Fingerprints of Berry phases in the bulk exciton spectrum of a topological insulator

Andrew A. Allocca,[1] Dmitry K. Efimkin,[2] and Victor M. Galitski[1]

[1]*Joint Quantum Institute and Condensed Matter Theory Center, Department of Physics, University of Maryland, College Park, Maryland 20742-4111, USA*

[2]*The Center for Complex Quantum Systems, The University of Texas at Austin, Austin, Texas 78712-1192, USA*

(Received 26 January 2018; published 27 July 2018)

We examine excitons formed in the bulk of a topological insulator as the system is tuned via a parameter between topological and trivial insulating phases, arguing that nontrivial topology has fingerprints in the spectrum of these excitons. The closely related hydrogen atom problem is well known to have a degeneracy due to a hidden symmetry, and the changes to the excitonic spectrum that we find can be understood as a result of breaking of this underlying symmetry due to the Berry phase. Furthermore, this phase is found to affect the spectrum in the topological parameter regime much more strongly than in the trivial regime. We first construct a semiclassical model of the system to develop qualitative intuition for the effects at play, then we move to a more robust numerical simulation of the full quantum system, working with the Bernevig-Hughes-Zhang model of a two-dimensional topological insulator.

DOI: 10.1103/PhysRevB.98.045430

I. INTRODUCTION

Since the discovery of topological insulators (TIs), much work has been done exploring how these new materials can be used to realize exotic new physical phenomena [1,2]. One of their defining features is the fact that these materials support robust conducting Dirac states on their surfaces, which themselves have been the focus of a great amount of research. From the perspective of exciton physics in particular, there have been investigations into the impact that exciton physics at the surface of topological insulators may have on the materials' optical properties [3,4], as well as the possibility of exotic interaction effects such as chiral excitons [5] or as a platform for potential realizations of excitonic condensation [6,7]. Another work examined the impact of exciton condensation on the quantum spin Hall effect [8].

Something that has been largely overlooked is how the topological nature of these materials manifests in the properties of the bulk. Far away from the surface, though global properties are different, the band structure of a topological insulator is qualitatively very similar to that of a trivial insulator or even a semiconductor with a large band gap. Consequently, optical and transport properties are naively expected to be similar as well, and indeed electrical conductivity through the bulk is exponentially small in the size of the gap for both trivial and topological insulators. Only a few studies have been done, however, exploring the effect of nontrivial topological character on other physical phenomena in the bulk. One such study examined the polarization properties of a two-dimensional (2D) topological model, concluding that features of the optical conductivity of this model, including a plasmon resonance absent in graphene or usual two-dimensional electron gasses (2DEGs), provide a way to identify its topological character via bulk measurements [9]. Another work found that phonon linewidths of bulk optical phonons contain information on band inversions in the electronic spectrum [10]. Here we add to this line of inquiry, investigating how the properties of excitons formed from the bulk bands of a topological insulator depend on the topological character.

We consider excitons in the bulk of a 2D model that can be continuously tuned between topologically trivial and nontrivial parameter regimes—the well-studied Bernevig-Hughes-Zhang (BHZ) model [11], developed to describe the band-inversion physics and resulting topological phase of 2D Hg(Cd)Te quantum wells. It is given by the Hamiltonian

$$H_{\mathrm{BHZ}}(\mathbf{p}) = \begin{pmatrix} \hat{h}_{\mathbf{p}} & 0 \\ 0 & \hat{h}^{*}_{-\mathbf{p}} \end{pmatrix},$$
$$\hat{h}_{\mathbf{p}} = \epsilon_p \hat{\mathbf{1}} + \mathbf{d}_{\mathbf{p}} \cdot \hat{\boldsymbol{\tau}}, \quad \mathbf{d}_{\mathbf{p}} = (Ap_x, -Ap_y, M_{\mathbf{p}}), \quad (1)$$

where $\epsilon_p = C - Dp^2$ is the electron-hole asymmetry, $M_{\mathbf{p}} = M - Bp^2$ is the momentum-dependent Dirac mass, $\hat{\boldsymbol{\tau}}$ is the vector of Pauli matrices, $\hat{\mathbf{1}}$ is the unit matrix, and A, B, C, D, and M are material parameters. The Hamiltonian is invariant under both time reversal and inversion, discussed in Appendix A. Importantly for our purposes, the masslike parameter M is related to the thickness of the quantum well and can be tuned between positive and negative values. Changing this sign changes the relative sign of the $p = 0$ and $p \to \infty$ limits of the mass term $M_{\mathbf{p}}$, corresponding to two topologically distinct phases. Though excitons have not been observed in the quantum wells described by this model, because of its simplicity it provides the ideal theoretical testbed for our analysis.

Important quantities to consider in the context of topological insulators are the Berry connection, the Berry curvature, and the resulting Berry phase. It has been well established that Berry physics can lead to a shift and splitting of otherwise degenerate exciton energy levels even in a system with trivial topological character [12,13], i.e., without the usual hallmarks of topological phases such as protected edge states and a nontrivial topological index such as the Chern number. It is therefore reasonable to expect that similar effects will be seen

 ©2018 American Physical Society

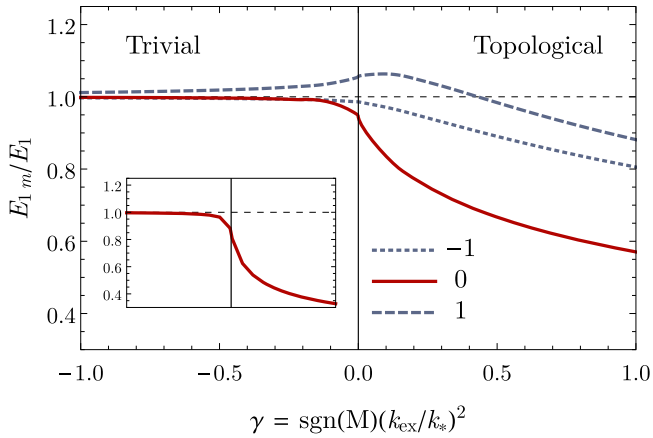

FIG. 1. The $n = 1$ energies for excitons formed from the particles in the upper BHZ block as a function of the dimensionless parameter $\gamma = \mathrm{sgn}(M)(k_{\mathrm{ex}}/k_*)^2 \propto M$, scaled by the $n = 1$ 2D hydrogen energy. The quantities k_* and k_{ex} are defined following Eq. (7). The sign of γ determines the topological phase, as labeled. The inset shows the single $n = 0$ state for the same range of γ, scaled by the $n = 0$ hydrogen energy. There is a clear qualitative difference in the behavior of the energy levels on either side of the transition, with a crossover between them. In the topological phase there is a large splitting of states due to Berry physics that is absent in the trivial phase. Furthermore, the $m = 0$ states change energy quickly above the topological transition. Energies are obtained using the effective fine-structure constant $\alpha = 0.4$.

in the topological phase of the BHZ model, which displays all of these features as a direct result of Berry physics.

We find that these expectations are indeed true, with key features of our main results given in Fig. 1; the hierarchy of exciton energy levels is drastically altered as one moves from the topologically trivial phase through the topological transition into the nontrivial phase. Within the topological phase, states with opposite orbital angular momentum are well split from each other and the $m = 0$ angular momentum state is pushed to a smaller binding energy than the rest, all of which would be degenerate in the absence of Berry curvature. In the trivial phase, however, all states are nearly degenerate, with the splitting decreasing the further one tunes away from the topological transition. Though there is no sharp feature at the topological transition itself that distinguishes these two regimes, as one might expect from a topological effect, this behavior can nonetheless be explained as arising from effects intimately tied to topological character.

It is worth mentioning for completeness that other phenomena could potentially lead to splitting of exciton states [14], such as nonparabolicity in the spectrum [15,16], screening of the Coulomb interaction in 2D systems [17,18], or the exchange interaction between particles and holes [19]. None of these, however, are present in the model we use to generate these results. Moreover, the effect we find is distinct and distinguishable since none of these other effects are sensitive to changes in topology, and so they would lead to splitting of states across the entire parameter range, not just on a single side of the topological transition.

In Sec. II we begin by presenting an intuitive understanding of the physics at play in this system, considering a semiclassical model as well as an effective Hamiltonian for Dirac-like systems. In Sec. III we formulate the full exciton problem and discuss further how topological effects will be manifested. In Sec. IV we discuss the methods used to numerically calculate the exciton spectra in the regimes of interest, and we present our main numerical results.

II. SEMICLASSICAL APPROACH

Before presenting the full quantum-mechanical exciton problem, we discuss its semiclassical counterpart. This not only gives a clear physical picture of the role of the Berry phase, but it also captures its effect on the electronic spectrum. The reason is that the semiclassical method applied to the usual 2D excitonic Coulomb problem reproduces the full spectrum exactly, and not just the structure of highly excited states. This remarkable result provides a fair assurance that our analysis here will provide useful insights into the problem at hand.

We start from the Lagrangian $L(\mathbf{r}_{\mathrm{e}}, \mathbf{r}_{\mathrm{h}}, \mathbf{p}_{\mathrm{e}}, \mathbf{p}_{\mathrm{h}})$ for the dynamics of interacting electron and hole wave packets [20–22] given by

$$L = \sum_{\alpha=\mathrm{e,h}} \left(\dot{\mathbf{r}}_\alpha \cdot \mathbf{p}_\alpha + \dot{\mathbf{p}}_\alpha \cdot \mathbf{A}_{\alpha,\mathbf{p}_\alpha} - E^\alpha_{\mathbf{p}_\alpha}\right) - V(\mathbf{r}_{\mathrm{e}} - \mathbf{r}_{\mathrm{h}}). \quad (2)$$

Here $\mathbf{r}_{\mathrm{e(h)}}$ is the location of the electron (hole) wave packet and $\mathbf{p}_{\mathrm{e(h)}}$ is its momentum. We approximate their dispersions as quadratic in the vicinity of band minima as $E^\alpha_{\mathbf{p}} = \alpha\,\epsilon_p + \mathbf{p}^2/2m$, where ϵ_p is the particle-hole asymmetry as defined after Eq. (1). $V(\mathbf{r}) = e^2/\epsilon r$ is the Coulomb interaction with dielectric constant ϵ. The function $\mathbf{A}_{\alpha,\mathbf{p}_\alpha} = i\langle \mathbf{p}, \alpha | \nabla_p | \mathbf{p}, \alpha \rangle$ is the Berry connection, calculated from the particle and hole states of the BHZ Hamiltonian (1). We consider only intrablock excitons with zero center-of-mass momentum $\mathbf{q}_{\mathrm{CM}} = \mathbf{0}$ since only they are optically active and are probed in experiments. The resulting Lagrangian for the relative motion of the electron and hole is given by

$$L = \dot{\mathbf{r}} \cdot \mathbf{p} + \dot{\mathbf{p}} \cdot \mathbf{A}_{\mathbf{p}} - \frac{\mathbf{p}^2}{2\mu} - V(\mathbf{r}), \quad (3)$$

where $\mu = m/2$ is the reduced electron-hole effective mass, and the corresponding energy is independent of electron-hole asymmetry $\epsilon_{\mathbf{p}}$ of the BHZ model; $\mathbf{A}_{\mathbf{p}} = \mathbf{A}_{\mathrm{e},\mathbf{p}} + \mathbf{A}_{\mathrm{h},-\mathbf{p}}$ is the Berry connection for the relative electron-hole motion, which for the BHZ model is found to be (see Appendix B)

$$\mathbf{A}_{\mathbf{p}} = -\left(s + \frac{M_{\mathbf{p}}}{|\mathbf{d}_{\mathbf{p}}|}\right)\frac{\hat{\mathbf{z}} \times \mathbf{p}}{p^2}, \quad (4)$$

where $s \equiv \mathrm{sgn}\, M/B$ with the corresponding Berry curvature

$$\Omega_{\mathbf{p}} = \nabla_p \times \mathbf{A}_{\mathbf{p}} = A^2 \frac{M + Bp^2}{|\mathbf{d}_{\mathbf{p}}|^3}\hat{\mathbf{z}}. \quad (5)$$

These two functions contain the topological information of two particle states within the model, with the integral of the Berry curvature over all momentum space giving the Chern number, a topological invariant that distinguishes topological and trivial phases. Note that the Berry connection is not gauge-invariant, changing by the divergence of a scalar function if the state vectors are transformed by multiplication with a momentum-

dependent phase, but the Berry curvature is invariant under such transformations.

The Euler-Lagrange equations obtained from (3) are given by

$$\dot{\mathbf{p}} = -\nabla_r V(\mathbf{r}), \quad \dot{\mathbf{r}} = \frac{\mathbf{p}}{\mu} + \nabla_r V(\mathbf{r}) \times \mathbf{\Omega}_\mathbf{p}. \tag{6}$$

The term containing the electron-hole Berry curvature is the anomalous velocity. Examining these equations in polar coordinates shows that the anomalous velocity contributes only to the angular motion of the exciton, with left-spinning and right-spinning states affected exactly oppositely, breaking the symmetry between them that is present in the absence of the Berry curvature. In particular, the anomalous velocity changes the usual expression detailing the conservation of angular momentum to

$$L_z = [\mathbf{r} \times \mathbf{p}]_z + \frac{\Phi_\mathbf{p}}{2\pi}\hat{\mathbf{z}}. \tag{7}$$

Here $\Phi_\mathbf{p}$ is the Berry phase acquired by traversing a circular trajectory with momentum $\mathbf{p}$. From this observation, one can anticipate that states with opposite angular momenta will not have the same energy in this system, unlike the case for the 2D hydrogen atom.

We can begin to get a more quantitative intuition for how topologically relevant physics comes into play through the function $\Phi_\mathbf{p}$. To do so, we compare two momentum scales. The first is the topological scale $k_* = \sqrt{|M/B|}$. In the topological regime, since M and B have the same sign, the momentum-dependent Dirac mass $M_\mathbf{k} = M - Bk^2$ changes sign at this momentum, making it the most relevant momentum scale in the context of topological effects. The Berry curvature reaches its maximum value near this momentum in the topological phase, and it is the point at which the function $\Phi_\mathbf{p}$ grows toward approximating the Chern number. In the trivial regime, though this is a well-defined momentum, there are no effects of note at this scale. The second scale is the characteristic momentum for excitonic physics, related to the inverse Bohr radius of the exciton, $k_{\text{ex}} = 1/a_B = \mu e^2/\epsilon$. In the topological phase, if the excitonic momentum is small compared to the topological scale, then the Berry phase term is likewise small. This is always the case in the trivial phase since the Berry curvature is small for all reasonable momenta (see Appendix B). If the ratio of these two scales becomes even moderately sized in the topological phase, however, then the Berry phase will become a nontrivial perturbation to the angular momentum. Though it is not immediately apparent how such a shift will affect the exciton spectrum, it is clear that any effect will only occur in the topological phase when k_{ex}/k_* becomes sufficiently large.

Additional insight can be gained by changing coordinates, rewriting the angular momentum as $L_z = (\mathbf{R} \times \mathbf{p})_z$. Here $\mathbf{R} = \mathbf{r} - \mathbf{A}_\mathbf{p}$ and $\mathbf{p}$ are the canonical coordinates of the problem, and the shift in the position coordinate is the momentum space equivalent of the Peierls substitution, which takes the Berry connection correctly into account. Using these coordinates, the equations of motion (6) can be derived from the effective Hamiltonian $H_{\text{eff}}(\mathbf{R}, \mathbf{p})$ given by

$$H_{\text{eff}} = \frac{\mathbf{p}^2}{2\mu} + V(\mathbf{R} + \mathbf{A}_\mathbf{p}). \tag{8}$$

Expanding in the Berry connection and taking into account its solenoidal distribution in momentum space, we get $H_{\text{eff}} = H_0 + \Delta H$, where H_0 is the Hamiltonian of the usual Coulomb problem,

$$H_0 = \frac{p_R^2}{2\mu} + \frac{L_z^2}{2\mu R^2} - \frac{e^2}{\epsilon R}, \tag{9}$$

with $p_R = \mathbf{p} \cdot \hat{\mathbf{R}}$, and ΔH is the correction due to the presence of the Berry curvature given by

$$\Delta H = \left(s + \frac{M_p}{|\mathbf{d}_p|}\right)\frac{e^2}{\epsilon R}\frac{L_z}{(Rp)^2}. \tag{10}$$

Examining this correction term, we see that it acts as a perturbation to the 2D hydrogen atom problem and will generically split energy levels with differing angular momentum, which we determined to be the effect of the Berry phase above.

The expression for ΔH we find here is a generalization of a correction derived previously using the Foldy-Wouthuysen transformation in the particular case of a constant Berry curvature [13,23]. The Foldy-Wouthuysen transformation produces a consistent quadratic approximation to a Hamiltonian with a linear dispersion at high energies. Because we have approximated the dispersion as purely quadratic in our semiclassical analysis thus far, we have missed the Darwin term, which has the form $H_{\text{Darwin}} = \frac{1}{4}\Omega_z \nabla^2 V(R)$, where $\Omega_z = \Omega_z(\mathbf{p} = \mathbf{0})$. This term gives an effective shift of the angular momentum by a value of $\frac{1}{2}$ in the perturbation to the hydrogen atom problem. This shift leads to an asymmetric splitting of states with opposite angular momentum as well as a shift for the $m = 0$ state, which, in this effective Hamiltonian picture, would otherwise remain unaffected by Berry physics.

III. EXCITONIC STATES

Excitons are two-particle electron-hole bound states formed due to Coulomb interactions. Only excitons with zero total momentum $\mathbf{q}_{\text{CM}} = \mathbf{0}$ are optically active and will be considered here. Excitonic states can be written

$$|X_{i,j}\rangle = \sum_\mathbf{k} C_\mathbf{k}^{(ij)}\, a^\dagger_{\mathbf{k},+,i} a_{\mathbf{k},-,j}|0\rangle. \tag{11}$$

Here $C_\mathbf{k}^{(ij)}$ is the wave function of the exciton in momentum space, $|0\rangle$ is the state with filled valence bands and empty conduction bands, and $a^\dagger_{\mathbf{k},\alpha,i}$ ($a_{\mathbf{k},\alpha,i}$) creates (destroys) the single-particle state $|\mathbf{k}, \alpha, i\rangle$, where $\alpha = \pm$ labels the band and $i = 1, 2$ labels the block of the BHZ Hamiltonian from which the state is taken. When rotated to the band basis, each of the two 2×2 blocks of the BHZ Hamiltonian produces a single conduction and valence band, hosting the electrons and holes that are the building blocks of excitons. When $i = j$, the electron and hole come from the same block (intrablock excitons), while the case of $i \neq j$ corresponds to an interblock exciton. In general, intrablock excitons are optically active while interblock excitons require some degree of inversion symmetry breaking to be accessible via optical means. We do not consider such symmetry breaking in our model, but we calculate interblock exciton energies nonetheless as a point of comparison.

The exciton wave function satisfies a Schrodinger-like equation in momentum space given by

$$2|\mathbf{d}_\mathbf{k}|C^{(ij)}_\mathbf{k} - \sum_{\mathbf{k}'} U_{\mathbf{k}-\mathbf{k}'}\mathcal{F}^{(ij)}_{\mathbf{k},\mathbf{k}'}C^{(ij)}_{\mathbf{k}'} = (E_g + E_X)C^{(ij)}_\mathbf{k}. \quad (12)$$

Here $E_g = 2|M|$ is the energy gap and $E_X < 0$ is the exciton binding energy. The screened Coulomb interaction is given by $U_\mathbf{q} = 2\pi e^2/\epsilon q$, with ϵ being the effective dielectric constant of the surrounding medium, and $2|\mathbf{d}_\mathbf{k}| = E_{k,+,i} - E_{k,-,j}$ is the two-particle free dispersion, independent of the particle-hole asymmetry. Finally, $\mathcal{F}^{(ij)}$ is a function resulting from the rotation from the original basis of the Hamiltonian to the band basis, and its importance will be discussed at length.

In our further analysis, we will approximate the two-particle dispersion with a constant parabolic dispersion,

$$2|\mathbf{d}_\mathbf{k}| \to 2|M| + \frac{k^2}{2\mu} \quad \text{with } \mu = \frac{|M|}{2A^2}. \quad (13)$$

This changes the features of the spectrum in the topological phase, which for the unmodified BHZ model develops a degenerate band minimum at a finite momentum for large enough $|M|$; in a more accurate approximation there is a value of M for which the effective mass near $k = 0$ changes sign within the topological phase. This alone can lead to large effects on excitonic properties but is completely unrelated to the topological transition. Since smooth deformations of the band structure leave topological properties unchanged, the above simplification is one way to remove this parametric dependence of the model on M while leaving topological properties intact. This allows us to more easily isolate the effect of topology alone.

Since this model has rotational symmetry, we can perform a full multipole decomposition, writing the exciton wave function as

$$C_\mathbf{k} = \sum_m C_m(k)e^{im\varphi_\mathbf{k}}, \quad (14)$$

where $\varphi_\mathbf{k}$ is the angle of $\mathbf{k}$ from the k_x axis. The eigenvalue equation itself becomes

$$\frac{k^2}{2\mu}C_m(k) - \sum_{k'} U^{\text{eff}}_m(k,k')C_m(k') = E_X C_m(k), \quad (15)$$

with U^{eff}_m, the effective interaction in the m channel, given by

$$U^{\text{eff}}_m(k,k') = \sum_{m'} U_{m-m'}(k,k')F_{m'}(k,k'). \quad (16)$$

Being the index related to rational invariance, m is a component of the angular momentum of the exciton, specifically the component related to the relative motion of its constituents (see Appendix A for the full angular momentum). It should be noted that the choice of the underlying spinor wave functions of electrons and holes is not unique (see Appendix B), and one can change them up to an arbitrary gauge transformation. Though gauge transformations leave all observables unchanged, they can in general uniformly shift the label m by any integer, making this label of excitonic states ambiguous and dependent on gauge choice. The gauge that we employ is chosen to reduce to the normal labeling of states for the 2D hydrogen atom in the limit $MB \to -\infty$, infinitely far into the trivial regime.

The function $\mathcal{F}^{(ij)}_{\mathbf{k},\mathbf{k}'} = \langle \mathbf{k},+,i|\mathbf{k}',+,i\rangle\langle \mathbf{k}',-,j|\mathbf{k},-,j\rangle$ results from the change to the band basis and is given by the overlaps of electron and hole spinor wave functions. We can explicitly write this function as

$$\begin{aligned}
\mathcal{F}^{(11)}_{\mathbf{k},\mathbf{k}'} &= e^{i(s-1)(\varphi_\mathbf{k}-\varphi_{\mathbf{k}'})}\cos^2\frac{\theta_\mathbf{k}}{2}\cos^2\frac{\theta_{\mathbf{k}'}}{2} \\
&\quad + e^{i(s+1)(\varphi_\mathbf{k}-\varphi_{\mathbf{k}'})}\sin^2\frac{\theta_\mathbf{k}}{2}\sin^2\frac{\theta_{\mathbf{k}'}}{2} \\
&\quad + 2e^{is(\varphi_\mathbf{k}-\varphi_{\mathbf{k}'})}\cos\frac{\theta_\mathbf{k}}{2}\cos\frac{\theta_{\mathbf{k}'}}{2}\sin\frac{\theta_\mathbf{k}}{2}\sin\frac{\theta_{\mathbf{k}'}}{2}, \\
\mathcal{F}^{(12)}_{\mathbf{k},\mathbf{k}'} &= \cos^2\frac{\theta_\mathbf{k}}{2}\cos^2\frac{\theta_{\mathbf{k}'}}{2} + \sin^2\frac{\theta_\mathbf{k}}{2}\sin^2\frac{\theta_{\mathbf{k}'}}{2} \\
&\quad + 2\cos\frac{\theta_\mathbf{k}}{2}\cos\frac{\theta_{\mathbf{k}'}}{2}\sin\frac{\theta_\mathbf{k}}{2}\sin\frac{\theta_{\mathbf{k}'}}{2}\cos(\varphi_\mathbf{k}-\varphi_{\mathbf{k}'}), \quad (17)
\end{aligned}$$

with $\mathcal{F}^{(11)} = \mathcal{F}^{(22)*}$, $\mathcal{F}^{(12)} = \mathcal{F}^{(21)}$, $\cos\theta_\mathbf{k} = M_\mathbf{k}/|\mathbf{d}_\mathbf{k}|$, and $s \equiv \text{sgn}\, M$. This is the only ingredient in the excitonic eigenvalue equation (12) that reflects the underlying topology, and it is qualitatively different in trivial and topological regimes. The topological information carried in these functions can be seen explicitly by considering their multipole expansions,

$$\begin{aligned}
\mathcal{F}^{(11)}_{\mathbf{k},\mathbf{k}'} &= \sum_{m=0}^{2} F_{sm}(k,k')e^{ism(\varphi_\mathbf{k}-\varphi_{\mathbf{k}'})}, \\
\mathcal{F}^{(12)}_{\mathbf{k},\mathbf{k}'} &= F_0(k,k') + F_{2s}(k,k') + F_s(k,k')\cos(\varphi_\mathbf{k}-\varphi_{\mathbf{k}'}),
\end{aligned} \quad (18)$$

and examining the behavior of the three functions F_{sm} in both the trivial and topological phases. These functions are plotted in Fig. 2.

Far into the trivial parameter regime, one sees that F_0 is approximately equal to 1 for all values of k, k', while the other functions are very small. Indeed in the limit $MB \to -\infty$, then $F_0 \to 1$ and $F_{m\neq 0} \to 0$, so $\mathcal{F} \to 1$ and the Schrödinger equation approaches exactly that for the 2D hydrogen atom.

A similar statement cannot be made in the topological parameter regime, with the behavior of $\mathcal{F}$ being fundamentally nontrivial for all values of the tuning parameter. In this regime, both F_0 and F_{-2} show nontrivial behavior as one or both of their arguments become large compared to the topological scale k_*. The remaining function F_{-1} does not display such a drastic change on either side of the topological transition, though in the topological phase it always reaches the value $1/2$ for $k = k' = k_*$. Note that these differences in the qualitative behavior of these functions are indeed tied directly to the topological character of the respective phases. There is a sudden transition between one behavior and the other as the tuning parameter passes through the topological transition, with the $(k, k') \to (\infty, \infty)$ limits of the functions F_0 and $F_{\pm 2}$ changing discontinuously at that point.

Both to gain further physical insight and to simplify eventual numerical integration, we rewrite the eigenvalue problem in a dimensionless form by scaling all momenta by the characteristic exciton momentum, $k_{\text{ex}} = 1/a_B = \mu e^2/\epsilon$. Since the Bohr radius is the most natural length scale in the problem, its inverse gives the most relevant momentum for excitonic physics. This rescaling naturally results in an equation with only two

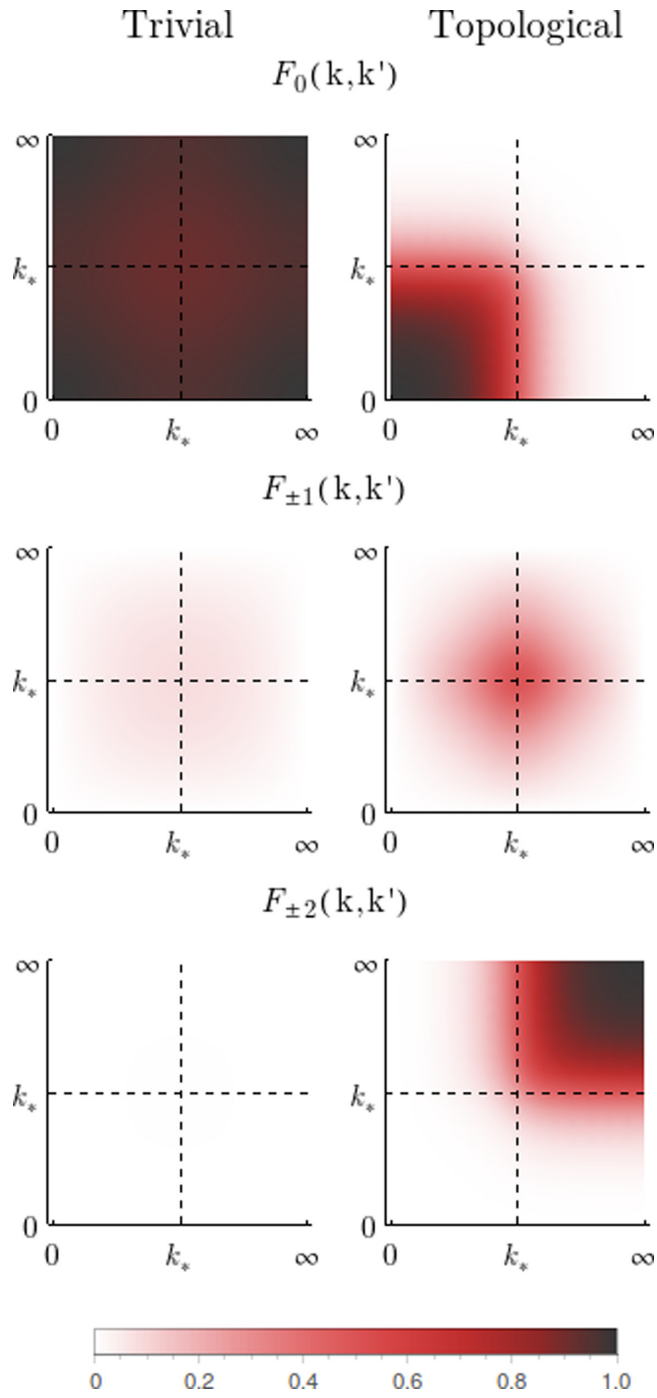

FIG. 2. The multipole coefficients of the overlap function $\mathcal{F}$ plotted as functions of their two arguments k and k', with the infinite domains $k, k' \in [0, \infty)$ projected onto a finite interval. The plots on the left show the typical behavior of the F's in the trivial phase, and those on the right show them in the topologically nontrivial phase. There is a clear and distinct difference in the qualitative behavior of these functions on either side of the transition, with a sudden jump from one behavior to the other at the transition itself.

dimensionless parameters: the relative fine-structure constant of the material $\alpha = e^2/\epsilon A$, and the quantity $\gamma = (k_{\text{ex}}/k_*)^2$, comparing the size of the excitonic and topological momentum scales, which we use as our tuning parameter.

As we first considered in Sec. II, we can understand how the topological nature of the system manifests itself in excitonic properties by considering the relative size of these momentum scales, i.e., the size of γ. First note that the small momentum features of the functions in Fig. 2 are very much alike on either side of the transition. Close to the transition where $|\gamma|$ is small and small momenta are most important, excitons in both sides of the topological transition should be qualitatively similar. Conversely, if $|\gamma|$ is not small, then the nontrivial features of the F functions near $k \sim k_*$ will be relevant in the topological phase, and excitons should behave quite differently depending on the sign of γ. From this we anticipate that our numerical analysis will not find a sharp feature in excitonic properties at the transition itself, instead seeing a gradual crossover between two behavior regimes.

Another way to see the effects of topology is to note how this exciton problem compares with the two-dimensional hydrogen atom. Just as for the 2D hydrogen atom, the eigenstates of the exciton problem in this model are labeled by two indices, $n = 0, 1, 2, \ldots$ and $m = 0, \pm1, \ldots, \pm n$, the principal and angular momentum quantum numbers.[1] For the 2D hydrogen atom, a hidden SO(3) symmetry [distinct from, but containing SO(2) rotational symmetry] ensures a perfect degeneracy between the $2n + 1$ angular momentum states for each n [24,25], with energies given by

$$E_n = -\frac{\mu e^4}{2\epsilon^2} \frac{1}{\left(n + \frac{1}{2}\right)^2}. \tag{19}$$

In our system, though n and m are still good quantum numbers, the nontrivial overlap function $\mathcal{F}$ breaks the SO(3) symmetry, mixing angular momentum channels of the Coulomb interaction and reorganizing the spectrum. The result is that the different angular momentum states for each energy level n will have their energies split from each other, as anticipated in Sec. II. Note that for interblock excitons, states with angular momentum differing by a sign must still be degenerate due to time-reversal symmetry, but those with different values of $|m|$ will generically be split.

In general, the breaking of this symmetry is ensured by the existence of any nonzero Berry curvature ($\mathcal{F} \neq 1$), even in a phase with a trivial Chern number. However, far enough into the topologically trivial phase, one can consider this symmetry breaking as just a small perturbation to the 2D hydrogen atom problem (i.e., $\mathcal{F} \approx 1 + \delta\mathcal{F}$ with $\delta\mathcal{F} \ll 1$), which only introduces a small splitting between the states. The same cannot be said of the topologically nontrivial phase, where the behavior of $\mathcal{F}$ is fundamentally nontrivial as well, as described above and in Fig. 2. For this case, the effect cannot be approximated as a small perturbation to the 2D hydrogen atom, so we can expect that the splitting between states will not necessarily be vanishingly small.

[1] Another common choice of quantum numbers for the 2D problem is $n_r = 0, 1, \ldots$ and $m = 0, \pm1, \pm2, \ldots$, related to our choice by $n = n_r + |m|$. For our purposes, n and m as defined in the text will prove more convenient.

IV. NUMERICAL ANALYSIS

We discretize the momentum in the integral Schrödinger equation according to a modified Gaussian quadrature method (with $N = 192$ points) that is designed to handle the divergence in the Coulomb potential at $\mathbf{k} = \mathbf{k}'$ [26]. Choosing a constant value $\alpha = 0.4$ for the effective fine-structure constant, we can then invert the resulting matrix equation to find the excitonic spectrum as a function of the parameter γ. We scale all energies that we calculate by the corresponding energies of the 2D hydrogen atom, i.e., with the same n, defining the effective mass in Eq. (19) the same way as in Eq. (13) so that the energy vanishes as $\gamma \to 0$.

Our main results are presented in Fig. 1, showing the $n = 0$ state and three $n = 1$ states for intrablock excitons.[2] In addition to these results, we also calculated the corresponding states for interblock excitons, finding similar effects. The qualitative behavior of the exciton energy levels in the topologically trivial and nontrivial regimes is immediately apparent, with the different angular momentum levels separating from each other quickly as a function of γ in the nontrivial regime, and converging to the 2D hydrogen energy moving deeper into the normal regime, as expected based on the discussed properties of the function $\mathcal{F}$. Furthermore, we find that intrablock exciton levels with opposite angular momentum split from each other, while corresponding levels in interblock excitons remain degenerate as ensured by symmetry. We also note that while there is a crossover between two behavior regimes, there is no sharp feature at the topological transition itself, again as anticipated.

The most notable behavior, seen in all cases, is the strong dependence of the $m = 0$ state on γ, which has considerably lower energy in the topological regime compared to the trivial regime. Indeed, for interblock excitons this is the primary feature we find. Since this effect is found in all cases, it must be caused by a different mechanism than that causing the splitting of opposite angular momentum states in the intrablock case, i.e., the Berry phase. In other words, it is an effect that is insensitive to time reversal and seems to be strongest for the cases of zero angular momentum. In particular, the Darwin term mentioned in Sec. II cannot be the only explanation since it is proportional to the Berry curvature and therefore cancels exactly in the interblock case. Unfortunately, we do not yet fully understand the effect leading to this phenomenon, but the onset of this energy shift at the transition indicates a topological origin, perhaps related to the quantum geometric tensor [27] or moments describing the distribution of Berry curvature in momentum space, e.g., the Berry curvature dipole [28].

V. CONCLUSIONS

By examining excitonic spectra in the bulk of a model with nontrivial topology, we have demonstrated that topology can in principle have strong manifestations in bulk physics. In particular, we have shown in the BHZ model that the degeneracy of 2D excitonic states that would exist in a system without a Berry curvature is broken due to the inclusion of such physical effects. In the trivial phase with Chern number 0 the splitting is small, with the Berry phase acting as a small perturbation to the 2D hydrogen atom problem. On the other side of the topological phase transition, however, the splitting is much greater since the effects of nontrivial topology can no longer be considered as just a small perturbation. Though there is no sharp feature precisely at the transition point, the difference in the behavior in the two phases can nevertheless be understood as a result of a change in the topological character. As the characteristic excitonic momentum scale becomes sizable compared to the scale associated with topological effects, the large momentum differences between the physics of the trivial and nontrivial phases becomes essential. The result is a dramatic reorganization of the excitonic spectrum, producing a hierarchy of states that is utterly distinct for values of the tuning parameter well into each of the two phases.

[2]The exciton binding energy is identically zero at $\gamma = 0$ where the gap closes, and the finite values of the curves in the figure at this point are due to our choice of scaling; the 2D hydrogen atom energy levels also vanish at this point, but the limit of the ratio is finite.

ACKNOWLEDGMENTS

We would like to thank Mehdi Kargarian for helpful discussions. This work was supported by NSF-DMR 1613029 (A.A.A.), the DOE-BES (DESC0001911), and the Simons Foundation (V.M.G.).

APPENDIX A: SYMMETRIES AND ANGULAR MOMENTUM

The BHZ Hamiltonian respects both time-reversal and inversion symmetry, with the two blocks of the Hamiltonian (1) mapping into each other under time reversal and remaining unchanged with inversion. This can be explicitly verified by representing the time reversal and inversion operators, respectively, as

$$\Theta = -i\hat{\sigma}_y K \otimes \hat{\mathbf{1}} \text{ and } P = \hat{\mathbf{1}} \otimes \hat{\tau}_z, \tag{A1}$$

and confirming that they commute with the Hamiltonian. Here K denotes complex conjugation. The set of single-particle eigenstates respects these symmetries as well, which one can straightforwardly verify, finding

$$\Theta|\mathbf{k}, \pm, i\rangle = \sum_j \epsilon_{ij}|-\mathbf{k}, \pm, j\rangle,$$
$$P|\mathbf{k}, \pm, i\rangle = \pm|-\mathbf{k}, \pm, i\rangle. \tag{A2}$$

Here the $\pm$ labels the conduction and valence bands, while i, j labels the block of the Hamiltonian that acts on the states.

In addition to these discrete symmetries, the system is also rotationally invariant so the total angular momentum is also a good quantum number. Since the system is two-dimensional, the total angular momentum is equivalent to its z component. The angular momentum of a particle has three components—spin, $S_z = \hat{\mathbf{1}} \otimes \hat{\sigma}_z/2$; atomic orbital, $K_z = \text{diag}(0, 1, 0, -1)$; and orbital, $L_z = \hat{\mathbf{1}}(\mathbf{r} \times \mathbf{p})_z$—so in total we have $J_z = S_z + K_z + L_z$, and a simple calculation confirms that $[J_z, H_{\text{BHZ}}] = 0$.

The eigenstates of H_{BHZ} are also eigenstates of J_z, and we can most easily compute the angular momentum of single-particle states at $\mathbf{k} = \mathbf{0}$, though the result must hold at all points

in k space, giving

$$J_z|\mathbf{k},\pm,i\rangle = (-1)^{i+1}\left[1 \mp \tfrac{1}{2}\,\mathrm{sgn}\,M\right]|\mathbf{k},\pm,i\rangle. \tag{A3}$$

Note that this value is simply either $\frac{1}{2}$ or $\frac{3}{2}$ up to a sign.

To consider excitons, we must add the Coulomb interaction to this single-particle Hamiltonian. With regard to symmetry, it is enough to note that the Coulomb interaction is also invariant under time reversal, inversion, and rotations, so the states of the interacting system must obey these symmetries as well. Let intrablock exciton eigenstates be labeled as $|n,m,i\rangle$, where n and m are two quantum numbers, and i labels the block from which we take the particle and hole constituents. In the center-of-mass frame, these exciton eigenstates are

$$|n,m,i\rangle = \sum_{\mathbf{k}} C^{(i)}_{nm\mathbf{k}}\, a^{\dagger}_{\mathbf{k},+,i} a_{\mathbf{k},-,i}|0\rangle. \tag{A4}$$

A straightforward calculation shows that

$$J_z|n,m,i\rangle = [m + (-1)^i\,\mathrm{sgn}\,M]|n,m,i\rangle \equiv j|n,m,i\rangle, \tag{A5}$$

where the second term in the eigenvalue is the sum of the spin and orbital angular momenta of the single-particle bands, Eq. (A3). We see here that m labels the part of the angular momentum interpreted classically as arising from the relative motion of the exciton's constituent particle and hole. Furthermore, it can be easily verified that time reversal acts in the expected way, simply flipping the sign of the angular momentum, $J_z\Theta|n,m,i\rangle = -j\,\Theta|n,m,i\rangle$.

APPENDIX B: BERRY PHYSICS IN THE BHZ MODEL

The spinor eigenstates corresponding to the two bands of the upper block of the Hamiltonian (1) are

$$|\mathbf{k},+\rangle = \begin{pmatrix} e^{i\frac{s+1}{2}\varphi_{\mathbf{k}}}\cos\frac{\theta_{\mathbf{k}}}{2} \\ e^{i\frac{s-1}{2}\varphi_{\mathbf{k}}}\sin\frac{\theta_{\mathbf{k}}}{2}\end{pmatrix},$$
$$|\mathbf{k},-\rangle = \begin{pmatrix} -e^{-i\frac{s-1}{2}\varphi_{\mathbf{k}}}\sin\frac{\theta_{\mathbf{k}}}{2} \\ e^{-i\frac{s+1}{2}\varphi_{\mathbf{k}}}\cos\frac{\theta_{\mathbf{k}}}{2}\end{pmatrix}, \tag{B1}$$

where $s \equiv \mathrm{sgn}\,MB$ and $\cos\theta_{\mathbf{k}} = M_{\mathbf{k}}/|\mathbf{d}_{\mathbf{k}}|$. The states for the other block can be generated from these by applying the time-reversal operator, discussed in Appendix A.

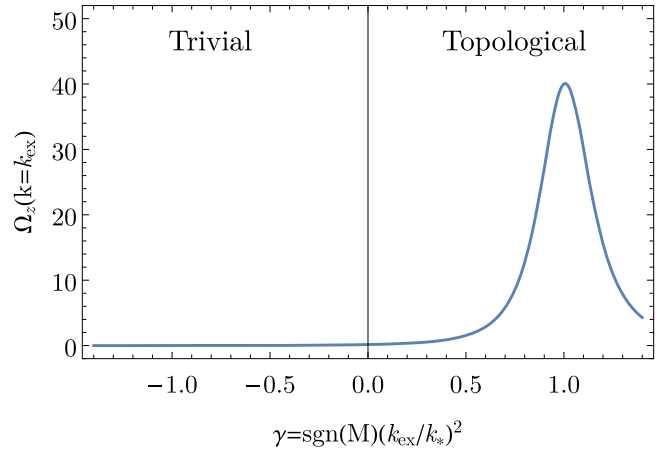

FIG. 3. A plot of the z component of the Berry curvature $\mathbf{\Omega}_+(\mathbf{k})$ as a function of momentum in both the topological (top) and trivial (bottom) phases. The Berry curvature $\mathbf{\Omega}_-(\mathbf{k})$ is simply related by a sign. The momentum is measured in units of the topological scale $|k_*| = \sqrt{|M/B|}$. In the topological regime, the Berry curvature is peaked near k_* and is positive for all values of the momentum, leading to a nonzero Chern number, while in the trivial regime it takes both positive and negative values producing a Chern number of 0.

The information about the topology is stored in the Berry connection of electrons defined in terms of these states as

$$\mathbf{A}_{\pm}(\mathbf{k}) = i\langle\mathbf{k},\pm|\nabla_k|\mathbf{k},\pm\rangle = \mp\frac{s+\cos\theta_{\mathbf{k}}}{2k^2}(\hat{\mathbf{z}}\times\mathbf{k}). \tag{B2}$$

The corresponding Berry curvature for the upper block is

$$\mathbf{\Omega}_{\pm}(\mathbf{k}) = \nabla_k \times A_{\pm}(\mathbf{k}) = \pm A^2\frac{M+Bk^2}{2d(k)^3}\hat{\mathbf{z}}. \tag{B3}$$

It is gauge-independent, and its integral gives the Chern number distinguishing topological phases. The momentum distribution of the Berry curvature in the topological and trivial regimes is presented in Fig. 3. Calculating the same quantities for the lower block of the Hamiltonian gives the same results up to overall signs.

When considering the interacting two-particle problem, we define the particle and hole states as

$$|\mathbf{k},e\rangle = |\mathbf{k},+\rangle, \quad |\mathbf{k},h\rangle = \mathcal{C}|\mathbf{k},-\rangle, \tag{B4}$$

using the particle-hole transformation $\mathcal{C} = K\hat{\sigma}_x$. With these definitions, we can define the Berry connections for particles and holes in the upper block analogously as in Eq. (B2) to find $\mathbf{A}_e(\mathbf{k}) = \mathbf{A}_+(\mathbf{k})$ and $\mathbf{A}_h(\mathbf{k}) = \mathbf{A}_-(\mathbf{k}) = -\mathbf{A}_e(\mathbf{k})$.

[1] M. Z. Hasan and C. L. Kane, Rev. Mod. Phys. **82**, 3045 (2010).
[2] X.-L. Qi and S.-C. Zhang, Rev. Mod. Phys. **83**, 1057 (2011).
[3] I. Garate and M. Franz, Phys. Rev. B **84**, 045403 (2011).
[4] D. K. Efimkin and Y. E. Lozovik, Phys. Rev. B **87**, 245416 (2013).
[5] D. K. Efimkin, Y. E. Lozovik, and A. A. Sokolik, Phys. Rev. B **86**, 115436 (2012).
[6] J. M. Blatt, K. W. Böer, and W. Brandt, Phys. Rev. **126**, 1691 (1962).
[7] B. Seradjeh, J. E. Moore, and M. Franz, Phys. Rev. Lett. **103**, 066402 (2009).
[8] D. I. Pikulin and T. Hyart, Phys. Rev. Lett. **112**, 176403 (2014).
[9] S. Juergens, P. Michetti, and B. Trauzettel, Phys. Rev. Lett. **112**, 076804 (2014).
[10] K. Saha, K. Légaré, and I. Garate, Phys. Rev. Lett. **115**, 176405 (2015).
[11] B. A. Bernevig, T. L. Hughes, and S.-C. Zhang, Science **314**, 1757 (2006).

[12] A. Srivastava and A. Imamoglu, Phys. Rev. Lett. **115**, 166802 (2015).
[13] J. Zhou, W.-Y. Shan, W. Yao, and D. Xiao, Phys. Rev. Lett. **115**, 166803 (2015).
[14] G. Wang, A. Chernikov, M. M. Glazov, T. F. Heinz, X. Marie, T. Amand, and B. Urbaszek, Rev. Mod. Phys. **90**, 021001 (2018).
[15] A. P. Silin and S. V. Shubenkov, Phys. Solid State **42**, 24 (2000).
[16] M. Trushin, M. O. Goerbig, and W. Belzig, Phys. Rev. B **94**, 041301 (2016).
[17] L. V. Keldysh, JETP Lett. **29**, 658 (1979).
[18] A. Chernikov, T. C. Berkelbach, H. M. Hill, A. Rigosi, Y. Li, O. B. Aslan, D. R. Reichman, M. S. Hybertsen, and T. F. Heinz, Phys. Rev. Lett. **113**, 076802 (2014).
[19] G. Bir and G. Pikus, *Symmetry and Strain-induced Effects in Semiconductors*, A Halsted Press Book (Wiley, New York, 1974).
[20] G. Sundaram and Q. Niu, Phys. Rev. B **59**, 14915 (1999).
[21] M.-C. Chang and Q. Niu, J. Phys.: Condens. Matter **20**, 193202 (2008).
[22] D. Xiao, M.-C. Chang, and Q. Niu, Rev. Mod. Phys. **82**, 1959 (2010).
[23] M. Trushin, M. O. Goerbig, and W. Belzig, Phys. Rev. Lett. **120**, 187401 (2018).
[24] X. L. Yang, M. Lieber, and F. T. Chan, Am. J. Phys. **59**, 231 (1991).
[25] D. G. W. Parfitt and M. E. Portnoi, J. Math. Phys. **43**, 4681 (2002).
[26] C. Y.-P. Chao and S. L. Chuang, Phys. Rev. B **43**, 6530 (1991).
[27] J. P. Provost and G. Vallee, Commun. Math. Phys. **76**, 289 (1980).
[28] I. Sodemann and L. Fu, Phys. Rev. Lett. **115**, 216806 (2015).