# Holographic treatment of boundary disorder in a topological insulator

Kun Woo Kim,[1] Roger S. K. Mong,[2,3] Marcel Franz,[4] and Gil Refael[2]

[1]*School of Physics, Korea Institute for Advanced Study, Seoul 130-722, Korea*

[2]*Department of Physics and Institute of Quantum Information and Matter, California Institute of Technology, Pasadena, California 91125, USA*

[3]*Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA*

[4]*Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z1*

The effect of boundary disorder on electronic systems is particularly interesting for topological phases with surface and edge states. Using exact diagonalization, it has been demonstrated that the surface states of a three-dimensional (3D) topological insulator survive strong surface disorder, and simply get pushed to a clean part of the bulk. Here we explore a method which analytically eliminates the clean bulk and reduces a $D$-dimensional problem to a Hamiltonian-diagonalization problem within the $(D-1)$-dimensional disordered boundary. This dramatic reduction in complexity allows the analysis of significantly bigger systems than is possible with exact diagonalization. We use our method to analyze a 2D topological spin-Hall insulator with nonmagnetic and magnetic edge impurities, and we calculate the disorder-induced redistribution of probability density (or local density of states) in the insulating bulk, as well as the transport effects of edge impurities. The analysis reveals how the edge recovers from disorder scattering as the disorder strength increases.

## I. INTRODUCTION

Solid-state systems inevitably contain impurities. The study of impurity effects in topological insulators is intensely pursued (for a review, see, e.g., Refs. [1–4]), especially since their surface states are expected to be fundamentally robust with respect to certain types of disorder. Impurity effects in topological insulators (TIs) are of particular importance, since their fabrication may require, for instance, a finite concentration of doping to place the Fermi energy within the band gap [5–7] or to induce superconductivity [8–10].

Impurities that are located at or near the TI surface are especially interesting. One relevant example is the treatment with $NO_2$ (see supplement of Ref. [5]) on the surface of as-grown topological insulator $Bi_{2-x}Cu_xSe_3$ which is necessary to prevent the surface band bending caused by the adsorption of residual atoms present in the vacuum chamber. As the manifestations of topological systems—such as the protected surface states or Majorana fermions—are localized near the surface or the edge of the system [11–15], the effect of surface impurities on the surface spectral and transport properties is of both theoretical and practical interest [16–18]. Impurity effects could be apparent through the local density of states (LDOS) and spectral functions. These can be efficiently probed by scanning tunneling spectroscopy, while the surface spectral function can be extracted from angle-resolved photoemission spectroscopy (ARPES) measurements. Even more directly, boundary impurities are expected to affect transport, which can be probed directly using a variety of methods.

The only general analysis of surface impurities was given recently, when an exact-diagonalization (ED) approach was applied to noninteracting TIs with surface disorder [16,19] (recent additional work was dedicated to surface Kondo effects in three-dimensional (3D) TIs [20,21]). This study found evidence for a crossover between a nearly ballistic response of the surface electrons at weak disorder, localization physics at intermediate disorder, and then a restored nearly ballistic

surface state hiding in the second layer when disorder is very strong. Further investigations of this phenomenon using ED are going to prove very challenging: the computational cost for analyzing the surface physics can be very high because the bulk of the system—although gapped—must be treated on equal footing with the surface degrees of freedom. The system sizes accessible to ED analysis are, therefore, rather small. In particular, we demonstrate here that when strong surface impurities are present in the system, the lower bound of system size required to clearly resolve the bulk electronic properties and their effect on the surface states becomes large in proportion to the impurity potential strength (at least in two dimensions). Treating sufficiently large systems using ED becomes computationally challenging in this limit and different techniques to address the problem are required. In this paper, we develop such a technique. We also emphasize the need to treat sufficiently large systems in order to distinguish system properties at the thermodynamic limit versus those of the "quantum-dot" regime, where finite-size effects dominate.

In this paper, we introduce a technique that allows us to efficiently extract the surface-state properties of surface-disordered TIs. We obtain properties such as the surface spectral function, LDOS, and transport properties of the surface channels by essentially "integrating out" the clean bulk degrees of freedom analytically and obtain the effective surface Hamiltonian describing a TI surface with arbitrarily strong impurities. This approach not only allows us to reduce the computational difficulty by one dimension (e.g., for a 3D TI with surface disorder, we only need to solve a 2D problem), but also allows us to map a strong disordered problem into a weak disordered one where perturbation theory is valid. By constructing a self-consistent transfer-matrix approach, we are able to recover the exact energies and wave functions of the surface states both at the disordered layer and in the remaining bulk layers.

Our method is applicable to any layered system at arbitrary dimension. The 2D case provides the simplest application of

the method, and we use it both to demonstrate its workings and also to explore one of its most interesting applications: the effects of edge impurities in 2D TIs. A number of numerical analyses of a 2D quantum spin-Hall system are available [22–24]. These do not study edge impurity effects, however.

Boundary impurity issues are natural concerns in experiments. In 2D topological phases, edge roughness, impurities, and fabrication faults are clearly an issue. In 3D, surface defects and contamination are also expected. Moreover, the response of TIs to boundary disorder and its interplay with topological boundary states is of theoretical interest, given the unique response of topological insulators. The applicability of our analysis of 2D TI phases is relevant not only to CdTe heterostructures. Experimental advances have been made in fabricating high-quality thin-film 3D TIs [25,26], which have the same effective Hamiltonian as a 2D TI [27]. The TI thin-film-based quantum anomalous Hall effect was already reported [28]. These developments potentially allow probing spectral functions in the bulk of a 2D TI directly.

The paper is organized as follows. In Sec. II, we first explain how to integrate out the clean bulk degrees of freedom. In Sec. III, we then introduce a generic model Hamiltonian of a 2D TI with impurities on one of its edges. Next, we report a series of results that clearly differentiate the 0D quantum-dot regime from the bulk regime (where finite-size errors are suppressed) and specify the lower bound on the system size to observe the latter for magnetic and nonmagnetic edge impurities. We conclude the section with a discussion of the surface properties. Lastly, in Sec. IV, the conductance through the 2D TI edge channels is computed. The latter provides complementary information to the spectral properties discussed in Sec. II. We conclude with a discussion in Sec. V.

## II. GENERAL FRAMEWORK: EFFECTIVE SINGLE-LAYER HAMILTONIAN

In this section, we provide a general derivation of our approach. Our goal is to exactly reduce diagonalization of a $D$-dimensional system with surface impurities to the diagonalization of an effective Hamiltonian describing just the $(D-1)$-dimensional surface. Taking the top surface to be disordered, we introduce a way to integrate out the clean layers from the bottom all the way to the top layer. This leaves us with a single-layer effective Hamiltonian which includes the impurity potential and a self-energy which accounts for the entire clean bulk.

### A. Layered Schrödinger equation and self-energy

We begin our analysis with the Schrödinger equation for layer parallel to the disordered surface (Fig. 1),

$$B\psi_{n-1} + [H_0 + V_{\text{imp}}\delta_{n,1} - E]\psi_n + B^\dagger\psi_{n+1} = 0, \quad (1)$$

where $\psi_0 = 0$, $\psi_n$ is a wave function on the $n$th layer parallel to the impurity surface, $H_{||}$ is an in-layer Hamiltonian, $V_{\text{imp}}$ is an impurity potential in the first layer, $B$ and $B^\dagger$ are hopping terms between layers, and $n > 0$ is the layer index. For notational convenience, we set $\psi_0 = 0$. For $n = N$, the last layer of the system, we can write exactly

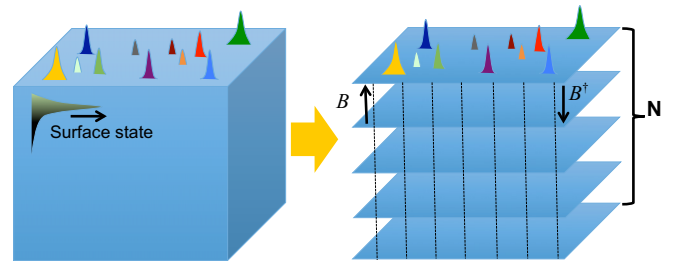$$B\psi_{N-1} + [H_0 - E]\psi_N = 0. \quad (2)$$



FIG. 1. (Color online) A pictorial description of a system with impurities on the surface only. We are interested in studying the local density of surface state as the impurity strength increases. Our strategy is to decompose the system into clean layers, coupled through matrices $B$ and $B^\dagger$, and a surface containing impurities, and then analytically integrate out the former.

Using the Schrödinger equation for the $n = N - 1$ layer, $B\psi_{N-2} + [H_0 - E]\psi_{N-1} + B^\dagger\psi_N = 0$, and substituting Eq. (2), we can "integrate" out the last ($N$th) layer,

$$B\psi_{N-2} + \left[H_0 - E + B^\dagger\frac{1}{E - H_0}B\right]\psi_{N-1} = 0. \quad (3)$$

Eliminating $\psi_N$ introduces for $\psi_{N-1}$ the effective potential $\Sigma_{N-1} = B^\dagger\frac{1}{E - H_0}B$. By repeating this process, we can integrate out all layers up to the first layer and the following recursion relation can be found:

$$\Sigma_n = B^\dagger\frac{1}{E - H_0 - \Sigma_{n+1}}B, \quad (4)$$

with a boundary condition $\Sigma_N = 0$. Recall that $B^\dagger$ is a hopping to the next layer and $B$ is a hopping to the prior layer. And the effective potential $\Sigma_n$ is obtained by sandwiching the Green's function in the $(n + 1)$th layer by $B^\dagger$ and $B$, describing a scattering process of hopping to the next layer, propagating, and hopping back to the original layer.

Let us next write an effective Hamiltonian in the top layer in the following way:

$$[E - H_0 - \Sigma_1]\psi_1 = V_{\text{imp}}\psi_1, \quad (5)$$

$$\left[B\Sigma_0^{-1}B^\dagger\right]\psi_1 = V_{\text{imp}}\psi_1, \quad (6)$$

where the recursion relation (4) is again used to further simplify the clean part of the Hamiltonian.

The recursive relation above appeared before in the literature in terms of Green's function [29] and Hamiltonian [30]. Here, our main result is the ability to analytically solve this relation for a finite system size $N$, and obtain from it an effective Hamiltonian of a semi-infinite size system. From our analysis, we are able to extract information about the LDOS at any layer in the system, starting with the impurity layer itself. We show the general formulation in the next section, and provide a concrete example of the procedure in Sec. III.

### B. "Holographic" mapping of the self-energy

The recursion relation can be straightforwardly solved by mapping the effective potential to a matrix $M$ which obeys the same Schrödinger equation as the layer wave functions,

$$BM_{n-1} + [H_0 - E]M_n + B^\dagger M_{n+1} = 0, \quad (7)$$

where the matrix $M_n$ has the same dimension as the Hamiltonian $H_{||}$, and is invertible by construction.

With Eq. (7), the recursion relation for the self-energy is easily solved,

$$\Sigma_n = B^\dagger M_{n+1} M_n^{-1}. \tag{8}$$

One can directly verify that this is a solution of the recursion relation for $M_n$ satisfying boundary condition $M_{N+1} = 0$. For a clean bottom surface, we can exactly construct $M_n$ for a system with finite thickness (the calculation is detailed in the Appendix).

The last step involves writing a closed-form equation for the wave function of the top (disordered) layered. Note that as $M_n$ is also a solution of the Schrödinger equation, an element of $M_n$ scales with eigenvalues of the transfer matrix of Schrödinger equation (7): $(M_n)_{ii'} \sim \rho_j^{N-n}$. Using the exact expression of $M_n$'s, we construct the left side of Eq. (6). Then, we obtain the Schrödinger equation expressed in terms of $M_n$'s,

$$\left[ B M_0 M_1^{-1} \right] \psi_1 = V_{\text{imp}} \psi_1. \tag{9}$$

This is the effective single-layer Hamiltonian. The left-hand side contains only elements from the clean part of the Hamiltonian and involves the self-energy from all subsequent layers; the right-hand side is simply the surface impurity potential operating on the top-layer wave function. $M_0$ and $M_1$ are a function of energy and one can find all eigenvalues of a system by finding the energies that satisfy $\det \left[ B M_0 M_1^{-1} - V_{\text{imp}} \right] = 0$. The surface wave function can be subsequently found from Eq. (9), which is identical to the result from exact diagonalization. To obtain a whole wave function in the layers beneath the top layer, we apply the transfer matrix, which is also obtained in terms of $M_n$, as shown in the next section.

### C. Transfer matrix of wave functions

The first-layer wave function can be exactly obtained from Eq. (9); therefore, the computational complexity is essentially reduced by one dimension. To obtain a full profile of the wave function in the subsequent layers, we construct an approach similar to the transfer-matrix approach in this section. We will use the term "transfer matrix" quite liberally in what follows.

$\Sigma_n$ plays the role of effective potential in the $n$th layer, induced by integrating out the $(n+1)$th layer up to the $N$th layer. In the Schrödinger equation (1), such a contribution is accounted for by the third term on the left side. Therefore, we have the following equality:

$$B^\dagger \psi_{n+1} = \Sigma_n \psi_n. \tag{10}$$

One can explicitly show this relation by the elimination method introduced in Eq. (3). The transfer matrix is conveniently expressed in terms of the $M_n$'s using Eq. (8),

$$\psi_{n+1} = \left[ M_{n+1} M_n^{-1} \right] \psi_n, \tag{11}$$

or, more generally, using the relation between the wave functions in the $m$ and $n$ layers,

$$\psi_n = \left[ M_n M_m^{-1} \right] \psi_m, \tag{12}$$

where the exact expression of $T_{n,m} \equiv M_n M_m^{-1}$ is known. Note that the expression for the transfer matrices is disorder free, which implies that the disordered wave function in the first

layer propagates into the subsequent layers, just as a clean wave function would. This, of course, makes sense since only the top layer contains impurities. A conventional transfer matrix constructed from the top surface, however, would always contain impurity potentials, and therefore the construction of the whole wave function would not be as straightforward.

### D. "Holographic" mapping of the impurity potential

For completeness, we address another question of interest: what is the effective impurity potential experienced by an electronic state in the bulk due to the surface impurity? This question can be answered in the same formalism introduced in earlier sections. To compute the effective impurity potential, we integrate out the first $(n - 1)$ layers. The recursion relation for effective potential is

$$V_{n+1} = B \frac{1}{E - H_0 - V_n} B^\dagger, \tag{13}$$

with boundary condition $V_1 = V_{\text{imp}}$. The "holographic" mapping helps us to analytically derive a scaling behavior of the effective potential,

$$V_n = B \tilde{M}_{n-1} \tilde{M}_n^{-1}, \tag{14}$$

where the $\tilde{M}_n$'s are similarly constructed to satisfy the Schrödinger equation for the individual layers, given by Eq. (7), and to be invertible. However, their boundary condition is different from the previous clean case. $\tilde{M}_n$ has to be constructed such that the following condition is satisfied:

$$V_{\text{imp}} = B \tilde{M}_0 \tilde{M}_1^{-1}. \tag{15}$$

Since $V_{\text{imp}}$ is a random matrix, it is nontrivial to determine $\tilde{M}_0$ in general. But, we know the object $\tilde{M}_n$ propagates just like a clean wave function. Therefore, it is possible to deduce the scaling of $(V_n)_{ij}$ with respect to layer index $n$, which includes the contribution from surface impurities as well as clean layers from the top to the $(n - 1)$th layer.

## III. APPLICATION TO A 2D TOPOLOGICAL INSULATOR

In the previous section, we introduced a general transfer-matrix framework for computing the full wave functions of layered systems with surface impurities. In this section, a 2D topological-insulator model [31] is employed to explicitly show how the local density of states can be computed in a system with edge impurities. The method developed in the previous section is not only applicable to topological insulators, but to any systems in arbitrary dimensions even without energy gap. We apply this method to study the interplay of edge disorder and the edge states of a 2D topological insulator. In addition to the relevance of this problem to the field, it allows us to demonstrate our method in a most transparent way.

Our main results are presented in Figs. 5–7, where we use the formalism developed earlier to compute the LDOS of the first and second layers of the TI, varying the disorder strength $W$.

### A. Model Hamiltonian

Consider the toy model of a 2D topological insulator. In momentum space,

$$
\begin{aligned}
H(\vec{k}) = {}& [m - 2b(2 - \cos k_x - \cos k_y)]\tau_z \\
& + A[\tau_x s_z \sin k_x + \tau_y \sin k_y],
\end{aligned} \tag{16}
$$

where $\tau_i$ is a Pauli matrix in an orbital basis and $s_i$ is a Pauli matrix in a spin basis. The lattice spacing is set to $a = 1$ such that the momenta $k_x$ and $k_y$ lie within the interval $[-\pi, \pi]$. To introduce an edge state, open boundary conditions are introduced in the $y$ direction, and periodic boundary conditions are applied to the $x$ direction. The intralayer Hamiltonian and the hopping term between layers described in Eq. (1) are

$$
\begin{aligned}
H_0 &= [m - 2b(2 - \cos k_x)]\tau_z + \tau_x s_z A \sin k_x, \\
B &= b\tau_z - i\frac{A}{2}\tau_y.
\end{aligned} \tag{17}
$$

The system is in the topological phase if the bands are inverted for some range of momentum: $\text{sign}(mb) > 0$. For this case, the dispersion of the top and bottom edge states is given by $E = \pm A \sin k_x$ [32]. For the numerical calculations presented below, we use $A = 1$, $m = 0.9$, and $b = 1$.

### B. A single-layer effective Hamiltonian

The system is equivalent to a set of parallel 1D wires coupled by a hopping matrix $B$ in a spin and orbital basis. We want to construct the matrix $M_n$ which is an essential building block of a single-layer Hamiltonian [Eq. (9)] and the transfer matrix [Eq. (12)]. To demonstrate the method, we construct $M_n(k_x = 0, E = 0, s_z = 1)$ here, and other $(k_x, E)$ cases will be shown in the Appendix. Note that because the clean Hamiltonian is diagonal in momentum and spin space, we only need to analyze the orbital space. The Schrödinger equation we need to solve is

$$
\left\{ \tau_z \left[ b\left(\frac{1}{\rho} + \rho\right) + m - 2b \right] - i\tau_y \frac{A}{2}\left(\frac{1}{\rho} - \rho\right) \right\} \psi_n = 0,
$$

where we use $\psi_{n+m} = \rho^m \psi_n$. This is one section of the Schrödinger equation at $k_x = 0$, $E = 0$, and $s_z = 1$. There are four transfer eigenvalues and corresponding eigenvectors. By taking the determinant of the terms in the square bracket, we get $\rho_1 = \lambda_1$, $\rho_2 = 1/\lambda_1$, $\rho_3 = \lambda_2$, and $\rho_4 = 1/\lambda_2$, where

$$
\begin{aligned}
\lambda_1 &= \frac{-(m - 2b) - \sqrt{m^2 - 4mb + A^2}}{A + 2b}, \\
\lambda_2 &= \frac{-(m - 2b) + \sqrt{m^2 - 4mb + A^2}}{A + 2b}.
\end{aligned}
$$

Here, $\lambda_{1,2}$ is chosen to be $|\lambda_{1,2}| < 1$ for $0 < m < 4b$ and corresponding eigenvectors are $\phi_1 = \phi_3 = |+\rangle$ and eigenvectors of $1/\lambda_{1,2}$ are $\phi_2 = \phi_4 = |-\rangle$, where

$$
|+\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad |-\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}. \tag{18}
$$

Considering only one spin section, this implies that a state $|+\rangle$ is localized at the top ($n = 1$) and a state $|-\rangle$ is localized at the bottom ($n = N$). The interlayer hopping operator can be expressed in terms of the eigenvectors:

$B = (b - A/2)|-\rangle\langle+| + (b + A/2)|+\rangle\langle-|$. For this given set of eigenvalues and vectors, we can construct an invertible $M_n$ in the following manner:

$$
M_{l+N+1} = \left(\rho_1^l - \rho_3^l\right)|+\rangle\langle+| - \left(\rho_2^l - \rho_4^l\right)|-\rangle\langle-|,
$$

where $l = n - N - 1$. $M_n$ is invertible for $n \neq N + 1$ and satisfies the homogeneous boundary condition $M_{N+1} = 0$. If $m^2 - 4mb + A^2 < 0$, eigenvalues are $\rho_1 = \lambda e^{i\theta}$ and $\rho_3 = \lambda e^{-i\theta}$ with $\lambda = |\frac{2b-A}{2b+A}|$. Thus, the effective Hamiltonian in the first layer is as follows see Eq. (9):

$$
\begin{aligned}
E - H_1^{\text{eff}} &= B M_0 M_1^{-1} \\
&= B\left[ \frac{\rho_1^{-N-1} - \rho_3^{-N-1}}{\rho_1^{-N} - \rho_3^{-N}} |+\rangle\langle+| \right. \\
&\quad \left. + \frac{\rho_2^{-N-1} - \rho_4^{-N-1}}{\rho_2^{-N} - \rho_4^{-N}} |-\rangle\langle-| \right] \\
&= (b - A/2)\frac{1}{\lambda}\frac{\sin[(N+1)\theta]}{\sin(N\theta)}|-\rangle\langle+| \\
&\quad + (b + A/2)\lambda\frac{\sin[(N+1)\theta]}{\sin(N\theta)}|+\rangle\langle-|, \tag{19}
\end{aligned}
$$

where $N$ is the number of parallel wires. The effective Hamiltonian contains off-diagonal elements only in the $|\pm\rangle$ basis.

Note that the effective single-layer Hamiltonian at $(k_x, E) = (0,0)$ depends on the width of the system, and if a wave function contains a component at $(k_x, E) = (0,0)$, it will also be system-size dependent. Because we cannot think of a localized wave function dependent of the system size for large enough $N$, we can say no eigenstate localized to an edge sits at $(k_x, E) = (0,0)$. More relevant Hamiltonian sections at zero energy will be $k_x \neq 0$, which is system-size independent in the large-$N$ limit. More generally, the Hamiltonian sectors not at $(k_x, E) = (2\pi l/N, A \sin k_x)$ are expressed in the following way:

$$
\begin{aligned}
\langle+|H_1^{\text{eff}}|+\rangle &= \frac{b + A/2}{1/r_3 - 1/r_1}\left[\rho_1^{-1} - \rho_3^{-1}\right], \\
\langle+|H_1^{\text{eff}}|-\rangle &= \frac{b + A/2}{1/r_3 - 1/r_1}\left[\rho_1^{-1}/r_1 - \rho_3^{-1}/r_3\right], \\
\langle-|H_1^{\text{eff}}|+\rangle &= \frac{b - A/2}{r_1 - r_3}\left[\rho_1^{-1}r_1 - \rho_3^{-1}r_3\right], \\
\langle-|H_1^{\text{eff}}|-\rangle &= \frac{b - A/2}{r_1 - r_3}\left[\rho_1^{-1} - \rho_3^{-1}\right].
\end{aligned} \tag{20}
$$

Here, $\rho_{1,3}$'s are the eigenvalues of the transfer matrix with magnitude smaller than unity, and $r_i = \langle+|\phi_i\rangle/\langle-|\phi_i\rangle$'s are the ratios of the overlaps between the transfer-matrix eigenstates corresponding to $\rho_i$ with the $|+\rangle$ and $|-\rangle$ states. We can see that as the $(k_x, E)$ approaches the on-shell condition, $r_i$ approaches zero and the wave functions have infinitesimal overlap with $|-\rangle$ since the $\langle-|H_1^{\text{eff}}|-\rangle$ component is huge. In other words, the Hamiltonian expressed in this way can be interpreted as a projection to the on-shell eigenstates.

With a set of impurities on the top wire (Fig. 2), to obtain eigenenergies we find energies where the determinant of the effective single-layer Hamiltonian is zero [see Eq. (9)].
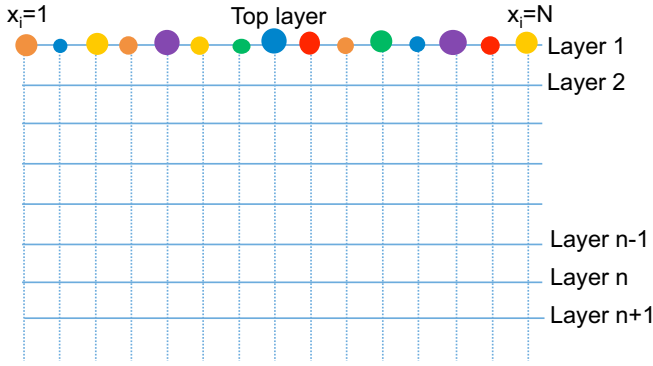
FIG. 2. (Color online) A 2D lattice model for a TI with edge impurities. Along the $x$ direction, we have a periodic boundary condition that $x_1 = x_{N+1}$, and along the $y$ direction, there are $N$ layers and the subscript $n$ indicates the $n$th layer along the $y$ direction. Different colors and sizes of dots on the top layer represent the random on-site impurity potentials. We study how the effect of the top edge impurities propagate down into the bulk layers.

In the strong impurity regime, we must use a large enough system size to correctly see the size-independent behavior of bulk electronic properties. Here, we distinguish the quantum-dot regime from the bulk regime by the dependence of physical observables on the system size. Figure 3 shows the histogram of the ratio of the local density of state in the first and the second wire for impurity strength $W = 20$ with increasing system sizes $N_x = 10, 20, 40, 80$. The series of histograms shows size dependence for $N_x \lesssim 40$; the histogram becomes Gaussian shape and size independent for $N_x \gtrsim 40$. Therefore, if one wants to numerically obtain physical observables in
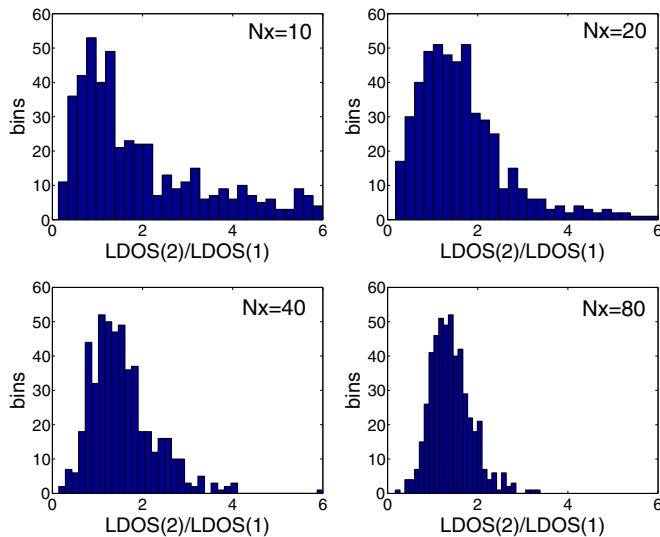


FIG. 3. (Color online) For disorder strength $W = 20$, histograms of the LDOS ratio for different system sizes $N_x = 10, 20, 40, 80$ are shown. When system size $N_x > 80$, the distribution converges and the ratio of the LDOS in the first and the second wire becomes system-size independent, and the average of the LDOS ratio in the thermodynamic limit can be estimated. On the other hand, system sizes $N_x < 40$ are in the quantum-dot regime and are not proper to compute the bulk electronic properties because of their system-size dependence.
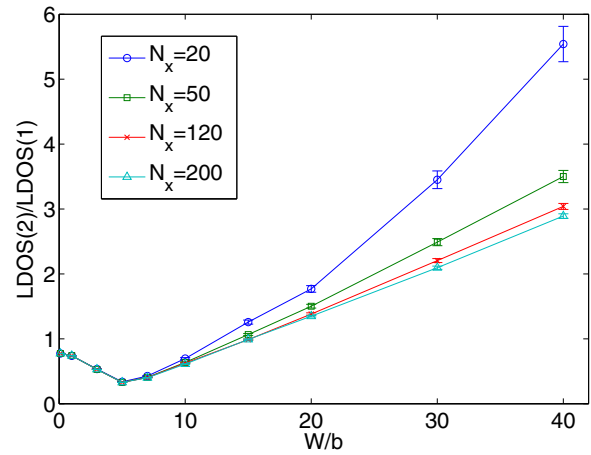


FIG. 4. (Color online) The LDOS ratio of the first and second layer is plotted over the impurity strength $W$ for system size $N_x = 20, 50, 120, 200$. The edge state at zero energy is primarily populated in the first layer from weak to moderate impurity strength, $W < 5$, and then the edge state moves to the second and following layers at strong impurity strength. The $N_x = 200$ curve shows the behavior in the thermodynamic limit as it becomes size independent, while the $N_x = 20$ curve shows quantum-dot behavior for $W > 5$, meaning the system size is not big enough to see bulk properties.

the thermodynamic limit, it is important to use a system size that is larger than $N_x = 80$ for $W = 20$ nonmagnetic edge impurities. The large-size requirement is less stringent for smaller impurity strengths, as evident from Fig. 4.

Figure 4 shows the ratio of the LDOS at zero energy with increasing impurity strength. This quantity tells us where the edge-state wave functions actually reside. A ratio below 1 indicates edge states rooted in the first layer. But ratios greater than 1 indicate edge states expelled to the second layer, which are therefore increasingly less immune to the disorder.

The edge states in weak and moderate disorder are mostly populated in the first layer (also see Fig. 5). This comes hand in hand with a spread of the edge function Fourier transform: it has broad support away from $k_x = 0$ due to impurity scattering. Once the impurity strength is comparable to or larger than the bandwidth, the edge state is populated less in the impurity layer, and it moves to the second and following layers. All curves in Fig. 4 show this behavior with a dip at $W = 5$. While we believe that the $N_x = 200$ curve properly describes the system-size–independent LDOS ratio in the thermodynamic limit, the $N_x = 20$ curve is only good for $W < 5$ and it begins to deviate from the $N_x = 200$ curve for strong impurity strength $W > 5$.

### C. Transfer matrix between single-layer wave functions

Once the wave function of the first layer, $\psi_1$, is obtained, we next propagate it to the subsequent layers to obtain a full profile of the state. This can be done by using the matrices $M_n$ as in Eq. (12). Let us write down the expression of the transfer matrix for the $(k_x, E) = (0, 0)$ case first from layer $m$ to layer $n$,

$$T_{n \leftarrow m} = M_n M_m^{-1}$$

$$= [\lambda^{n-m}|+\rangle\langle+| + \lambda^{m-n}|-\rangle\langle-|]\frac{\sin(N + 1 - n)\theta}{\sin(N + 1 - m)\theta},$$
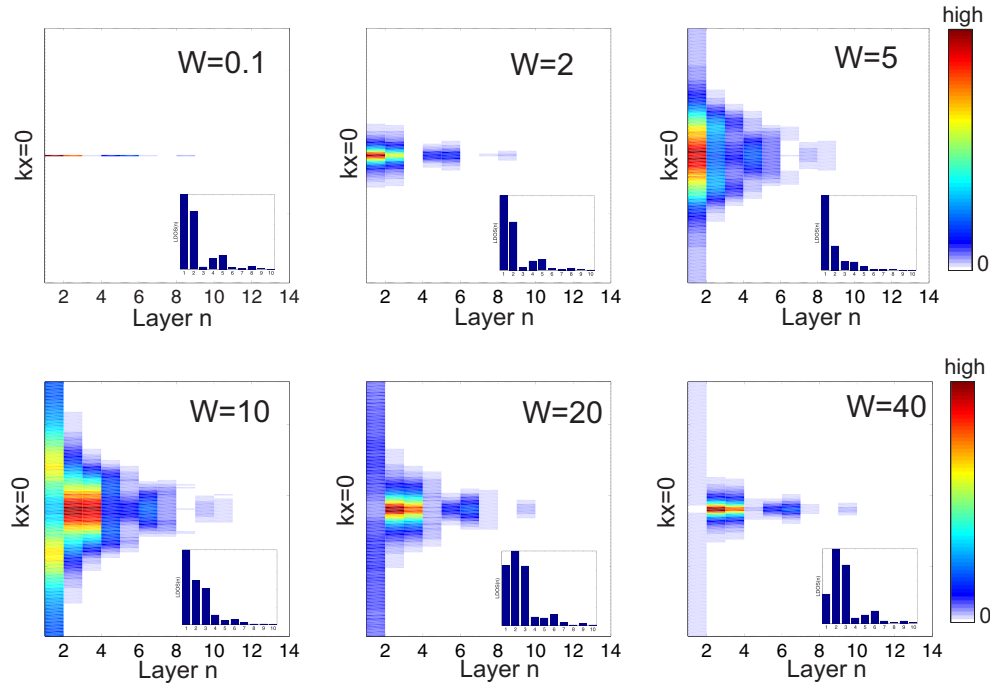
FIG. 5. (Color online) The disorder-averaged LDOS at $E = 0$ in wire index $n$ and momentum space $k_x$ of the edge state at zero energy is plotted for increasing impurity strength. Starting from the clean edge state where the LDOS is only at $k_x = 0$, the LDOS is spread out in momentum space, then shifted to the second layer. Insets: The $k_x$-integrated LDOS as a function of layer $n$. Large enough system size $N_x = 180$ is chosen such that the averaged LDOS is size independent. 500 disorder realizations are averaged.

where $\rho_{1,3} = \lambda e^{\pm i\theta}$ is used as before with $\lambda < 1$, and $N$ is the number of layers. We can see that the $|+\rangle$ component exponentially decays from the top surface towards the other end ($n > m$), while $|-\rangle$, if it is present, exponentially increases. Thus, it is apparent that $|+\rangle$ is a state localized to the top edge, while $|-\rangle$ is localized to the bottom edge. However, note that the transfer matrix contains an oscillating term dependent of the system size $N$, just as in the effective single-layer Hamiltonian, given by Eq. (19). It implies that if there is a $|+\rangle$ component in the wave function at exactly $(k_x, E) = (0,0)$, its oscillating part is dependent on the number of layers, which does not make sense in the physical picture where $N$ is much larger than the localization length of the edge state. Therefore, we can say that the $|+\rangle$ component at $(k_x, E) = (0,0)$ must be vanishingly small as the system size is increased.

The transfer matrix from the first layer to the $n$th layer for a general $(k_x, E)$ is expressed in the following way:

$$\langle +|T_{n\leftarrow 1}|+\rangle = \frac{1}{r_1 - r_3}\left[\rho_1^{n-1} r_1 - \rho_3^{n-1} r_3\right],$$

$$\langle +|T_{n\leftarrow 1}|-\rangle = \frac{1}{r_1 - r_3}\left[\rho_1^{n-1} - \rho_3^{n-1}\right],$$

$$\langle -|T_{n\leftarrow 1}|+\rangle = \frac{1}{1/r_3 - 1/r_1}\left[\rho_1^{n-1} - \rho_3^{n-1}\right], \quad (21)$$

$$\langle -|T_{n\leftarrow 1}|-\rangle = \frac{1}{1/r_3 - 1/r_1}\left[\rho_1^{n-1}/r_1 - \rho_3^{n-1}/r_3\right],$$

with $r_i = \langle +|\phi_i\rangle / \langle -|\phi_i\rangle$.

We apply this transfer matrix to the first-layer wave function to obtain wave functions in the bulk layers. In Fig. 5, the disorder-averaged probability density in momentum and layer basis $P(k_x, n) = |\psi_n(k_x)|^2$ is plotted for shown impurity strength $W$. In the weak disorder regime, $W = 0.1$, where impurity strength is much smaller than the energy gap, the probability density is concentrated near $k_x = 0$ at zero energy. As the impurity strength increases, the probability density gains width in momentum space and its weight is shifted to the second layer. While this trend is quite strong already with $W = 10$, in strong impurity regime $W = 40$—which is much larger than the bandwidth—the zero-energy wave function is completely absent in the first layer, but occupies the subsequent layers in a narrow range of momentum space. This indicates that the wave function has been pushed to the next layer and behaves as if the system is clean.

This behavior of the local density of states is shown for nonmagnetic edge impurities, which cannot affect the transport properties of helical edge states in 2D topological insulators. Therefore, the modification of LDOS should be discussed separately from the change of transport nature, at least in 2D. For a strip geometry, the transport is studied for both nonmagnetic and magnetic edge impurities in the following sections.

The widths of probability density $P(k_x, n) = |\psi_n(k_x)|^2$ in the first and second layer at zero energy are plotted in Fig. 6. The width in momentum space is indicative of how disordered the edge state is due to impurities. In the weak and strong disorder limit, the wave function behaves like a clean system in the LDOS shape, as shown in Fig. 5. With strong disorder, the width of $P(k_x, n = 1)$ saturates near 0.3, although it carries little weight in that limit: $P(k_x, n = 1) \ll 1$. Meanwhile, the width of $P(k_x, n = 2)$ increases and decreases again as the impurity strength is varied, peaking at around the bandwidth of the system ($W \sim 8$).
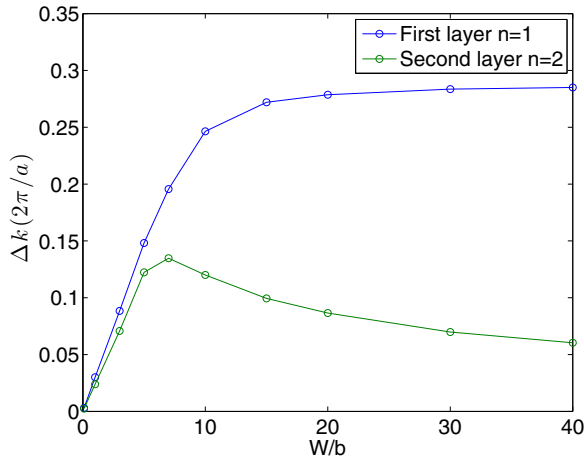
FIG. 6. (Color online) For system size $N_x = 120$, the width of LDOS in the first and the second wire in momentum space $k_x$ is plotted as a function of disorder strength $W$ for nonmagnetic impurities. In the clean-system limit, the edge-state LDOS $P(k_x)$ is concentrated at $k_x = 0$ and its width is infinitesimally small. As the disorder strength increases, the LDOS spreads in momentum space. However, for the stronger disorder, the first-layer LDOS takes on all possible momenta and its width saturates, while the second-layer LDOS become concentrated around $k_x = 0$ again.

### D. Magnetic edge impurities

The same calculation was repeated for a system with magnetic edge impurities. We simply needed to extend the Hamiltonian to have two spin sections and introduce random magnetic impurities, $V(x_i) = \vec{V}_i \cdot \vec{s}$, where three-component random variable $\vec{V}_i = (V_i^x, V_i^y, V_i^z)$. We found that to simulate the bulk regime for $W = 20$, the system size needs to be at least $N_x = 400$ as opposed to $N_x = 120$ for the nonmagnetic edge impurity case. In other words, the lower bound of the system size to see the thermodynamic properties is much larger and it becomes computationally challenging even for 2D system. Figure 7 shows the width of the LDOS distribution $P(k_x,n)$ for
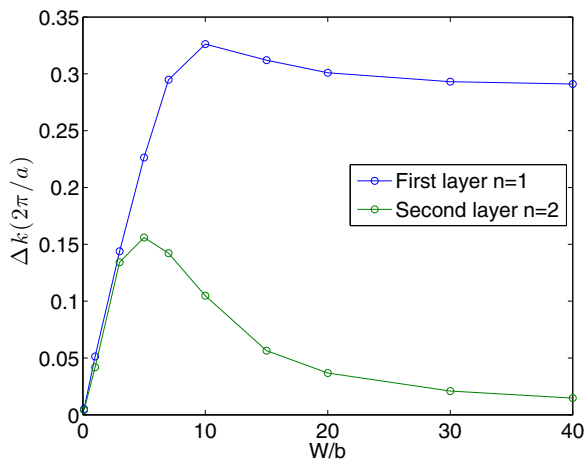


FIG. 7. (Color online) LDOS momentum-space width in the first and second layers with *magnetic* impurities, at system size $N_x = 120$. The overall behavior is the same with the case of a nonmagnetic edge impurity case (Fig. 6), but much larger systems are required to see the size-independent physical properties in the bulk regime.

$n = 1, 2$. The data is qualitatively similar to the nonmagnetic case, which demonstrates the universality of the result.

## IV. TRANSPORT BEHAVIOR

The local density of states discussed in the last section can be probed by angle-resolved photoemission spectroscopy [8]. The transport along the edge states, however, provides additional information independent of the local density of states. For instance, in the case of nonmagnetic impurities, edge modes cannot backscatter and their conductance remains quantized at the value of the clean system, despite the local density of states associated with it changing its support between the layers. To clarify the transport nature of the system with magnetic and nonmagnetic impurities along the edge, in this section we study conductance of the systems using the Landauer-Büttiker method.

Imagine a system where two semi-infinite leads are connected to a disordered region at the ends $x = 1$ and $x = N_x$. Landauer and Büttiker [33–35] related the conductance with the transmission coefficient through the disordered region, $g = \frac{e^2}{h} T_{N1}$, where $g$ is conductance and $T_{N1}$ is a transmission coefficient through the disordered region from site 1 to site $N$. Using linear response formalism, Fisher and Lee [36] expressed the transmission coefficient in terms of Green's functions:

$$g = \frac{e^2}{h} \text{Tr}[\Gamma_L G_{N1} \Gamma_R G_{N1}^\dagger], \qquad (22)$$

where $\Gamma_L = i(\Sigma_L - \Sigma_L^\dagger)$, $G_{N1}$ is a Green's function from site 1 to $N$ renormalized by the presence of the leads, and $\Sigma_L$ is a self-energy of the semi-infinite left lead. Each term in this formula can be computed recursively, such that the conductance of a long system can be obtained with a reasonable computation effort. A good review of the detailed calculation can be found in Ref. [37].

### A. Nonmagnetic impurities case

Consider the 2D topological insulator system introduced earlier with nonmagnetic impurities along the top edge. Because the Hamiltonian is diagonal in a spin basis without magnetic impurities, we can consider transport in just one spin sector. When the chemical potential is in the energy gap, neither backscattering nor scattering into the bulk is possible. Therefore, the conductance must remain quantized even in the presence of edge impurities. This is indeed what our calculation shows. Indeed, when the disorder is nonmagnetic, the transport behavior does not reflect the development of the LDOS.

### B. Magnetic impurities case

The two opposite-spin, counterpropagating, chiral edge modes couple as soon as magnetic edge impurities are introduced. As a result, transport through the disordered edge is suppressed, while the transport through the clean edge remains unaffected. Therefore, we expect the total conductance to rapidly approaches $e^2/h$ when introducing and increasing magnetic edge disorder.

However, in the strong-impurity quantum-dot limit, $W \gg N_x$, we found that the conductance recovers its clean-system value of $2e^2/h$. In this regime, the impurities are strongly
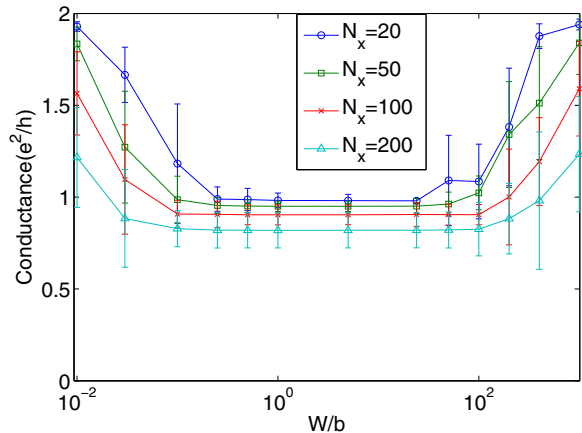
FIG. 8. (Color online) Edge channel conductance vs impurity strength when magnetic edge impurities are present. Four system sizes, $N_x = 20, 50, 100, 200$, are used. The edge channel conductance drops to $e^2/h$ initially. The conductance recovers to $2e^2/h$ as the system enters the quantum-dot regime at strong magnetic impurity strength. This reflects the nonmonotonic behavior of the localization length of the edge mode as disorder increases.

bound to electrons at energies far away from the Fermi energy and they play negligible roles in the transport at the Fermi energy. Put another way, strong disorder pushes the edge modes to the next layer where they effectively become weak scatters. This behavior is clear in Fig. 8, which shows the conductance vs the disorder strength for different system sizes. We can see that in the intermediate range of impurity strength, the conduction through the disordered top edge is significantly suppressed due to magnetic impurities, while the conductance recovers up to $2e^2/h$ value in the strong impurity limit. We note that in the thermodynamic limit $N_x \to \infty$, the conductance is always $e^2/h$ for any disorder strength, illustrated in Fig. 9. Our calculation reflects the nonmonotonic dependence of the localization length of the scattered edge mode on disorder
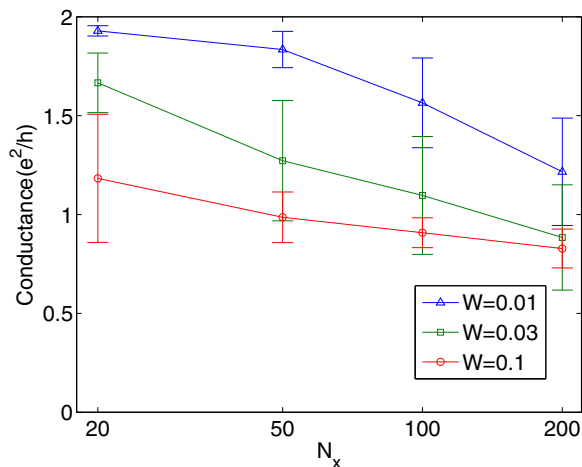


FIG. 9. (Color online) Disorder-averaged conductance with increasing system size $N_x$ for three impurity strengths $W$. The conductance exponentially decreases with system size as anticipated in the magnetic edge impurity case.

strength, which is consistent with the expulsion of the LDOS from the disordered first layer.

## V. SUMMARY AND CONCLUSION

The effects of boundary disorder in systems with boundary (edge or surface) states are the focus of our work. This problem is of fundamental as well as technological importance, given that topological quantum wells and bulk topological insulators are susceptible to boundary roughness and pollution. Such manifestations of disorder are clearly going to affect the behavior of the boundary states in these topological materials. Previously, an exact-diagonalization analysis showed that the response of a 3D topological insulator to surface disorder is nonmonotonic: First the surface states become diffusive and their conductance is suppressed. At a finite disorder strength, however, the surface states mean free path recovers and they reconstitute at the disorder-free second layer of the TI [16].

In this paper, we developed a formalism that reduces solving a bulk $D$-dimensional Hamiltonian to a surface-only diagonalization problem. This, in principle, enables an exact-diagonalization analysis of much bigger bulk systems than previously possible. Roughly speaking, our method constitutes a systematic integrating out and elimination of the bulk degrees of freedom layer by layer. We show how to carry out this procedure with a technique reminiscent of transfer-matrix methods. From the resulting surface-only diagonalization problem, we are able to reconstruct the wave functions of the bulk layers and study the effect of disorder on them. Our method is generally valid for any system composed of coupled layers but becomes particularly useful for topological systems whose surface states are fundamentally dependent on the existence of the gapped bulk.

We used our method to study 2D topological insulators in the case of either magnetic or nonmagnetic edge impurities. As the strength of the edge impurities increases, we found that edge states near zero energy become more concentrated in the disordered layer up to the moderate impurity strength (i.e., comparable to the bandwidth), and then the edge states are gradually expelled to the second layer. The ratio of the local density of states in the first two layers in Fig. 4 shows this behavior for different system sizes. Furthermore, the width of the edge state in momentum space in the second layer reaches a maximum at moderate impurity strength and then narrows at stronger impurity strength; despite the strong disorder, the edge states have momentum restored to being a good quantum number. This nonmonotonic response to the surface disorder is equally true for magnetic and nonmagnetic disorder (see Figs. 6 and 7).

The transport properties, however, show a sharp contrast between the magnetic and nonmagnetic cases: nonmagnetic edge impurities does not affect the edge transport properties, while magnetic edge impurities immediately induce a finite localization length due to backscattering, and initially suppress the edge conductance (see Fig. 8). In our numerical simulations that include magnetic impurities, however, we found that the conductance then recovers to the clean-system values when the impurity-induced localization length exceeds the system size at strong disorder.

These finding do not imply that the edge states completely decouple at large but finite disorder. Rather, it is an indication that the localization length of the spin-orbit locked modes exhibits a nonmonotonic localization behavior as a function of disorder. In our simulations, perfect conductance will be recovered when the localization length exceeds the system width. This is precisely the regime we nicknamed the quantum-dot regime. The localization length inferred from transport calculations must be proportional to the inverse of the average width of the edge LDOS in momentum space, for momenta parallel to the edge. Therefore, our transport results are yet another manifestation of the reconstitution of the edge mode at the second layer, which is disorder free. In the limit of infinitely strong disorder, we expect the localization length of the reconstituted edge state to diverge and translational invariance is regained. We emphasize that these results could not have been obtained from an analysis of effective Hamiltonians that contain an effective theory of the edge states alone. Such a model would exhibit backscattering-induced Anderson localization monotonously for all disorder strengths.

In future work, we intend to apply our method to the 3D topological insulator with surface disorder. As we emphasize throughout, one must use large enough systems in order to explore the bulk properties in the thermodynamic limit (see Figs. 3 and 4). This makes 3D topological systems with moderate to strong surface impurities harder to study using the ED, as the cost of calculation increases. Our analytic approach will then be very useful, since it reduces the computational cost dramatically.

## APPENDIX: DERIVATION OF THE EFFECTIVE HAMILTONIAN AND THE TRANSFER MATRIX

The construction of the object $M$ in Sec. II is a central element for further computation in any examples. In this appendix, we show how the object $M$ in Sec. III was constructed for given momentum $k_x$ and energy $E$. We work in the momentum space since the clean Hamiltonian is diagonalized in momentum space. But, as the layer degree of freedom is integrated out, the effective Hamiltonian is dependent of energy nontrivially, as can be seen in the recursive Eq. (4). Therefore, the dispersion relation cannot be immediately deduced from the single-layer effective Hamiltonian. Rather, we compute the quantity $\det[E - H_1^{\text{eff}} - V_{\text{imp}}]$ in momentum space as varying energy $E$ such that the determinant is zero. In this way, we find eigenenergies and then eigenstates of a surface-disordered system.

The object $M$ satisfies the layered Schrodinger equation (7). Thus, the construction is convenient in terms of the wave functions of the Schrödinger equation (1). The most general expression will be

$$M_{l+N+1} = \sum_{i,j=1,2,3,4} c_{ij} \rho_i^l |\phi_i\rangle\langle\phi_j|, \qquad (A1)$$

with the vanishing boundary condition at the last layer $M_{N+1} = 0$. There are different ways to build $M$ and we introduce one way for on-shell $E = A \sin k_x$ and off-shell $E \neq A \sin k_x$ states.

### 1. On-shell states, $E = A \sin k_x$

Considering only one spin sector of the Hamiltonian, as in Sec. III B, the eigenvalues and eigenvectors of the transfer matrix over the layer are straightforwardly obtained from the Schrödinger equation with the replacement $\psi_{n-1} = \frac{1}{\rho}\psi_n$ and $\psi_{n+1} = \rho\psi_n$,

$$B\frac{1}{\rho}\psi_n + [H_0(k_x) - E]\psi_n + B^\dagger \rho\psi_n = 0. \qquad (A2)$$

Solving a $2 \times 2$ matrix equation, we obtain four eigenvalues $\rho_{i=1,2,3,4}$ with eigenvectors $\phi_{i=1,2,3,4}$. For on-shell states, we immediately know that two eigenvectors are $|+\rangle$,

$$\phi_1 = \begin{pmatrix} \alpha_1 \\ \beta_1 \end{pmatrix}, \ \ \phi_2 = \begin{pmatrix} \alpha_2 \\ \beta_2 \end{pmatrix}, \ \ \phi_3 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \ \ \phi_4 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

with eigenvalues $|\rho_{i=3,4}| < 1$ so that $\phi_{j=3,4}$ are physically localized states at the top layer, $n = 1$. This is to satisfy the vanishing boundary condition of the clean system with two edge-localized wave functions. To construct the object $M$, we need not only the vanishing boundary condition, but also $M$ must be invertible. To do that, consider the decomposition of $\phi_{i=1,2}$ into $|+\rangle$ and $|-\rangle$:

$$\phi_{j=1,2} = \begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} = \frac{A_j}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{B_j}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}. \qquad (A3)$$

From this set of eigenvectors, we can construct two copies of $|-\rangle$'s behaving differently over the layers. Specifically,

$$|-_{(j,l)}\rangle = \frac{1}{B_j}\phi_j - \frac{A_j}{B_j}\phi_l, \qquad (A4)$$

where $j = 1, 2$ and $l = 3, 4$. Then, we can assume the following construction of the object $M$:

$$M_{N+1} = [|-_{(1,3)}\rangle - |-_{(2,4)}\rangle]\langle-| + [\phi_3 - \phi_4]\langle+|,$$

which is zero. For $n \in [0, N]$, $M_n$ is of course nonzero as eigenvalues of the eigenvectors are different, and $M_n$ is generally invertible. For the $n$th layer,

$$\langle+|M_n|+\rangle = \rho_3^{n-N-1} - \rho_4^{n-N-1}, \qquad (A5a)$$

$$\langle+|M_n|-\rangle = \frac{A_1}{B_1}\left[\rho_1^{n-N-1} - \rho_3^{n-N-1}\right]$$
$$+ \frac{A_2}{B_2}\left[\rho_4^{n-N-1} - \rho_2^{n-N-1}\right], \qquad (A5b)$$

$$\langle-|M_n|+\rangle = 0, \qquad (A5c)$$

$$\langle-|M_n|-\rangle = \rho_1^{n-N-1} - \rho_2^{n-N-1}. \qquad (A5d)$$

From this, the construction of the transfer matrix and the effective single-layer Hamiltonian is the following

$(T_{n \leftarrow m} = M_n M_m^{-1})$:

$$\langle + | T_{n \leftarrow m} | + \rangle = \frac{\rho_3^{n-N-1} - \rho_4^{n-N-1}}{\rho_3^{m-N-1} - \rho_4^{m-N-1}}, \tag{A6a}$$

$$\langle + | T_{n \leftarrow m} | - \rangle = \frac{\frac{A_1}{B_1}\rho_1^{n-N-1} - \frac{A_2}{B_2}\rho_2^{n-N-1}}{\rho_1^{m-N-1} - \rho_2^{m-N-1}} + \frac{\left(\rho_3^{n-N-1} - \rho_4^{n-N-1}\right)\left(\frac{A_2}{B_2}\rho_2^{m-N-1} - \frac{A_1}{B_1}\rho_1^{m-N-1}\right)}{\left(\rho_3^{m-N-1} - \rho_4^{m-N-1}\right)\left(\rho_1^{m-N-1} - \rho_2^{m-N-1}\right)}, \tag{A6b}$$

$$\langle - | T_{n \leftarrow m} | + \rangle = 0, \tag{A6c}$$

$$\langle - | T_{n \leftarrow m} | - \rangle = \frac{\rho_1^{n-N-1} - \rho_2^{n-N-1}}{\rho_1^{m-N-1} - \rho_2^{m-N-1}}. \tag{A6d}$$

Expressing $B = (b + A/2)|+\rangle\langle -| + (b - A/2)|-\rangle\langle +|$, the effective single-layer Hamiltonian is

$$\langle + | E - H_1^{\text{eff}} | + \rangle = (b - A/2)\langle -| T_{0 \leftarrow 1} | + \rangle, \tag{A7a}$$

$$\langle + | E - H_1^{\text{eff}} | - \rangle = (b - A/2)\langle -| T_{0 \leftarrow 1} | - \rangle, \tag{A7b}$$

$$\langle - | E - H_1^{\text{eff}} | + \rangle = (b + A/2)\langle +| T_{0 \leftarrow 1} | + \rangle, \tag{A7c}$$

$$\langle - | E - H_1^{\text{eff}} | - \rangle = (b + A/2)\langle +| T_{0 \leftarrow 1} | + \rangle. \tag{A7d}$$

One can see that the effective Hamiltonian is still system-size dependent as the $H_1^{\text{eff}}(k_x = 0, E = 0)$ case computed in the main text. Plus, $E - H_1^{\text{eff}}$ contains no zero eigenvalues, implying that any finite system cannot have the energy dispersion $E = A \sin k_x$ due to finite-size effects. Therefore, the expression of the Hamiltonian for on-shell states is not useful for actual computation. Rather, we need the Hamiltonian expression of off-shell states for a finite-size system with surface impurities, for which we find an analytic expression of the effective single-layer Hamiltonian and let $N \to \infty$ for the semi-infinite limit.

### 2. Off-shell states, $E \neq A \sin k_x$

For a system with surface impurities, eigenstates are not described by the clean-system dispersion $E = A \sin k_x$; rather, the state has a mix of different momentum components in each given energy. More concretely, if the size of the system along the periodic boundary condition is $N_x$, the exact edge-state dispersion of the clean system discussed in Sec. III is $E = A \sin(2\pi l / N_x)$ with integer $l$. Only at those discrete set of energies do we have two eigenvectors parallel to $|+\rangle$ and the discussion in the previous section applies. Except for those on-shell points, we have the following general set of eigenvectors:

$$\phi_1 = \begin{pmatrix} \alpha_1 \\ \beta_1 \end{pmatrix}, \quad \phi_2 = \begin{pmatrix} \alpha_2 \\ \beta_2 \end{pmatrix}, \quad \phi_3 = \begin{pmatrix} \alpha_3 \\ \beta_3 \end{pmatrix}, \quad \phi_4 = \begin{pmatrix} \alpha_4 \\ \beta_4 \end{pmatrix}, \tag{A8}$$

with eigenvalues $\rho_{j=1,2,3,4}$. Without loss of generality, let us say $|\rho_{j=2,4}| > 1$. Each eigenvector can be written as the sum of $|+\rangle$ and $|-\rangle$, like Eq. (A3). Then, we similarly construct

two pairs of $|+\rangle$ and $|-\rangle$ by the superposition of $\phi_{j=1,2,3,4}$,

$$|+_{(j,l)}\rangle = \frac{1}{\frac{A_j}{B_j} - \frac{A_l}{B_l}}\left(\frac{1}{B_j}\phi_j - \frac{1}{B_l}\phi_l\right),$$

$$|-_{(j,l)}\rangle = \frac{1}{\frac{B_j}{A_j} - \frac{B_l}{A_l}}\left(\frac{1}{A_j}\phi_j - \frac{1}{A_l}\phi_l\right). \tag{A9}$$

Next, we can construct the object $M_n$ satisfying the vanishing boundary condition at $n = N + 1$,

$$M_{N+1} = (|-_{(1,2)}\rangle - |-_{(3,4)}\rangle)\langle -| + (|+_{(1,2)}\rangle - |+_{(3,4)}\rangle)\langle +|, \tag{A10}$$

where we intentionally split $\phi_2$ and $\phi_4$ in each term so that no terms vanish in the $N \to \infty$ limit. Explicitly, the components are

$$\langle +| M_{l+N+1} |+\rangle = \frac{1}{\frac{A_1}{B_1} - \frac{A_2}{B_2}}\left[\frac{A_1}{B_1}\rho_1^l - \frac{A_2}{B_2}\rho_2^l\right]$$
$$- \frac{1}{\frac{A_3}{B_3} - \frac{A_4}{B_4}}\left[\frac{A_3}{B_3}\rho_3^l - \frac{A_4}{B_4}\rho_4^l\right], \tag{A11a}$$

$$\langle -| M_{l+N+1} |+\rangle = \frac{1}{\frac{A_1}{B_1} - \frac{A_2}{B_2}}\left[\rho_1^l - \rho_2^l\right] - \frac{1}{\frac{A_3}{B_3} - \frac{A_4}{B_4}}\left[\rho_3^l - \rho_4^l\right], \tag{A11b}$$

$$\langle +| M_{l+N+1} |-\rangle = \frac{1}{\frac{B_1}{A_1} - \frac{B_2}{A_2}}\left[\rho_1^l - \rho_2^l\right] - \frac{1}{\frac{B_3}{A_3} - \frac{B_4}{A_4}}\left[\rho_3^l - \rho_4^l\right], \tag{A11c}$$

$$\langle -| M_{l+N+1} |-\rangle = \frac{1}{\frac{B_1}{A_1} - \frac{B_2}{A_2}}\left[\frac{B_1}{A_1}\rho_1^l - \frac{B_2}{A_2}\rho_2^l\right]$$
$$- \frac{1}{\frac{B_3}{A_3} - \frac{B_4}{A_4}}\left[\frac{B_3}{A_3}\rho_3^l - \frac{B_4}{A_4}\rho_4^l\right]. \tag{A11d}$$

We are interested in the behavior of edge states near the top ($n = 1$). Thus, in the limit of $N \to \infty$, $l = n - N - 1 \to -\infty$

and the terms with $(\frac{\rho_{j=1,3}}{\rho_{j=2,4}})^l$ dominate,

$$\langle+|T_{n\leftarrow m}|+\rangle = \frac{\frac{A_1}{B_1}\rho_1^{n-m} - \frac{A_3}{B_3}\rho_3^{n-m}}{\frac{A_1}{B_1} - \frac{A_3}{B_3}}, \quad \langle+|T_{n\leftarrow m}|-\rangle = \frac{\rho_1^{n-m} - \rho_3^{n-m}}{\frac{A_1}{B_1} - \frac{A_3}{B_3}},$$

$$\langle-|T_{n\leftarrow m}|+\rangle = \frac{\rho_3^{n-m} - \rho_1^{n-m}}{\frac{B_1}{A_1} - \frac{B_3}{A_3}}, \quad \langle-|T_{n\leftarrow m}|-\rangle = \frac{\frac{B_3}{A_3}\rho_3^{n-m} - \frac{B_1}{A_1}\rho_1^{n-m}}{\frac{B_1}{A_1} - \frac{B_3}{A_3}}. \tag{A12}$$

And the expression for the Hamiltonian is following Eqs. (A7a)–(A7d).

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

[2] M. Z. Hasan and J. E. Moore, Annu. Rev. Condens. Matter Phys. **2**, 55 (2011).

[3] M. Franz and L. Molenkamp, *Topological Insulators*, Vol. 6 (Elsevier, New York, 2013).

[4] X.-L. Qi and S.-C. Zhang, Rev. Mod. Phys. **83**, 1057 (2011).

[5] D. Hsieh, Y. Xia, D. Qian, L. Wray, J. H. Dil, F. Meier, J. Osterwalder, L. Patthey, J. G. Checkelsky, N. P. Ong, A. V. Fedorov, H. Lin, A. Bansil, D. Grauer, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Nature (London) **460**, 1101 (2009).

[6] Y. Chen, J. Analytis, J.-H. Chu, Z. Liu, S.-K. Mo, X.-L. Qi, H. Zhang, D. Lu, X. Dai, Z. Fang *et al.*, Science **325**, 178 (2009).

[7] Y. Xia, D. Qian, D. Hsieh, L. Wray, A. Pal, H. Lin, A. Bansil, D. Grauer, Y. Hor, R. Cava *et al.*, Nat. Phys. **5**, 398 (2009).

[8] L. A. Wray, S.-Y. Xu, Y. Xia, D. Hsieh, A. V. Fedorov, Y. S. Hor, R. J. Cava, A. Bansil, H. Lin, and M. Z. Hasan, Nat. Phys. **7**, 32 (2011).

[9] S. Sasaki, M. Kriener, K. Segawa, K. Yada, Y. Tanaka, M. Sato, and Y. Ando, Phys. Rev. Lett. **107**, 217001 (2011).

[10] T. H. Hsieh and L. Fu, Phys. Rev. Lett. **108**, 107005 (2012).

[11] V. Mourik, K. Zuo, S. Frolov, S. Plissard, E. Bakkers, and L. Kouwenhoven, Science **336**, 1003 (2012).

[12] A. Das, Y. Ronen, Y. Most, Y. Oreg, M. Heiblum, and H. Shtrikman, Nat. Phys. **8**, 887 (2012).

[13] J. Alicea, Rep. Prog. Phys. **75**, 076501 (2012).

[14] C. W. J. Beenakker, Annu. Rev. Condens. Matter Phys. **4**, 113 (2013).

[15] S. R. Elliott and M. Franz, Rev. Mod. Phys. **87**, 137 (2015).

[16] G. Schubert, H. Fehske, L. Fritz, and M. Vojta, Phys. Rev. B **85**, 201105 (2012).

[17] F. Pientka, G. Kells, A. Romito, P. W. Brouwer, and F. von Oppen, Phys. Rev. Lett. **109**, 227006 (2012).

[18] D. Rainis, L. Trifunovic, J. Klinovaja, and D. Loss, Phys. Rev. B **87**, 024515 (2013).

[19] R. Queiroz and A. P. Schnyder, Phys. Rev. B **91**, 014202 (2015).

[20] A. K. Mitchell, D. Schuricht, M. Vojta, and L. Fritz, Phys. Rev. B **87**, 075430 (2013).

[21] L. Fritz and M. Vojta, Rep. Prog. Phys. **76**, 032501 (2013).

[22] P. Tang, P. Chen, W. Cao, H. Huang, S. Cahangirov, L. Xian, Y. Xu, S.-C. Zhang, W. Duan, and A. Rubio, Phys. Rev. B **90**, 121408 (2014).

[23] X. Qian, J. Liu, L. Fu, and J. Li, Science **346**, 1344 (2014).

[24] J.-J. Zhou, W. Feng, C.-C. Liu, S. Guan, and Y. Yao, Nano Lett. **14**, 4767 (2014).

[25] T. Zhang, P. Cheng, X. Chen, J.-F. Jia, X. Ma, K. He, L. Wang, H. Zhang, X. Dai, Z. Fang, X. Xie, and Q.-K. Xue, Phys. Rev. Lett. **103**, 266803 (2009).

[26] Y. Zhang, K. He, C.-Z. Chang, C.-L. Song, L.-L. Wang, X. Chen, J.-F. Jia, Z. Fang, X. Dai, W.-Y. Shan *et al.*, Nat. Phys. **6**, 584 (2010).

[27] C.-X. Liu, H. J. Zhang, B. Yan, X.-L. Qi, T. Frauenheim, X. Dai, Z. Fang, and S.-C. Zhang, Phys. Rev. B **81**, 041307 (2010).

[28] C.-Z. Chang, J. Zhang, X. Feng, J. Shen, Z. Zhang, M. Guo, K. Li, Y. Ou, P. Wei, L.-L. Wang *et al.*, Science **340**, 167 (2013).

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

[30] G. W. Bryant, Phys. Rev. B **35**, 5547 (1987).

[31] B. A. Bernevig, T. L. Hughes, and S.-C. Zhang, Science **314**, 1757 (2006).

[32] K.-I. Imura, A. Yamakage, S. Mao, A. Hotta, and Y. Kuramoto, Phys. Rev. B **82**, 085118 (2010).

[33] R. Landauer, IBM J. Res. Dev. **1**, 223 (1957).

[34] R. Landauer, Philos. Mag. **21**, 863 (1970).

[35] M. Büttiker, Phys. Rev. B **38**, 12724 (1988).

[36] D. S. Fisher and P. A. Lee, Phys. Rev. B **23**, 6851 (1981).

[37] G. Metalidis and P. Bruno, Phys. Rev. B **72**, 235304 (2005).