

# Topological insulators beyond the Brillouin zone via Chern parity

Andrew M. Essin<sup>1</sup> and J. E. Moore<sup>1,2</sup>

<sup>1</sup>*Department of Physics, University of California, Berkeley, California 94720, USA*

<sup>2</sup>*Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*

(Received 10 May 2007; revised manuscript received 8 August 2007; published 10 October 2007)

The topological insulator is an electronic phase stabilized by spin-orbit coupling that supports propagating edge states and is not adiabatically connected to the ordinary insulator. In several ways it is a spin-orbit-induced analog in time-reversal-invariant systems of the integer quantum Hall effect (IQHE). This paper studies the topological insulator phase in disordered two-dimensional systems, using a model graphene Hamiltonian introduced by Kane and Mele [Phys. Rev. Lett. **95**, 226801 (2005)] as an example. The nonperturbative definition of a topological insulator given here is distinct from previous efforts in that it involves boundary phase twists that couple only to charge, does not refer to edge states, and can be measured by pumping cycles of ordinary charge. In this definition, the phase of a Slater determinant of electronic states is determined by a Chern parity analogous to Chern number in the IQHE case. Numerically, we find, in agreement with recent network model studies, that the direct transition between ordinary and topological insulators that occurs in band structures is a consequence of the perfect crystalline lattice. Generically, these two phases are separated by a metallic phase, which is allowed in two dimensions when spin-orbit coupling is present. The same approach can be used to study three-dimensional topological insulators.

DOI: [10.1103/PhysRevB.76.165307](https://doi.org/10.1103/PhysRevB.76.165307)

PACS number(s): 73.43.-f, 85.75.-d, 73.20.At, 73.20.Fz

## I. INTRODUCTION

Considerable theoretical and experimental effort has been devoted to the quest for an intrinsic spin Hall effect<sup>1-4</sup> that would allow generation of spin currents by an applied electric field. Interesting mechanisms for such spin current generation make use of spin-orbit coupling, which breaks the SU(2) spin symmetry of free electrons but not time-reversal symmetry. A dissipationless type of intrinsic spin Hall effect was predicted<sup>5,6</sup> to arise in materials that have an electronic energy gap. This “quantum spin Hall effect” (QSHE) in certain materials with time-reversal symmetry has a subtle relationship to the integer quantum Hall effect, in which time-reversal symmetry is explicitly broken by a magnetic field.

In a system with unbroken time-reversal symmetry, a dissipationless charge current is forbidden, but a dissipationless transverse spin current is allowed, of the form

$$\mathcal{J}_j = \alpha \epsilon_{ijk} E_k. \quad (1)$$

The current on the left is a spin current and  $\epsilon$  is the fully antisymmetric tensor. Note that a spin current requires two indices, one for the direction of the current and one for the direction of angular momentum that is transported. The constant of proportionality  $\alpha$  depends on the specific mechanism: for example, the (dissipative) extrinsic D’yakonov-Perel mechanism<sup>7</sup> predicts a small  $\alpha$  that depends on impurity concentration. The QSHE builds on the construction by Haldane<sup>8</sup> of a lattice “Chern insulator” model, with broken time-reversal symmetry but without net magnetic flux, that shows a  $\nu=1$  integer quantum Hall effect (IQHE). The simplest example of a QSHE is obtained by taking two copies of Haldane’s model, one for spin-up electrons along some axis and one for spin-down electrons. Time-reversal symmetry can be maintained if the effective IQHE magnetic fields are opposite for the two spin components. Then an applied electric field generates a transverse current in one

direction for spin-up electrons, and in the opposite direction for spin-down electrons. There is no net charge current, consistent with time-reversal symmetry, but there is a net spin current. However, models like this in which one component of spin is perfectly conserved are both unphysical, since realistic spin-orbit coupling does not conserve any component, and not very novel, since for each spin component the physics is exactly the same as Haldane’s model and the spin components do not mix.

More subtle physics emerges when one asks how the QSHE appears in more realistic band structures. Remarkably, band insulators of noninteracting two-dimensional electrons with spin-orbit coupling divide into two classes: the “ordinary insulator,” which, in general, has no propagating edge modes and no spin Hall effect, and the “topological insulator,” which has stable propagating edge modes and a generic spin Hall effect, although the amount of spin transported [the coefficient  $\alpha$  in Eq. (1)] is nonuniversal. These phases are associated with a  $\mathbb{Z}_2$ -valued topological invariant (an “oddness” or “evenness” in the language of parity<sup>41</sup>) in the same way that IQHE phases are associated with an integer-valued topological invariant. For explicitness, consider the model of graphene introduced by Kane and Mele.<sup>5</sup> This is a tight-binding model for independent electrons on the honeycomb lattice (Fig. 1). The spin-independent part of the Hamiltonian consists of a nearest-neighbor hopping, which alone would give a semimetallic spectrum with Dirac nodes at certain points in the two-dimensional (2D) Brillouin zone, plus a staggered sublattice potential whose effect is to introduce a gap:

$$H_0 = t \sum_{\langle ij \rangle \sigma} c_{i\sigma}^\dagger c_{j\sigma} + \lambda_v \sum_{i\sigma} \xi_i c_{i\sigma}^\dagger c_{i\sigma}. \quad (2)$$

Here,  $\langle ij \rangle$  denotes nearest-neighbor pairs of sites,  $\sigma$  is a spin index,  $\xi_i$  alternates sign between sublattices of the honeycomb, and  $t$  and  $\lambda_v$  are parameters.

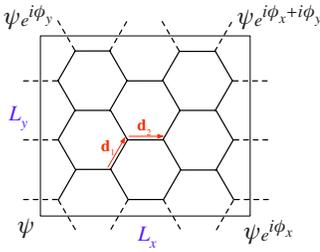


FIG. 1. (Color online) The honeycomb lattice on which the tight-binding Hamiltonian resides. For the two sites depicted, the factor  $\nu_{ij}$  of Eq. (3) is  $\nu_{ij} = -1$ . The phases  $\phi_{x,y}$  describe twisted boundary conditions, introduced in Eq. (10).

The insulator created by increasing  $\lambda_v$  is an unremarkable band insulator. However, the symmetries of graphene also permit an “intrinsic” spin-orbit (SO) coupling of the form

$$H_{SO} = i\lambda_{SO} \sum_{\langle\langle ij \rangle\rangle_{\sigma_1\sigma_2}} \nu_{ij} c_{i\sigma_1}^\dagger s_{\sigma_1\sigma_2}^z c_{j\sigma_2}. \quad (3)$$

Here,  $\nu_{ij} = (2/\sqrt{3})\hat{\mathbf{d}}_1 \times \hat{\mathbf{d}}_2 = \pm 1$ , where  $i$  and  $j$  are next-nearest neighbors, and  $\hat{\mathbf{d}}_1$  and  $\hat{\mathbf{d}}_2$  are unit vectors along the two bonds that connect  $i$  to  $j$ . Including this type of spin-orbit coupling alone would not be a realistic model. For example, the Hamiltonian  $H_0 + H_{SO}$  conserves  $s^z$ , the distinguished component of electron spin, and reduces for fixed spin (up or down) to Haldane’s model.<sup>8</sup> Generic spin-orbit coupling in solids should not conserve any component of electron spin. However, the unusual phase generated when  $H_{SO}$  is strong turns out to survive, with subtle changes, once the spin-orbit coupling is made more realistic, as we now review.

The topological insulator phase created when  $|\lambda_{SO}| \gg |\lambda_v|$  is quite different from the ordinary insulator that appears when  $|\lambda_v| \gg |\lambda_{SO}|$  (here, we assume that there is an energy gap between the lower and upper band pairs in which the Fermi level lies). The former has counterpropagating edge modes and shows the QSHE, while the latter does not.<sup>5</sup> Does this phase exist for more realistic spin-orbit coupling? The spin component  $s^z$  is no longer a good quantum number when the Rashba spin-orbit coupling is added:

$$H_R = i\lambda_R \sum_{\langle ij \rangle_{\sigma_1\sigma_2}} c_{i\sigma_1}^\dagger (s_{\sigma_1\sigma_2} \times \hat{\mathbf{d}}_{ij})_z c_{j\sigma_2}, \quad (4)$$

with  $\mathbf{d}_{ij}$  the vector from  $i$  to  $j$ , and  $\hat{\mathbf{d}}_{ij}$  the corresponding unit vector. Note that Rashba spin-orbit coupling is not intrinsic to graphene but generated by inversion-symmetry breaking in the out-of-plane direction.<sup>9</sup> The Rashba coupling is a standard form that is believed to be a reasonable model for the dominant spin-orbit coupling in adsorbed graphene.

The topological insulator survives but is strongly modified in the presence of the Rashba term. For a general 2D band structure with  $s^z$  conserved, there are many phases labeled by an integer  $n$ , as in the IQHE: if spin-up electrons are in the  $\nu = n$  state, then spin-down electrons must be in the  $\nu = -n$  state by time-reversal symmetry, where the sign indicates that the direction of the effective magnetic field is reversed. Once  $s^z$  is not conserved, as when  $\lambda_R \neq 0$ , there are

only two insulating phases, the ordinary and topological insulators. A heuristic definition of the topological insulator, without reference to any particular spin component or the spin Hall effect, is as a band insulator that is required to have gapless propagating edge modes at the sample boundaries. The decoupled  $\nu = \pm n$  cases with  $s^z$  conserved are adiabatically connected, once  $s^z$  is not conserved, to the ordinary insulator for even  $n$  and to the topological insulator for odd  $n$ . A review of how these two cases emerge as the only possibilities in 2D follows in Sec. II.

It should be clarified that the intrinsic spin-orbit coupling is now believed to be quite weak in graphene,<sup>10–12</sup> so that the topological insulator is unlikely to be realized. However, the same topological insulator phase is now believed to exist for realistic spin-orbit coupling in other materials such as HgTe.<sup>13</sup> We choose to study the graphene model introduced by Kane and Mele because it is the first and simplest model showing a transition between ordinary and topological insulators. It is the simplest possible model in that it has four spin-split bands, which is the minimum number required for the nontrivial phase to exist.<sup>14</sup> For this reason, it has received the most attention in the other studies<sup>15,16</sup> to which our results will be compared. It is straightforward to generalize the approach presented here for the graphene model to another material with more complicated spin-orbit coupling, and the same qualitative results are expected to apply.

It is not obvious at first glance how to generalize the topological insulator phase to finite, noncrystalline systems, rather than band structures, as when the parameters of the Hamiltonian  $H = H_0 + H_{SO} + H_R$  are drawn from a random distribution. The first approach was in terms of a spin Chern number<sup>15</sup> similar to the Chern integer in finite IQHE systems, but there is now agreement that for a clean band structure the only invariants are of  $\mathbb{Z}_2$  type, rather than integer type.<sup>14,17,18</sup> Two equivalent definitions of the appropriate  $\mathbb{Z}_2$  invariant for a finite disordered system, in the simple case when the disorder splits all degeneracies other than Kramers degeneracies, are as follows (the full definition is given and compared to previous work in the following section). The finite system can be considered as a unit “supercell” of a large 2D lattice. A large, finite supercell gives many bands, but each pair of bands connected by time reversal (Kramers pair) can be assigned its own  $\mathbb{Z}_2$  invariant.<sup>14</sup> The phase of the supercell system, if the Fermi level lies in a gap, is then identified by adding up all the invariants (mod 2). Alternatively, a direct definition of the phase in the finite system can be given that is related to the notion of “ $\mathbb{Z}_2$  pumping.”<sup>17</sup> Real charge is pumped as the flux through the system is taken from 0 to  $hc/2e$  (half the usual flux quantum that appears in IQHE pumping); we show that in the topological insulator, any pumping cycle, properly defined, pumps an odd number of electron charges, while for the ordinary insulator, any cycle pumps an even number of charges.

We implement this definition numerically using an explicit algorithm introduced by Fukui and Hatsugai<sup>19</sup> for computing  $\mathbb{Z}_2$  topological invariants on a Brillouin zone. The topological insulator phase is robust to disorder: while different realizations of disorder assign different “Chern parities” to individual subbands, it is found that the total for occupied subbands is always “odd” for a wide range of pa-

rameters, which in our definition indicates a topological insulator. In the IQHE, a pair of bands of opposite Chern numbers can annihilate as the strength of disorder is increased; in the QSHE, two band pairs that both have odd Chern parity can annihilate, i.e., become two even-parity band pairs. If the topological insulator can be destroyed by band annihilation, then there are extended (i.e., topologically nontrivial) states with an arbitrarily small gap; it may be the case that for some range of parameters, there are extended states at the Fermi level even in the thermodynamic limit, indicating a metallic phase.

In the IQHE, there is only a single energy with extended states rather than a range of energies, and hence, no metallic phase. We find the phase diagram of the graphene model with on-site disorder, and in the presence of nonzero Rashba coupling, find evidence for a metallic phase intervening between ordinary and topological insulators. The existence of the metallic phase can be understood from work on 2D localization in the symplectic universality class,<sup>20</sup> in which time-reversal symmetry is unbroken but spin-orbit coupling is present. It is found that extended states can be stable against this disorder over a nonzero range in energy, unlike in the orthogonal class, in which there are no extended states, or in the unitary (integer quantum hall) class, in which there are extended states only at isolated energies. The result of a metallic phase can be understood as indicating that, although the  $\mathbb{Z}_2$  topological invariant allows one to distinguish two kinds of insulators when the Fermi level has no extended states, this invariant does not modify the standard picture of bands of extended states in the symplectic universality class.

Recent work by Obuse *et al.*<sup>21</sup> obtains a phase diagram and critical exponents using a network model for the spin quantum Hall effect that is similar to the Chalker-Coddington network model<sup>22</sup> for the IQHE (see also Onoda *et al.*<sup>16</sup> for a quasi-one-dimensional study of localization in the Kane-Mele Hamiltonian with disorder). Our results on the phase diagram are consistent with these works, although our method is unable to generate large enough system sizes to confirm the exponents found for the phase transitions. To understand how the network and Chern-parity approaches complement each other, consider the IQHE: while the phenomenological network approach to the IQHE is valuable both to find the critical indices precisely and to identify the minimal necessary elements of a theory for the transition, Chern-number studies remain important for studies of effects such as the floating of extended states,<sup>23–25</sup> where knowledge of the topological properties of a state is required. The network model gives more accurate information about the phase transitions but, if only the localization length is probed, does not distinguish the different phases in bulk. A more technical difference between the two approaches is discussed at the end of Sec. II.

Section II reviews how the topological insulator phase in perfect crystals arises from a parity-valued topological invariant of the band structure, similar to the Thouless–Kohmoto–Nightingale–den Nijs<sup>26</sup> (TKNN) integers in the IQHE. It then gives two mathematically equivalent definitions of the topological insulator phase in disordered systems based on Chern parity. One definition simply considers a finite disordered system as a supercell of an infinite lattice

system, while the other is based on closed charge pumping cycles driven by the application of flux to a finite periodic system. Section III reviews the Fukui-Hatsugai algorithm<sup>19</sup> adapted to the numerical computation of these invariants in disordered systems, then computes the phase diagram of the Kane-Mele graphene model<sup>5</sup> with on-site disorder. The conclusions of our study for general 2D disordered systems are summarized in Sec. IV. While there is a three-dimensional version of the QSHE<sup>14,27–29</sup> that is less directly connected to the IQHE and has interesting localization behavior, we will restrict our attention to two dimensions except for some comments in the final section.

## II. CHERN PARITIES FOR DISORDERED NONINTERACTING ELECTRON SYSTEMS

### A. Definition of the $\mathbb{Z}_2$ invariant in clean systems

We review one definition of the  $\mathbb{Z}_2$  invariant of a band pair in a 2D band structure, then explain its generalization to noncrystalline systems. With the Hamiltonian

$$H = H_0 + H_{SO} + H_R \quad (5)$$

defined in Eqs. (2)–(4), or any periodic, single-electron Hamiltonian, a Berry connection  $\mathcal{A}$  can be defined on the Brillouin zone (BZ) from the periodic part  $u(\mathbf{k})$  of a Bloch state  $\psi_{\mathbf{k}} = u(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}}$ . For a single nondegenerate band  $j$ ,  $\mathcal{A}_j(\mathbf{k}) = -i\langle u_j | \nabla_{\mathbf{k}} | u_j \rangle$ . This Berry connection serves as a potential for the Berry field strength  $\mathcal{F} = (\nabla_{\mathbf{k}} \times \mathcal{A})_z$ , which, when integrated over the BZ, returns an integer, called a TKNN integer in the context of the IQHE:<sup>26</sup>

$$\begin{aligned} n_j &= -\frac{1}{2\pi} \int_{\text{BZ}} \mathcal{F}_j d^2k \\ &= \frac{i}{2\pi} \int_{\text{BZ}} \left[ \left\langle \frac{\partial u}{\partial k_x} \middle| \frac{\partial u}{\partial k_y} \right\rangle - \left\langle \frac{\partial u}{\partial k_y} \middle| \frac{\partial u}{\partial k_x} \right\rangle \right] d^2k. \end{aligned} \quad (6)$$

When bands touch, only the total TKNN integer of the bands is well defined,<sup>30</sup> in which case it makes sense to generalize  $\mathcal{A}$  as

$$\mathcal{A}_j = -i(\langle u_{j1} | \nabla_{\mathbf{k}} | u_{j1} \rangle + \langle u_{j2} | \nabla_{\mathbf{k}} | u_{j2} \rangle) = -i \text{tr} \mathbf{u}_j^\dagger \nabla_{\mathbf{k}} \mathbf{u}_j \quad (7)$$

for the degenerate Bloch functions  $u_{j1}$  and  $u_{j2}$ . In the second equation,  $\mathbf{u}_j = (u_{j1}, u_{j2})$ , where the Bloch functions are viewed as column vectors, i.e.,  $\mathbf{u}_j$  is a matrix. This compact notation follows Fukui and Hatsugai.<sup>18</sup> This generalization is always needed in the case of a time-reversal-invariant system with half-odd integer spin, or fermionic statistics, because such a system has “Kramers degeneracies” at certain time-reversal-invariant momenta (see, for example, Ref. 31). Briefly, time-reversal invariance requires that

$$\Theta \mathcal{H}(-\mathbf{k}) \Theta^{-1} = \mathcal{H}(\mathbf{k}), \quad (8)$$

where  $\mathcal{H}(\mathbf{k})$  is the Bloch Hamiltonian and  $\Theta = -i\sigma_y K$  is the action of time reversal ( $K$  performs complex conjugation and  $\sigma_y$ , the usual Pauli matrix, acts on spin indices). At momenta for which  $-\mathbf{k} = \mathbf{k}$ ,  $\mathcal{H}(\mathbf{k})$  commutes with  $\Theta$ , but there are no single-electron eigenstates of  $\Theta$  (time reversal will always

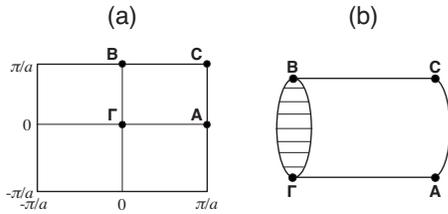


FIG. 2. (a) A two-dimensional Brillouin zone. Note that any such Brillouin zone, including that for graphene, can be smoothly deformed to a torus. The labeled points are time-reversal-invariant momenta. (b) The effective Brillouin zone (EBZ). The horizontal lines on the boundary circles  $\partial(\text{EBZ})$  connect time-reversal-conjugate points, where the Hamiltonians are related by time reversal and so cannot be specified independently.

flip the spin), so there must be a degenerate pair of energy eigenstates at each such momentum. Once we take these degeneracies into account, however, the TKNN integers for the band pairs vanish in a time-reversal-invariant band structure (see, e.g., Ref. 14). Instead there is a  $\mathbb{Z}_2$  invariant associated with a band pair in a time-reversal-invariant 2D Fermi system.<sup>6</sup>

Fu and Kane<sup>17</sup> give the following formula for the  $\mathbb{Z}_2$  topological invariant in terms of the Bloch functions of the clean system:

$$D = \frac{1}{2\pi} \left[ \oint_{\partial(\text{EBZ})} dk \cdot \mathcal{A} - \int_{\text{EBZ}} d^2k \mathcal{F} \right] \text{ mod } 2. \quad (9)$$

The notation EBZ stands for effective Brillouin zone,<sup>14</sup> which describes one-half of the Brillouin zone together with appropriate boundary conditions. Since the BZ is a torus, the EBZ can be viewed as a cylinder, and its boundary  $\partial(\text{EBZ})$  as two circles, as in Fig. 2(b). While  $\mathcal{F}$  is gauge invariant,  $\mathcal{A}$  is not, and different (time-reversal invariant) gauges can change the boundary integral by an even amount.

A proof of the existence and  $\mathbb{Z}_2$  nature of the topological invariant for multiple bands and some intuition for Eq. (9) can be obtained<sup>14</sup> by considering time-reversal-invariant band structures as maps from the EBZ to the space of Bloch Hamiltonians, following work on the IQHE by Avron *et al.*<sup>30</sup> The Berry field strength can be written in terms of the (gauge invariant) projection operator onto the band pair rather than the (gauge dependent) wave functions. In this approach, the ambiguity by an even integer that is crucial to obtain a  $\mathbb{Z}_2$  rather than a  $\mathbb{Z}$  invariant corresponds to the many different ways in which the circles that form EBZ boundaries in Fig. 2(b) can be contracted to make the EBZ into a sphere. The boundary integrals in Eq. (9) just calculate the contribution to the Chern number from these ‘‘contractions.’’ On this sphere, Chern integers are well defined for each nondegenerate band pair, but the different ways of contracting the boundaries cause the resulting integers to differ by even numbers. An explicit numerical implementation of the Fu-Kane formula (9) was given by Fukui and Hatsugai<sup>19</sup> and will be reviewed in Sec. III.

What happens when disorder is added? We now explain how the definition of the topological insulator generalizes to

disordered systems. Just as TKNN integers of a band structure give rise to Chern integers in a disordered system of finite size, so do the  $\mathbb{Z}_2$  invariants of band pairs become Chern parities.

### B. Topological insulator phase for Slater determinants via Chern parity

The TKNN integers generalize in the presence of disorder and interactions to the Chern number of the many-particle wave function.<sup>32</sup> A natural question is how disorder and interactions modify the  $\mathbb{Z}_2$  invariants of band pairs in spin-orbit-coupled 2D band structures. All derivations of the  $\mathbb{Z}_2$  invariants of clean systems depend on Fermi statistics in some way: for example, the existence of Kramers degeneracies and the related fact that the time-reversal operator squares to  $-1$  both depend on Fermi statistics. A many-fermion wave function describing an even number of fermions does not behave in the same way as single-fermion wave functions under time reversal. Hence, given only the many-fermion wave function, it does not seem likely that there is a generalization of the  $\mathbb{Z}_2$  invariant.

However, for the particular case of many-fermion wave functions that are single Slater determinants of single-particle wave functions, the invariant can be generalized as we now show. While the assumption of a single Slater determinant limits the treatment of interactions to the Hartree-Fock level, there is no requirement that the wave functions in the Slater determinant be Bloch states. As a result, the topological insulator and QSHE can be defined and studied for any disorder strength.

Niu *et al.*<sup>32</sup> and Avron and Seiler<sup>33</sup> showed that for disordered quantum Hall systems, there exists a generalization of the TKNN invariant defined for clean systems.<sup>26</sup> They introduce generalized periodic boundary conditions and find an invariant Chern number, similar in form to the TKNN invariant, on the space of boundary phases. Consider a finite system of noninteracting electrons with boundary conditions that are periodic up to phases  $\phi_x, \phi_y$ , as shown in Fig. 1: this is equivalent to putting magnetic fluxes  $\Phi_{x,y} = \phi_{x,y} \Phi_0 / 2\pi$  through the two noncontractible circles on the torus ( $\Phi_0 = hc/e$  is the magnetic flux quantum). As motivation, think of the finite system as a (possibly very large) unit cell of a lattice system. Then in order to determine the phase of this lattice system, instead of integrating over  $k$  to do the integrals in the Fu-Kane formula (9), we integrate over the boundary phases, which introduces offsets to the wave vectors. We now carry out this procedure, show that it reproduces the band-structure result for clean systems, and then discuss a physical picture and its relation to previous definitions.

Consider the single-particle wave functions of a lattice Hamiltonian such as the graphene model on a finite lattice of size  $L_x \times L_y$  (see Fig. 1). Instead of the physical boundary conditions  $\psi(\mathbf{x} + \mathbf{L}_{x,y}) = \psi(\mathbf{x})$  for a single-particle wave function  $\psi$ , introduce the boundary phases, or ‘‘twists,’’  $\phi = (\phi_x, \phi_y)$  via

$$\psi(\mathbf{x} + \mathbf{L}_x) = e^{i\phi_x} \psi(\mathbf{x}), \quad \psi(\mathbf{x} + \mathbf{L}_y) = e^{i\phi_y} \psi(\mathbf{x}). \quad (10)$$

Then a unitary transformation of the form

$$\chi(\mathbf{x}) = e^{-i(\phi_x x/L_x + \phi_y y/L_y)} \psi(\mathbf{x}) \quad (11)$$

will transfer the twist angles to the Hamiltonian. Under the change of basis (11), Kane and Mele's model Hamiltonian<sup>5</sup>  $H=H_0+H_{SO}+H_R$  (see Sec. I) becomes (suppressing spinor indices)

$$\begin{aligned} H \rightarrow \mathcal{H}(\boldsymbol{\phi}) = & \sum_{\langle ij \rangle} c_i^\dagger [t + i\lambda_R(\mathbf{s} \times \hat{\mathbf{d}}_{ij})_z] c_j e^{-i\boldsymbol{\Delta} \cdot \mathbf{d}_{ij}} \\ & + i\lambda_{SO} \sum_{\langle\langle ij \rangle\rangle} v_{ij} c_i^\dagger s^z c_j e^{-i\boldsymbol{\Delta} \cdot \mathbf{d}_{ij}} + \sum_i (\lambda_v \xi_i + w_i) c_i^\dagger c_i, \end{aligned} \quad (12)$$

where  $\boldsymbol{\Delta}=(\phi_x/L_x, \phi_y/L_y)$ ,  $\mathbf{d}_{ij}$  is still the vector  $i \rightarrow j$ , and we have added a random term for on-site disorder,  $w_i$ , drawn from the Gaussian distribution of zero mean and standard deviation  $\sigma_w$ . It is now clear that under time reversal,

$$\Theta \mathcal{H}(-\boldsymbol{\phi}) \Theta^{-1} = \mathcal{H}(\boldsymbol{\phi}), \quad (13)$$

since the extra phase factors in  $\mathcal{H}$  change sign under complex conjugation.<sup>42</sup> Since this directly parallels Eq. (8), the  $\mathbb{Z}_2$  invariant  $D$  of the Brillouin zone passes directly to twist space:

$$D_\phi = \frac{1}{2\pi} \left[ \oint_{\partial(\text{ETZ})} d\boldsymbol{\phi} \cdot \mathcal{A} - \int_{\text{ETZ}} d^2\boldsymbol{\phi} \mathcal{F} \right] \text{ mod } 2, \quad (14)$$

with ETZ for effective twist zone, i.e.,  $\text{ETZ}=\{\boldsymbol{\phi} | 0 \leq \phi_x \leq \pi, -\pi < \phi_y \leq \pi\}$ . Note that there is an independent Chern parity  $D_\phi$  for each Kramers-degenerate band pair separated from the rest of the spectrum by a gap at all  $\boldsymbol{\phi}$ : for such an isolated pair,  $\mathcal{A}$  and  $\mathcal{F}$  are defined as in Eq. (7) and after, with  $\mathbf{k} \rightarrow \boldsymbol{\phi}$  and  $u \rightarrow \chi$ .

In order to make contact with the band-structure definition, we note that if there is no disorder in the Hamiltonian (i.e.,  $\sigma_w=0$ ), there are discrete translational symmetries within the  $L_x \times L_y$  supercell that induce additional non-Kramers degeneracies at some points in twist space. With such degeneracies, only the total Chern parity of all the degenerate states is well defined. Disorder, as discussed in the following section, breaks all degeneracies resulting from translational invariance, leaving only separated band pairs, each of which has its own Chern parity. We now discuss to what extent Chern parities can be connected to observable quantities in a finite system.

### C. Charge pumping cycles in time-reversal-invariant systems

The total Chern number in a finite IQHE system can be interpreted as measuring the number of charges pumped when the flux through one noncontractible circle on the torus increases adiabatically by one flux quantum. Briefly, one of the boundary phases corresponds to this driving flux, and the average over the other can be shown to yield the pumped charge.<sup>34</sup> The idea of  $\mathbb{Z}_2$  pumping suggested by Fu and Kane<sup>17</sup> is the following: in a finite cylinder with boundaries, the operation of increasing the phase  $\phi_x$  (in the periodic direction, around the cylinder) from 0 to  $\pi$ , corresponding to a magnetic flux of one-half flux quantum through the cylinder,

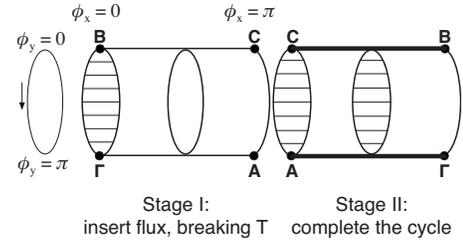


FIG. 3. Graphical representation of charge pumping cycle for Chern parities. The first stage takes place on the ETZ [as in Eq. (14)], and the flux  $\phi_x$  increases adiabatically from 0 to  $\pi$ . In the second stage, the Hamiltonian at  $(\phi_x=\pi, \phi_y)$  is adiabatically transported through the space of Hamiltonians to return to the Hamiltonian at  $(\phi_x=0, \phi_y)$ . The difference between the second stage and the first is that at every step of the second stage, the Hamiltonians obey the time-reversal conditions required at  $\phi_x=0$  or  $\phi_x=\pi$ . The bold lines indicate paths along which all Hamiltonians are time-reversal invariant, and the disk with horizontal lines indicates, as before, how pairs of points in the second stage are related by time reversal.

has the following effect in the topological insulator. The values  $\phi_x=0$  and  $\phi_x=\pi$  are special because, unlike general values, they are consistent with time-reversal invariance. At these special fluxes, there are gapless states at the Fermi level that are localized near the edges because of the bulk gap. Fermi statistics requires that these states lie in Kramers doublets. If at zero flux the Kramers doublet at one edge is partially occupied (has one state occupied), then the operation of changing the flux changes its occupancy to either double or zero occupancy. Since total charge is conserved, this requires a flow of  $\mathbb{Z}_2$  from one boundary to the other: the final state differs from the initial state.

We now give an alternate definition of the topological insulator in term of *cyclic* pumping of ordinary charge. This definition is mathematically equivalent to the definition of Sec. II B based on treating the finite system as a supercell. In order to describe a closed pumping cycle, we need to add a second stage to the process of increasing boundary phase  $\phi_x$  from 0 to  $\pi$ : although both these phases are consistent with time-reversal invariance, physical properties, like the occupancy of an edge doublet, are not identical at these different values of  $\phi_x$  (even at the same  $\phi_y$ ). Hence, the charge pumped in this process is not automatically well defined. The only requirement on the second stage is essentially that it return the system to its original state *without applying a time-reversal-breaking flux*. While this definition precisely reproduces the supercell definition above, it should be noted that it is slightly different from pumping in the IQHE since, in some cases, the second stage requires changing the system's Hamiltonian and not just the boundary phase.

Although the number of charges pumped is dependent not only on the first stage but on the second stage as well, whether this is an even or odd number is entirely determined by the first stage, as we now show. A closed pumping cycle is shown in Fig. 3. The original physical system's ETZ is the first stage of the cycle: the Hamiltonians are functions of  $\phi_x$  from 0 to  $\pi$  and  $\phi_y$  from  $-\pi$  to  $\pi$ , with time-reversal constraints that act on the boundary circles at  $\phi_x=0$  and  $\phi_x=\pi$ .

If  $\phi_x$  takes one of these values, then the Hamiltonian at  $\phi_y$  is time-reversal conjugate to that at  $\phi'_y = -\phi_y$ . The second stage can be any continuous change of the Hamiltonians that takes the  $\phi_x = \pi$  system back to the  $\phi_x = 0$  system and always satisfies the conjugacy condition between  $\phi_y$  and  $\phi'_y$ . This is the key difference between the second stage and the first stage: for intermediate values  $0 < \phi_x < \pi$ , there is no such conjugacy condition. The physical interpretation is that the second stage should be possible without introducing flux through the first noncontractible circle. The Fukui-Hatsugai algorithm<sup>18</sup> reviewed in the following section can be used to “choose a gauge.”

Now the torus shown in Fig. 3 has one Chern integer for each isolated band pair. Summing over occupied bands gives the amount of charge pumped in the cycle. Although this integer charge depends on the second stage, its parity is solely determined by the first stage, i.e., the physical system. In particular, for the ordinary insulator, there is some closed cycle that pumps zero charge, while for the topological insulator, there is some closed cycle that pumps unit charge. These results follow from the same proof as for the band-structure case in Ref. 14: one shows that the differences in resulting Chern integers between any two second stages are even. The pumping definition gives a different physical picture for the contractions introduced there; instead of contracting the EBZ to a sphere, here the ETZ is contracted to a torus by adding the second stage. The technical reason that these two constructions are equivalent is that, since the appropriate spaces of Hamiltonians are contractible (i.e., have  $\pi_1 = 0$ ), the two closed manifolds, the torus and the sphere, both have the same topological invariants, namely, integer-valued Chern numbers.

The topological insulator in disordered systems has been studied previously by locating the transition between topological and ordinary insulators as a point or region where the localization length of single-electron eigenstates diverges. The existence of the topological insulator is inferred from the existence of this transition region (or, alternately, from the edge states in the topological insulator phase). In principle, this approach could give a different result from ours, in that our definition probes the existence not just of extended states but specifically of extended states that contribute to the pumping of charge or, alternately, that can give rise to edge states. The same distinction arises in the quantum Hall effect, where looking for extended states of nonzero Chern number is a more direct probe of quantum Hall physics than considering the inverse participation ratio, for example, which would detect extended states of zero Chern number in addition to topological states. However, as in the quantum Hall case, we find that the phase boundaries from our Chern-parity definition are consistent with those obtained from calculations of the localization length.<sup>16,21</sup>

### III. GRAPHENE MODEL AND NUMERICS

#### A. Phase diagram of the disordered graphene model

It is useful to review some general expectations before applying the definitions of the previous section to study a disordered version of the graphene model. In two-

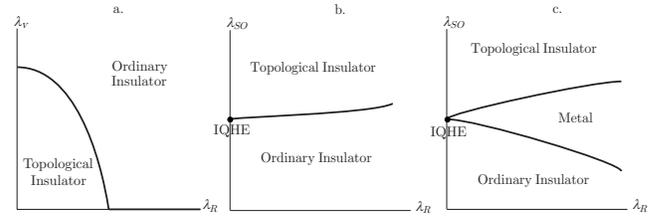


FIG. 4. Schematic phase diagram. (a) Phase diagram for a clean system, with fixed  $t$  and  $\lambda_{SO} \neq 0$  [after Kane and Mele (Ref. 6)]. (b) Again a clean system, now with fixed  $t$  and  $\lambda_v \neq 0$ . (c) The expected form of the phase diagram at nonzero disorder (we run all simulations at fixed  $\lambda_v$ ). The phase boundary in (b) opens up into a metallic phase, closing only when  $\lambda_R = 0$ , where there should be an IQHE transition.

dimensional systems with time-reversal invariance and no spin-orbit coupling, even very weak disorder will localize electron wave functions, so that these systems do not ever conduct in the thermodynamic limit. In the presence of a magnetic field, as in the IQHE, there are isolated energies with extended states, but no finite-width range of energies with extended states, and hence, no true metallic phase. Hikami *et al.* showed<sup>20</sup> that disordered two-dimensional systems *with spin-orbit coupling* can, nevertheless, support a metallic phase, referred to as a “symplectic metal.” They calculated ladder diagrams for the problem of random potential scattering in two dimensions in order to obtain a renormalization-group equation for the resistance, and found that, in the presence of strong spin-orbit scattering without magnetic scattering (so that the system is time-reversal invariant), the resistance flows to zero, indicating extended states and a metallic phase. In the presence of strong magnetic scattering (i.e., the IQHE), in contrast, the resistance does not flow at that order, and at next order grows (see also Ref. 35).

In the case studied here, any metallic phase induced by disorder would presumably appear in a region around the parameter set that closes the (clean) gap, as depicted schematically in Fig. 4. At  $\lambda_R = 0$ , the  $z$  component of spin is a good quantum number ( $s^z$  commutes with  $H$ ), so the system reduces to two copies of the Haldane model,<sup>8</sup> which has a quantum Hall plateau transition with no metallic phase. Hence, when  $\lambda_R = 0$ , there should be a direct transition between insulators even with disorder.

#### B. Lattice implementation

For numerical work, we use the algorithm of Fukui and Hatsugai, which we review here.<sup>19</sup> The formula (9) for  $D$  requires a gauge choice for the Hamiltonian eigenstates at each  $\phi$  on the two boundaries of the half-torus  $0 \leq \phi_x \leq \pi$ ,  $-\pi < \phi_y \leq \pi$ . That is, the “field strength”  $\mathcal{F}$  is gauge invariant, but the gauge potential  $\mathcal{A}$  is not. The eigenstates form Kramers pairs related by time reversal, and the gauge choice must respect this constraint.

Now, at the time-reversal-invariant points  $\phi = (0, 0)$ ,  $(0, \pi)$ ,  $(\pi, 0)$ , and  $(\pi, \pi)$ , the solid points in Fig. 5, the spectrum is degenerate, with two states at each energy. The gauge

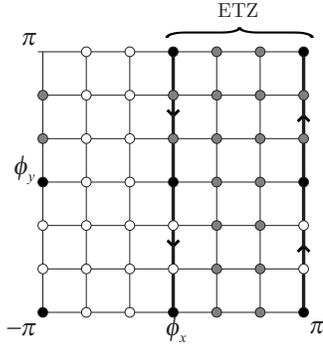


FIG. 5. Twist space. The bold lines indicate the boundaries of the “effective twist zone,” the region we integrate (or sum) over to calculate the Chern parity. The arrows indicate the direction to perform the sum over the boundary terms, and the lattice sites in gray indicate those for which the Hamiltonian eigenvectors are independently specified. That is, time-reversal symmetry determines the eigenvectors on the white sites once those at the gray sites are found.

condition requires that  $\Theta$  interchange the two with a phase factor  $e^{\pm i\pi/2}$  (so that  $\Theta^2 = -1$  as required for single-fermion states). Numerical diagonalization will not, in general, return eigenvectors that obey this condition, but we can force them to do so as follows: choosing one of the two members of each Kramers pair at energy  $\varepsilon_{2n-1} = \varepsilon_{2n}$  and calling that vector  $\chi_{2n-1}$ , we discard the other and replace it by

$$\chi_{2n} = \Theta \chi_{2n-1}. \quad (15)$$

On the rest of the boundary, eigenvectors can be chosen freely on  $0 < \phi_y < \pi$ . On  $-\pi < \phi_y < 0$ , the algorithm takes

$$\chi_n(-\phi) = \Theta \chi_n(\phi). \quad (16)$$

In summary, the algorithm leaves alone the results of numerical diagonalization at all the gray points in Fig. 5, and by hand, enforces the gauge condition on the rest of  $\partial(\text{ETZ})$ .<sup>43</sup>

With the eigenstates fixed at each point on the ETZ, we follow Fukui and Hatsugai, and construct U(1) parallel transporters on the links as

$$U_x(\phi) = \frac{g_x}{|g_x|}, \quad g_x = \det \chi^\dagger(\phi) \chi(\phi + \hat{x}), \quad (17)$$

and  $U_y$  similarly. Like  $\mathbf{u}$  in Eq. (7),  $\chi$  is a matrix built from occupied state vectors, and  $\hat{x}$  translates by one link in the  $\phi_x$  direction. In the continuum limit,  $g$  should approach a pure phase, but for nonzero lattice spacing it will, in general, have  $|g| \leq 1$ , since the occupied subspace of interest will not embed in the total Hilbert space in the same way at every lattice point.

In the end, the only retained information will be the variation of the relative phases (hence the definition of  $U$ ), which can be captured by choosing a lattice constant so small that the phase field varies slowly over one link. However, the scale of variation will presumably differ for different disorder realizations, and we would like a way to diagnose this and throw out those realizations for which fast variation makes the calculation unreliable. The phase is periodic, and

if it were to wind through  $2\pi$  over the distance of one link, the algorithm would miss this fact, so the relative phase will not provide a good diagnostic. As a proxy, we choose to cull disorder realizations that result in small determinants in most simulations, since a small overlap between adjacent occupied eigenspaces indicates rapid variation. Of course, this filtering could introduce a selection bias into the results; these effects are within the statistical uncertainty of our analysis (in particular, within the error bars of Fig. 9).

Associated with the transporter on each link is a gauge potential  $A_{x,y} = \log U_{x,y}$ . This  $A$  is imaginary, and the logarithm is defined to return the branch  $A/i \in (-\pi, \pi)$ . Associated with the transport around each plaquette is a flux

$$F(\phi) = \log U_x(\phi) U_y(\phi + \hat{x}) U_x^{-1}(\phi + \hat{y}) U_y^{-1}(\phi), \quad (18)$$

again satisfying  $F/i \in (-\pi, \pi)$ . With these definitions in hand, the lattice  $\mathbb{Z}_2$  invariant corresponding to  $D_\phi$  (14) is

$$D_L = \frac{1}{2\pi i} \left[ \sum_{l \in \partial(\text{ETZ})} A_y - \sum_{\square \in \text{ETZ}} F \right] \text{ mod } 2. \quad (19)$$

Of course, the sum over the boundary should have the same orientation as the corresponding contour integral in Eq. (14); in Fig. 5, this means the sum on the left boundary should carry a minus sign, following the arrows. As mentioned previously, this formalism has the desirable property of guaranteeing that  $D_L = 0$  or 1, but using too coarse a mesh can return the wrong value.

### C. Numerical results

As noted after Eq. (14), the number of nondegenerate Kramers pairs in a disordered system will generically be extensive, and we can use Eq. (19) to calculate the Chern parity of each pair separately. Figure 6 shows the results of such calculations on a  $4 \times 6$  lattice to present a picture of the three phases: normal insulator, symplectic metal, and topological insulator. In all phases, there are Kramers pairs with  $D_\phi = 1$  in the lower half of the energy spectrum, as indicated by the bars. However, in the normal insulator ( $\lambda_{SO}$  well below the transition), there are an *even* number of such pairs in the occupied half of all realizations, so that the overall Chern parity is even. The presence of a very small number of realizations (2 of 215 for a  $6 \times 8$  lattice) with odd parity indicates either the tail of the disorder-broadened transition or, more likely, that the weak filter applied for these simulations failed to catch all realizations with rapidly varying link variables.

When  $\lambda_{SO}$  is large, almost all realizations have an odd number of  $\mathbb{Z}_2$ -odd Kramers pairs (as with the small- $\lambda_{SO}$  case, two realizations do not follow the rule), and in the region near the transition of the clean system, there are instances of both types. The presence of disorder causes the “gap” to close at different values of  $\lambda_{SO}$  for different realizations, and also at different energies. The latter fact means that extended states are present throughout a finite spread of energies, while the former means that the metallic state exists over a finite region of parameter space.

For an integer quantum Hall system, Yang and Bhatt<sup>25</sup> have shown how to extract the localization length exponent  $\nu$

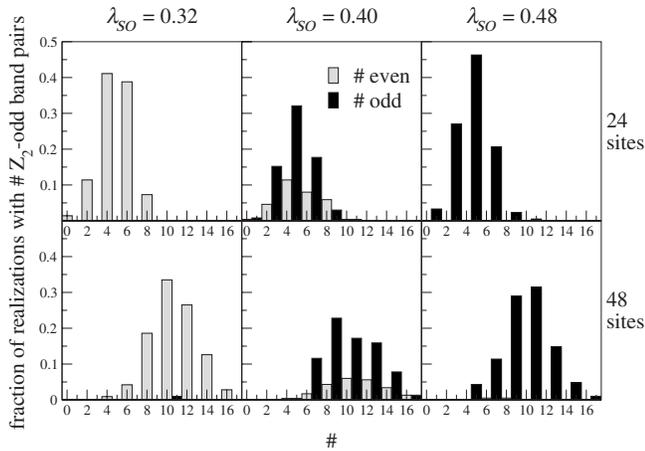


FIG. 6. Distribution of Chern parities for band pairs. The bar heights represent the fraction of disorder realizations (out of  $\sim 200$  trials) that have a given number ( $\#$ ) of band pairs with  $D_L=1$  in the occupied (half-filled) subspace. Those with an even number ( $\#$ ) will have an overall  $D_L=0$ , those with an odd number will have overall  $D_L=1$ . All these simulations were done with  $t=-1$ ,  $\lambda_R=\lambda_v=1$ , and  $\sigma_w=0.3$ . Reading across,  $\lambda_{SO}$  increases and the system transitions from all realizations having even parity at small  $\lambda_{SO}$  to odd parity at large  $\lambda_{SO}$ . Reading down, doubling the system size doubles the total number of Kramers pairs and roughly doubles the number of  $Z_2$ -odd pairs.

from such calculations, in their case the Chern numbers of Landau sublevels. Specifically, sublevels with nonzero Chern number contain extended states, which should only occur at isolated energies in the IQHE. Therefore, the number  $N_c$  of such sublevels should decrease as the system size  $N_s$  increases, and, in fact,  $\langle N_c \rangle \propto N_s^{1-1/2\nu}$ , where  $\langle \rangle$  indicates an average over disorder realizations. A similar approach for the quantum spin Hall (QSH) system here should reveal  $\langle N_D \rangle \propto N_s$ , where  $N_D$  is the number of  $Z_2$ -odd Kramers pairs, since we expect a stable metallic band of energies in the thermodynamic limit (as observed by Obuse *et al.*<sup>21</sup> and Onoda *et al.*<sup>16</sup>). With enough data and large enough systems, finite-size-induced broadening of the edges of this band should also make  $\nu$  accessible via a subleading term in the scaling. We find that larger system sizes require a finer mesh in twist space for these Kramers-pair-resolved simulations to return stable results, so that the requirements quickly outstrip our resources. Nevertheless, comparison of the two rows in Fig. 6 indicates that the location of the mean roughly doubles. This is what would be expected for the middle case given the above considerations; the mean for the topological insulator (the right-hand panels in Fig. 6) should grow more slowly, as in the case studied by Yang and Bhatt.

The total phase of the system, given by the total Chern parity, is more relevant to possible measurements than the Chern parity of each Kramers pair. The former maintains its meaning if we consider the ground state wave function of the many-electron system, formed as a Slater determinant of the single-electron states we use here. There is also a computational benefit to calculating  $D_\phi$  for the whole occupied subspace rather than for individual pairs—at larger system sizes, and also at stronger disorder, the link variables for the half-

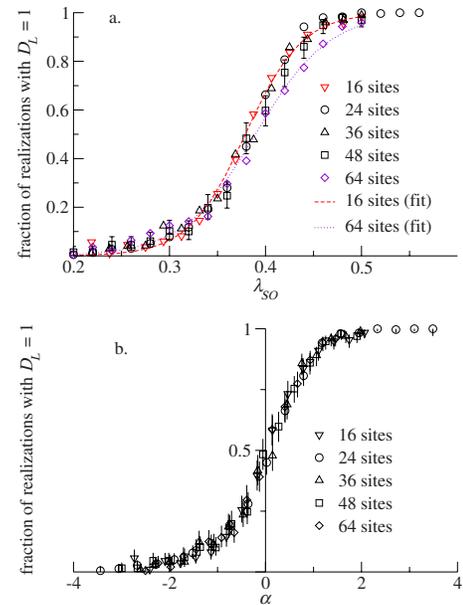


FIG. 7. (Color online) (a) At finite  $\lambda_R$ , the “metallic” region persists as the system size grows, and even broadens in the case shown here ( $t=-1$ ,  $\lambda_R=\lambda_v=\sigma_w=1$ ). We identify the metallic region as those values of  $\lambda_{SO}$  for which some, but not all, disorder realizations have  $D_L=1$ . As explained in the text, the fits to the simulation data have the form  $(\tanh \alpha + 1)/2$ , with  $\alpha = m(\lambda_{SO} - \lambda_{SO}^*)$ . (b) The scaling collapse of the data in (a), based on the best fit  $m$  and  $\lambda_{SO}^*$  for each system size. The error bars represent 95% confidence intervals assuming a binomial distribution of outcomes for each  $\lambda_{SO}$ .

filled subspace vary much more slowly than those for the individual Kramers pairs, making the calculation more robust. For these reasons, the remaining plots in this paper depict only the Chern parity of the half-filled system.

To show that a metallic region of nonzero extent in parameter space exists in the thermodynamic limit, we need to verify that the mixed-phase region does not shrink to zero as we increase the system size. Figure 7(a) shows that as we make the system larger, the transition region certainly does not get narrower, and, in fact, the largest system size seems to have the broadest transition.

We can quantify the scaling of the metallic region’s width with system size by assuming a simple one-parameter scaling form for the curves in Fig. 7(a) and defining the width of the curve to be proportional to the reciprocal of the maximum slope: width  $\sim 1/\text{slope}$ . With sufficient data, one could expand the approximately linear region near the middle of the transition in a power series with a few coefficients as fit parameters. Since our simulation data are limited, we opt instead to assume the form  $[\tanh \alpha(s) + 1]/2$ , which has roughly the right shape. If  $\alpha = m(\lambda_{SO} - \lambda_{SO}^*)$ , then  $m$  is exactly the maximum slope we want. Figure 7(b) plots the data versus the best-fit scaling variable  $\alpha$  for each system size; the points appear to fill out a smooth curve, justifying the scaling hypothesis. The best fit for  $m$  and  $\lambda_{SO}^*$  is determined by minimizing a weighted  $\chi^2$  statistic.<sup>36</sup> In particular, we assume that the results ( $D_L = \pm 1$ ) of independent simulations at a fixed parameter set are distributed binomially, and that for each

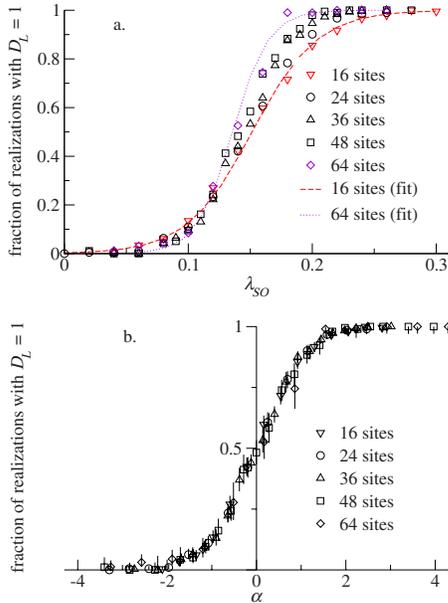


FIG. 8. (Color online) (a) At  $\lambda_R=0$ , the metallic region gets narrower as the system size increases. As in Fig. 7,  $t=-1$  and  $\lambda_v = \sigma_w=1$ . (b) The scaling variable  $\alpha$  again comes from fitting to a tanh, and error bars represent 95% confidence intervals.

system size there are the same number of trials at each value of  $\lambda_{SO}$  (which is roughly true). In that case, the variance of the distribution can be estimated as  $\sigma^2 \propto p(1-p)$ ,<sup>37</sup> where  $p$  is the fraction of disorder realizations returning  $D_L=1$ , and then  $1/\sigma^2$  is an appropriate weight for the statistical test. The error bars in Figs. 7 and 8 are also assigned based on a binomial model of the data.<sup>36</sup>

By contrast, at  $\lambda_R=0$ , the Hamiltonian (5) reduces to two copies of the Haldane model and so should exhibit the quantum Hall plateau transition, which looks like a step function at zero temperature in the thermodynamic limit. In Fig. 8(a), the width of the transition region shrinks as the system size grows, consistent with the prediction.

More quantitatively, Pruisken<sup>38</sup> has shown in a renormalization-group framework that the functional form of the crossover for the IQHE looks like

$$p(L, B) = f(\alpha), \quad \alpha \propto L^{1/\nu}(B - B^*), \quad (20)$$

where  $L$  is the linear size of the finite sample,  $B$  is the applied magnetic field, and  $\nu$  is again the localization length exponent. The function  $p$  could be either the longitudinal or transverse conductivity in the IQHE. Therefore,  $p(\infty, B^* \pm \varepsilon) = f(\pm\infty)$ , i.e., the transition is sharp in the thermodynamic limit. In our system, the analogous parameter to  $B$  is  $\lambda_{SO}$ : in the Haldane model, the spin-orbit coupling breaks time-reversal invariance locally, like  $B$  does in the IQHE. The width of the transition region in  $B$  is governed by the way the Landau level energies respond to changing  $B$ , and the width of the transition region in our model (for fixed  $\lambda_v$ ) is determined by the response of the gap to changing  $\lambda_{SO}$ . Since both responses are linear, we expect that the appropriate scaling variable will be  $\alpha \propto L^{1/\nu}(\lambda_{SO} - \lambda_{SO}^*)$ .

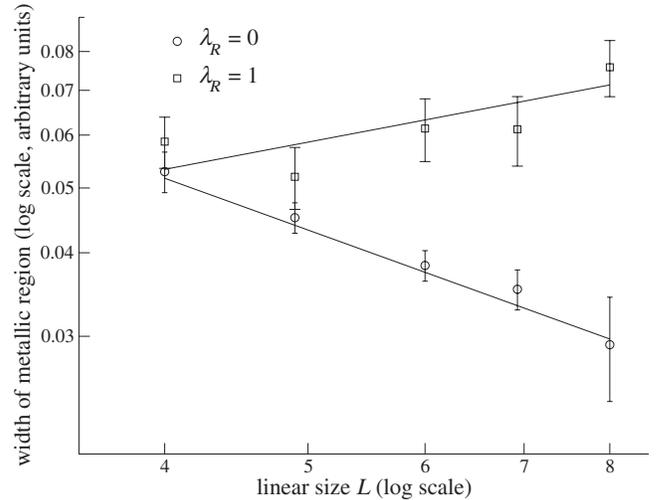


FIG. 9. Width of the metallic transition region as a function of linear size  $L$ . The width is determined by the reciprocal of the maximum slope of the simulation data in Figs. 7 and 8, and the linear size is taken to be the square root of the number of sites.

This form would allow us to extract the exponent  $\nu$  from the scaling of the maximum slope with system size for large systems (there are corrections at small system sizes). Again making the fit to a tanh described above, Fig. 8(b) shows the scaling form, and Fig. 9 shows the scaling of width with linear system size. In particular, a regression gives  $1/\nu = 0.78 \pm 0.03$ , to be compared with the accepted value of  $1/\nu \approx 0.42$ . That is far from good agreement, but the observed scaling should not be taken as implying a new universality class. (For reference, the network-model work by Obuse *et al.*<sup>21</sup> found  $1/\nu \approx 0.37$ , and Onoda *et al.*<sup>16</sup> recently found a value  $1/\nu \approx 0.63$ .) First, there should be finite-size corrections to the simple scaling assumed for the small systems considered here. Second, the scaling form is, in principle, different for different geometries, and the simulations were done for systems of varying aspect ratio (1–1.5). Nevertheless, it is clear that the qualitative behavior at  $\lambda_R=0$  is as expected, showing no metallic phase, and the behavior at  $\lambda_R=1$  is consistent with the presence of a metallic region.

Finally, by varying  $\lambda_R$  and noting the  $\lambda_{SO}$  values that mark the edges of the transition region for each  $\lambda_R$ , we can map out the phase diagram of the Hamiltonian (5) on a system of fixed size, as in Fig. 10. The widths obtained this way are in rough agreement with the scaling analysis outlined above, which returns the behavior of the width and not its normalization. As the finite-size scaling results in Figs. 7 and 8 show, this phase diagram overestimates the width of the metallic phase at  $\lambda_R=0$ , which is really zero. Together, these simulations confirm the expectation of Fig. 4 within the accuracy of our computational methods.

The same numerical methods could be used to study different forms of spin-orbit coupling quantitatively. The phase diagram of the model we have studied is reasonably simple in that, when the spin-independent part of the Hamiltonian is fixed and the spin-orbit couplings are not too strong, the intrinsic SO coupling pushes the system toward the topological insulator, and the Rashba coupling pushes the system

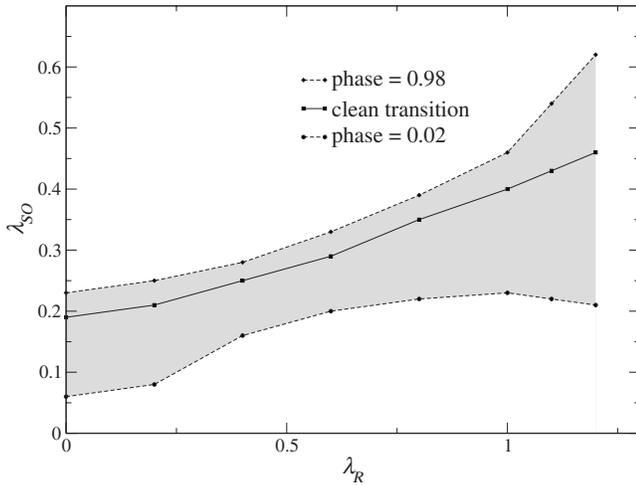


FIG. 10. Approximate width of metallic region for a  $4 \times 6$  lattice (fixed  $\lambda_v = \sigma_w = 1$ ). The dashed curves indicate the parameter values at which 98% and 2% of the disorder realizations have  $D_L = 1$ . This underestimates the true width of the metallic region but hopefully avoids some amount of the inevitable error due to small system size. This diagram should offer a reasonable approximation to the thermodynamic (infinite system) phase diagram away from  $\lambda_R = 0$ . At  $\lambda_R = 0$ , there should be no metallic phase but a sharp transition between the two insulating phases in the thermodynamic limit. Given the plots in Fig. 7, it appears that finite-size effects also reduce the width of the metallic phase at large  $\lambda_R$ .

toward the ordinary insulator. For more complicated forms of the spin-orbit coupling, the same analysis could be carried out to find the phase diagram, but it is not always possible to parametrize spin-orbit coupling with two parameters as for the graphene model. Also note that in an experiment, modifying the system using a perturbation such as a gate voltage is likely to modify both the spin-orbit and spin-independent parts of the Hamiltonian, so that a purely spin-orbit phase diagram would be insufficient. However, an important and potentially universal feature of the graphene phase diagram is that, *unless there is an extra conservation law* such as the  $s^z$  conservation when Rashba coupling is absent, a metallic region always appears between ordinary and topological insulators in the presence of disorder. The existence of this metallic region is an important difference between the spin Hall effect and the integer quantum Hall effect.

#### IV. SUMMARY

Previous work<sup>5,6,14,17,39</sup> defined a  $\mathbb{Z}_2$  topological invariant in infinite lattices that is similar to the TKNN invariant for the integer quantum Hall effect.<sup>26</sup> In disordered systems with boundaries, Fu and Kane<sup>17</sup> defined a topological invariant in terms of pumping of the occupancy of Kramers-degenerate

edge states. We have given a definition of a topological invariant valid for disordered systems without boundary, i.e., without appeal to edge states. The ‘‘Chern parity’’ can be thought of as describing either a finite system with boundary phases or an arbitrarily large supercell in an infinite lattice system with well defined wave vector. A physical effect of Chern parity is that it determines whether the amount of charge pumped in a certain type of closed pumping cycle is even or odd. The ordinary and topological insulator phases can be distinguished by this invariant as long as many-body effects do not prevent description of the ground state as a Slater determinant. Chern parity in the spin quantum Hall effect is the natural generalization of Chern number in the integer quantum Hall effect.

In a disordered system, the only degeneracies expected to survive are Kramers degeneracies at time-reversal-invariant values of the boundary phases. In this case, each pair of states related by Kramers degeneracies can be assigned its own Chern parity, and the overall Chern parity of all occupied state pairs determines the observable phase. The lattice algorithm for  $\mathbb{Z}_2$  topological invariants laid out by Fukui and Hatsugai<sup>19</sup> allows numerical identification of the topological insulator phase in disordered QSH systems. Implementing this algorithm for the specific graphene model Hamiltonian of Kane and Mele<sup>6</sup> with added on-site disorder, we observe the ordinary and topological insulator phases in simulations. While the number of odd pairs (state pairs with odd Chern parity) varies with the disorder realization, there is an even number of odd pairs in the ordinary insulator, and an odd number of odd pairs in the topological insulator.

We find that a metallic phase opens up between the two insulating phases for generic spin-orbit coupling. This agrees with the prediction of Hikami *et al.*<sup>20</sup> that spin-orbit coupling can protect a 2D metallic phase from disorder, and confirms the simulation results of Onoda *et al.*<sup>16</sup> and Obuse *et al.*<sup>21</sup> The methods in this paper could, in principle, be used to study the three-dimensional case and confirm the argument in Ref. 40 that only one of the four invariants of a band structure<sup>14,27,28</sup> is stable to disorder. While there is now strong evidence<sup>21</sup> that the phase transitions in the 2D QSHE are, except for special points, in the previously studied symplectic metal-insulator class, there is as yet no numerical study of the phase transitions in three-dimensional topological insulators.

#### ACKNOWLEDGMENTS

The authors thank L. Balents, H. Haggard, D.-H. Lee, C. L. Kane, and A. Vishwanath for useful comments, and acknowledge support from NSF DMR-0238760. Numerical work took place on a cluster donated by the IBM SUR program.

<sup>1</sup>S. Murakami, N. Nagaosa, and S.-C. Zhang, *Science* **301**, 1348 (2003).

<sup>2</sup>J. Sinova, D. Culcer, Q. Niu, N. A. Sinitsyn, T. Jungwirth, and A. H. MacDonald, *Phys. Rev. Lett.* **92**, 126603 (2004).

<sup>3</sup>Y. K. Kato, R. C. Myers, A. C. Gossard, and D. D. Awschalom, *Science* **306**, 1910 (2004).

<sup>4</sup>J. Wunderlich, B. Kaestner, J. Sinova, and T. Jungwirth, *Phys. Rev. Lett.* **94**, 047204 (2005).

<sup>5</sup>C. L. Kane and E. J. Mele, *Phys. Rev. Lett.* **95**, 226801 (2005).

<sup>6</sup>C. L. Kane and E. J. Mele, *Phys. Rev. Lett.* **95**, 146802 (2005).

<sup>7</sup>M. I. D'yakonov and V. I. Perel, *JETP Lett.* **13**, 467 (1971).

<sup>8</sup>F. D. M. Haldane, *Phys. Rev. Lett.* **61**, 2015 (1988).

<sup>9</sup>Y. A. Bychkov and É. I. Rashba, *JETP Lett.* **39**, 78 (1984).

<sup>10</sup>D. Huertas-Hernando, F. Guinea, and A. Brataas, *Phys. Rev. B* **74**, 155426 (2006).

<sup>11</sup>H. Min, J. E. Hill, N. A. Sinitsyn, B. R. Sahu, L. Kleinman, and A. H. MacDonald, *Phys. Rev. B* **74**, 165310 (2006).

<sup>12</sup>Y. Yao, F. Ye, X.-L. Qi, S.-C. Zhang, and Z. Fang, *Phys. Rev. B* **75**, 041401(R) (2007).

<sup>13</sup>B. A. Bernevig, T. L. Hughes, and S.-C. Zhang, *Science* **314**, 1757 (2006).

<sup>14</sup>J. E. Moore and L. Balents, *Phys. Rev. B* **75**, 121306(R) (2007).

<sup>15</sup>D. N. Sheng, Z. Y. Weng, L. Sheng, and F. D. M. Haldane, *Phys. Rev. Lett.* **97**, 036808 (2006).

<sup>16</sup>M. Onoda, Y. Avishai, and N. Nagaosa, *Phys. Rev. Lett.* **98**, 076802 (2007).

<sup>17</sup>L. Fu and C. L. Kane, *Phys. Rev. B* **74**, 195312 (2006).

<sup>18</sup>T. Fukui and Y. Hatsugai, *Phys. Rev. B* **75**, 121403(R) (2007).

<sup>19</sup>T. Fukui and Y. Hatsugai, *J. Phys. Soc. Jpn.* **76**, 053702 (2007).

<sup>20</sup>S. Hikami, A. I. Larkin, and Y. Nagaoka, *Prog. Theor. Phys.* **63**, 707 (1980).

<sup>21</sup>H. Obuse, A. Furusaki, S. Ryu, and C. Mudry, *Phys. Rev. B* **76**, 075301 (2007).

<sup>22</sup>J. T. Chalker and P. D. Coddington, *J. Phys. C* **21**, 2665 (1988).

<sup>23</sup>R. B. Laughlin, *Phys. Rev. Lett.* **52**, 2304 (1984).

<sup>24</sup>D. E. Khmel'nitskii, *Phys. Lett.* **106A**, 182 (1984).

<sup>25</sup>K. Yang and R. N. Bhatt, *Phys. Rev. Lett.* **76**, 1316 (1996).

<sup>26</sup>D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs, *Phys. Rev. Lett.* **49**, 405 (1982).

<sup>27</sup>R. Roy, arXiv:cond-mat/0607531 (unpublished).

<sup>28</sup>L. Fu, C. L. Kane, and E. J. Mele, *Phys. Rev. Lett.* **98**, 106803 (2007).

<sup>29</sup>B. A. Bernevig and H.-D. Chen, arXiv:cond-mat/0611766 (unpublished).

<sup>30</sup>J. E. Avron, R. Seiler, and B. Simon, *Phys. Rev. Lett.* **51**, 51 (1983).

<sup>31</sup>J. J. Sakurai, *Modern Quantum Mechanics* (Addison-Wesley, 1994), Chap. 4.4.

<sup>32</sup>Q. Niu, D. J. Thouless, and Y.-S. Wu, *Phys. Rev. B* **31**, 3372 (1985).

<sup>33</sup>J. E. Avron and R. Seiler, *Phys. Rev. Lett.* **54**, 259 (1985).

<sup>34</sup>Q. Niu and D. J. Thouless, *Phys. Rev. B* **35**, 2188 (1987).

<sup>35</sup>S. Hikami, *Prog. Theor. Phys. Suppl.* **107**, 213 (1992).

<sup>36</sup>W.-M. Yao *et al.*, *J. Phys. G* **33**, 1 (2006).

<sup>37</sup>G. D'Agostini, *Rep. Prog. Phys.* **63**, 1383 (2003).

<sup>38</sup>A. M. M. Pruisken, *Phys. Rev. Lett.* **61**, 1297 (1988).

<sup>39</sup>R. Roy, arXiv:cond-mat/0604211 (unpublished).

<sup>40</sup>L. Fu and C. L. Kane, *Phys. Rev. B* **76**, 045302 (2007).

<sup>41</sup>Throughout "parity" is used to denote oddness or evenness, rather than a spatial inversion eigenvalue.

<sup>42</sup>Note that this means  $\mathcal{H}$  is not generically time reversal invariant. Indeed, there are only four boundary conditions that respect time reversal, namely, those for which  $\psi$  picks up a real phase upon translation around each cycle of the honeycomb lattice. These four boundary conditions correspond to the TRIM of the clean system in the calculation of  $D$ .

<sup>43</sup>The points  $(\phi_x, \pi)$  and  $(\phi_x, -\pi)$  are not physically distinct; as pointed out earlier, a gauge transformation relates  $\mathcal{H}(\phi_x, -\pi)$  to  $\mathcal{H}(\phi_x, \pi)$  as given, for example. We, therefore, impose  $\mathcal{H}(\phi_x, -\pi) \equiv \mathcal{H}(\phi_x, \pi)$  in the calculation. Alternatively, we could introduce boundary terms along  $\phi_y = \pi, -\pi$  to compensate for the discrepancy.