The geometry of an atomic lattice and its chemical structure determine the electronic properties of a solid. All the band structures of solids can be classified according to their topological properties. If these properties are nontrivial, we observe such extraordinary phenomena as insulators with conducting surfaces, currents running without dissipation or particles which are their own antiparticles. In this course we aim to introduce the topology as a tool for condensed matter physics, study its consequences in systems of various symmetries and dimensionality, and explore the experimental signatures of nontrivial topology.

I. PREFACE

This course is based on two books: Mikio Nakahara’s “Geometry, topology and physics”\(^1\) provides the mathematical foundation, B. Andrei Bernevig’s “Topological insulators and topological superconductors” provides the physical interpretation. There are also several other books and scripts, where the ideas of the authors help to understand the topic in question in more ways than one. The ones which I used most are

- Shun-Qing Shen “Topological insulators - Dirac equation in condensed matters”
- Joel E. Moore “Notes for the MIT minicourse on topological phases”, [http://socrates.berkeley.edu/~jemoore/Moore_group,_UC_Berkeley/Welcome_to_the_Moore_group,_Department_of_Physics,_UC_Berkeley_files/mitcourse.pdf](http://socrates.berkeley.edu/~jemoore/Moore_group,_UC_Berkeley/Welcome_to_the_Moore_group,_Department_of_Physics,_UC_Berkeley_files/mitcourse.pdf)

II. EQUIVALENCE RELATIONS

In this course we will use several powerful tools. The first one is the correspondence between the physical parameters or quantum numbers of the system and the Hilbert space of quantum states. This correspondence is a mapping, defined by the system’s Hamiltonian.

---

\(^1\) The German translation is available as an ebook from the university library.
The second tool is the equivalence, under certain conditions of the interesting objects whose properties we do not know, and the familiar objects, whose properties we know well. This equivalence, in the context of topology, is called homeomorphism, which means that the two objects can be continuously deformed into each other.

Let us start with maps.

Definitions:
Let $X$, $Y$ be sets. A map or mapping is the rule by which we assign to elements $x \in X$ the elements $y \in Y$. It is denoted by $f : X \to Y$ and $f : x \to f(x)$. The set $X$ is the domain of the map, the set $Y$ the range of the map. The set of all $y$’s such that $y = f(x)$ for some $x \in X$ is called the image of the map. For a given $y$ belonging to the image of the map, the inverse image of $y$ is defined as $f^{-1}(y) = \{x \in X | f(x) = y\}$.

Note that a map is only fully defined when both its domain and range are specified. For instance, consider the following maps:

$$
\begin{align*}
  f : x &\to \exp(x)  &\text{if } &X, Y = \mathbb{R}, &y = -1 &\text{has no inverse image} \\
  & &\text{if } &X, Y = \mathbb{C}, &f^{-1}(y) = \{(2n + 1)\pi i | n \in \mathbb{Z}\} \\
  f : x &\to \sin(x)  &\text{if } &X, Y = \mathbb{R}, &\text{the range of the map is } [-1, 1] \\
  & &\text{if } &X, Y = \mathbb{C}, &\text{the range of the map is the whole complex plane.}
\end{align*}
$$

Maps can have, among other, the following properties:

(i) A map $f : X \to Y$ is injective (or one to one) if $x \neq x'$ implies $f(x) \neq f(x')$.

(ii) A map $f : X \to Y$ is surjective (or onto) if each $y \in Y$ is the image of some $x \in X$.

(iii) If a map is both injective and surjective, it is bijective.

Examples:

$f : \mathbb{R} \to \mathbb{R}$ defined by $f : x \to ax$ ($a \in \mathbb{R}\{0\}$) is bijective,

$f : \mathbb{R} \to \mathbb{R}$ defined by $f : x \to x^2$ is neither injective nor surjective,

$f : \mathbb{R} \to \mathbb{R}$ defined by $f : x \to \exp(x)$ is injective but not surjective,

$f : \mathbb{R} \to \mathbb{R}$ defined by $f : x \to \sin(x)$ is neither injective nor surjective, but if we restrict its domain and range so that $f : [-\pi/2, \pi/2] \to [-1, 1]$, it is bijective.

Definitions: Given two maps $f : X \to Y$ and $g : Y \to Z$, the composite map of $f$ and $g$ is a map $g \circ f : X \to Z$ defined by $g \circ f(x) = g(f(x))$. A diagram of maps is called commutative if any composite maps between a pair of sets do not depend on how they are composed. For instance, $f, g : \mathbb{R} \to \mathbb{R}$, where $f : x \to x + n, \ n \in \mathbb{Z}$ and $g : x \to \text{frac}(x) := x - \lfloor x \rfloor$ form a commutative diagram of two maps, illustrated below.

The identity map $\text{id}_X : X \to X$ is defined by $\text{id}_X(x) = x$. If $f : X \to Y$ defined by $f : x \to f(x)$ is bijective, there exists an inverse map $f^{-1} : Y \to X$ such that $f^{-1} \circ f : x \to x$, which is also bijective. The maps $f$ and $f^{-1}$ satisfy $f \circ f^{-1} = \text{id}_Y$ and $f^{-1} \circ f = \text{id}_X$. Conversely, if $f : X \to Y$ and $g : Y \to X$ satisfy $f \circ g = \text{id}_Y$ and $g \circ f = \text{id}_X$, then $f$ and $g$ are bijections.

Examples:

$f : \mathbb{R} \to (0, \infty)$ is a bijection defined by $f(x) = \exp(x)$. The inverse map is $f^{-1} : x \to \ln x$.

Let $M$ be an element of the general linear group $\text{GL}(n, \mathbb{R})$, whose matrix representation is given by $n \times n$ matrices with non-vanishing determinant. Then $M : \mathbb{R}^n \to \mathbb{R}^n, x \to Mx$ is bijective and the inverse map is given by the inverse of the matrix $M$, $M^{-1}$. (If $\det M = 0$, $M$ would be neither injective nor surjective.)

The $n$-dimensional Euclidean group $E^n$ is made of an $n$-dimensional translation $a : x \to x + a \ (x, a \in \mathbb{R}^n)$ and an $O(n)$ rotation, $R : x \to Rx, R \in O(n)$. A general element $(R, a)$ of $E^n$ acts on $x$ by $(R, a) : x \to Rx + a$. The product is defined by $(R_2, a_2) \circ (R_1, a_1) : x \to R_2(R_1x + a_1) + a_2$, i.e. $(R_2, a_2) \circ (R_1, a_1) = (R_2R_1, R_2a_1 + a_2)$. The inverse maps are $a^{-1} : x \to x - a, R^{-1} : x \to R^{-1}x, (R, a)^{-1} : x \to R^{-1}x - R^{-1}a - a$. 
Definitions: Suppose that the sets $X, Y$ are endowed with some algebraic structures (product, addition...). If $f : X \to Y$ preserves these structures, it is called a homomorphism. For instance, if the structure in question is a product, then $f : ab \to f(a)f(b)$. Note that $ab$ is a product in $X$, while $f(a)f(b)$ is a product in $Y$. If a homomorphism is bijective, it is called an isomorphism and $X$ is said to be isomorphic with $Y$, $X \cong Y$.

The elements of a set can be in a relation, which is also a mapping: $\sim : X \times X \to \{\text{True}, \text{False}\}$. The examples of relations are $>, <, =, \in$. If the relation satisfies the three following conditions:

(i) $a \sim a$ (reflective)

(ii) if $a \sim b$ then $b \sim a$ (symmetric)

(iii) if $a \sim b$ and $b \sim c$ then $a \sim c$ (transitive),

the relation is called an equivalence relation. One example of such a relation is introduced by the division modulo $2$: $x \sim y$ if $(x - y)_{\text{mod} 2} = 0$.

Given a set $X$ and an equivalence relation $\sim$, we can partition $X$ into mutually disjoint sets called equivalence classes. A class $[a]$ is made of all the elements $x \in X$ such that $x \sim a$: $[a] = \{ x \in X | x \sim a \}$

The set of all equivalence classes is called the quotient space, denoted by $X/ \sim$. The element $a$ (or any element in $[a]$) is called a representative of $[a]$.

Examples:

The division modulo 2 separates $\mathbb{Z}$ into even and odd integers. We can choose the representatives to be 0 and 1. Then the quotient space $\mathbb{Z}/\text{mod} 2$ is isomorphic with the cyclic group $\mathbb{Z}_2$, whose algebra is defined by $0 + 0 = 0, 0 + 1 = 1 + 0, 1 + 1 = 0$. If the division is modulo $n$, the quotient space $\mathbb{Z}/\text{mod} n$ is isomorphic with $\mathbb{Z}_n$.

Another important quotient space is $\mathbb{R}$, where we have impose the equivalence of all points which are separated by $2\pi n$, where $n \in \mathbb{Z}$ ($x \sim y$ if there exists such $n \in \mathbb{Z}$ that $y = x + 2\pi n$). The class $[x]$ is the set $\{...x - 2\pi, x, x + 2\pi,...\}$. All classes in $\mathbb{R}$ have their representatives in the interval $[0, 2\pi)$, which is the quotient space $\mathbb{R}/ \sim$. We can picture it as sketched below.

\[ \mathbb{R} \text{ with the equivalence relation} \]

\[ \mathbb{R}/ \sim \]

\[ \mathbb{S}^1 \]

Note that the points $x = \varepsilon$ and $x = 2\pi - \varepsilon$ lie far from each other on $\mathbb{R}$, but close in the quotient space. The notion of the closeness of points is one of the central ones in topology.

Similar equivalence relation can be defined on $\mathbb{R}_2$, and they yield several quotient spaces which are either very interesting, or very useful, or both.

(i) Let $(x_1, y_1) \sim (x_2, y_2)$ if $x_2 = x_1 + 2\pi n$, where $n \in \mathbb{Z}$. The resulting quotient space is a cylinder.

---

2 Prove that two classes $[a]$ and $[b]$ satisfy either $[a] = [b]$ or $[a] \cap [b] = \emptyset$. 

---
(ii) Let \((x_1, y_1) \sim (x_2, y_2)\) if \(x_2 = x_1 + 2\pi n \land y_2 = y_1 + 2\pi m\), where \(n, m \in \mathbb{Z}\). The resulting quotient space is a torus, denoted \(T^2\). We will encounter it very often, as it is the Brillouin zone of a 2D square lattice.

(iii) Let \((x_1, y_1) \sim (x_2, y_2)\) if \(x_2 = x_1 + 2\pi n \land y_2 = (−1)^n y_1\), where \(n \in \mathbb{Z}\). The resulting quotient space is called a Möbius strip or Möbius band. It is non-orientable - it has only one surface, and is in fact a subspace of all other non-orientable spaces\(^3\).

(iv) Let \((x_1, y_1) \sim (x_2, y_2)\) if \(x_2 = x_1 + 2\pi n \land y_2 = (−1)^n y_1 + 2\pi m\), where \(n, m \in \mathbb{Z}\). The quotient space is a so-called Klein bottle, which is actually impossible to embed in our \(\mathbb{R}^3\) space without it intersecting itself. In higher dimensions the Klein bottle has no intersections with itself - just as a twisted 1D loop has a self-intersection if it is embedded in \(\mathbb{R}^2\), but in \(\mathbb{R}^3\) we can lift one of the loop’s sides above the plane.

---

\(^3\) You are invited to try to cut the Möbius band along the central line or at one third and two thirds of the strip’s height. Interesting figures result.
(v) Let \((x_1, y_1) \sim (x_2, y_2)\) if \(x_2 = (-1)^m x_1 + 2\pi n\) \& \(y_2 = (-1)^n y_1 + 2\pi m\), where \(n, m \in \mathbb{Z}\). The resulting quotient space is the real projective plane, denoted \(RP^2\), also called a cross-cap.

The cross-cap is self-intersecting in \(\mathbb{R}^3\). It can be visualized as a hemisphere where the antipodal points on the equator are identical. The cross-cap represents also the set of directions which a non-oriented rod (i.e. not an arrow, which has a head and a tail) can take. Because the rod is not oriented, the directions differing by \(\pi\) correspond to the same position of the rod.

The twisted quotient spaces, the M"obius band, the Klein bottle and the cross-cap are so unusual that they keep inspiring artists of any ilk. See below a doubly twisted architectonic M"obius band, a glass-blown Klein bottle and a crocheted cross-cap.

Design: Vincent Callebaut, source: Vincent Callebaut Architectures


Design: Matthew Wright, http://mrwright.name/crochet/
III. TOPOLOGICAL SPACES AND INVARIANTS

There are several definitions of topological spaces. Here I give two: one which emphasizes the closeness of points (the neighbourhood definition), one which emphasizes their separability (the open sets definition).

Definitions:

Open set definition. Let $X$ be any set and $\mathcal{T}$ denote a certain collection of subsets of $X$. The pair $(X, \mathcal{T})$ is a topological space if $\mathcal{T}$ satisfies the following requirements:

(i) $\{\emptyset, X\} \in \mathcal{T}$
(ii) the union (finite or infinite) of any collection of subsets from $\mathcal{T}$ also belongs to $\mathcal{T}$
(iii) the intersection of a finite number of subsets from $\mathcal{T}$ also belongs to $\mathcal{T}$.

$\mathcal{T}$ is then said to give a topology to $X$.

Neighbourhood (Hausdorff) definition. Let $X$ be a set and $N$ a function assigning to each element from $X$ (henceforth called a point) a non-empty collection of subsets of $X$. The elements $n \in N(x)$ are called neighbourhoods of $x$ with respect to $N$. $N$ is called a neighbourhood topology if the axioms below are satisfied:

(i) if $n \in N(x)$, $x \in n$ (each point belongs to all its neighbourhoods)
(ii) if $m \in X$ and contains a neighbourhood of $x$, $m \in N(x)$ (a superset of a neighbourhood of $x$ is again a neighbourhood of $x$)
(iii) the intersection of two neighbourhoods of $x$ is a neighbourhood of $x$
(iv) any neighbourhood $n$ of $x$ contains a neighbourhood $m$ of $x$ such that $n$ is a neighbourhood of each point in $m$.

Then $(X, N)$ is called a topological space.

Examples:

Let $X = \{A, B, C\}$. We can define several $\mathcal{T}$’s:

\[
\begin{align*}
\mathcal{T}_1 &= \{\emptyset, X\} \\
\mathcal{T}_2 &= \{\emptyset, A, X\} \\
\mathcal{T}_3 &= \{\emptyset, A, \{A, B\}, X\} \\
\mathcal{T}_4 &= \{\emptyset, \{A, B\}, \{B, C\}, X\}
\end{align*}
\]

The sets $\mathcal{T}_1$ and $\mathcal{T}_2$ define indeed topologies. The set $\mathcal{T}_3$ does not, because it violates the third axiom of the open set definition, $B = \{A, B\} \cap \{B, C\} \notin \mathcal{T}_3$.

There are several special ways in which we can introduce a topology. If $\mathcal{T} = \{\emptyset, X\}$, we have a trivial topology - there is only one neighbourhood and all points belong to it. If $\mathcal{T}$ contains all subsets of $X$ (including the intersections of infinite collection of open sets), the topology is discrete. The usual topology on $\mathbb{R}$ contains all open intervals $(a, b)$ and their unions $(a, b)$ can be $-\infty, +\infty$. By extension, on $\mathbb{R}^n$ the usual topology contains all open intervals $(a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_n, b_n)$ and their unions.

The set $X$ can be endowed with a mapping $d : X \times X \rightarrow \mathbb{R}$, such that

i) $d(x, y) = d(y, x)$
ii) $d(x, y) \geq 0$, and $d(x, y) = 0$ if and only if $x = y$
iii) $d(x, y) + d(y, z) \leq d(x, z)$ (triangle inequality).
This map is then called a **metric** and it gives $X$ a topology defined as the collection of all open discs $U_\varepsilon(x) = \{ y \in X \mid d(x, y) < \varepsilon \}$. This is the **metric topology**. The topology on $X$ may be inherited from its superset; if $(X, T)$ is a topological space, we can define a **relative topology** on $X' \subseteq X$, given by $T' = \{ U_i \in T \mid U_i \in X' \}$. In this way the sphere $S^n$, defined as $\{ x \in \mathbb{R}^{n+1} \mid d(x, 0) = 1 \}$, inherits the topology defined by the metric on $\mathbb{R}^{n+1}$.

When are two topological spaces equivalent, and what does “equivalence” mean in this context? The central idea in topology is that of **homeomorphism**. A map $f : X \to Y$ is said to be **continuous** if the inverse images of open sets in $Y$ are open sets in $X$. (Note that this relation is not symmetric: the image of an open set in $X$ does not have to be open, consider e.g. $f : \mathbb{R} \to \mathbb{R}$ such that $f(x) = x^2$. It is continuous, and the image of the open set $(-\varepsilon, \varepsilon)$ is $[0, \varepsilon^2]$ which is neither open nor closed.)

If the continuous map $f : X \to Y$ has an inverse map $f^{-1} : Y \to X$, which is also continuous, it is called a **homeomorphism**. It defines an equivalence relation, dividing all topological spaces into equivalence classes.

This is the source of the common statement that “topology is the science of objects made of rubber” - the equivalence classes consist of spaces which can be continuously deformed into each other, without cutting or pasting. Two examples are shown below - the first one is the famous coffee cup/doughnut equivalence. The second is more subtle: two linked rings (a so-called Hopf link) are homeomorphic to two unlinked rings. We can continuously deform these two objects into each other only when we embed them into $\mathbb{R}^4$, which requires some stretching of the imagination. We can also define the homeomorphism directly.

![a cup is homeomorphic with a doughnut](image1)

![linked rings are homeomorphic with unlinked rings](image2)

If we do not know the homeomorphism though, in general there is no way to determine whether two spaces are homeomorphic. We can say what needs to be conserved when deforming one space to the other, i.e. define **topological invariants**. It does not help us to make a positive statement that two spaces are homeomorphic, because even if some invariants are the same, others may be different, and we can never be certain that we know all of them. What we can say instead, is that if we find a topological invariant which is different in two spaces, they are certainly not homeomorphic.

**Definitions:**

*Connectedness.* A topological space $X$ is **connected** if it cannot be written as a disjoint sum of other spaces. It is **arcwise connected** if for any two points $x, y \in X$ we can define a continuous mapping $f : [0, 1] \to X$ such that $f(0) = x$ and $f(1) = y$. For all spaces with which we will be concerned in this course the two definitions will be equivalent.

A **loop** is a map $f : [0, 1] \to X$ such that $f(1) = f(0)$. If any loop in $X$ can be continuously shrunk to a point, $X$ is **simply connected**. For instance: $\mathbb{R} \setminus \{0\}$ is disconnected. $\mathbb{R}^2 \setminus \{0\}$ is connected, but not simply connected, and $\mathbb{R}^3 \setminus \{0\}$ is simply connected.

*Compactness.* A set $A \subseteq X$ is called **closed** if its complement $X \setminus A$ is open. The smallest closed set containing $A$ is called a **closure** of $A$ and denoted $\overline{A}$. The largest open set contained in $A$ is called the **interior** of $A$ and denoted $A^\circ$, while $A \setminus A^\circ$ is the **boundary**. In $\mathbb{R}^n$ the set is compact if it is closed and bounded. This is an important property - in fact, we need it for most of our band structure calculations, when we apply periodic boundary conditions on an infinite bulk lattice, turning it effectively from an $\mathbb{R}^n$ into $T^n$. Note, however, that the procedure of compactification changes a topological invariant - $\mathbb{R}^n$ and $T^n$ are not homeomorphic. We are allowed to work on a $T^n$ instead of $\mathbb{R}^n$, because our $\mathbb{R}^n$ has an equivalence relation $\sim$, the set of translations by lattice vectors. Now, the quotient space $\mathbb{R}^n / \sim$ is homeomorphic to $T^2$.

Basing on these two properties we can already identify some pairs of spaces which are not homeomorphic.
1. The closed interval $[0, 1]$ is not homeomorphic to the open interval $(0, 1)$ even though both are simply connected, because one is compact and the other not.

2. Interestingly, an open interval is homeomorphic with the whole $\mathbb{R}$. As an example consider $X = (-\pi/2, \pi, 2), Y = \mathbb{R}$ and the map $f : X \to Y, f(x) = \tan(x)$. This map is continuous and bijective, and has an inverse $f^{-1}(y) = \arctan(y)$, which is also continuous and bijective given our choice of $X$ and $Y$. Thus boundedness is not a topological invariant.

3. The circle $S^1$ and the closed interval $[0, 1]$ are not homeomorphic, even though both are compact and connected, because the latter is simply connected while the former isn’t.

4. A sphere $S^2\setminus\{N\}$ (without the North Pole) and the plane $\mathbb{R}^2$ are homeomorphic, where the homeomorphism is defined by the stereographic projection (see below). Both are simply connected and non-compact.

If we close the sphere by adding the North Pole and close $\mathbb{R}^2$ by adding $\infty$, we obtain again two homeomorphic spaces, $S^2$ and $\mathbb{R}^2 \cup \{\infty\}$, this time both doubly connected and compact. This procedure is called one-point compactification.

The properties defined above are topological invariants, but there are also others, which are more quantifiable.

One of them is based on the equivalence of maps. Let $f, g : X \to Y$ be continuous maps. If we can define a continuous map $F : X \times [0, 1] \to Y$ such that $F(x, 0) = f(x)$ and $F(1) = g(x)$, the map is said to be homotopic to $g$ (denoted $f \sim g$) and $F$ is the homotopy between $f$ and $g$. If there exist continuous maps $f : X \to Y$ and $g : Y \to X$ such that $f \circ g \sim \text{id}_Y$ and $g \circ f \sim \text{id}_X$, $Y$ is said to be of the same homotopy type as $X$. The map $f$ is then the homotopy equivalence, and $g$ the homotopy inverse.

Example:

An interval $X = (0, 1)$ is of the same homotopy type as a point $Y = \{0\}$. We can define $f : X \to Y, f(x) = 0$ and $g : Y \to X, g(y) = 1/2$. The composite map $g \circ f = \text{id}_Y$, and $f \circ g \sim \text{id}_X$, because we can find a continuous map $F : X \times [0, 1] \to X$, given by $F(x, a) = ax + (1 - a)/2$ and indeed $F(x, 0) = 1/2$ and $F(x, 1) = x$.

If two spaces $X$ and $Y$ are homeomorphic, they are also of the same homotopy type, but the reverse is not true. For instance, a disc is of the same homotopy type as a point (disc can be continuously shrunk to a point), but the two are not homeomorphic.

Nevertheless, “of the same homotopy type” defines an equivalence relation in the set of topological spaces, albeit weaker than the homeomorphism does.

The mapping $f : [0, 1] \to X$ defines a path in $X$. If on top of that $f(0) = f(1)$, the path is a loop. The concatenation of paths defines a group structure in the space of all paths on $X$, with the neutral element being the constant path (or constant loop) which maps $[0, 1]$ onto a single point. The relation of homotopy defines the equivalence classes of loops which can be continuously deformed into each other. The set of all the equivalence classes defines the fundamental (or first) homotopy group of $X$, $\pi_1(X)$.

Examples:

1. All loops in $\mathbb{R}^2$ can be continuously shrunk to a point, therefore $\pi_1(\mathbb{R}^2) = 0$.

2. All loops on $S^1$ fall into the homotopy classes defined by the number of times they encircle $S^1$, called their winding number, as illustrated below. The black paths are some arbitrary loops, and the blue ones are the representants of their homotopy class.
The paths which do not encircle $S^1$ are homotopic with a constant loop and have winding number $w = 0$. Those which encircle it once ($w = 1$) are all in the same homotopy class, and they cannot be deformed to those which encircle it twice ($w = 2$). The first homotopy group of $S^1$ is then $\pi_1(S^1) = \mathbb{Z}$.

(Juan-Diego has a nice analytical insight here: if we parametrize each path as $\exp(i\omega t)$, with $t \in [0, 1]$, the loops are closed only if $\omega = 2\pi n$, with $n \in \mathbb{Z}$.)

3. The closed paths on the real projective plane, shown below, fall into two homotopy classes: those which cross the equator ($\sim \beta$) and those which don’t ($\sim \alpha$). Those which cross the equator twice ($\sim \gamma$), or indeed any even number of times, can be again shrunk to a point and thus are homotopic to $\alpha$. The first homotopy group of $\mathbb{R}P^2$ is $\pi_1(\mathbb{R}P^2) = \mathbb{Z}_2$.

If two spaces $X, Y$ are homeomorphic, their first homotopy groups are equal. Again, the reverse statement is not true. Also higher homotopy groups can be defined, consisting of the equivalence classes of closed surfaces and hypersurfaces. We’ll return to the homotopy groups when we arrive at the classification of topological defects (if we have time).

**Euler characteristic.** Let us now restrict ourselves to $\mathbb{R}^3$ and the polyhedra embedded into it. A polyhedron is a closed object consisting of $N_F$ faces, $N_E$ edges and $N_V$ vertices - we shall call them all simplexes. Note that the boundary of a simplex is a simplex one level lower, e.g. the boundary of a face is a set of edges.

Let $X$ be a figure in $\mathbb{R}^3$ which is homeomorphic to a polyhedron $K$. The **Euler characteristic** of $X$ is then defined as

$$\chi(X) = N_F(K) - N_E(K) + N_V(K).$$

The Poincaré-Alexander theorem states that the Euler characteristic $\chi(X)$ does not depend on what polyhedron $K$ we choose to calculate it, as long as $K$ is homeomorphic with $X$.

**Examples:**

1. The Euler characteristic of a point is $\chi(\text{point}) = 0 - 0 + 1 = 1$. The Euler characteristic of a line is $\chi(\text{line}) = 0 - 1 + 2 = 1$. This example shows that the equality of Euler characteristics does not imply the homeomorphism between sets.

2. The Euler characteristic of a cube is $\chi(\text{cube}) = 6 - 12 + 8 = 2$. The Euler characteristic of a sphere can be found from our quotient-space picture: the sphere is homeomorphic to a square patch with the edges identified as shown below. (The edges of the same colour become one edge, the vertices of the same colour become one vertex, and there is only one face.)
\( \chi(S^2) = 1 - 2 + 3 = 2 \), which is indeed the same as that of the cube. That was the content of the original Euler’s theorem: for every polyhedron homeomorphic to \( S^2 \), \( N_F - N_E + N_V = 2 \).

3. We can apply the same method to finding the Euler characteristic of a cylinder (diagram shown below).

\( \chi(\text{cylinder}) = 1 - 3 + 2 = 0 \), which is actually the same as that of the Möbius strip. The Euler characteristic does not know about the orientability of the surface in question.

4. Let us take the torus \( T^2 \). We can continuously deform it into a polyhedron \( K \) shown below, with \( \chi(T^2) = \chi(K) = 12 - 24 + 12 = 0 \).

We can also calculate it from the geometric representation of the torus as a quotient space of \( \mathbb{R}^2 \) in the previous lecture. The four vertices of the square patch are identified into one, out of four edges only two remain, and the patch has one face. Then \( \chi(T^2) = 1 - 2 + 1 = 0 \).

5. In the same way we can calculate the \( \chi \) of a Klein bottle (\( \chi(\text{Klein bottle}) = 1 \)) and of a real projective plane (\( \chi(\mathbb{R}P^2) = 1 \)).

The connected sum \( X \# Y \) of two spaces \( X \) and \( Y \) is constructed by removing a disc from each of the two spaces and connecting the created holes by a cylinder, as shown below for two spheres.
The Euler characteristic of a connected sum is given by
\[ \chi(X \# Y) = \chi(X) + \chi(Y) - 2. \]

This is easy to prove, if we deform the cylinder to a trigonal prism, which we are allowed to do. In the process of removing the discs (now triangles) we have added to each space three vertices and three edges, while removing one face. By connecting them with the prism we have added three edges and three faces. In the end
\[ \chi(X \# Y) = (\chi(X) - 1 - 3 + 3) + (\chi(Y) - 1 - 3 + 3) + 3 - 3 = \chi(X) + \chi(Y) - 2. \]

The Euler characteristic can also be defined for smooth surfaces, and there it is given by the Gauss-Bonnet theorem as the integral of the local curvature over the whole surface. We leave now the domain of pure topology and enter that of differential manifolds. Since the differential geometry is usually treated in the courses on electrodynamics and field theory, I’ll assume most of the basic definitions to be known.

Let us consider a smooth (i.e. infinitely differentiable) closed surface \( M \) of arbitrary shape. At every point \( p \) on the surface we can define a tangent plane \( T_p \). In the local system of Cartesian coordinates on the plane the equation of the surface is given by
\[ z = z(x, y), \]
with \( z(0, 0) = 0 \) since \( (x, y)_p = (0, 0) \). The equation of the surface in local coordinates can be approximated (locally) as
\[ z(x, y) \approx \frac{1}{2} \begin{pmatrix} x \\ y \end{pmatrix} \begin{pmatrix} \frac{\partial^2 z}{\partial x^2} & \frac{\partial^2 z}{\partial x \partial y} \\ \frac{\partial^2 z}{\partial y \partial x} & \frac{\partial^2 z}{\partial y^2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \]

(2)

The Hessian matrix contains only the second derivatives because the tangency of the plane requires \( \frac{\partial z}{\partial x}|_p = \frac{\partial z}{\partial y}|_p = 0 \). The matrix is real and symmetric. It can be diagonalized and has two eigenvalues \( \lambda_1, \lambda_2 \) which correspond to two eigendirections in the local coordinate system. The magnitude of \( \lambda_i \) is the inverse radius of curvature and the sign tells whether the surface curves towards the positive \( z \) direction in our local system, or away from it. The Gaussian curvature \( \Omega_p \) is the determinant of the Hessian at the point \( p \). In general it can be positive, negative, or zero. The Gauss-Bonnet theorem states that
\[ \frac{1}{2\pi} \int_M d\sigma \Omega_p = 2(1 - g) = \chi(M). \]

(3)

The number \( g \) is the so-called genus of the surface, equivalent to the number of “handles” in the surface\(^4\).

Example:
Let us consider a sphere \( S^2 \) with radius \( R \). The equation of the sphere in the local coordinates of the tangent plane at point \( p \) is
\[ z = R - \sqrt{R^2 - x^2 - y^2} \approx \frac{x^2 + y^2}{2R}, \]
and the Hessian matrix is
\[ H_p = \begin{pmatrix} 1/R & 0 \\ 0 & 1/R \end{pmatrix}. \]

We obtain then \( \Omega_p = 1/R^2 \), independent of \( p \). Its integral over the sphere is
\[ \int_0^{2\pi} d\varphi \int_0^\pi d\theta R^2 \sin^2 \theta \frac{1}{R^2} = 4\pi = 2\pi(2 - 0). \]

The genus of the sphere is confirmed to be 0.

\(^4\) The Euler characteristic can also be defined for surfaces with a boundary. We calculate it by first capping off each boundary component with a disc, and subtracting the number of discs \( (\chi(\text{disc}) = 1) \) from the Euler characteristic of the closed surface.


IV. GEOMETRIC PHASE

The Euler characteristic was our first topological invariant which could be expressed in terms of an integral over our topological space. We are going to see now another, directly related to physics, and it will be directly related to the so-called geometric or Berry phase.

The original 1984 paper of Michael Berry (now Sir Michael Berry) described the evolution of a quantum system under an adiabatic variation of its parameters. An adiabatic evolution means that the system at every moment in time is in an eigenstate of the current Hamiltonian. The variation in time happens through a manipulation of the parameters which define the Hamiltonian - those parameters may be the electric field, the magnetic field, the pressure, the deformation of the atomic lattice and many others. In the following we are going to use a generic set of parameters \( R = R(t) \). We introduce a so-called snapshot basis, defined by

\[
|\psi_n(t)\rangle = \exp(i\gamma_n(t)) \exp\left\{-\frac{i}{\hbar} \int_{t_0}^{t} dt' \varepsilon_n(t')\right\} |n(R(t))\rangle.
\]

The adiabatic evolution along a closed loop in the parameter space brings the state to the same as the initial one but with a possible difference in phase.

\[
\langle \psi_n(t) | = \exp(i\gamma_n(t)) \exp\left\{-\frac{i}{\hbar} \int_{t_0}^{t} dt' \varepsilon_n(t')\right\} |n(R(t))\rangle.
\]

The second phase factor is the usual dynamical phase, the first phase factor is the geometric phase. Using the Schrödinger equation

\[
\frac{i\hbar}{\partial t} |\psi_n(t)\rangle = H(R(t)) |\psi_n(t)\rangle,
\]

we obtain

\[
\text{LHS: } \quad \frac{i\hbar}{\partial t} |\psi_n(t)\rangle = -\hbar \gamma_n(t) |\psi_n(t)\rangle + \varepsilon_n(t) |\psi_n(t)\rangle + i\hbar \exp(i\gamma_n(t)) \exp\left\{-\frac{i}{\hbar} \int_{t_0}^{t} dt' \varepsilon_n(t')\right\} \frac{\partial}{\partial t} |n(R(t))\rangle,
\]

\[
\text{RHS: } \quad H(R(t)) |\psi_n(t)\rangle = \varepsilon_n(t) |\psi_n(t)\rangle,
\]

\[
\Rightarrow \quad \gamma_n(t) = i \langle n(R(t)) | \frac{\partial}{\partial t} |n(R(t))\rangle = i \langle n(R) | \nabla_R |n(R)\rangle \frac{\partial R}{\partial t},
\]

where the whole dependence on time is now included in the parameters \( R \). The integration over the cycle \( C \) in the parameter space yields the total Berry phase

\[
\gamma_n(C) = i \int_{t_0}^{t_C} dt' \langle n(R) | \nabla_R |n(R)\rangle \frac{\partial R}{\partial t'} = i \oint_C dR \langle n(R) | \nabla_R |n(R)\rangle.
\]

The integrand in this formula is the so-called Berry connection,

\[
\mathcal{A}_n(R) = i \langle n(R) | \nabla_R |n(R)\rangle.
\]

The meaning of a “connection” is exactly what it says. It provides us with the information on how the objects change when they are transported from one location on our manifold to another. One type of a connection is the parallel transport, which tells us how to transport vectors from one point on our surface to another so that they remain parallel. The recipe, in a nutshell, is “connect the two points with a geodesic line and transport the vector in such a way that the angle between the vector and the geodesic is constant”. On some surfaces, such as that of a cylinder, it is possible to transport the vector on any closed path so that its final direction is parallel to the initial one. On the surface of a sphere this is possible only when the closed path is a geodesic one, i.e. a great circle, as shown below.
The connection $A_n(R)$ is gauge-dependent. If we transform $|n(R)\rangle \rightarrow e^{i\xi_n(R)} |n(R)\rangle$,
\[ A_n(R) \rightarrow i \langle n(R)| e^{-i\xi_n(R)} \nabla R e^{i\xi_n(R)} |n(R)\rangle = -\nabla R \xi_n(R) \langle n(R) | n(R) \rangle + i \langle n(R) | \nabla R | n(R) \rangle = A_n(R) - \nabla R \xi_n(R). \]
The total Berry phase changes by $\zeta(R(T)) - \zeta(R(0))$. If we impose that the phase must be smooth and single-valued, this phase difference must be an integer multiple of $2\pi$. This implies that the Berry phase is defined only modulo $2\pi$, i.e. it has an integer ambiguity.

From the conservation of the norm it follows that the Berry connection, as well as the Berry phase, is purely imaginary:
\[ \Im \langle n(t) | n(t) \rangle = \langle \partial_t n(t) | n(t) \rangle + \langle n(t) | \partial_t n(t) \rangle \rightarrow (\partial_t n(t) | n(t)) = \langle n(t) | \partial_t n(t) \rangle^* = - \langle n(t) | \partial_t n(t) \rangle \in i\mathbb{R}. \]

Basing on our connection, we can define in analogy to electrodynamics a gauge field tensor, called Berry curvature
\[ \Omega_{\mu\nu}(R) = \frac{\partial}{\partial R^\mu} A^\nu_n(R) - \frac{\partial}{\partial R^\nu} A^\mu_n(R) = i \left[ \left( \frac{\partial n}{\partial R^\mu} \right) \frac{\partial n}{\partial R^\nu} - \left( \frac{\partial n}{\partial R^\nu} \right) \frac{\partial n}{\partial R^\mu} \right]. \]
In contrast to the Berry connection, Berry curvature is gauge-invariant, which is easily checked. In the particular case of $\mathbb{R}^3$ we can define the Berry field vector, $\Omega^\mu_n(R) = \nabla_R \times A_n(R)$, which is connected to the Berry curvature tensor by $\Omega_{\mu\nu}(R) = \epsilon_{\mu\nu\lambda}(\Omega^\lambda_n(R))$. Using Stokes theorem we can also express Berry phase as a surface integral
\[ \gamma_n(C) = \oint_C dR \cdot A_n(R) = \iint_{S_C} dR^\mu \wedge dR^\nu \frac{1}{2} \Omega_{\mu\nu}^n(R) = \iint_{S_C} dS \cdot \Omega^n(R), \]
where $S_C$ is any surface whose boundary is the loop $C$.

The Berry connection and Berry curvature can be recast in a form which does not require the differentiation of the snapshot basis, which is especially a boon when our basis states have been obtained numerically. Since the phase $\gamma_n(C)$ is purely imaginary, we can write it as
\[ \gamma_n(C) = -\Im \iint_{S_C} dS \cdot \nabla_R \times |n \rangle \langle \nabla_R n| = -\Im \iint_{S_C} dS \cdot (\nabla_R |n\rangle \times |\nabla_R n\rangle) = -\Im \iint_{S_C} dS \cdot \sum_{m \neq n} \langle \nabla_R n | m \rangle \times \langle m | \nabla_R n \rangle, \]
\[ \Longleftrightarrow \langle n | \nabla_R n \rangle \text{ is imaginary}. \]

From the Schrödinger equation we have
\[ \langle m | H \nabla_R | n \rangle = \langle m | \nabla_R (H | n \rangle) - \langle m | (\nabla_R H) \rangle | n \rangle, \]
\[ E_m \langle m | \nabla_R n \rangle = E_n \langle m | \nabla_R n \rangle + \nabla_R \langle m | n \rangle - \langle m | (\nabla_R H) \rangle | n \rangle, \]
\[ \Rightarrow \langle m | \nabla_R n \rangle = \frac{\langle m | (\nabla_R H) \rangle | n \rangle}{E_n - E_m}, \]
and in consequence
\[ \Omega^n(R) = \Im \sum_{m \neq n} \frac{\langle m | (\nabla_R H) | m \rangle \times \langle m | (\nabla_R H) | n \rangle}{(E_n - E_m)^2}. \]

We have just obtained a very useful result - we have shifted the differentiation with respect to $R$ from the eigenstates to the Hamiltonian itself. Since the dependence of the Hamiltonian on $R$ is known, this will simplify considerably any further calculations of the Berry effects.

A consequence of the above formula is that the sum of all Berry curvatures of all eigenstates vanishes for every $R$,
\[ \sum_n \Omega_{\mu\nu}^n(R) = 0. \]

\section*{A. Two-level system}

The simplest system in which the Berry phase may appear is a two-level one. Its Hamiltonian is a $2 \times 2$ Hermitian matrix, and as such it can be written as a linear combination of Pauli matrices,
\[ H = h(R) \cdot \sigma = \begin{pmatrix} h_z & h_x - i h_y \\ h_x + i h_y & -h_z \end{pmatrix} = h \begin{pmatrix} \cos \theta & \sin \theta e^{-i \phi} \\ \sin \theta e^{i \phi} & -\cos \theta \end{pmatrix}. \]
where we have parametrized our Hamiltonian with $h = h(\sin \theta \cos \varphi, \sin \theta \sin \varphi, \cos \theta)$. The eigenvalues are $\pm h(R)$, and the eigenvectors can be chosen as

$$u_+ = \left( \cos \frac{\theta}{2} e^{-i\varphi}, \sin \frac{\theta}{2} \right), \quad u_- = \left( \sin \frac{\theta}{2} e^{-i\varphi}, -\cos \frac{\theta}{2} \right).$$

These vectors are well defined everywhere except at $\theta = \pi$ (the south pole) for $u_+$ and $\theta = 0$ (north pole) for $u_-$. Since the upper component of these eigenvectors vanishes, the phase $\varphi$ is ill-defined.

Let us calculate Berry connection and curvature.

$$A^+_{\theta} = i(\cos \frac{\theta}{2} e^{i\varphi}, \sin \frac{\theta}{2}) \cdot \left( -\frac{1}{2} \sin \frac{\theta}{2} e^{-i\varphi} \right) = 0, \quad A^+_{\varphi} = i(\cos \frac{\theta}{2} e^{i\varphi}, \sin \frac{\theta}{2}) \cdot \left( -i \cos \frac{\theta}{2} e^{-i\varphi} \right) = \cos^2 \frac{\theta}{2},$$

$$A^-_{\theta} = i(\sin \frac{\theta}{2} e^{i\varphi}, -\cos \frac{\theta}{2}) \cdot \left( \frac{1}{2} \cos \frac{\theta}{2} e^{-i\varphi} \right) = 0, \quad A^-_{\varphi} = i(\sin \frac{\theta}{2} e^{i\varphi}, -\cos \frac{\theta}{2}) \cdot \left( -i \sin \frac{\theta}{2} e^{-i\varphi} \right) = \sin^2 \frac{\theta}{2},$$

and the non-vanishing components of the Berry curvature are

$$\Omega^+_{\theta\varphi} = \partial_{\theta} A^+_{\varphi} - \partial_{\varphi} A^+_{\theta} = 2 \cos \frac{\theta}{2} \left( -\sin \frac{\theta}{2} \right) \frac{1}{2} = -\frac{1}{2} \sin \theta,$$

$$\Omega^-_{\theta\varphi} = \partial_{\theta} A^-_{\varphi} - \partial_{\varphi} A^-_{\theta} = 2 \sin \frac{\theta}{2} \left( \cos \frac{\theta}{2} \right) \frac{1}{2} = \frac{1}{2} \sin \theta.$$

When we calculate the Berry phase over a closed path $C$, we can express it as

$$\gamma_{\pm}(C) = \iint_{S_C} dS \cdot \Omega_{\pm}(R) = \pm \frac{1}{2} \iint_{S_C} \sin \theta d\theta d\varphi.$$
When we express the Berry phase in terms of a surface integral, we have the freedom of choosing the surface. The loop $C$ is the boundary of both $S_1$ and $S_2$ shown below.

![Diagram showing the boundary of $S_1$ and $S_2$](image)

Using the result above, we know that the integral of the Berry phase over the whole sphere, which spans the solid angle of $4\pi$, should be $\mp 2\pi$. Then

$$\gamma^+(S_1) - \gamma^+(S_2) = -2\pi \quad \rightarrow \quad \gamma^+(S_1) = \gamma^+(S_2) - 2\pi,$$

and we encounter again the modulo $2\pi$ ambiguity of the Berry phase. The integral of the Berry curvature over the whole parameter space is

$$\gamma^\pm(S^2) = (\mp 1) 2\pi.$$

In later lectures we will identify the $\mp 1$ as the Chern number of the $\pm$ bands.

**Note on the interpretation of the Berry curvature.** It is tempting to try and imagine the Berry curvature as a sort of generalized Gaussian curvature, with a tangential plane at each point of our parameter surface, equipped with a local basis represented as two orthogonal vectors. The association with the parallel transport enhances this temptation even more.

The method which we apply in both cases is indeed similar - we take a parameter space $B$ (base space), for each point in $B$ we find another space $F$ and we assign one to another, constructing local products $B \times F^5$. For the Gaussian curvature $B$ was the sphere and $F$ was the local tangential plane. For the Berry curvature the parameter surface was again a sphere, but $F$ is instead a whole Hilbert space for the Hamiltonian defined by the values of its parameters at a particular point. We can certainly try to find a pictorial representation for the local Hamiltonian eigenstates on the sphere, but we must remember that this picture is only a metaphor for the real thing. And as usual with metaphors, we can push them too far. For instance, the parallel transport on the sphere over a closed path turns both vectors spanning the local tangential plane in the same direction. The parallel transport of the eigenstates $u_+$ and $u_-$ over the same loop in the parameter space “rotates” them in the opposite directions (Berry phases have opposite signs).

**B. Geometric phase in optics: Pancharatnam’s phase**

The geometric phase which Berry described appears so often throughout many domains of physics, that it had already been noticed several times. What Berry did was to provide a mathematical context (that of topology) in which all these remarkable occurrences can be seen as different manifestations of the same principle. At the time of writing his famous 1984 paper he knew of two such occurrences, the Aharonov-Bohm effect and the so-called dynamical Jahn-Teller effect in molecular physics. After the publication he learned that thirty years earlier, in 1956, an Indian physicist, S. Pancharatnam, had a very similar insight. His interest was sparked by the interference patterns which he saw when looking through a biaxial absorbing crystal, iolite. The patterns were strongest when he inserted a linear polarizer in front of and behind the crystal, which also makes the geometrical interpretation clearer. The experiment is sketched on the left below. The first polarizer sets the beam to a polarization $A$. The second polarizer sets it to the polarization $D$, but the beams recombining on our retinas were propagating through different environments in the anisotropic medium of iolite; say, the beam coming from the right went through a sequence of polarizations $A \rightarrow B \rightarrow D$, but the beam coming from the left through a sequence $A \rightarrow C \rightarrow D$. The resulting pattern must have looked similar to what you can see in the photo of the sky, taken by Berry through a sandwich of two linear polarizers with an overhead transparency in between. The acetate from which the transparency was made, is indeed an anisotropic medium.

---

5 In a more technical language, the space $F$ is called a **fibre**, and the base space, together with the set of all the fibres and the rules by which we assign them, is called a **fibre bundle**.
Pancharatnam used a Poincaré sphere to parametrize the path which each beam traced. For monochromatic light, travelling in the $z$ direction, the electric field vector lies in the $xy$ plane and the state of polarization is described by

$$d = (d_x, d_y), \quad |d| = 1, \quad d_x, d_y \in \mathbb{C}.$$  

We can combine both polarizations in a spinor $|\psi\rangle = (\psi_+, \psi_-)$, with $\psi_{\pm} = (d_x \pm id_y)/\sqrt{2}$. The polarization matrix is

$$H(r) = r \cdot \sigma = \begin{pmatrix} z & x - iy \\ x + iy & -z \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta e^{-i\phi} \\ \sin \theta e^{i\phi} & -\cos \theta \end{pmatrix}.$$  

Each point $r_A = (\theta_A, \varphi_A)$ defines a matrix $H$, whose eigenvector with + sign is the polarization $|A\rangle$. At the poles $\theta = 0$ and $\theta = \pi$ we have $\psi_- = 0$ and $\psi_+ = 0$, respectively. These two points denote the two circular polarizations. On the equator, $\theta = \pi/2$, $|\psi_+| = |\psi_-|$ and these points correspond to different directions of the linear polarization. Everywhere else the polarization is elliptical.

Pancharatnam found a way in which to determine the phase difference between two beams with different polarizations, or in other words he found the Pancharatnam connection. His reference point was the situation in which the intensity of the combined beam was maximum:

$$\langle (|A| + |B|)(|A\rangle + |B\rangle) \rangle = 2 + 2|\langle A |B \rangle| \cos(\text{ph}(A, B)).$$

With this definition of phase difference he was able to determine the relative phase accumulated by the two beams travelling on the paths $ABD$ and $ACD$ on the Poincaré sphere (above on the right). This phase difference turned out to be half of the solid angle which the loop $ABCDA$ encompassed when viewed from the centre of the sphere.

C. Berry phase in the Bloch bands

In solid state physics one of the most prominent dependencies of the electronic Hamiltonian is that on the crystal momentum. It can also be manipulated by electromagnetic fields, or by deforming the lattice. The Hamiltonian can
be written in several forms, and we’ll begin by the most basic one, 
\[ H = \frac{p^2}{2m} + V(r), \quad \text{where } V(r + a) = V(r), \]
with \( a \) the Bravais lattice vectors. The Bloch theorem states that the eigenstates of such a Hamiltonian fulfill the condition
\[ \psi_{nq}(r + a) = e^{iq \cdot a} \psi_{nq}(r). \]

We have now a \( q \)-independent Hamiltonian and \( q \)-dependent boundary conditions. Since differentiating the Hamiltonian is usually easier than differentiating the eigenstates, we can apply a gauge transformation \( H(q) = e^{-iq \cdot r} H e^{iq \cdot r} \) and obtain
\[ H(q) = \frac{(\hat{p} + \hbar q)^2}{2m} + V(r), \]
with a transformed eigenstate \( u_{nq}(r) = e^{-iq \cdot r} \psi_{nq}(r) \). The boundary condition which \( u_{nq} \) obeys is strict,
\[ u_{nq}(r + a) = u_{nq}(r), \]
and the \( u_{nq} \) functions have the same periodicity as the atomic lattice. The Brillouin zone is the parameter space of the transformed Hamiltonian \( H(q) \), with \( |u_{nq}\rangle \) as the basis functions. Its topology is that of a torus; since the states \( |\psi_n(q)\rangle \) and \( |\psi_n(q + G)\rangle \) satisfy the same boundary condition, they may differ by at most a phase. We may keep our torus unclosed by setting no condition on the phase (that can sometimes be convenient), or we can close the torus by fixing \( |\psi_n(q)\rangle = |\psi_n(q + G)\rangle \). Consequently, the relation holding for \( |u_n(q)\rangle \) is
\[ u_{nq}(r) = e^{iG \cdot r} u_{nq + G(r)}. \]

This choice of gauge is called the periodic gauge.

If the crystal momentum \( q \) is forced to vary, the Bloch state will pick up a Berry phase
\[ \gamma_n = \oint_C dq \cdot \langle u_n(q) | i \nabla_q | u_n(q) \rangle. \]

The Berry phase has a meaning only over a closed path, but Berry connection and curvature are defined at every point in the parameter space. In particular, the Berry curvature of the energy bands is
\[ \Omega_n(q) = \nabla_q \times \langle u_n(q) | i \nabla_q | u_n(q) \rangle. \]

The integral of the Berry curvature over the Brillouin zone is a so-called Zak’s phase,
\[ \gamma_n = \oint_{BZ} dq \cdot \langle u_n(q) | i \nabla_q | u_n(q) \rangle. \]

D. **Adiabatic current**

We want now to find out about the response of our system to an external perturbation. We’ll assume that the perturbation is adiabatic and cyclic in time, though the requirement of time periodicity can later be dropped.

Let us consider a general state \( |\psi(t)\rangle \), obeying the usual Schrödinger equation \( i\hbar \partial_t |\psi(t)\rangle = H(t) |\psi(t)\rangle \). Using the snapshot basis, we can express it in terms of Hamiltonian eigenstates as
\[ |\psi(t)\rangle = \sum_n \exp \left\{ -\frac{i \hbar}{\epsilon_n(t')} \right\} a_n(t) |n(t)\rangle. \]

Inserting it into the Schrödinger equation we obtain
\[ \sum_n \dot{a}_n(t) \exp \left\{ -\frac{i \hbar}{\epsilon_n(t')} \right\} |n(t)\rangle = - \sum_n a_n(t) \exp \left\{ -\frac{i \hbar}{\epsilon_n(t')} \right\} |\partial_t n(t)\rangle \]

Acting on both sides with \( |n'(t)| \) we arrive at
\[ \dot{a}_{n'}(t) = - \sum_n a_n(t) \exp \left\{ -\frac{i \hbar}{\epsilon_n(t')} (\epsilon_n(t') - \epsilon_{n'}(t')) \right\} |n'(t) |\partial_t n(t)\rangle. \]
We shall now choose a certain gauge, called the parallel transport gauge, which will make our calculations easier. The gauge corresponds to taking

$$|\tilde{n}(t)| \to |\tilde{n}(t)| = \exp \left\{ i \int_0^t \langle \tilde{n}(t) | \partial t \tilde{n}(t) \rangle \right\} |\tilde{n}(t)|,$$

or equivalently

$$\langle \tilde{n}(t) | \partial t \tilde{n}(t) \rangle = \langle \tilde{n}(\mathbf{R}) | \nabla_{\mathbf{R}} \tilde{n}(\mathbf{R}) \rangle \frac{\partial \mathbf{R}}{\partial t} \approx 0.$$

The last factor $\partial \mathbf{R}/\partial t$ may be considered as an adiabaticity parameter, and as such ought to be as small as possible. We shall use it for the expansion of our wave functions.

By adopting the parallel transport gauge we have sacrificed the idea that our wave function ought to be single-valued. The phase of $|\tilde{n}(t)|$ depends now on the way by which $t$ was reached, so the value of the wave function is path-dependent. After the cyclic evolution we shall recover again a Berry phase $\gamma_n(C)$, but it will now arise from the multi-valuedness of the wave function, not from the Berry connection. If our final expressions are gauge-invariant though, we can use them in any convenient basis.

If the system starts in the state $|\tilde{n}(0)\rangle$, we have $a_n = 1$ and $a_{n'} \neq n = 0$. Because of the parallel transport gauge which we adopted, $\dot{a}_n(0) = 0$, but $\dot{a}_n(0) \neq 0$. At $t = dt$, up to the first order in the adiabaticity parameter $\partial \mathbf{R}/\partial t$ we’ll have still $a_n(dt) = 1$, but $a_{n'}(dt)$ is

$$\dot{a}_{n'}(dt) = -\langle \tilde{n}'(dt) | \partial t \tilde{n}(dt) \rangle \exp \left\{ -i \int_0^t \kappa (\varepsilon_n(t') - \varepsilon_{n'}(t')) \right\}.$$

One ansatz for the $a_{n'}$ which would obey the differential equation for $\dot{a}_{n'}(t)$, is

$$a_{n'}(t) = -i \hbar \langle \tilde{n}' | \partial t \tilde{n} \rangle \exp \left\{ -i \hbar \int_0^t \kappa (\varepsilon_n(t') - \varepsilon_{n'}(t')) \right\}.$$

Differentiating it over time, we obtain

$$\dot{a}_{n'}(t) = \left\{ -i \hbar \frac{\partial \langle \tilde{n}' | \partial t \tilde{n} \rangle}{\varepsilon_n - \varepsilon_{n'}} + i \hbar \frac{\langle \tilde{n}' | \partial t \tilde{n} \rangle}{\varepsilon_n - \varepsilon_{n'}} \partial t (\varepsilon_n - \varepsilon_{n'}) - \langle \tilde{n}(t) | \partial t \tilde{n}(t) \rangle \right\} \exp \left\{ -i \hbar \int_0^t \kappa (\varepsilon_n(t') - \varepsilon_{n'}(t')) \right\}.$$

The first two terms are of the second order in adiabaticity on account of $\partial t (\tilde{n}' | \partial t \tilde{n})$ and $\partial t (\varepsilon_n - \varepsilon_{n'})$. The last term is indeed the right side of the equation for $\dot{a}_{n'}(t)$. The state which started out as $|n(0)\rangle$ becomes then $|\tilde{\psi}(t)\rangle$, given by

$$|\tilde{\psi}(t)\rangle = \exp \left\{ -i \hbar \int_0^t \kappa (\varepsilon_n(t')) \right\} |\tilde{n}(t)\rangle - i \hbar \sum_{n' \neq n} \langle \tilde{n}'(t) | \partial t \tilde{n}(t) \rangle \frac{\langle \tilde{n}'(t) | \partial t \tilde{n}(t) \rangle}{\varepsilon_n(t) - \varepsilon_{n'}(t)}.$$

Let us now return to our Bloch Hamiltonian. The system we will investigate is a 1D (for ease of calculation) band insulator, so that while integrating over our parameter surface we won’t encounter any degeneracies. The perturbation is time-dependent, slow (for adiabaticity) and periodic $(H(t + T) = H(t))$ so that the parameter surface can be closed. If the time-dependent Hamiltonian still has the translational symmetry of the crystal, its instantaneous eigenstates have the Bloch form

$$|\psi_n(q, t)\rangle = e^{iqx} |u_n(q, t)\rangle.$$

The state which started out as $|u_n(q, 0)\rangle$ will become after a time $t$

$$|u_n(q, t)\rangle = |\tilde{u}_n(q, t)\rangle - i \hbar \sum_{n' \neq n} |\tilde{u}_{n'}(q, t)\rangle \frac{\langle \tilde{u}_{n'}(q, t) | \partial t \tilde{u}_n(q, t) \rangle}{\varepsilon_n - \varepsilon_{n'}}.$$

The velocity operator associated with $|u_n(q)\rangle$ is

$$v_q = \frac{i}{\hbar} [H(q, t), r] = \frac{i}{2m} [(\vec{p} + \hbar \mathbf{q})^2, r] = \frac{i}{2m\hbar} (\hat{p} + \mathbf{q})[\hat{p} + \mathbf{q}, r] + [\hat{p} + \mathbf{q}, \hat{p} + \mathbf{q}, r] = \frac{1}{m} (\hat{p} + \mathbf{q}) = \frac{\partial H(q, t)}{\hbar \partial q}.$$

The average velocity in a state $|u_n(q)\rangle$ is, to the first order in adiabaticity,

$$v_n(q, t) = \langle u_n(q, t) | \frac{\partial H(q, t)}{\hbar \partial q} | u_n(q, t) \rangle = \langle \tilde{u}_n \rangle \frac{\partial H(q, t)}{\hbar \partial q} | u_n \rangle.$$

---

6 As in the Problem 2.3, where in the tangential gauge the boundary condition incorporated the Aharonov-Bohm flux directly, but in the radial gauge only through the singularity at $x = 0 = 2\pi$. 
\[-i \sum_{n' \neq n} \frac{1}{\varepsilon_n - \varepsilon_{n'}} \left\{ \langle \tilde{u}_n | \partial_q H | \tilde{u}_{n'} \rangle \langle \tilde{u}_{n'} | \partial_t \tilde{u}_n \rangle - \langle \tilde{u}_{n'} | \partial_q H | \tilde{u}_n \rangle \langle \tilde{u}_n | \partial_t \tilde{u}_{n'} \rangle \right\} \,.

Using the identity \( \sum_{n'} \langle \tilde{u}_{n'} | \tilde{u}_{n'} \rangle = 1 \), and

\[
\langle \tilde{u}_n | \partial_q H | \tilde{u}_{n'} \rangle = \langle \tilde{u}_n | \partial_q (H | \tilde{u}_{n'} \rangle) - \langle \tilde{u}_n | H \partial_q | \tilde{u}_{n'} \rangle = \langle \tilde{u}_n | \partial_q (\varepsilon_{n'} | \tilde{u}_{n'} \rangle) - \varepsilon_n \langle \tilde{u}_n | \partial_q \tilde{u}_{n'} \rangle
\]

we obtain

\[
\nu_n(q) = \frac{\partial \varepsilon_n}{\hbar \partial q} - i \sum_{n' \neq n} \left\{ \langle \tilde{u}_n | \partial_q \tilde{u}_{n'} \rangle \langle \tilde{u}_{n'} | \partial_t \tilde{u}_n \rangle - \langle \tilde{u}_{n'} | \partial_q \tilde{u}_n \rangle \langle \tilde{u}_n | \partial_t \tilde{u}_{n'} \rangle \right\}
\]

\[
\frac{\partial \varepsilon_n}{\hbar \partial q} - i \left\{ \langle \partial_q \tilde{u}_n | \partial_t \tilde{u}_n \rangle - \langle \partial_t \tilde{u}_n | \partial_q \tilde{u}_n \rangle - \langle \tilde{u}_n | \partial_q \tilde{u}_n \rangle \langle \tilde{u}_n | \partial_q \tilde{u}_n \rangle + \langle \tilde{u}_n | \partial_q \tilde{u}_n \rangle \langle \tilde{u}_n | \partial_q \tilde{u}_n \rangle \right\}
\]

\[
\frac{1 \partial \varepsilon_n}{\hbar \partial q} - \Omega_n^q.
\]

The first term on the right hand side is the usual one for the velocity; the second part is an additional, anomalous velocity, responsible for the off-diagonal response in the various types of quantum Hall effect, which we shall see later.