Photonic Topological Insulators: A Beginner’s Introduction
Dia’aaldin J. Bisharat, Robert J. Davis, Yun Zhou, Prabhakar R. Bandaru, and Daniel F. Sievenpiper

The control and manipulation of electromagnetic (EM) waves has reached a new level with the recent understanding of topological states of matter. These metamaterials have the potential to revolutionize many areas in traditional EM design, from highly robust cavities to small-footprint waveguides. Much of the past literature has been on the cutting edge of condensed-matter physics, but there are now ample opportunities to explore their usage for practical microwave and optical devices.

To assist the beginner, in this tutorial, we give a basic introduction to the essential concepts of topological phenomena in EM systems, including geometric phases, topological invariants, pseudospin states, and the integer/valley/spin quantum Hall effects (QHEs). Our focus is on engineered photonic topological insulators (PTIs) in 2D systems. We highlight methods for characterizing such structures and how they result in unique waveguiding properties. In addition, we provide recipes on how to realize PTIs using photonic crystals (PhCs) and metasurfaces, examine differences among different types of PTIs, and discuss the limitations and advantages of some of the existing enabling platforms.

Digital Object Identifier 10.1109/MAP.2021.3069276
Date of current version: 2 June 2021

EDITOR’S NOTE

“Everything should be made as simple as possible,” said Einstein. Unfortunately, today’s scientific knowledge is often presented in an obscure fashion, and the maelstrom of unfiltered information assailing us daily does not favor clarity. Electromagnetics is no exception to this trend. Luckily, a platform such as IEEE Antennas and Propagation Magazine represents a fantastic opportunity to disseminate knowledge via paths that deviate from common routes and, hence, to provide alternative explanations that may demystify previously inaccessible concepts. The “Electromagnetic Perspectives” column aims to leverage this opportunity with articles that include the following features:

- historical contextualization
- maximal simplification (without compromising accuracy)
- pedagogical creativity
- first-principles methodology.

Authors feeling that they may provide a contribution in the spirit of this column are welcome to contact me at christophe.caloz@polymtl.ca.

Topology (from the Greek τόπος, meaning “place or location,” and λόγος, signifying “study”) is a field of mathematics that deals with geometrical properties and spatial relations that are preserved under continuous deformations. It has many ramifications in science and engineering, one of them being topological insulators. Topological insulators were first reported as a quantum mechanical effect in electronics, but it has been recently transposed to photonics, where it has become a very popular topic. This article by Bisharat et al. offers an introduction and a broad picture of the field that, I hope, are of great interest to the readers of the magazine.

INTRODUCTION

In much the same way as PhCs applied the ideas of solid-state physics to photons [1], i.e., EM waves, the new field of PTIs [2] finds its origins in the world of electronic systems. In electronic TIs, electrons propagate along certain directions only on the exterior of the system. This explains part of the name: it is an “insulator” insomuch as it acts...
like a regular electrical insulator within the bulk of a material. “Topological,” on the other hand, comes from the global topology of the energy band structure since it can be categorized by an integer (the “topological invariant”) that does not depend on the fine details of the system (see Figure 1). The occurrence of electrical current on the surface of TIs—and how it responds to changes in energy—is tied directly to the mathematical concept of a topological invariant (see Figure 2) rather than minor changes to the material, as in ordinary materials.

TIs found their start in the 1980s with the discovery of the QHE in a 2D electron gas when subjected to periodic potentials and external magnetic fields [3], [4]. As in the normal Hall effect, applying a magnetic field causes the electrons to spin in cyclotron orbits, with their frequencies being determined by the strength of the $B$ field. When the material is strongly confined to 2D and cooled to very low temperatures, the quantization of the energy of these orbits becomes relevant, with the difference between the allowed energies becoming very large as the field strength increases. When the strength of the $B$ field varies enough to permit or remove an energy level, there will be a sudden jump in the transverse conductivity by an exact multiple of the fundamental constants.

Hence, the QHE shows that conductivity is fundamentally discrete [5]. Importantly, it was found [6] that this discrete behavior could be explained by a special phase (called the geometric or Berry phase, detailed in the “Geometric Phase” section) that each electron accumulates as it orbits in cyclotron motion in reciprocal ($k$) space.

How does topology relate to this? As it turns out, the discrete nature of the conductance is highly robust to deformations to the bulk of the material, and it can be shown that the added geometric phase responsible for the quantization is tied directly to the mathematical framework of topological invariants (see Figure 2) [3], [6], [7]. This has some important consequences: it gives us a simple means to classify materials (i.e., bandgap materials) by calculating their topological invariant (which is a property of the bulk material), and it results in the technologically useful effects that TIs offer.

Materials that have an invariant of zero are “trivial,” and they act the same as ordinary materials. If the invariant is nonzero, however, then the effects of the geometric phase become relevant, and “nontrivial” effects can be observed. One of the most startling effects is what happens at the edge between a nontrivial material and a trivial material (or another nontrivial material with a different invariant), where a highly robust transport mode can exist [5].

These special modes, called edge modes, exist within the bandgap of the nontrivial material, and they can be explained by the sudden change in the invariant across the boundary (e.g., going from one to zero). Even more remarkable is that electrons moving along these boundaries must do so in one direction only, with no possibility of scattering back in the other direction (illustrated in Figure 1) [8]. These edge modes are the corollary of the quantization of conductivity in the QHE.

In repeated experiments, these edge states are observed regardless of the impurities in different material samples [9]. Since the invariant is resistant to a wide range of distortions to the material, the edge states are said to be topologically “protected”—i.e., guaranteed to exist so long as the invariant stays the same [5]. This is of technological importance due to the potential to reduce power consumption by eliminating sources of loss as well as simplify manufacturing by increasing defect tolerances. These discoveries led to the Nobel Prize in Physics being awarded to Thouless, Haldane, and Kosterlitz in 2016.

These systems with topological behavior are a consequence of the wave nature of the electrons, not specifically their quantum interactions [10]. As a result, it is possible to construct classical

![Figure 1](image1.png)

**FIGURE 1.** (a) Normal versus (b) topologically protected transport. The normal case has backscatter at sharp corners and defects, whereas the topological one does not.

![Figure 2](image2.png)

**FIGURE 2.** Topology concerns quantities that are preserved under continuous deformations of objects. A transformation is “continuous” if it does not cause any sharp cuts or tears in the object. The number of holes in a closed surface is an example of a topological invariant since a hole cannot be added or removed continuously: a torus can be stretched and pulled into a coffee cup shape, but not a sphere. Electrical conductance in TIs is also determined by a topological invariant, called the Chern number ($C$), where the object is an energy band in the Brillouin zone, and the “holes” are determined by the accumulation of the Berry phase.
wave systems with analogous properties to their electronic counterparts. This opens the door to a vast range of theoretical proposals and experimental demonstrations. Replacing the electron with a photon (along with a reinterpretation of some quantities), we arrive at PTIs, which demonstrate many of the same features of TIs [2] and are the primary subject of this tutorial (see “Side Note 1”).

In this tutorial, we outline the basic concept of the geometric phase and extend it to periodic systems in which topological properties emerge. We focus on 2D PhCs and showcase the physical implications of Chern numbers and topological transitions that can arise in such systems. In addition, we discuss the formation of degeneracy (Dirac) points in PhCs and then the various mechanisms to introduce topological phases that make different types of PTIs. Finally, we discuss recent developments in and future perspectives on this emerging field.

**GEOMETRIC PHASE**
The key idea behind topological effects in all areas is the geometric phase, a universal concept that emerges when a parameter describing a system is gradually varied in a closed cycle [11]. This phase was first proposed in 1956 by Pancharatnam for the propagation of light through a sequence of polarizers [12] and was later generalized by Berry for quantum mechanics [13]. Many phenomena in physics can be attributed to the geometric phase, from the mechanical Foucault pendulums [14] to the polarization in helical waveguides [15].

Any wave possesses an amplitude (call it \( E_0 \)) and an ordinary phase \( \phi \) at a given position and time, \( E(r,t) = E_0 (r) e^{i\phi} \). When the values of \( r \) and \( t \) are slowly changed from \( r_0 \) and \( t_0 \) to distinct intermediate values \( r_1 \) and \( t_1 \), then smoothly changed back to \( r_0 \) and \( t_0 \), we would intuitively expect that the initial value of \( E(r,t) \) would be exactly the same as the final value. However, there are some physically important cases where this intuition fails, as in the case shown in Figure 3.

If we take a tangent vector (the red arrows) and slide it along a path on a sphere (the black arrows), returning it to the starting position, it will no longer be pointing in the same direction. Hence, the starting value no longer matches the final value. Upon its return to the north pole, the orientation of the vector is rotated by an angle \( \phi \), which, in this situation, is equal to \( \pi/2 \). Note that this is the case only because the path is a closed loop: if the tangent retraced the same path to return to the start (enclosing no area), the orientations would match, and \( \phi \) would be zero. In the closed-path case, we can consider the angle to be an added phase, the geometric phase, that causes the initial and final values to differ.

This phase is geometric because it corresponds to a geometric area (\( \Omega \), shaded in blue) of the parameter space that the path encloses.

If we think of the sphere in Figure 3(a) as the sphere of constant wave vector \( \mathbf{k} \) and the vector as the electric field (E-field), this parallel transport explains the change in polarization in the helical and bent waveguides [Figure 3(b)] [16]. In the example of a bent waveguide, if polarized light gradually changes direction from \( z \) to \( -x \), then from \( -x \) to \( y \), and finally back to \( z \), the wave will pick up a geometric phase that is added to the complex exponential form of the E-field. The “path,” in this case, is the path traced over the \( \mathbf{k} \) sphere as the wave vector, corresponding to the direction, is changed. In this situation, the extra phase (in the form of the polarization direction) can be attributed to the E-field always being perpendicular to the direction of propagation; therefore, as the direction of propagation is changed, the polarization must necessarily be altered, despite the propagation direction eventually returning to the starting value. Note that this effect is possible only due to the existence of

---

**SIDE NOTE 1**

Although the field of PTIs originated from the electronic version, there are some fundamental differences between the two. The most significant is that photons are bosons, whereas electrons are fermions. This difference manifests itself in the ways that different symmetries (like time reversal, written as an operator \( T \)) can change how a system behaves.

Specifically, time reversal for fermions has the relationship \( T^2 = -1 \), whereas for bosons, it is \( T^2 = +1 \). A more practical concern is that absorption in the medium can be an issue in photonics. Nevertheless, many of the most technologically relevant features of TIs (including backscatter immunity) can still be found in photonic systems, so long as care is taken to distinguish the circumstances in which they can exist.

---

**FIGURE 3.** (a) The geometric phase from parallel transport. (b) The polarization (shown as the vector E-field inset at different locations) in a bent circular waveguide shows that, as the propagation path is varied and then returned to its initial state, there can be a phase shift to the polarization state, which is due to the geometric phase. (Source: [16]).
two degenerate modes in the waveguide (from the circular symmetry).

**BERRY PHASE, CONNECTION, AND CURVATURE**

A geometric phase can emerge due to the gradual variation of a state in many types of parameter spaces, including the momentum space of a periodic system, like those of a PhC [17]. For any path that traverses an allowed band of a periodic system and does not intersect with any other band, the wave vector \( \mathbf{k} \) (Bloch momentum) varies in closed loops due to the lattice periodicity, where \( \mathbf{k} - \mathbf{k}_0 \equiv \mathbf{k}_a \).

In a 2D crystal, \( \mathbf{k} \) traverses the surface of a torus geometry, which bounds the entire Brillouin zone (BZ) (see Figure 4). Many of the most important topological properties appear in such systems and are a simple platform to understand how they emerge.

The literature on TIs employs a great deal of terminology, most of which merely refers to a few mathematical constructions that assist in characterizing when a topological invariant is nontrivial. Fortunately, most of these constructions have parallels to standard EM theory and provide straightforward methods to numerically calculate topological features of real systems.

Consider a lattice described by a general eigenvalue problem in momentum space:

\[
H(k) \cdot \psi_n(k) = \lambda_n(k) \psi_n(k),
\]

where \( \lambda_n(k) \) is the eigen energy, and \( \psi_n(k) \) is the normalized eigen wave function of \( H(k) \) (often called the Hamiltonian in the literature) at each \( k \) for the \( n \)th band, which can be determined via Bloch’s theorem. In the following, we will make use of the shorthand notation of the inner product,

\[
\langle \psi_n(k) | \psi_m(k) \rangle = \int \psi_n^*(k) \psi_m(k) dk,
\]

where \( \psi_n(k) \) and \( \psi_m(k) \) are wave functions at the same \( k \) point. The notation \( \langle \alpha | \beta \rangle \) represents the inner product of the wave functions \( \alpha \) and \( \beta \), whereas \( \langle \alpha | \nabla_k | \beta \rangle \) denotes the inner product of \( \alpha \) and \( \nabla_k \beta \).

To calculate the total Berry phase, we need a means to add up the phase contributions from each small change to the wave function. The phase shift between two \( \psi_n \) states infinitesimally separated by \( \Delta k \) can be represented by their inner product [16], expanded as a low-order Taylor series as

\[
\begin{align*}
\langle \psi_n(k) | \psi_n(k + \Delta k) \rangle &= 1 + \Delta k \langle \psi_n(k) | \nabla_k | \psi_n(k) \rangle \\
&= \exp[-i \Delta k \cdot A_n(k)].
\end{align*}
\]

Here, we can see that \( \Delta k \cdot A_n(k) \) is the phase shift over \( \Delta k \), and \( A_n(k) \) is the rate of change of the phase shift (see “Side Note 2”). \( A_n(k) \) is called the Berry connection or Berry vector potential:

\[
A_n(k) = i \langle \psi_n(k) | \nabla_k | \psi_n(k) \rangle.
\]

Therefore, the Berry phase for the \( n \)th band is defined as the integral of \( A_n(k) \) over some closed path \( l \) in the \( k \) space:

\[
\phi_n = \int_l A_n(k) \cdot dk.
\]

The path \( l \) is simply a smooth curve of values over the BZ, such as the blue and red lines shown in Figure 4(b). If we know what a given wave function looks like in the BZ, we could use (5) to calculate the Berry phase for that path. However, there is a catch: the Berry connection \( A_n(k) \) is not uniquely defined. If a phase change \( \zeta(k) \) is added to the eigen wave function \( \psi_n(k) \), where \( \zeta(k) \) is a periodic function with \( \zeta(k_{\text{mod}}) = \zeta(k_{\text{orig}}) + 2\pi m_i \), the new wave function \( e^{i\zeta(k)} \psi_n(k) \) is still an eigen wave function to \( H(k) \). The Berry connection is then transformed as

\[
A_n(k) \rightarrow A_n(k) - (\partial \zeta(k)/\partial k) \zeta(k),
\]

where it changes its formula with a different choice of \( \zeta(k) \).

The Berry phase, on the other hand, is invariant modulo \( 2\pi \):

\[
\oint \frac{\partial}{\partial k} A_n(k) \cdot dk - \oint \frac{\partial}{\partial k} \zeta(k) \cdot dk = 2\pi n_i.
\]

This can also be understood qualitatively. As the wave vector \( \mathbf{k} \) slowly travels around the loop of a band, the wave function \( \psi_n(k) \) eventually returns to where it starts and picks up a phase of a multiple of \( 2\pi \), with most systems picking up zero [11]. Since the Berry connection depends on how we set up the calculation, yet we know that the Berry phase should not, it is helpful (especially for numerical

**FIGURE 4.** The (a) BZ can be considered as a (b) torus by taking each periodic boundary (red and blue) and connecting them together.

**SIDE NOTE 2**

The notation \( \langle \alpha | \beta \rangle \) represents the inner product of the wave functions \( \alpha \) and \( \beta \), whereas \( \langle \alpha | \nabla_k | \beta \rangle \) denotes the inner product of \( \alpha \) and \( \nabla_k \beta \).
purposes) to define a quantity that will be invariant to any arbitrary phase \( \zeta(k) \) that we may add.

The Berry curvature, or Berry flux, a quantity that is invariant under such transformation, can be constructed by taking the curl of the Berry connection:

\[
\Omega_\kappa(k) = \nabla_k A_\kappa(k). \tag{7}
\]

Then, using Stokes’s theorem, the Berry phase can be rewritten as the integral of the Berry curvature:

\[
\phi_\kappa = \oint_S d\mathbf{l} \cdot \Omega_\kappa(k). \tag{8}
\]

where the integration is over the surface bounded by the path \( l \) [16].

**SIDE NOTE 3**

Why is the Chern number always an integer? A simple explanation comes from comparing the equations with those of the magnetic field. The Berry curvature has the same form as the magnetic field, where the Berry phase can be thought of as a magnetic flux:

\[
\Omega_\kappa = \nabla_\kappa \times A_\kappa \Rightarrow B = \nabla_\kappa \times A. \tag{S1}
\]

Since the Chern number is just the integral of the Berry curvature, this is the \( k \)-space version of integrating the magnetic field:

\[
C_\alpha = \int \Omega_{\kappa} \, dk \Rightarrow \int B \, dr. \tag{S2}
\]

From Gauss’ law for \( B \) fields, we know that doing so will always give zero, unless there exists a magnetic monopole in the integration area. In such a case, the integral will give an integer multiple of monopole charges. In contrast to the \( B \) field, the Berry phase can contribute “monopoles” to the Berry curvature, the number of which is the Chern number [16]. Hence, the Chern number must be an integer. Further proofs can be found in [2] and [7].

**TOPOLOGY IN 2D PhCs**

The previous section dealt with the general concepts of the geometric phase in periodic media, regardless of the physical setting (i.e., electronic, photonic, and so on). To make this explicit, here, we show how this theory can be specialized for 2D EM systems. For EM waves, the eigenvalue problem space is described by the macroscopic Maxwell equations. For nonbiantisotropic materials in 2D, the magnetic field can be eliminated—for simplicity, for treating transverse magnetic (TM) modes, given by the \( E_z \)-field alone—and Maxwell’s equations can be recast in a compact form as

\[
\nabla \times [\mu^{-1}(r) \nabla E_z(r)] = \omega^2 \varepsilon(r) E_z(r), \tag{9}
\]

where \( \omega \) is the angular frequency; \( E_z(r) \) is the \( z \)-component of the E-field (hereafter, we will drop the \( z \) subscript); and \( \mu(r) \) and \( \varepsilon(r) \) are the magnetic permeability and dielectric permittivity tensors, respectively. Note that we are ignoring dispersive effects for now, but further analysis shows this is valid in many cases [16]. By applying Bloch’s theorem to (9), the eigen wave function can be obtained in the form of E-field \( E_{n,k}(r) \), assuming a periodicity of the material parameters [11].

Since the eigenvalue problem involves the dielectric permittivity \( \varepsilon(r) \) on the right-hand side of (9), the inner product of two eigen wave functions \( E_{n,k}(r) \) and \( E_{n,k'}(r) \) can be written as

\[
\langle E_{n,k} | E_{n,k'} \rangle = \int d^2r E_{n,k}^*(r) E_{n,k'}(r), \tag{10}
\]

where \( \ast \) denotes complex conjugation. The Berry connection then takes the form of

\[
A_{\kappa}(k) = i \langle E_{n,k} | \nabla_k | E_{n,k} \rangle = i \int d^2r E_{n,k}^*(r) \varepsilon(r) \nabla_k E_{n,k}(r). \tag{11}
\]

As in the general case, \( E_{n,k} \) is normalized such that \( \langle E_{n,k} | E_{n,k} \rangle = 1 \). Subsequently, the Berry curvature and phase can be written as discussed earlier.

**CHERN NUMBER**

With the aid of the Berry curvature, we can calculate the Berry phase that a given EM mode may acquire for a 2D PhC lattice. As mentioned in the “Introduction” section, the relationship to topology comes in the form of an invariant tying a nonzero Berry phase to the edge modes. For EM systems, this invariant is called the **Chern number**, after Chinese American mathematician Shiing-Shen Chern.

The Chern number always takes an integer value (see “Side Note 3”). When it is nonzero, the 2D photonic system is said to be topologically nontrivial. The Chern number of the nth band of a 2D lattice is simply the Berry phase over the full BZ:

\[
C_n = \frac{1}{2\pi} \int_{BZ} d^2k \Omega_n(k_x,k_y). \tag{12}
\]

where, in 2D, the Berry curvature only has two terms:

\[
\Omega_n(k_x,k_y) = \frac{\partial A^n_{x}}{\partial k_y} - \frac{\partial A^n_{y}}{\partial k_x}. \tag{13}
\]

where \( A^n \) is the Berry connection for the nth mode:

\[
A^n_{x} = i \int d^2r E_{n,k}^*(r) \varepsilon(r) \frac{\partial A_{n,k}(r)}{\partial k_x}, \tag{14}
\]

\[
A^n_{y} = i \int d^2r E_{n,k}^*(r) \varepsilon(r) \frac{\partial A_{n,k}(r)}{\partial k_y}. \tag{15}
\]

When calculated for an arbitrary polarized band over the whole BZ, the Chern number, expressed by \( C_n \), takes a nonzero value only when the time-reversal symmetry (TRS) is broken for the lattice [2]. The most common case when this happens is if a magnetic field is applied (the Faraday and Kerr effects). In such cases, the system is often called a **Chern insulator** or **Chern PTI**.

However, there are a few ways to observe topological effects, even when the TRS is retained [4]. Such systems are still reciprocal (i.e., they cannot form true isolators), but they can display immunity to certain types of backscatter and act as robust polarization filters [18].

One popular version of time-reversal-symmetric PTIs is the “valley” PTI,
which associates different directions (Γ to K and Γ to K') in k-space with “valley Chern numbers.” To calculate the valley Chern number C_v, the integral in (12) is simply performed over only one half of the BZ, such that two ordinarily identical high-symmetry points (K and K') are separated. This gives two different values of C_v, one for each half of the BZ. Added together, they will equal the normal Chern number (zero for reciprocal systems), but, considered separately, they can be nonzero under special cases [19].

The other major reciprocal PTI is the “spin” PTI, which associates the handedness of a circularly polarized mode or other combinations of modes with a “spin Chern number” C_s [20]. In general, these spins/polarizations are constructed by a superposition of multiple bands at the same frequency (see “Side Note 4”). Each polarization (right- and left-handed) corresponds to its own value of C_s [20, 21]. As such, the E-field in the inner product definition must be replaced with the field associated with a given polarization.

NUMERICAL CALCULATION OF THE CHERN NUMBER

In calculations of the Chern number for simulations or experiments, we need to discretize the continuous 2D BZ into a lattice, as shown in Figure 5. The discretization is for a square BZ (for square lattices), but the same methods will work on triangular lattices, where the usual hexagonal BZ is shifted to form a rhombus [23, 24]. The Chern number can be written as [23, 24]

\[ C_n = \frac{1}{2\pi} \sum_{k_x, k_y} \Omega_n(k_x, k_y) \Delta k_x \Delta k_y, \]  

(16)

where

\[ \Omega_n(k_x, k_y) \Delta k_x \Delta k_y = (A_n^{\text{xx}}(k_x + \Delta k_x, k_y) - A_n^{\text{xx}}(k_x, k_y)) \Delta k_x \]
\[ - (A_n^{\text{xy}}(k_x, k_y + \Delta k_y) - A_n^{\text{xy}}(k_x, k_y)) \Delta k_y. \]

(17)

Since Δk_x is small,

\[ A_n^{\text{xx}}(k_x, k_y) \Delta k_x \approx i \ln \langle E_n(k_x, k_y) | E_n(k_x + \Delta k_x, k_y) \rangle. \]

(38)

If we number the four vertices of a small cell as 1, 2, 3, and 4 in a clockwise direction, as shown in Figure 5, the integral of the Berry curvature over the grid can then be written as

\[ \Omega_n(k_x, k_y) \Delta k = -i (\ln \langle E_{n,1} | E_{n,2} \rangle + \ln \langle E_{n,2} | E_{n,3} \rangle + \ln \langle E_{n,3} | E_{n,4} \rangle + \ln \langle E_{n,4} | E_{n,1} \rangle) \]
\[ = -i \ln \langle E_{n,1} | E_{n,2} \rangle \langle E_{n,2} | E_{n,3} \rangle \langle E_{n,3} | E_{n,4} \rangle \langle E_{n,4} | E_{n,1} \rangle. \]

(19)

Here,

\[ \langle E_{n,p} | E_{n,q} \rangle = \sum_{w,m} E_{n,p}(w, m) \epsilon(w, m) E_{n,q} \delta_{w, \Delta s}, \]

(20)

where p, q = (1, 2, 3, 4) denotes the four vertices; (w, m) indicates the (w, m)th discretized cell in the real space; Δk = Δk_x, Δk_y; and Δs is the area of the discretized lattice in the real space. Equation (19) shows that the integral of the Berry curvature over each small cell in the BZ can be obtained by taking the inner products of the eigen E-fields at adjacent vertices in a clockwise order, as illustrated by the inset in Figure 5.

Substituting (19) into the summation in (16), we get a discrete approximation of the Chern number. It can be shown that this approximation converges to the (continuous variable) Chern number at the limit Δk(0) → 0 [25]. Fortunately, it also rapidly converges, often as coarse a grid as 24 × 24 cells enough for accurate determination of the Chern number [23, 25].

The spin Chern number can be computed by separating the two distinct spin eigenmodes (generally polarization based) and performing the Chern number calculation on each [20]. For non-trivial spin PTIs, this will result in two identical values, each being the negative of the other [22, 23]. For valley Chern number calculations, only half of the BZ is integrated in (20) to account for the contribution of a finite region in the momentum space that corresponds to specific high-symmetry points in the BZ [23] (see “Side Note 5”).

From a practical perspective, the Chern number gives a straightforward means of checking whether a given system has edge states and, therefore, if it will be robust to various forms of...
disorder. We have provided a collection of general-purpose MATLAB functions that perform the various steps, available via a public repository [26]. It is worth noting that these numerical methods are not the only option to determine the nontriviality of a system, with another powerful technique being the Green’s function approach (which also simplifies the analysis of degenerated bands) [27]. In the proceeding sections, we show example calculations of each PTI type (Chern, valley, and spin) using this code, with the eigenmode data simulated via the Ansys high-frequency structure simulator.

The Chern number describes the topology of a band and characterizes the most fascinating and technologically relevant phenomena: topologically “protected” edge states. These edge states appear at the interface between two structures with unequal Chern numbers.

Unlike traditional photonic waveguides, with a trivial edge state between two ordinary insulators (with a Chern number of zero), the nontrivial edge waveguide formed by these two topologically inequivalent structures (at least one structure is of a nonzero Chern number) would be immune to defects and backscattering. This is because, when two domains with different topological invariants are connected directly to form an interface, a topological phase transition must happen at the interface [2] (see “Side Note 6”).

Essentially, the differing topologies mean that the respective bands in each bandgap material cannot be continuously transformed into one another. Transforming one into the other requires the frequency gap to close at the interface and then reopen on the other side. This phase transition gives rise to the gapless edge states at the interface. To accommodate the jump in the Chern number’s integer value, e.g., from 1 to 0, 1 to –1, and so on, the number of gapless edge modes turns out to be the difference of the Chern numbers across the interface [28]. This is known as the bulk-edge correspondence [2].

These gapless modes are tied to the bulk Chern numbers, so they are robust and must always exist, regardless of the specific shape of the boundary (unlike conventional waveguides) [2]. It is worth stressing that these modes are distinct from those found in standard PhC waveguides (which can also possess high robustness [29]), with the primary difference being that their immunity to certain forms of scattering is a global property of the bulk rather than any specific arrangement of PhC cells.

**CHERN PTIs**

In general, PhCs and other periodic structures have a zero Chern number [2]. To engineer one, we need to focus on two key steps: 1) finding a degenerate point between the bands and 2) breaking a symmetry that opens a bandgap near that point (see “Side Note 7”). This section details how the simplest type of PTI, the Chern PTI, is constructed and demonstrates the exciting features it has for practical designs. This type of structure is a direct emulation of the QHE discussed previously [30].

The first step, finding degeneracies, relates to the abrupt nature of the Chern number: a material can change its Chern number (a topological phase transition) only when two or more photonic bands are degenerate at a point. This is part of the reason for the robustness of edge modes, as any small change to the structure that keeps the bandgap open in the bulk does not affect the mode. Finding a degeneracy in PhCs is common, but the second step, opening a bandgap via a broken symmetry, places some restrictions on the degeneracies that are useful for making a PTI [2].

The simplest type of degenerate point for PTIs is a linear crossing of two bands, often called a Dirac cone in the literature [4]. Such a crossing can be made via a PhC in a honeycomb lattice, where the degenerate point will always occur at the K(K) high-symmetry point in the BZ [1]. To obtain a nontrivial PhC, it is, however, not necessary to form a linear-type degeneracy, as any other type (e.g., quadratic [31] or accidental [32]) will also work.

To see how this works for a real device, we use the example of
Wang et al. [33], which was later developed into the first experimental demonstration of a Chern PTI [34]. First, to create the initial degenerate point, we select a square lattice of circular rods [chosen to be made of yttrium–iron–garnet, (YIG), for reasons soon explained] and tune the geometry to find a quadratic crossing of the second and third bands at the M point, shown in Figure 6(a). Note that there is also a degeneracy at the Γ point for the third and fourth bands, but we focus on the quadratic M point here.

Now that we have our degeneracy, we must break a symmetry that opens a complete bandgap near it. The chief symmetries present in the system are the TRS (where running time backward does not affect the response) and SIS (where flipping the coordinate axes maintains the shape and orientation of the unit cell).

Breaking either will induce a bandgap, but only breaking the TRS will cause a nonzero accumulation of the Berry phase over the whole BZ and, so, will result in the desired edge modes [24]. Following Wang et al.’s approach, the TRS can be broken by applying a static magnetic field perpendicular to the 2D plane. Doing so induces an anisotropy to the magnetic permeability of the YIG with the form

\[ \mathbf{\mu} = \begin{bmatrix} \mu & i k & 0 \\ -i k & \mu & 0 \\ 0 & 0 & \mu_0 \end{bmatrix}. \]  \quad (21)

Here, \( \kappa \) represents the effect of the \( z \)-directed dc magnetic field and is zero when the field is turned off. For a 1600-G magnetic field [33], the values at 4.28 GHz are \( \kappa = 12.4 \mu_0 \) and \( \mu = 14 \mu_0 \), with \( \mu_0 \) being the vacuum permeability in meter–kilogram–seconds. Breaking the TRS by turning on the magnetic field opens a bandgap near the degenerate point for the second and third bands at the M point, as shown in Figure 6(b).

To confirm that the opened bandgap is indeed topologically nontrivial, we can observe the behavior of the Berry curvature for the various bands, shown in Figure 7(a) and (b). We can see that the second mode has a very large contribution to the Berry curvature right at the M point and, likewise, for the X point of the third mode. Integrating over each separately, we find the Chern number of the lowest band to be zero, while the next three are \(-1, 2, \) and \(1 \), indicating the existence of edge modes within the upper bandgaps.

An important thing to note is that, while the Chern number \( C_n \) is associated with each band \( n \) of the bulk, edge modes are associated with the bandgaps between them. To differentiate this, we often speak of the “gap Chern number” \( C_{gap} = \sum_n < n, C_n \) which is just the sum of the Chern numbers of all bands below a given bandgap with upper band \( n_u \) [16]. Hence, to observe edge modes, we need to operate within a bandgap between two materials with different gap Chern numbers, with the net number of modes being their difference:

\[ N_{modes} = C_{gap} = [C_{gap, 1} - C_{gap, 2}]. \]

For this example, that implies the lowest bandgap will have no edge modes (or, more precisely, the net number of rightward modes equals the net number of leftward modes [35]), while the second and third bandgaps will.

The magnetized YIG model displays all of the telltale signs of a Chern PTI, and, as such, we can construct a wide range of devices that exploit its nonreciprocal and highly robust nature. One such demonstration is an isolating transmission line with two 90° bends, shown in Figure 7(c) and (d) (compare with Figure 1). Similar to ferrite-based magnetic isolators, the device is nonreciprocal for any EM mode inside the nontrivial bandgaps, but there are a few important and highly attractive features:

- Unlike a traditional isolator, EM energy is not simply routed to a lumped element load and dissipated locally as heat [34]. Instead, the influence of the Berry phase results in an edge mode only for a single direction of propagation, with no allowed modes in the opposite direction. Hence, any energy sent the opposite way will either be reflected or decay exponentially into the bulk in the same manner as a trivial PhC [Figure 7(d)].
- The directionality of the mode is determined by the direction of the

**FIGURE 6.** A band diagram for the square lattice of (a) unmagnetized and (b) magnetized YIG rods in air. The inset in (a) shows the BZ path, and the inset in (b) represents a unit cell with \( r = 0.11a \).
bias magnetic field, so flipping its direction will also flip the allowed propagation direction.

With the lack of backward modes, a source of backscatter, like the shown 90° bends, will force the energy around corners with negligible losses. This will occur so long as the strength of the scatterer is not greater than the size of the nontrivial bandgap, provided the scatterer is nonmagnetic [2]. Likewise, any small defect, like the three missing rods, will not lead to scattering.

Being essentially a distributed device, the level of isolation and insertion loss can be tuned by varying the length and shape of the structure. This example is for a 2D system, which can be experimentally emulated via a parallel-plate waveguide structure, with the separation between the plates being very thin, ensuring only TM modes can propagate. This platform makes it easy to analyze but is less straightforward to integrate into normal EM and photonic systems. However, there are numerous studies and ongoing work to create Chern PTIs for more practical settings [36]. A major breakthrough for this line of research was the development of the valley and spin PTIs (detailed in the following sections), which remove the requirement of the external magnetic field (see “Side Note 8”).

**VALLEY PTIs**

Although the Chern PTI has many advantages, the requirement for magneto-optical materials and external magnetic fields places limits on the practical applications. The question then arises: can we construct a PTI with similar features of robustness to disorder or sharp turns while still being reciprocal? The answer turns out to be yes, with some limitations. There is another type of PTI made of passive materials that exploits an inherent degree of freedom of hexagonal lattices that can be used to mimic similar phenomena for robust edge state propagation, although the level of robustness depends on the types of disorder considered.

**SIDE NOTE 8**

Although the example of a Chern PTI studied here concerns periodic structures with discrete translational symmetry, it has been recently shown that Chern numbers can also be defined for continuous media, such as a homogenous magnetized plasma [S1]. In addition, other platforms, including arrays of coupled waveguides, can be used to emulate the effect at optical frequencies [38].

**Reference**

[S1] M. G. Silveirinha, “Chern invariants for continuous media,” Phys. Rev. B, vol. 92, no. 12, p. 125153, Sept. 2015. doi: 10.1103/PhysRevB.92.125153.
Specifically, in a hexagonal/graphene-like lattice, the angular rotation of the E-fields at the high-symmetry point $K$ or $K'$ generates an intrinsic magnetic moment, which is called the valley degree of freedom [19]. The term valley is used owing to the shape of the dispersion near the $K$ ($K'$) point, which, in a triangular lattice, is a deep dip, or a sharp peak, both of which are referred to as valleys. Just as the Chern PTI emulates the QHE, the valley PTI is a model for the “quantum valley Hall effect,” studied in graphene-like materials.

To design a valley PTI, we can start from a graphene-like PhC, which possesses the “Dirac”-like degenerate point at the $K$ ($K'$) point. Such a lattice can be constructed by a unit cell containing two rods of equal radius (A and B sites), shown in the inset to Figure 8(a). Like when constructing a Chern PTI, a symmetry must be broken to lift the degeneracy. In a valley PTI, a controllable bandgap can be achieved by differentiating the A and B rods in the unit cell, thus breaking the inversion symmetry. As we show through examples, a graphene-like PhC that lacks inversion symmetry exhibits opposite Berry curvatures at the $K$ ($K'$) point as shown in Figure 8(a).

We then break the inversion symmetry by shrinking the A rod slightly ($r_A = 0.19a$). This lifts the degeneracy and opens a complete bandgap around it, as shown in Figure 8(b). Note that we can tune this bandgap by tuning the dimensions of A and/or B. The more different A and B are, the larger the bandgap.

As shown in Figure 9, the in-plane E-field distribution of the first and second bands at the $K$ valley are accompanied by an energy flux (i.e., time-averaged Poynting vectors) rotating in either a clockwise or counterclockwise manner. In accordance with the TRS, we also find that the field profile at the $K'$ valley exhibits the reversed direction.
of energy flux. The flux vortex’s center corresponds to a singular point of the phase (here, the out-of-plane E-field $E_z$), carrying an orbital angular momentum (OAM) with its sign depending on the vortex direction [39].

This vortex can be considered as an “artificial magnetic field”-like effect that replaces the role of the real magnetic field of the Chern PTI. Meanwhile, inverting the orientation of the unit cells in the plane (i.e., swapping the A and B lattice sites) results in identical band structures but opposite signs of the OAM at the K and K’ valleys. Importantly, the frequency order of the OAM states at each valley is also inverted, which indicates a topological phase transition.

To further validate the nontrivial topological character of the bands, we can numerically calculate the Berry curvature, as shown in Figure 10(a) and (b). The spike at the K point results in a Berry phase of $\pi$, while there is a $-\pi$ Berry phase accumulation at K’. Integrating over half of the BZ [or near K (K’)], we get the valley Chern number of $+(1/2)$ for K and $-(1/2)$ for K’. If the A and B sites are exchanged, we get the valley Chern number of $-(1/2)$ for K and $+(1/2)$ for K’ (see “Side Note 9”).

Furthermore, we can see that the signs of both flip from the lower band to the higher band. This indicates that valley-polarized topological edge states exist within the bandgap at an interface between structures with opposite unit cell orientations (between A–B and B–A). The interface will, therefore, have A sites adjacent to A sites or B sites adjacent to B sites. The number of edge modes at each valley, in accordance with the bulk-edge correspondence, is determined by the difference of $C_\nu$ above and below the bandgap: $C_\nu^K - C_\nu^\Gamma = 1$, $C_\nu^{K’} - C_\nu^{\Gamma} = -1$, where each value corresponds to a single valley (K or K’). The differing signs here could be interpreted as the sign of the group velocity of the forward- and backward-propagating modes at the two valleys [19].

We can again build, essentially, the same bent waveguiding model as the Chern PTI to demonstrate the robustness of the valley structure, shown in Figure 10(c) and (d). The valley PTIs also have their own features (see “Side Note 10”):

- Unlike the Chern PTIs (and spin PTIs), where topological edge modes can be formed at the interface between nontrivial and trivial lattices, the valley edge modes exist only at the interface between two valley PTIs with opposite valley Chern numbers ($(1/2)$ and $-(1/2)$).
- Therefore, when constructing a valley waveguide, there must always be a pair of complementary valley PhCs.
- Since the valley edge modes are coupled to K or K’ valleys, these...
edge modes are the most robust in the directions where the valleys are clearly defined (i.e., the Γ-to-K or Γ-to-K’ directions), indicating that they will preserve their unidirectional-ality only when sharp turns are 120°. In contrast to Chern PTIs, defects that scatter valleys into each other (breaking the C₆ᵥ symmetry, for example) will weaken the edge mode and give lower robustness [40].

- Valley states can couple to spin-type modes in certain PTIs [41]. These spin–valley coupled edge states can be used to form “valley splitters” that are not restricted by the orientation of the unit cell [41].

- Unlike many spin PTI designs, it is simple to construct planar or nearly planar valley PTIs, making them an attractive choice for integration with normal silicon [42] or metal-on-insulator designs.

**SPIN PTIs**
The last major PTI type, the spin PTI, is, in a practical sense, similar to the valley type, being reciprocal while still possessing highly robust features. A full example of a recently demonstrated planar metallic PTI [44], which has a number of attractive features for integration with traditional microwave systems [45], is provided in the additional online materials. Spin PTIs are also readily adaptable to dielectric platforms suitable for optical bands [46].

**CONCLUSIONS AND FUTURE OUTLOOK**
In this tutorial, we have given an overview of the concepts, mathematics, and implementations of PTIs. The central idea that relates the topics together is the geometric phase, which lies at the heart of both theory and physical realization. This concept is readily applied to a 2D PhC, which is a simple platform to engineer topological modes. Computations of the Berry curvature, Chern number, and other topological invariants for a given design illuminate how the geometric phase influences the system and can be readily calculated with numerical tools to aid in design.

The three most common formulations of PTIs, the Chern, valley, and spin PTIs, all represent different strategies to achieve the effects of topologically protected modes, lending considerable flexibility to their usage. In all cases, such devices possess remarkable robustness to a wide class of disorder, which could enable much greater fabrication tolerances for applications like extremely robust integrated optical waveguides [38]. Likewise, their immunity to backscattering off of sharp bends has the potential to shrink device footprints by eliminating the gradual bends or careful engineering needed at the edges to overcome losses or higher-order mode mixing when a turn is required in a waveguide [2], [36].

For the case of the Chern PTI devices, there is a large effort to deploy topologically protected lasers [47], with many recent studies realizing arbitrarily shaped optical cavities immune to disorder [48]. Further uses can be seen in isolators [34] and circulators [37]. In the magnetic field-free valley and spin implementations, there is potential to use such devices in place of traditional transmission lines or waveguides, with the added benefits of sharp bend immunity and robustness to disorder [40], [44].

For microwave devices, many of these are functionally similar to traditional metallic structures, and, as such, could be integrated into standard systems with relative ease [40]. At optical frequencies, topologically protected designs could enable features like spin-selective filtering [43] and unidirectional polarization control [19] beyond general robustness.

**ONLINE SUPPLEMENTARY MATERIAL**
This article has supplementary downloadable material available at https://doi.org/10.1109/map.2021.3069276, provided by the authors.

**AUTHOR INFORMATION**
**Dia’aaldin J. Bisharat** (dbisharat@ucsd.edu) is a postdoctoral researcher at the City University of New York Graduate Center, New York, New York, 10016, USA. Previously, he was with the University of California, San Diego, La Jolla. His research interests include novel approaches for guiding electromagnetic waves. He is a Member of IEEE.

**Robert J. Davis** (rjdavis@ucsd.edu) is a Ph.D. student at the University of California, San Diego, La Jolla, California, 92093, USA. His research interests include applying new concepts from condensed-matter physics to electromagnetic problems. He is a Student Member of IEEE.

**Yun Zhou** (yuz421@eng.ucsd.edu) is a Ph.D. student at the University of California, San Diego, La Jolla, California, 92093, USA. Her research interests include phononic topological insulators. She is a Fellow of IEEE.

**Prabhakar R. Bandaru** (prbandaru@ucsd.edu) is a professor at the University of California, San Diego, La Jolla, California, 92093, USA. His research interests include electrochemical energy storage, control of thermal energy, and fluid flow at the nanoscale. He is a Fellow of IEEE.

**Daniel F. Sievenpiper** (dsievenpiper@eng.ucsd.edu) is a professor at the University of California, San Diego, La Jolla, California, 92093, USA. His research interests include...
