction $U$. In this large disorder regime, the system is dominated by sites where $U_j$ is particularly large in magnitude, favoring a trivial charge ordered state.

**Numerical results.** We corroborate our analytic results by calculating the phase diagram numerically. We compute the entanglement spectrum which is degenerate in a topological phase [37] [see the Supplemental Material (SM) [38]]. Specifically, we classify the phase as topological if the finite-size splitting of entanglement eigenvalues is less than some constant $c$. This criterion has been shown to be robust against disorder [28].

Figure 1 shows the phase diagram of the disordered, interacting Kitaev chain, as given by the analytic expression (11) as well as from the degeneracy of the entanglement spectrum for a system of size $N = 800$. We have set $c = 0.1$ and verified that the transition is sharp enough to be insensitive to this arbitrary choice when averaged over 500 disorder realizations. Finite-size effects lead to a slight suppression of the topological phase in the numerically calculated values, due to cases where the Majorana decay length is comparable to the system size. The transition to the trivial phase at strong disorder is also less sharp and finite-size errors are amplified. In the SM [38] we demonstrate that these results for the symmetric region are robust to a disordered hopping amplitude.

Away from the symmetric region, we perform ED of the interacting, disordered Kitaev chain (1). In Fig. 2 we show the ground-state energy gap $\Delta E$ as a function of the mean interaction $U$ in the clean $\sigma_U = 0$ and disordered $\sigma_U \neq 0$ case, for uniform $\mu_j = \mu$ as well as disordered chemical potential $\mu_j \in [−\sqrt{3}\sigma_U, \sqrt{3}\sigma_U]$. For $\mu_j = 0$, zero modes exist for both $|U| > t$ and $|U| < t$, so the gap vanishes away from the transition. However, the accidental $|U| > t$ zero mode is destroyed by a nonzero chemical potential, while the topological mode $|U| < t$ persists. This is the case for uniform $\mu \neq 0$ [Figs. 2(a) and 2(b)] and disordered $\sigma_\mu \neq 0$ [Figs. 2(c) and 2(d)] chemical potentials.

By comparing to the exact results for the symmetric region, it is apparent that the topological phase boundary can be identified even for a small number of sites, and that the transition point depends smoothly and weakly on the chemical potential. This explicitly demonstrates that the ground-state degeneracy of the topological phase is stable for a finite chemical potential, while that of the trivial phase is not. We therefore conclude that the qualitative aspects of the phase diagram (Fig. 1) are robust away from the symmetric region. In particular, as seen from Figs. 2(b) and 2(d), the phase transition occurs for an interaction strength greater than in the clean case $U = t$ for all $\mu$ and $\sigma_\mu$.

**Discussion.** Our work provides a unique insight into the effects of disorder and interactions on topological order. The Kitaev chain in the symmetric region is equivalent to two copies of conventional noninteracting chains [see Eq. (6)]. We have shown that topological order of the original system is related to that of the first of these copies (subsystem 1), for which the interaction plays the role of the chemical potential. As a consequence, our analytic condition for the topological phase, Eq. (10), has a similar form as that found for noninteracting, disordered Kitaev chains [28]. However, while the phase diagram of Ref. [28] captures the competition between Anderson insulating and superconducting phases driven by a chemical potential, we here study transitions driven by interaction. While disorder and interactions can separately degrade the topological phase, their combination can be less detrimental.

Our explicit expression for the many-body Majorana mode (8) is a rare example of an analytical expression of the general form in Eq. (3), albeit with the simplification that most of the coefficients $\{a\}$ are zero. This is due to the integrability of the system in the symmetric region. Nevertheless, one can see how higher-order multiparticle contributions occur at higher order in $U$, as expected from perturbation theory. In particular, we show explicitly that in the expansion of the many-body Majorana operator, terms with $(2n + 1) \gamma$ operators are proportional to $U^n$ [34].

While our analytical results are restricted to the fine-tuned point, $t_j = \Delta_j$ and $\mu_j = 0$, our ED results demonstrate that the qualitative aspects of our findings hold more generally.

**Conclusions and outlook.** We have been able to calculate analytically the topological phase boundary of a class of interacting, disordered Kitaev chains as a function of mean interaction and disorder strength (Fig 1), demonstrating that moderate amounts of disorder in the interactions can enhance the topological phase into the strongly interacting regime $U > t$.

Our work represents a step in utilizing the exact solution of Ref. [30] which enabled us to uncover an interacting part of the symmetric Kitaev chain phase diagram. Having understood the topological properties of the system, we can exploit it further to address a number of experimentally relevant questions. Specifically, how do interactions alter the
zero-bias conductance peak [22] or topological Josephson current [39]?

The analytic tractability of the symmetric Kitaev chain also holds great promise of studying the effects of interactions and disorder in a number of other situations. In particular, we suggest looking at the nature of localized states in the disordered chain in the context of many-body localization [40]. Additionally, the nonlocal nature of the transformation is likely to affect entanglement dynamics and out-of-equilibrium phenomena [41,42]. Finally, our exactly soluble interacting chain with disorder will provide a benchmark point for numerical methods such as density matrix renormalization group (DMRG) [43].

Acknowledgments. J.K. is supported by the Marie Curie Programme under EC Grant Agreement No. 703697. A.N. holds a University Research Fellowship from the Royal Society and acknowledges support from the Winton Programme for the Physics of Sustainability.