Computing topological invariants without inversion symmetry

Alexey A. Soluyanov[*] and David Vanderbilt[†]
Department of Physics and Astronomy, Rutgers University, Piscataway, New Jersey 08854-0849, USA

(Dated: March 1, 2011)

We consider the problem of calculating the weak and strong topological indices in noncentrosymmetric time-reversal ($\mathcal{T}$) invariant insulators. In 2D we use a gauge corresponding to hybrid Wannier functions that are maximally localized in one dimension. Although this gauge is not smoothly defined on the two-torus, it respects the $\mathcal{T}$ symmetry of the system and allows for a definition of the $\mathbb{Z}_2$ invariant in terms of time-reversal polarization. In 3D we apply the 2D approach to $\mathcal{T}$-invariant planes. We illustrate the method with first-principles calculations on GeTe and on HgTe under [001] and [111] strain. Our approach differs from ones used previously for noncentrosymmetric materials and should be easier to implement in *ab initio* code packages.

PACS numbers: 73.43.Cd, 03.65.Vf, 71.20.Nr, 71.70.Ej

I. INTRODUCTION

A series of theoretical developments starting in 2005, showing that non-magnetic insulators admit a topological $\mathbb{Z}_2$ classification in two dimensions (2D)[1,2] and then in three dimensions (3D),[3,4] has sparked enormous interest, especially after numerous realizations of such systems were confirmed both theoretically[5–9] and experimentally.[10–13] These developments, nicely summarized in some recent reviews,[14–16] have essentially given rise to a new subfield of condensed-matter physics, with the topology of the band structure now regarded as a fundamental characteristic of the electronic ground state for semiconductors and insulators.

The $\mathbb{Z}_2$ classification divides time-reversal ($\mathcal{T}$) invariant band insulators into two classes: ordinary ($\mathbb{Z}_2$-even) insulators that can be adiabatically converted to the vacuum (or to each other) without a bulk gap closure, and "topological" ($\mathbb{Z}_2$-odd) ones that cannot be so connected (although they can be adiabatically connected to each other). Even and odd phases are separated by a topological phase transition, and the bulk gap has to vanish at the transition point, at least in a non-interacting system.[17,18] The $\mathbb{Z}_2$-odd states are characterized by the presence of an odd number of Kramers pairs of counter-propagating edge states in 2D, or by an odd number of Fermi loops enclosing certain high-symmetry points of the surface band structure in 3D.

In view of all this, there is an obvious motivation to develop simple yet effective methods for computing the topological indices of a given material. For centrosymmetric crystals, a convenient method was introduced in Ref. 6, where it was shown that the knowledge of the parity eigenvalues of the electronic states at only four $\mathcal{T}$-invariant momenta in 2D (or eight of them in 3D) is sufficient to compute the topological characteristics of a given material. This approach is limited to centrosymmetric systems, however, and the calculation of the $\mathbb{Z}_2$ invariant for noncentrosymmetric insulators is not so trivial.

One possible approach, suggested in Ref. 19, is based on the existence of a topological obstruction to choosing a smooth gauge that respects the $\mathcal{T}$ symmetry in the $\mathbb{Z}_2$-odd case. For the implementation of this method, a gauge must be chosen on the boundary of half of the Brillouin zone (BZ) in such a way as to respect $\mathcal{T}$ symmetry, which involves acting with the time-reversal operator on one of the states from each Kramers pair to construct the other. Although this method has been implemented in the *ab initio* framework[20,21], its implementation is basis-set dependent and involves the application of a unitary rotation to the computed eigenvectors when fixing the gauge, which may be tedious when there are many occupied bands and basis states.

Another existing method[7] relies on the fact that the system will necessarily be in the $\mathbb{Z}_2$-even (normal) state in the absence of spin-orbit (SO) coupling. In this method, the strength of the SO coupling is artificially tuned from $\lambda_{SO} = 0$ (no SO coupling) to $\lambda_{SO} = 1$ (full SO coupling), and a closure of the band gap at some intermediate coupling strength is taken as evidence of an inverted band structure. However, a closure of the band gap in the course of tuning λ_{SO} to full strength is a necessary, but not a sufficient, condition for a topological phase transition. Therefore, in order to determine whether the system is really in the topologically nontrivial phase, a first-principles calculation of the surface states is carried out in order to count the number of Dirac cones at the surface of the candidate material. Such a calculation, although illustrative, is quite demanding in terms of computational resources.

In summary, existing methods have some shortcomings, and it would be very useful to develop a simple and effective method that would use the electronic wavefunctions, as obtained directly from the diagonalization procedure, to determine the desired topological indices.

In this paper we develop a method for computing $\mathbb{Z}_2$ invariants that meets these criteria, and which is easy to implement in the context of *ab initio* code packages. The method is based on the concept of time-reversal polarization[22] (TRP), but implemented in such a way that a visual inspection of plotted curves is not required in order to obtain the topological indices. Instead, all the indices can be obtained directly as a result of an automated calculation. We describe the method, and then verify it using centrosymmetric Bi and Bi_2Se_3 as illustrative test examples before applying it to the more difficult cases of noncentrosymmetric GeTe and strained HgTe.

The paper is organized as follows. In Sec. II we start by reviewing the formalism of TRP in the context of the $\mathbb{Z}_2$ spin pump in one dimension (1D), emphasizing its relation to the charge centers of Wannier functions. We then discuss the numerical implementation of these ideas to 2D and 3D cases, and suggest a simple numerical procedure for calculating the $\mathbb{Z}_2$ invariant in noncentrosymmetric $\mathcal{T}$-invariant systems in Sec. III. We further illustrate this method with *ab initio* calculations in Sec. IV, and present some concluding remarks in Sec. V.

II. $\mathbb{Z}_2$ INVARIANT VIA WANNIER CHARGE CENTERS

In this section we review the notion of TRP and the definition of the $\mathbb{Z}_2$ invariant in terms of TRP derived in Ref. 22. The definition arises by virtue of an analogy between a 2D $\mathcal{T}$-invariant insulator and a $\mathcal{T}$-symmetric pumping process in a 1D insulator. We further reformulate this definition in terms of Wannier charge centers, setting the stage for the numerical method discussed in the next section.

A. Review of time reversal polarization

Fu and Kane[22] considered a family of 1D $\mathcal{T}$-invariant bulk-gapped Hamiltonians $H(t)$ parametrized by a cyclic parameter t (i.e., $H[t+T]=H[t]$) subject to the constraint

$$H[-t]=\theta H[t]\theta^{-1}, \tag{1}$$

where θ is the time-reversal operator. This can be understood as a pumping cycle, with t being the pumping parameter. The constraint guarantees that the Hamiltonian is $\mathcal{T}$-invariant at the points $t=0$ and $t=T/2$, while the $\mathcal{T}$-symmetry is broken at the other values of the parameter. If we also limit ourselves to Hamiltonians having unit period, so that H is invariant under $x\to x+1$, then the eigenstates may be represented by the periodic parts $|u_{nk}\rangle = e^{-ikx}|\psi_{nk}\rangle$ of the Bloch states $|\psi_{nk}\rangle$. At $t=0$ and $t=T/2$ the Hamiltonian is time-reversal invariant and the eigenstates come in Kramers pairs, being degenerate at $k=0$ and $k=\pi$.

Since the system is periodic in both k and t, the $|u_{nk}\rangle$ functions are defined on a torus. Moreover, the system must also be physically invariant under a gauge transformation of the form

$$|\tilde{u}_{nk}\rangle = \sum_m U_{mn}|u_{mk}\rangle \tag{2}$$

where $U(k,t)$ expresses the $\mathcal{U}(\mathcal{N})$ gauge freedom to choose $\mathcal{N}$ representatives of the occupied space at each (k,t). We adopt a gauge that is continuous on the half-torus $t\in[0,T/2]$ and that respects $\mathcal{T}$ symmetry at $t=0$ and $T/2$ in the sense of Fu and Kane,[22] i.e.,

$$\begin{aligned} |u^{I}_{\alpha,-k}\rangle &= -e^{i\chi_{\alpha k}}\theta|u^{II}_{\alpha k}\rangle, \\ |u^{II}_{\alpha,-k}\rangle &= e^{i\chi_{\alpha,-k}}\theta|u^{I}_{\alpha k}\rangle. \end{aligned} \tag{3}$$

Here the occupied states $n=1,...,\mathcal{N}$ have been relabeled in terms of pairs $\alpha=1,...,\mathcal{N}/2$ and elements I and II within each pair. Note that Eq. (3) is a property which is not preserved by an arbitrary $\mathcal{U}(\mathcal{N})$ transformation. It allows the Berry connection

$$\mathcal{A}(k)=i\sum_n\langle u_{nk}|\partial_k|u_{nk}\rangle \tag{4}$$

to be decomposed as

$$\mathcal{A}(k)=\mathcal{A}^I(k)+\mathcal{A}^{II}(k) \tag{5}$$

where

$$\mathcal{A}^S(k)=i\sum_\alpha\langle u^S_{\alpha k}|\partial_k|u^S_{\alpha k}\rangle \tag{6}$$

and $S=I,II$. Having chosen a gauge that obeys these conventions at $t=0$ and $T/2$ and evolves smoothly for intermediate t,[23] the "partial polarizations"[22]

$$P^S_\rho=\frac{1}{2\pi}\oint dk\mathcal{A}^S(k) \tag{7}$$

can be defined such that their sum is the total charge polarization[24]

$$P_\rho=\frac{1}{2\pi}\oint dk\mathcal{A}(k)=P^I_\rho+P^{II}_\rho. \tag{8}$$

Note that the total polarization is defined only modulo an integer (the quantum of polarization) under a general $U(\mathcal{N})$ gauge transformation, while the "partial polarization" is not gauge invariant at all. A quantity that *is* gauge-invariant is the change in total polarization during the cyclic adiabatic evolution of the Hamiltonian, and using Eq. (1) it follows that

$$P_\rho(T)-P_\rho(0)=C \tag{9}$$

where C is the first Chern number, an integer topological invariant corresponding to the number of electrons pumped through the system in one cycle of the pumping process.[25] For a $\mathcal{T}$-invariant pump that satisfies the conditions of Eq. (1), C must be zero.

In order to describe the $\mathbb{Z}_2$ invariant of a $\mathcal{T}$-symmetric system in a similar fashion, the "time reversal polarization" was introduced as[22]

$$P_\theta=P^I_\rho-P^{II}_\rho. \tag{10}$$

Then the integer $\mathbb{Z}_2$ invariant can be written as

$$\Delta=P_\theta(T/2)-P_\theta(0)\mod 2. \tag{11}$$

To summarize, the $\mathbb{Z}_2$ invariant is well defined via Eq. (11) when the gauge respects $\mathcal{T}$-symmetry at $t = 0$ and $T/2$ and is continuous on the torus between these two parameter values. Note, however, that while such a gauge choice is possible on the half-torus even for the $\mathbb{Z}_2$-odd case (Δ=1), it can only be extended to cover the full torus continuously in the $\mathbb{Z}_2$-even case (Δ=0).[26]

B. Formulation in terms of Wannier charge centers

Let us now rewrite Eq. (11) in terms of the Wannier charge centers (WCCs). By definition the Wannier functions (WFs) belonging to unit cell R are

$$|Rn\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} dk e^{-ik(R-r)}|u_{nk}\rangle, \tag{12}$$

Adopting again the notation of Eq. (3), we note that at $t = 0$ and $t = T/2$ the WFs will also appear in pairs obeying

$$\bar{x}^I_\alpha = \bar{x}^{II}_\alpha \mod 1. \tag{13}$$

This follows because $\mathcal{T}$ maps the space spanned by WF $|R\alpha, I\rangle$ and its periodic images into the space spanned by $|R\alpha, II\rangle$ and its periodic images. Since we have also insisted on the gauge being continuous for $t \in [0, T/2]$, it is possible to follow the evolution of each WCC during the half-cycle. Taking into account that $\sum_\alpha \bar{x}^s_\alpha = (1/2\pi)\oint_{\mathrm{BZ}} \mathcal{A}^S$ for $S = I, II$, Eq. (11) yields

$$\Delta = \sum_\alpha \left[\bar{x}^I_\alpha(T/2) - \bar{x}^{II}_\alpha(T/2)\right] - \sum_\alpha \left[\bar{x}^I_\alpha(0) - \bar{x}^{II}_\alpha(0)\right]. \tag{14}$$

Since the gauge is assumed to be smooth, the evolution of the charge centers must also be smooth. Being defined in this way, Δ is clearly a mod-2 quantity, and as shown in Ref. 22 it represents the desired $\mathbb{Z}_2$ invariant.

However, if the gauge breaks $\mathcal{T}$ symmetry or it is not continuous in the half-cycle, Eq. (14) no longer defines a topological invariant. A discontinuity in the gauge in the process of the half cycle can change Δ by 1, so the mod-2 property is lost. Breaking $\mathcal{T}$ in the gauge choice means that the corresponding centers are not necessarily degenerate at $t = 0$ and $t = T/2$. In fact, Δ can even take non-integer values in this case.

The above argument implies that in order to compute the $\mathbb{Z}_2$ invariant via Eq. (14), one needs a gauge that satisfies both $\mathcal{T}$-invariance and continuity on the half-torus. We now argue that the gauge that corresponds to 1D maximally localized WFs at each t has the desired properties. The criterion introduced in Ref. 27 for constructing the maximally localized WFs was that the gauge choice should provide the minimum possible quadratic spread $\Omega = \sum_n[\langle 0n|\mathbf{r}^2|0n\rangle - \langle 0n|\mathbf{r}|0n\rangle^2]$. In 1D, the maximally localized WFs constructed according to this criterion are eigenstates of the position operator $\hat{X}$ in the band subspace.[27,28] Since this operator commutes with θ, its eigenvalues will be doubly degenerate and its eigenstates will come in Kramers pairs at $t = 0$ and $T/2$.

Let us briefly discuss how to construct this gauge on a k-mesh $k_{j+1} = k_j + \Delta k$ by carrying out a multiband parallel-transport construction along the Brillouin zone.[27] At a given value of t, starting from $k = 0$ one constructs overlap matrices $M^{(k_j,k_{j+1})}_{mn} = \langle u_{mk_j}|u_{nk_{j+1}}\rangle$ in such a way that they are Hermitian. This can be done in a unique way by means of the singular value decomposition $M = V\Sigma W^\dagger$, where Σ is positive real diagonal and V and W are unitary matrices. With this decomposition a unitary rotation of the states at k_{j+1} by $WV^\dagger$ leaves $M^{(k_j,k_{j+1})}$ Hermitian. Repeating this procedure, one finds that states $|\psi_{nk}\rangle$ at $k = 2\pi$ are related to those at $k = 0$ by a unitary rotation Λ, whose eigenvalues $\lambda_n = e^{-i\bar{x}_n}$ give the 1D maximally-localized WCCs $\bar{x}_n$. The corresponding eigenvectors can be used to define a gauge that is continuous on the half-torus $0 \le t \le T/2$.

Now, having established a particular gauge choice in which Eqs. (11) and (14) are valid, it is straightforward in principle to obtain the $\mathbb{Z}_2$ invariant. In practice, when working on a discrete mesh of t values when many bands are present, it may not be entirely straightforward to enforce the continuity with respect to t. In the next section we present a simple and automatic numerical procedure that is robust in this respect, and use it to illustrate the calculation of the $\mathbb{Z}_2$ invariants for several materials of interest.

III. NUMERICAL IMPLEMENTATION

The method outlined above, in which the WCCs obtained with the 1D maximally-localized gauge are used to compute the $\mathbb{Z}_2$ invariant via Eq. (14), can be implemented by plotting the WCCs at each point on the t mesh and then visually tracking the evolution of each WCC, as we describe next in Sec. III A. However, we find that a more straightforward and more easily automated approach is to track the *largest gap* in the spectrum of WCCs instead. This gives rise to our proposed method, which is described in Sec. III B.

A. Tracking WCC locations

Let us first interpret Eq. (14) in terms of the winding of the WCCs around the BZ during the half-cycle $t \in [0, T]$. Since the WCCs are defined modulo 1, one can imagine the $\bar{x}_n$ living on a circle of unit circumference, as illustrated in the left panels of Fig. 1. During the pumping process, the WCCs migrate along this circle. The system will be in the $\mathbb{Z}_2$-odd state (δ=1) if and only if the WCCs flow during the pumping cycle in such a way as to wind around the circle an odd number of times.

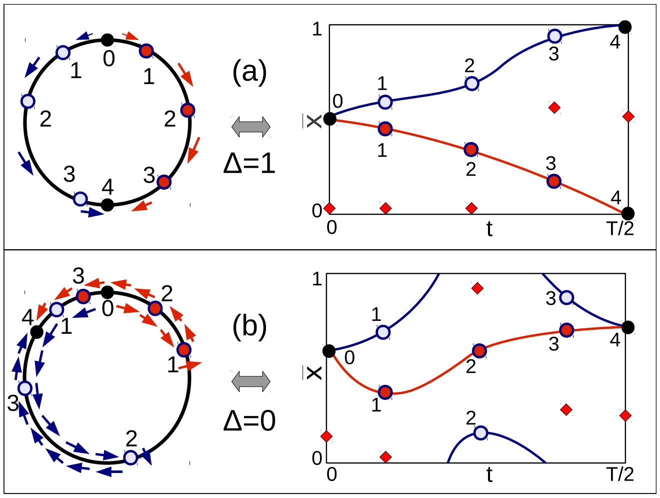

FIG. 1. Sketch of evolution of Wannier charge centers (WCCs) $\bar{x}$ vs. time t during an adiabatic pumping process. Regarding $\bar{x} \in [0, 1]$ as a unit circle and $t \in [0, T/2]$ as a line segment, the cylindrical $(\bar{x}, t)$ manifold is represented via a sequence of circular cross sections at left, or as an unwrapped cylinder at right. Each red rhombus marks the middle of the largest gap between WCCs at given t. (a) $\mathbb{Z}_2$ insulator; WCCs wind around the cylinder. (b) Normal insulator; WCCs reconnect without wrapping the cylinder.

Consider, for example, the case of only two occupied bands, as sketched in Fig. 1. The top panel shows the $\mathbb{Z}_2$-odd case; the blue and red arrows show the evolution of the first and second WCC from t_0 $(=0)$ to t_4 $(=T/2)$, and they meet in such a way that the unit circle is wrapped exactly once. Correspondingly, as shown in the right-hand part of the figure, the WCCs "exchange partners" during the pumping process.[22] For the $\mathbb{Z}_2$-even case shown in the bottom panel, by contrast, the unit circle is wrapped zero times, and no such exchange of partners occurs.

There is a problem, however. In Fig. 1 we show the evolution of the WCCs as the solid blue and red curves in the panels at right; with this connectivity information at hand, it is trivial to determine the winding number. However, in practice one will typically have the WCC values only on a discrete mesh of t points, in which case the connectivity can be far from obvious. Certainly one cannot simply make the arbitrary branch cut choice $\bar{x}_n \in [0, 1]$, sort the $\bar{x}_n$ in increasing order, and use the resulting indices to define the paths of the WCCs. This would, for example, give an incorrect evolution from t_1 to t_2 in Fig. 1(b), since one WCC passes through the branch cut in this interval, apparently jumping discontinuously from the "top" to the "bottom" of the unwrapped rectangle at right. (A similar jump happens again near t_3.)

One possible approach is to increase the t-mesh density until, by visual inspection, the connectivity becomes obvious. This can work up to a point, but it can become very difficult to implement in the case of many WCCs. As we shall see below, in first-principles calculations one can easily have 10-30 occupied bands, not just two. In this case, it is typical for some of the WCCs to cluster rather closely together during part of the evolution in t. If this clustering happens near the artificial branch cut, it can become very difficult to determine the connectivity from one t to the next, even if a rather dense mesh of t values is used. Moreover, an algorithm of this kind is difficult to automate. For these reasons, we find that the direct approach of plotting the evolution of the WCCs is not a very satisfactory algorithm for obtaining the topological indices, at least in the case of a large number of occupied bands.

B. Tracking gaps in the WCC spectrum

Here we propose a simple procedure that overcomes the above obstacles, allowing the $\mathbb{Z}_2$ invariant to be computed in a straightforward fashion. The main idea is to concentrate on the *largest gap between WCCs*, instead of on the individual WCCs themselves. This approach can be implemented without reference to any branch cut in the determination of the $\bar{x}_n$, allowing the $\mathbb{Z}_2$ invariant to be determined from the flow of WCCs on the cylindrical $(\bar{x}, t)$ manifold directly.

As in Fig. 1, we again consider a set of M circular sections of the cylinder that correspond to the pumping parameter values $t^{(m)} = T(m-1)/2M$, where $m \in [0, M]$. At each t_m we define $z^{(m)}$ to be the center of the largest gap between two adjacent WCCs on the circle. (If two gaps are of equal size, either can be chosen arbitrarily.) For definiteness we choose $z^{(m)} \in [0, 1)$, but as we shall see shortly, the branch choice is immaterial. In the continuous limit $M \to \infty$, $z(t)$ takes the form of a series of path segments on the surface of the cylinder, with discontinuous jumps in the $\bar{x}$ direction at certain critical parameter values t_j. Our algorithm consists in counting the number of WCCs jumped over at each t_j, and summing them all mod 2. As becomes clear from an inspection of Fig. 1 and similar examples of increasing complexity, the WCCs exchange partners during the evolution from t=0 to $T/2$ only if this sum is odd, so that this sum determines the $\mathbb{Z}_2$ invariant of the system.

The approach generalizes easily to the case of discrete $z^{(m)}$. Let Δ_m be the number of WCCs $\bar{x}_n^{(m+1)}$ that appear between gap centers $z^{(m)}$ and $z^{(m+1)}$, mod 2. As we shall see below, this can be computed in a manner that is independent of the branch cut choices used to determine the $\bar{x}_n^m$ and $z^{(m)}$. Then the overall $\mathbb{Z}_2$ invariant is just

$$\Delta = \sum_{m=0}^{M} \Delta_m \mod 2. \tag{15}$$

This argument is illustrated in the right-hand panels of Fig. 1 for the two band-case and $M = 4$. The rectangles represent the surface of the cylinder in the parameter space, and should be regarded as glued along the longer sides. The circles correspond to $\bar{x}_n^{(m)}$ values, while each red rhombus represents the center $z^{(m)}$ of the largest gap between $\bar{x}_n^{(m)}$ values. In Fig. 1(a) there is one jump that occurs between m=2 and m=3, in which one WCC is jumped over; thus, $\Delta_m = 0$ except for $\Delta_2 = 1$, giving Δ=1. In Fig. 1(b), on the other hand, there are two jumps, once between m=1 and m=2 and again between m=2 and m=3, so that $\Delta_1 = \Delta_2 = 1$ and $\Delta = 0$ (mod 2).

We now show how the Δ_m can be computed straightforwardly in a manner that is insensitive to the branch-cut choices made in determining the $\bar{x}_n^m$ and $z^{(m)}$. We use the fact that the directed area of a triangle defined by angles ϕ_1, ϕ_2, and ϕ_3 on the unit circle is[29]

$$g(\phi_1,\phi_2,\phi_3) = \sin(\phi_2-\phi_1)+\sin(\phi_3-\phi_2)+\sin(\phi_1-\phi_3). \tag{16}$$

Therefore the sign of $g(\phi_1,\phi_2,\phi_3)$ tells us whether or not ϕ_3 lies "between" ϕ_1 and ϕ_2 in the sense of counterclockwise rotation. Identifying $\phi_1 = 2\pi z^{(m)}$, $\phi_2 = 2\pi z^{(m+1)}$ and $\phi_3 = 2\pi\bar{x}_n^{(m+1)}$, it follows that

$$(-1)^{\Delta_m} = \prod_{n=1}^{\mathcal{N}} \mathrm{sgn}\left[g(2\pi z^{(m)}, 2\pi z^{(m+1)}, 2\pi\bar{x}_n^{(m+1)})\right], \tag{17}$$

where sgn(x) is the sign function. The Δ_m defined in this way is precisely the needed count of WCCs jumped over, mod 2, in evolving from m to $m+1$.

As a last detail, we discuss the case of possible degeneracies between the three arguments of $g(\phi_1,\phi_2,\phi_3)$. First, note that $z^{(m+1)} = \bar{x}_n^{(m+1)}$ is impossible, since $z^{(m+1)}$ is by definition in a gap between $\bar{x}_n^{(m+1)}$ values. If the mesh spacing in t is fine enough, then by continuity we expect that $z^{(m)} = \bar{x}_n^{(m+1)}$ will also be unlikely. It is recommended to test whether these values ever approach within a threshold distance, and restart the algorithm with a finer t mesh if such a case is encountered; two cases of this kind are discussed later in Sec. IV. Finally, it can happen that $z^{(m)} = z^{(m+1)}$. In this case, the signum function (which technically assigns value 0 to argument 0) should be replaced in Eq. (17) by a function that returns s whenever $z^{(m)} = z^{(m+1)}$, where s is chosen once and for all to be either +1 or −1. Since the same degeneracy appears in every term of the product over $\mathcal{N}$ factors in Eq. (17), where $\mathcal{N}$ is even, the choice of s is arbitrary as long as it is applied consistently.

The above-described algorithm, based on Eqs. (15-17), constitutes one of the principal results of the present work. The implementation of this algorithm is straightforward, and allows for an efficient and robust determination of the $\mathbb{Z}_2$ invariant even when many bands are present, and even for only moderately fine mesh spacings. In Sec. IV, we will demonstrate the successful application of this approach to the calculation of the strong and weak topological indices of some real materials.

C. Application to 2D and 3D $\mathcal{T}$-invariant insulators

As pointed out in Ref. (22), the pumping process discussed above for a 1D system is the direct analogue of a 2D $\mathcal{T}$-invariant insulator, i.e., one whose Hamiltonian is subject to the condition $H(-\mathbf{k}) = \theta^{-1}H(\mathbf{k})\theta$. To see this, let $\mathbf{k} = \sum_i k_i\mathbf{b}_i/2\pi$, where $\mathbf{b}_1$ and $\mathbf{b}_2$ have been chosen as primitive reciprocal lattice vectors. Then we can let k_1 and k_2 play the roles of k and t respectively. Just as $H(k,t)$ displays $\mathcal{T}$ symmetry at t=0 and T, so $H(k_1,k_2)$, regarded as the Hamiltonian $H(k_1)$ of a fictitious 1D system for given k_2, is $\mathcal{T}$-invariant at $k_2 = 0$ and π. The Wannier functions of the effective 1D system can be understood as "hybrid Wannier functions" that have been Fourier transformed from k space to r space only in direction 1, while remaining extended in direction 2. The topological $\mathbb{Z}_2$ invariant of the 2D system can therefore be determined straightforwardly by applying the approach outlined above.

A topological phase of a 3D $\mathcal{T}$-symmetric insulator is described by one strong topological index ν_0 and three weak indices ν_1, ν_2, and ν_3.[3,4,30] These indices may be understood as follows. Again letting $\mathbf{k} = \sum_i k_i\mathbf{b}_i/2\pi$, there are eight $\mathcal{T}$-invariant points $\Gamma_{(n_1,n_2,n_3)}$, where $n_i = 0$ or 1 denotes $k_i = 0$ or π respectively. These eight points may be thought of as the vertices of a parallelogram in reciprocal space whose six faces are labeled by n_1=0, n_2=0, n_3=0, n_1=1, n_2=1, and n_3=1. On any one of these six faces, the Hamiltonian $H(\mathbf{k})$, regarded as a function of two k variables, can be thought of as the Hamiltonian of a fictitious 2D $\mathcal{T}$-symmetric system, and the argument of the previous paragraph can thus be applied to each of these six faces separately. The three weak indices $\nu_{i=1,2,3}$ are defined to be the $\mathbb{Z}_2$ invariants associated with the three surfaces n_1=1, n_2=1, and n_3=1.[3] These weak indices obviously depend on the choice of reciprocal lattice vectors. The strong index ν_0 is the sum (mod 2) of the $\mathbb{Z}_2$ invariants of the n_j=0 and n_j=1 faces for any one of the j (implying some redundancy among the six indices); it is also a $\mathbb{Z}_2$ quantity, but is independent of the choice of reciprocal lattice vectors.[3,4]

Thus, a complete topological classification in 3D, given by the index $\nu_0;(\nu_1\nu_2\nu_3)$, can be obtained by applying our analysis to each of these six faces in the 3D Brillouin zone. Note that in general, this determines the strong index ν_0 with some redundancy, providing a check on the internal consistency of the method. However, symmetry considerations often play a role. For systems having a 3-fold symmetry axis, for example, one typically needs to compute the $\mathbb{Z}_2$ index on only two faces, as we shall see below.

IV. APPLICATION TO REAL MATERIALS

In this section we discuss the application of the above-described method to real materials. First, we illustrate the validity of the approach for centrosymmetric Bi and

Bi_2Se_3, where weak and strong indices may alternatively be computed directly from the parities of the occupied Kramers pairs at the eight $\mathcal{T}$-invariant momenta.[6] We then apply the method to noncentrosymmetric crystals of GeTe and strained HgTe, showing that the first is a trivial insulator, while the latter is a strong topological insulator under both positive and negative strains along [001] and under positive strain along [111].

The calculations were carried out in the framework of density-functional theory[31] using the local-density approximation with the exchange and correlation parametrized as in Ref. 32. We used HGH pseudopotentials[33] with semicore 5d-states included for Hg, while for all other elements only the s and p valence electrons were explicitly included. The calculations were carried out using the ABINIT code package[34,35] with a $10 \times 10 \times 10$ **k**-mesh for the self-consistent field calculations and a 70 Ha planewave cutoff. The spin-orbit interaction was included in the calculation via the HGH pseudopotentials. Note that the overlap matrices $M_{mn}^{(k_j,k_{j+1})}$ defined in Sec. II B, are the same as those needed for the calculation of the electric polarization[24] or the construction of maximally-localized Wannier functions,[27] and are thus readily available in many standard *ab initio* code packages including ABINIT.

A. Centrosymmetric materials

We start by illustrating the method with the examples of Bi and Bi_2Se_3. Although Bi is a semimetal, its ten lowest-lying valence bands are separated from higher ones by an energy gap everywhere in the BZ, so in this case the topological indices describe the topological character of a particular group of bands. Since this is not the occupied subspace of an insulator, these topological indices are not "physical," but it is still of interest to compute them and compare with methods based on the parity eigenvalues.[6] According to the latter approach, the group of ten lowest-lying bands of Bi was shown to be topologically trivial.[6] Bi_2Se_3, on the other hand, is a true insulator, and the parity approach demonstrated that it is a strong topological insulator.[7]

Bi and Bi_2Se_3 both belong to the rhombohedral space group $R\bar{3}m$ (#166), which has a 3-fold rotational axis. Thus, it is enough to compute only one weak $\mathbb{Z}_2$ index, say for $n_1 = 1$, since all three of them are equal by symmetry. To get the strong index, one just needs to compute just one more of the $\mathbb{Z}_2$ invariants, say for n_1=0.

Our results for Bi, obtained with the lattice parameters used in previous studies,[36] are presented in Fig. 2. Panels (a) and (b) show the determination of the $\mathbb{Z}_2$ invariant at n_1=0 and n_1=1 respectively, with k_2 treated as the pumping parameter (like t) for an effective 1D system with wavevector k_3. The k_2 axis was initially discretized into ten equal intervals ($m = 1, ..., 10$) running from 0 to π, but for reasons discussed below an extra point (number 10 on the horizontal axis of the plot) was inserted midway in the last segment to make a total of eleven m values in Panel (b). As noted above, we are treating a group of ten valence bands labeled by n, so we have an array of WCC values $\bar{x}_n^{(m)}$ whose values are indicated by the black circles in the plot. These form Kramers pairs at k_2=0 and π, but not elsewhere. Each red rhombus indicates the center $z^{(m)}$ of the largest gap between adjacent $\bar{x}_n^{(m)}$ values, as discussed in Sec. III.

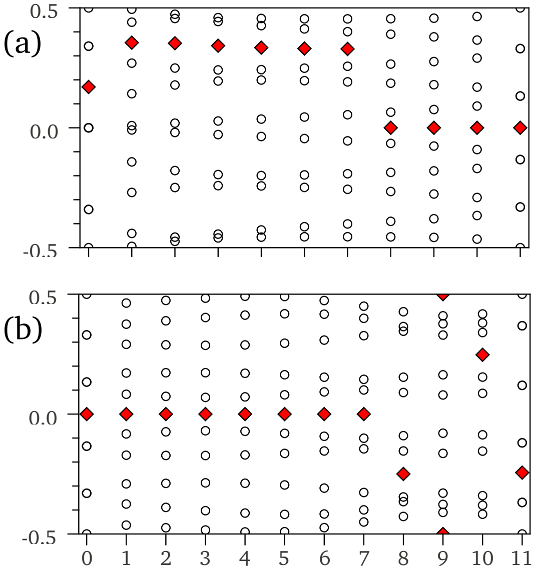

FIG. 2. Evolution of Bi WCCs $\bar{x}_n$ (circles) in the r_3 direction vs. k_2 at (a) k_1=0; (b) k_1=π. Red rhombus marks midpoint of largest gap. k_2 is sampled in ten equal increments from 0 to π, except that an extra point is inserted midway in the last segment in panel (b) (see text).

Looking first at Fig. 2(a), we see that the gap center jumps over one WCC at m=1, and then over three WCCs at m=7, for a total of four, which is even. In Fig. 2(b) we get a total of $2+7+3+4 = 16$ jumps, which is again even. The visual determinations of the number of jumped bands is confirmed by the application of the automated procedure of Eqs. (15-17). Thus, both $\mathbb{Z}_2$ indices are 0, and the 3D index is 0; (000), indicating a normal band topology as anticipated.[6,19]

We now discuss the above-mentioned insertion of one extra k_2 point in Fig. 2(b). This was necessary because the gap center $z^{(9)}$ at $k_2 = 0.9\pi$ had almost the same value as one of the WCC values at $k_2 = \pi$ (now labeled as '11' on the horizontal axis), making it ambiguous whether or not that x_n value should be counted as one of the ones that has been jumped over. To resolve this difficulty, we included an extra step at $k_2 = 0.95\pi$ (now labeled as '10' on the horizontal axis). The reason for the fast motion of the WCC in this case is that the minimum gap to the next higher (eleventh) band becomes rather small near $k_2 = \pi$.

Note that the detection of this kind of problem does not have to be done by visual inspection, but can be au-

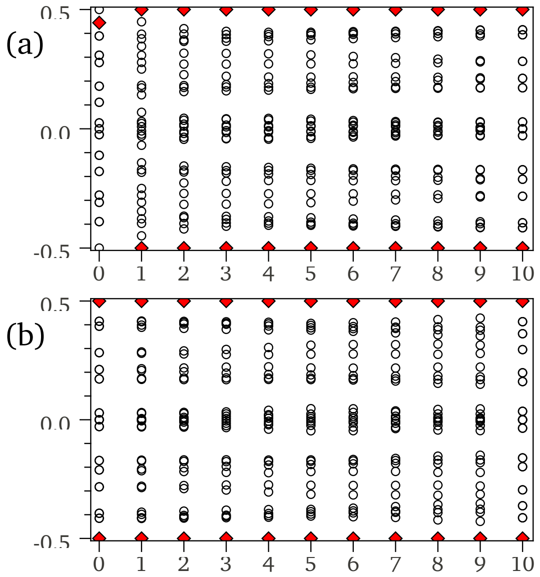

FIG. 3. Evolution of Bi_2Se_3 WCCs $\bar{x}_n$ (circles) in the r_3 direction vs. k_2 at (a) k_1=0; (b) k_1=π. Red rhombus marks midpoint of largest gap. k_2 is sampled in ten equal increments from 0 to π.

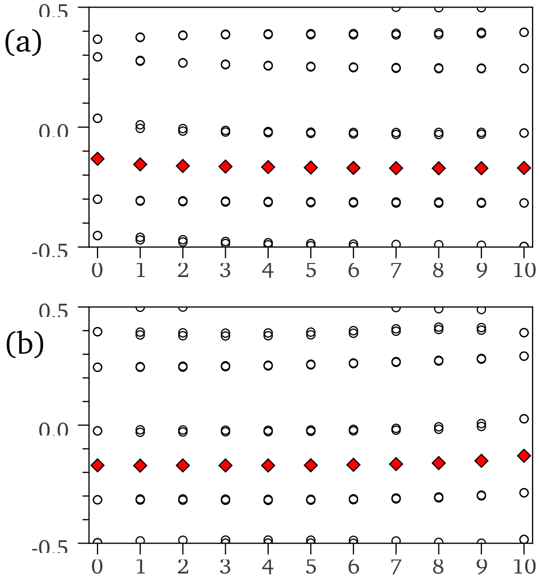

FIG. 4. Evolution of GeTe WCCs $\bar{x}_n$ (circles) in the r_3 direction vs. k_2 at (a) k_1=0; (b) k_1=π. Red rhombus marks midpoint of largest gap. k_2 is sampled in ten equal increments from 0 to π.

tomated in the context of Eqs. (15-17). As already mentioned in Sec. III B, we simply test whether any $\bar{x}_n^{(m+1)}$ approaches within a certain threshold of $z^{(m)}$ (mod 1); if so, we flag the interval in question for replacement by a finer mesh. Still, it is recommended to choose a mesh that is fine enough so that this threshold is rarely encountered, with a finer mesh recommended in cases where the minimum band gap is small.[37]

The analysis of the same n_1=0 and n_1=1 faces for the 28 WCCs of Bi_2Se_3 is illustrated in Fig. 3. The experimental lattice parameters[38] were used. Here there are no jumps over WCCs except for a single one in the very first step in the top panel (n_1=0). It follows that the topological index is 1; (000), in accord with previous studies.[7]

B. Noncentrosymmetric materials

We now proceed to systems without inversion symmetry, which are the principal targets of our method since an analysis based on parity eigenvalues is not possible.

GeTe belongs to the rhombohedral $R3m$ space group (#160) and has no inversion symmetry, although like Bi and Bi_2Se_3 it has a 3-fold rotational symmetry, so that only two reciprocal-space faces have to be studied. The experimental lattice parameters[39] were used, and the evolution of the 10 WCCs is presented in Fig. 4 following similar conventions as for Bi and Bi_2Se_3. For both faces Eq. (17) gives a trivial $\mathbb{Z}_2$ index, with the center of the largest gap making no jumps, so that GeTe is in the topologically trivial state 0; (000). This result could have been anticipated from the fact that the spin-orbit interaction in GeTe is weak, as reflected in the approximate pairwise degeneracy of the WCCs throughout the evolution.

Finally, let us consider the more interesting case of strained HgTe. In the absence of strain this is a zero-band-gap material. Any anisotropic strain breaks the four-fold symmetry at Γ, making it possible that the gap might open. Based on an adiabatic continuity argument, HgTe was predicted to be a strong topological insulator under compressive strain in the [001] direction.[6] This was later verified with tight-binding calculations.[5,40] Application of our approach to HgTe under uniaxial strain also confirms that HgTe is a strong topological insulator, with index 1; (000), under both positive and negative[6] 2% strains along the [001] direction (not shown). This means that although the positive-strain and negative-strain states are separated by a gap closure at zero strain, there is no topological phase transition associated with this gap closure.

We also studied strains in the [111] direction. Under compressive strains of −2% and −5% the system becomes metallic and the direct band gap vanishes, so that no topological index can be associated with the occupied space. Under tensile strain of +2% we find that HgTe becomes a narrow-gap semiconductor with an indirect energy gap of $E_g = 0.054\,\text{eV}$, while for +5% strain it becomes metallic. Even at +5% strain, however, the lowest 18 bands remain separated from higher ones by an energy gap at all $\mathbf{k}$, so that, as for Bi, one can still assign a topological index to this isolated group of bands. The computed band structures for both cases are illustrated in Fig. 5 along lines connecting the high symmetry points

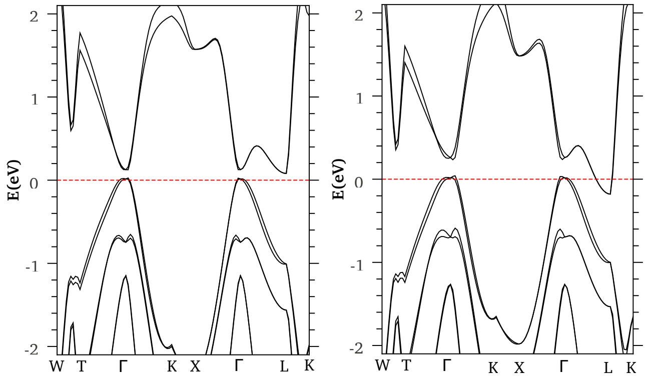

FIG. 5. Band structure along high-symmetry lines of the undistorted FCC structure for HgTe under tensile strain in the [111] direction. (a) +2% strain. (b) +5% strain.

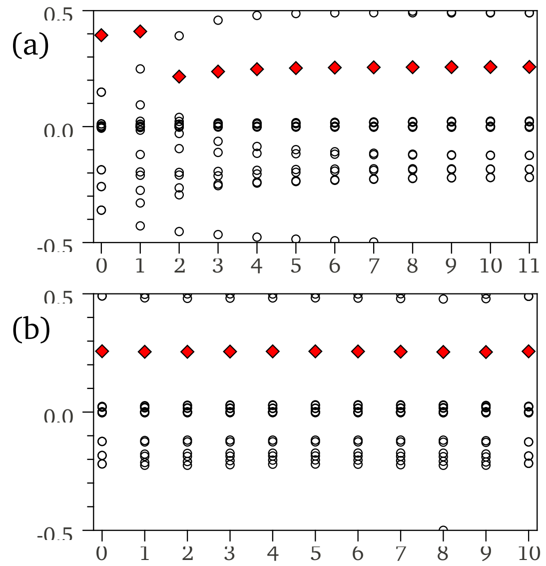

FIG. 6. Evolution of WCCs for HgTe under +2% strain in the [111] direction. WCCs $\bar{x}_n$ (circles) in the r_3 direction are plotted vs. k_2 at (a) k_1=0; (b) k_1=π. Red rhombus marks midpoint of largest gap. k_2 is sampled in ten equal increments from 0 to π, except that an extra point is inserted midway in the first segment in panel (a) (see text).

of the undistorted FCC structure.

The space group of [111]-strained HgTe is rhombohedral $R3m$ (#166), the same as for GeTe, so that again only two $\mathbb{Z}_2$ indices need to be calculated. The results of our WCC analysis for the case of +2% strain are shown in Fig. 6. We find $\mathbb{Z}_2$=1 and $\mathbb{Z}_2$=0 for n_1=0 and n_2=1 respectively, so that the topological class is 1; (000). The behavior in Panel (b) is rather uninteresting, since the gap is large everywhere on the n_2=1 face. However, in Panel (a) we again find an example of a rapid change of WCCs with k_2, which was repaired by inserting an extra point (the one now labeled '1' on the horizontal axis) at $k_2 = 0.05\pi$. Actually, we anticipated the need for this denser sampling for small k_2 from the fact that the zero-strain gap closure occurs at Γ, so that a delicate dependence on **k** near the BZ center was expected.

V. SUMMARY AND CONCLUSIONS

We have proposed a new approach for calculating topological invariants in $\mathcal{T}$-invariant systems. The method is based on following the evolution of hybrid Wannier charge centers, and is very general, being easily applicable in both tight-binding and DFT contexts. The needed ingredients are the same as those needed for the calculation of the electric polarization or the construction of maximally-localized Wannier functions, and are thus readily available in standard code packages. The present algorithm is relatively inexpensive, however, because the analysis is confined to a small number of 2D slices of the 3D Brillouin zone. The method is easily automated and remains robust even when many bands are present. We hope that our method can help to make the search for topological phases in noncentrosymmetric materials a routine task, and that it will lead to further progress in this rapidly developing field.

Note: In the final stages of preparing this manuscript, we became aware of independent work by Yu *et al.*[41] that is closely related. These authors carry out a similar analysis based on WCCs, but without the automated analysis described in our Sec. III.

VI. ACKNOWLEDGMENTS

This work was supported by NSF Grant DMR-1005838. We thank D. R. Hamann for useful discussions.

---

[*] alexeys@physics.rutgers.edu

[†] dhv@physics.rutgers.edu

[1] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 226801 (2005).
[2] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 146802 (2005).
[3] L. Fu, C. L. Kane, and E. J. Mele, Phys. Rev. Lett. **98**, 106803 (2007).
[4] J. E. Moore and L. Balents, Phys. Rev. B **75**, 121306 (2007).
[5] B. A. Bernevig, T. L. Hughes, and S.-C. Zhang, Science **314**, 1757 (2006).
[6] L. Fu and C. L. Kane, Phys. Rev. B **76**, 045302 (2007).
[7] H. Zhang, C.-X. Liu, X.-L. Qi, X. Dai, Z. Fang, and S.-C. Zhang, Nat. Phys. **5**, 438 (2009).

[8] S. Chadov, X. L. Qi, J. Kuebler, G. H. Fecher, C. Felser, and S. C. Zhang, Nature Materials **9**, 541 (2010).

[9] M. Klintenberg arXiv:1007.4838.

[10] M. Konig, S. Wiedmann, C. Brune, A. Roth, H. Buhmann, L. W. Molenkamp, X.-L. Qi, and S.-C. Zhang, Science **318**, 766 (2007).

[11] Y. L. Chen, J. G. Analytis, J.-H. Chu, Z. K. Liu, S.-K. Mo, X. L. Qi, H. J. Zhang, D. H. Lu, X. Dai, Z. Fang, S. C. Zhang, I. R. Fisher, Z. Hussain, and Z.-X. Shen, Science **325**, 178 (2009).

[12] D. Hsieh, D. Qian, L. Wray, Y. Xia, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Nature **452**, 970 (2008).

[13] Y. Xia, D. Qian, D. Hsieh, L. Wray, A. Pal, H. Lin, A. Bansil, D. Grauer, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Nat. Phys. **5**, 398 (2009).

[14] M. Z. Hasan and C. L. Kane, Rev. Mod. Phys. **82**, 3045 (2010).

[15] X.-L. Qi and S.-C. Zhang, Physics Today **63** (2010).

[16] X.-L. Qi and S.-C. Zhang arXiv:1008.2026.

[17] A. P. Schnyder, S. Ryu, A. Furusaki, and A. W. W. Ludwig, Phys. Rev. B **78**, 195125 (2008).

[18] A. Kitaev, AIP Conf. Proc. **1134**, 22 (2009).

[19] T. Fukui and Y. Hatsugai, J. Phys. Soc. Jpn. **76**, 053702 (2007).

[20] D. Xiao, Y. Yao, W. Feng, J. Wen, W. Zhu, X.-Q. Chen, G. M. Stocks, and Z. Zhang, Phys. Rev. Lett. **105**, 096404 (2010).

[21] W. Feng, D. Xiao, J. Ding, and Y. Yao, Phys. Rev. Lett. **106**, 016402 (2011).

[22] L. Fu and C. L. Kane, Phys. Rev. B **74**, 195312 (2006).

[23] Since we do not constrain the gauge of the 1D system to obey any particular symmetries at intermediate t, it is always possible to perform the unitary mixing at intermediate t in such a way that the pair of "bands" belonging to the same α at t=0 also belong to the same α at t=$T/2$.

[24] R. D. King-Smith and D. Vanderbilt, Phys. Rev. B **47**, 1651 (1993).

[25] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs, Phys. Rev. Lett. **49**, 405 (1982).

[26] A. A. Soluyanov and D. Vanderbilt, Phys. Rev. B **83**, 035108 (2011).

[27] N. Marzari and D. Vanderbilt, Phys. Rev. B **56**, 12847 (1997).

[28] S. Kivelson, Phys. Rev. B **26**, 4269 (1982).

[29] This follows from the fact that the directed area of the triangle defined by vertices z_j in the complex plane is $\mathrm{Im}[z_1^* z_2 + z_2^* z_3 + z_3^* z_1]$; specializing this to $z_j = \exp(i\phi_j)$ yields Eq. (16).

[30] R. Roy, Phys. Rev. B **79**, 195322 (2009).

[31] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[32] S. Goedecker, M. Teter, and J. Hutter, Phys. Rev. B **54**, 1703 (1996).

[33] C. Hartwigsen, S. Goedecker, and J. Hutter, Phys. Rev. B **58**, 3641 (1998).

[34] X. Gonze *et al.*, Z. Kristallogr. **220**, 558 (2005).

[35] X. Gonze *et al.*, Computer Phys. Comm. **180**, 2582 (2009).

[36] X. Gonze, J.-P. Michenaud, and J.-P. Vigneron, Phys. Rev. B **41**, 11827 (1990).

[37] In the vicinity of a small gap, it is also advisable to reduce the mesh spacing along the k-point strings used for the parallel transport construction.

[38] J. Phys. Chem. Solids **15**, 13 (1960).

[39] A. Onodera, I. Sakamoto, Y. Fujii, N. Mori, and S. Sugai, Phys. Rev. B **56**, 7935 (1997).

[40] X. Dai, T. L. Hughes, X.-L. Qi, Z. Fang, and S.-C. Zhang, Phys. Rev. B **77**, 125319 (2008).

[41] R. Yu, X. L. Qi, A. Bernevig, Z. Fang, and X. Dai arXiv:1101.2011.