Adiabatic Invariance, the Geometric Phase,
and the Born-Oppenheimer Approximation

A time-dependent system is said to be *adiabatic* if the time-dependence is slow. First we will say a few words about adiabatic systems in classical mechanics. A simple example is a pendulum with a slowly varying length. Let L be the length of the pendulum, so that the frequency is given by

$$\omega = \sqrt{\frac{g}{L}}. \tag{1}$$

When we say that the length is slowly varying, mean that L does not vary much on during a single cycle of oscillation, or

$$\frac{\dot{L}}{L} \ll \omega. \tag{2}$$

We can write $L = L(\epsilon t)$ to indicate this slow time dependence, where we think of ϵ as a small parameter. Since ω is a function of L, we also have $\omega = \omega(\epsilon t)$. Then for small angles θ of oscillation, the equation of motion of the pendulum is

$$\ddot{\theta} + \omega(\epsilon t)^2 \theta = 0, \tag{3}$$

which is that of a harmonic oscillator with slowly varying frequency. It is interesting that a treatment of this equation when ϵ is small is mathematically the same as the treatment of the one-dimensional Schrödinger equation when $\hbar$ is small. That is, the method is WKB theory.

More generally, an adiabatic system in classical mechanics is (generally an oscillator) with slowly time-dependent parameters. Neither the frequency nor the energy of the oscillator is conserved in the adiabatic process, but there is an approximate conserved quantity,

$$I = \frac{1}{2\pi} \oint p \, dq, \tag{4}$$

which is the action of the oscillator. The action is not exactly conserved, but it is conserved to higher accuracy as the changes in the parameters are carried out more slowly. The action is what is called an *adiabatic invariant*. As we will see below, the situation is similar in quantum mechanics; the energy is not conserved, but there is an approximate conservation law, which is that of the quantum number (in a sense to be made precise below). This makes sense, because we know that according to the Bohr-Sommerfeld quantization rule, the quantum number is proportional to the classical action.

Let us consider a quantum Hamiltonian with a slow time-dependence, which we denote by

$$H = H(\epsilon t) = H(\tau), \tag{5}$$

where $\tau = \epsilon t$. The ϵ serves to remind us that the time-dependence is slow. It also has a more formal mathematical meaning, as follows. Let us suppose the Hamiltonian depends on certain parameters, which we can imagine are under experimental control. One can think of the electric and magnetic fields with which an atom or spinning particle interacts. Therefore by changing these parameters, we can change the Hamiltonian, say, from some initial value H_0 to some final value H_1. Suppose we change the Hamiltonian from H_0 to H_1, passing through some specified sequence of intermediate Hamiltonians, in some time interval $t = 0$ to $t = t_1 = T$. At the end of this time we can make various physical measurements on our system. Now let us repeat exactly the same procedure, except we carry out the changes in the Hamiltonian more slowly than before by some scale factor ϵ, so that the changes take place over the time interval $t = 0$ to $t = t_1 = T/\epsilon$. The initial and final Hamiltonians are still the same, as is the sequence of intermediate Hamiltonians, but the time taken to get to the final Hamiltonian is longer by a factor of $1/\epsilon$. Then we can study how various quantities of interest, such as the probability of finding the system in some state $|n\rangle$, scale with ϵ as $\epsilon \to 0$. If a quantity scales as some power of ϵ, say, ϵ^a, then we will say that that quantity is $O(\epsilon^a)$. For example, the elapsed time t_1 is $O(1/\epsilon)$. For brevity, we will refer to a time of order $1/\epsilon$ as "a long time."

We imagine that the final Hamiltonian H_1 differs from the initial Hamiltonian by some amount which is of order unity, in the usual sense of that phrase in physics. For example, if we are changing magnetic fields, perhaps the final magnetic field is twice as strong as the initial one. But the final Hamiltonian is also order unity in the limiting sense defined in the preceding paragraph, since H_1 is independent of ϵ and therefore scales as $\epsilon^0 = 1$.

We will be interested in solving the time-dependent Schrödinger equation,

$$H(\epsilon t)|\psi(t)\rangle = i\hbar \frac{\partial}{\partial t}|\psi(t)\rangle, \tag{6}$$

over a long time interval. Although the time-dependent Schrödinger equation is not separable in time, nevertheless there is some advantage in introducing the instantaneous eigenkets and eigenvalues of the Hamiltonian. That is, we write,

$$H(\tau)|n(\tau)\rangle = E_n(\tau)|n(\tau)\rangle, \tag{7}$$

where naturally the eigenvalues E_n and eigenkets $|n\rangle$ depend on τ since H itself does. The notation $|n(\tau)\rangle$ is somewhat illogical, since it is the ket and not the quantum number that is a function of τ, but we use it anyway since the more logical notation $|n\rangle(\tau)$ is awkward. Notice that since H changes by an amount which is order unity in the (long) time interval of interest, so also do the eigenvalues and eigenkets; for example, we may expect some final energy eigenfunction to look rather different from the initial eigenfunction. The label n is the same, of course. The eigenvalues E_n are also expected to change by an amount which is of order unity.

We will assume that the spectrum of H is discrete and nondegenerate for all τ of interest. The nondegeneracy assumption in particular seems a strong one, since even if H is nondegenerate at $t = 0$, we might expect the energy levels E_n, which after all are functions of τ and are expected to

change by an amount which is of order unity, to cross one another in the course of time. Such a crossing of course gives rise to a degeneracy. Nevertheless, we will pursue the consequences of the nondegeneracy assumption because it leads to the simplest analysis of the adiabatic process, and consider later what happens if levels become degenerate or nearly degenerate in the course of time.

Since the eigenkets $|n\rangle$ form a complete set, we can without loss of generality expand the exact solution $|\psi(t)\rangle$ in terms of them. We write,

$$|\psi(t)\rangle = \sum_n c_n(t)|n(\tau)\rangle. \tag{8}$$

Substituting this into the Schrödinger equation (6), we easily find the equation of evolution for the unknown coefficients c_n,

$$\dot{c}_n = -\frac{i}{\hbar}E_n(\tau)c_n(t) - \epsilon \sum_m \langle n|\frac{d}{d\tau}|m\rangle c_m. \tag{9}$$

This equation is exact. In the second term, the operator $d/d\tau$ is not an operator in the usual sense (it is not a mapping of the ket space onto itself), but rather the matrix element shown is interpreted as the scalar product of the bra $\langle n|$ with the ket $d|m\rangle/d\tau$.

To solve Eq. (9) at lowest order, we neglect the second term on the right hand side, which is $O(\epsilon)$. Then we have

$$c_n(t) = e^{i\phi_n(t)} c_n(0), \tag{10}$$

where

$$\phi_n(t) = -\frac{1}{\hbar} \int_0^t E_n(\epsilon t') \, dt'. \tag{11}$$

We see that in this approximation, the final coefficient $c_n(t)$ differs from the initial coefficient $c_n(0)$ by a pure phase factor, which means that the probability P_n of finding the system in energy eigenstate $|n\rangle$ is independent of time:

$$P_n = |c_n(t)|^2 = |\langle n(\tau)|\psi(t)\rangle|^2 = \text{const.} \tag{12}$$

As a special case of interest, suppose that at $t = 0$ the system was in energy eigenstate N, that is,

$$|\psi(0)\rangle = |N(0)\rangle. \tag{13}$$

Then at the final time we have,

$$|\psi(t)\rangle = e^{i\phi_N(t)} |N(\tau)\rangle, \tag{14}$$

which we can express by saying that the quantum number is conserved. We emphasize again that neither the energy E_N nor the eigenket $|N\rangle$ is conserved. If we think in terms of wave functions, we see that as the Hamiltonian changes, the wave function continuously distorts so as to remain proportional to an eigenfunction of H. The conclusions contained in Eqs. (10), (12) and (14) are referred to as the *adiabatic theorem*.

The phase $\phi_n(t)$, defined in Eq. (11), is called the "dynamical phase." Obviously if the Hamiltonian were truly time-independent, the dynamical phase would be $\phi_n = -E_n t/\hbar$. As it is, Eq. (11) expresses ϕ_n as the time integral of the instantaneous frequency $E_n/\hbar$, which is no surprise.

It is of interest to go to the next order of approximation in solving Eq. (9), for at least three reasons. First, the at the next order we will discover the conditions of validity of the adiabatic theorem. Second, the next order term leads us to the interesting subject of Berry's phase, and the associated gauge theory. Finally, it is easy to see that the next order correction is needed for various interference experiments, in which we study the interference of the final state $\psi(t)$ with some fixed state [perhaps a copy of the initial state $\psi(0)$]. To understand this interference even at a qualitative level, we must know the phase of the final state through terms that are of order unity, that is, with an error that is small compared to 2π. But Eq. (11) shows that the dynamical phase is $O(1/\epsilon)$, since it is the integral of a quantity that is order unity over a long time. Therefore we need the next term in the expansion of the phase, which is $O(1)$.

To go to the next order, we rewrite Eq. (10), not as the solution of an approximate equation, but as an exact change of variables. That is, we write

$$c_n(t) = e^{i\phi_n(t)} b_n(t), \tag{15}$$

which serves to define a new set of coefficients b_n. We substitute this into Eq. (11) and find,

$$\dot{b}_n(t) = -\epsilon \sum_m \langle n|\frac{d}{d\tau}|m\rangle e^{i(\phi_m - \phi_n)} b_m. \tag{16}$$

This equation is still exact. If we integrate both sides of this equation over a long time interval, then we might at first expect the integral of the right hand side to be of order unity, since the integrand is $O(\epsilon)$ and the time interval is $O(1/\epsilon)$. However, the off-diagonal terms on the right hand side (the terms $m \neq n$) involve the phase factor $\exp[i(\phi_m - \phi_n)]$, which oscillates many times in the long time interval in question. Therefore in the integration over the long time interval, these oscillations prevent the off-diagonal terms from accumulating. Instead, the integral of the off-diagonal terms gives an answer that is itself $O(\epsilon)$, since each oscillation nearly cancels the previous one. On the other hand, the diagonal terms $m = n$ do not have the oscillating phase factor, so the integral of them does accumulate over a long time to give a result that is of order unity.

This argument justifies the neglect of the off-diagonal terms in Eq. (16), which leaves us with

$$\dot{b}_n = -\epsilon \langle n|\frac{d}{d\tau}|n\rangle b_n. \tag{17}$$

We integrate this to find

$$b_n(t) = e^{i\gamma(t)} b_n(0), \tag{18}$$

where the new phase $\gamma(t)$ is defined by

$$\gamma(t) = i\epsilon \int_0^t \langle n|\frac{d}{d\tau'}|n\rangle \, dt' = i \int_0^\tau \langle n|\frac{d}{d\tau'}|n\rangle \, d\tau'. \tag{19}$$

The phase $\gamma(t)$ is called the "geometric phase" or "Berry's phase." Notice that it is of order unity. Notice also that γ is real, for if we take the normalization condition $\langle n|n \rangle = 1$ and differentiate with respect to time, we find

$$\langle n|\dot{n} \rangle + \langle \dot{n}|n \rangle = \langle n|\dot{n} \rangle + \langle n|\dot{n} \rangle^* = 0, \tag{20}$$

which shows that $\dot{\gamma} = i\langle n|\dot{n} \rangle$ is purely real.

Combining these results, we can express the solution to the original problem, including both the zeroth and first order approximations, in the form,

$$c_n(t) = e^{i[\phi(t)+\gamma(t)]} c_n(0), \tag{21}$$

which shows that even at this order, the probability of finding the system in a given eigenstate is independent of time. In fact, one can show that with the proper understandings, the probability of finding the system in a given eigenstate is independent of time to all orders in the expansion parameter ϵ. This does not mean that the probability is really constant, merely that the amount by which the probability changes goes to zero faster than any power of ϵ as $\epsilon \to 0$. [This conclusion is true if the Hamiltonian is an infinitely differentiable function of time, perhaps an analytic function. If H has only a finite number of time derivatives, then the change in the probability goes to zero as some power of ϵ.]

Now we can see what happens if degeneracies or near-degeneracies develop in the course of time. An exact degeneracy, say, $E_n = E_m$ for some $n \neq m$, would mean that there was an off-diagonal term in Eq. (16) that did not have the rapidly varying phase factor, at least over some time interval. This would mean that we could not neglect all the off-diagonal terms, and that the integration of Eq. (16) over time would introduce a coupling between states n and m. That is, the term in question would induce transitions between the two states. In practice, the same is true for near degeneracies, since in practice we always work with some finite time interval, and we never really take the limit $\epsilon \to 0$.

We can easily understand these transitions from the standpoint of time-dependent perturbation theory. For let us suppose that the time-dependence of the Hamiltonian involves a characteristic range of frequencies, out to some value ω_H. We write this schematically as

$$\omega_H \sim \frac{\dot{H}}{H}. \tag{22}$$

Note that ω_H is $O(\epsilon)$ by our assumptions; the time-dependence of the Hamiltonian is slow. On the other hand, the resonance frequency between states n and m is $(E_n - E_m)/\hbar$, and if this comparable to ω_H, then we expect the time-dependence of H to induce transitions. Thus, we can say that the condition for the validity of the adiabatic theorem is

$$\frac{\dot{H}}{H} \ll \frac{|E_n - E_m|}{\hbar}. \tag{23}$$

This should be true of all states n and m of interest; for example, if we choose the initial conditions (13), we can set $n = N$ and $m = N \pm 1$, since it is the nearest neighbors to eigenvalue E_N that are in greatest danger of violating the condition (23).

We turn now to the gauge structure that is inherent in adiabatic theory. To begin, we shift emphasis slightly, and imagine that the time-dependence of H takes place through a set of parameters $\mathbf{R} = (R_1, R_2, \ldots)$ that themselves are slow functions of time, so that $H = H(\mathbf{R})$ and $\mathbf{R} = \mathbf{R}(\tau)$. A useful example to keep in mind is that of a spin in a slowly varying magnetic field, for which

$$H = -\boldsymbol{\mu} \cdot \mathbf{B}(\tau). \tag{24}$$

In this example, we can identify the parameters $\mathbf{R}$ with the magnetic field $\mathbf{B}$, so that parameter space is the same as "magnetic field space," and is 3-dimensional. By the condition (22), the time dependence of this Hamiltonian is slow enough for the adiabatic theorem to hold if it is slow in comparison to the spin precession frequency, $\mu B/\hbar$.

More generally, we imagine some parameter space of some dimensionality with coordinates $\mathbf{R}$, and we imagine finding the energy eigenvalues and eigenkets as functions of $\mathbf{R}$, according to

$$H(\mathbf{R})|n(\mathbf{R})\rangle = E_n(\mathbf{R})|n(\mathbf{R})\rangle. \tag{25}$$

We think of the energy eigenvalues and eigenkets as fields over parameter space (scalar and ket fields, respectively). We assume that the eigenvalues are discrete and nondegenerate over some region of parameter space of interest; this implies that the eigenvalues are smooth functions of $\mathbf{R}$ (assuming the dependence of H on $\mathbf{R}$ is smooth). Later we worry about what happens at degeneracies. Likewise, we assume the eigenkets $|n(\mathbf{R})\rangle$ can be defined in a smooth and continuous manner in the region of parameter space of interest. This will be possible if the region in question is contractible and contains no degeneracies, as we assume. (But Born-Oppenheimer theory gives rise to some interesting applications where these assumptions are not met.) Of course, the eigenkets are only defined to within a phase factor by Eq. (25); when we write $|n(\mathbf{R})\rangle$, we assume some phase conventions have been chosen for each value of $\mathbf{R}$, in a smooth manner.

Now the history of the Hamiltonian under the adiabatic process can be described by a curve $\mathbf{R} = \mathbf{R}(\tau)$ in parameter space, which starts at some point $\mathbf{R}_0$ and ends and some point $\mathbf{R}_1$. The curve is fixed in parameter space, but is traversed ever more slowly as $\epsilon \to 0$. Quantities such as the eigenkets change along the curve according to

$$\frac{d}{d\tau}|n(\mathbf{R})\rangle = \nabla|n(\mathbf{R})\rangle \cdot \frac{d\mathbf{R}}{d\tau}, \tag{26}$$

where $\nabla = \partial/\partial\mathbf{R}$. Then we can write the geometric phase [see Eq. (19)] in terms of the line integral of a certain vector field in parameter space,

$$\gamma = i\int_0^\tau \langle n|\nabla|n\rangle \cdot \frac{d\mathbf{R}}{d\tau'} \, d\tau' = \int_{\mathbf{R}_0}^{\mathbf{R}_1} \mathbf{A}(\mathbf{R}) \cdot d\mathbf{R}, \tag{27}$$

where

$$\mathbf{A}(\mathbf{R}) = i\langle n(\mathbf{R})|\nabla|n(\mathbf{R})\rangle. \tag{28}$$

In this equation, the vector field $\mathbf{A}(\mathbf{R})$ is associated with the state n in question, and sometimes we might prefer to write $\mathbf{A}_n(\mathbf{R})$ to emphasize this. The line integral in Eq. (27) is taken along

the given curve in parameter space (the history of the Hamiltonian). The vector field $\mathbf{A}(\mathbf{R})$ is the vector potential or *gauge potential* for the problem, and is analogous to the vector potential in ordinary electromagnetic theory, as suggested by the notation. We see that all dependence on the time parameterization has dropped out in the final expression in Eq. (27) for the geometric phase, and that this phase is independent of the rate at which the curve in parameter space is traversed, as long as it is slow enough for the adiabatic theorem to be valid. It is for this reason that this phase is called "geometric."

The gauge transformations in this theory are changes in the phase conventions for the eigenkets $|n(\mathbf{R})\rangle$. Such a change in phase conventions will be indicated by

$$|n(\mathbf{R})\rangle \to |\tilde{n}(\mathbf{R})\rangle = e^{ig(\mathbf{R})}|n(\mathbf{R})\rangle, \tag{29}$$

here the tilde indicates the new phase conventions. Of course there is no physics in these phase conventions, so all physical results must be gauge-invariant. This is just as in electromagnetic theory. The gauge scalar g in Eq. (29) is allowed to be a function of $\mathbf{R}$, because there is no reason why we cannot change the phase in a different manner for different parameter values. The $\mathbf{R}$-dependence of the gauge scalar is what qualifies these gauge transformations as "local." In gauge theories, only local gauge transformations are at all nontrivial or interesting. Of course, for a given eigenket $|n(\mathbf{R})\rangle$ (for a given value of $\mathbf{R}$ and a given phase convention), there is no meaning to talking about the "phase" of the eigenket in any absolute sense. For example, if the ket is associated with a wave function $\psi_n(\mathbf{r}) = \langle\mathbf{r}|n\rangle$, then ψ_n is in general a complex function of position $\mathbf{r}$, with different phases at different points. On the other hand, the relative phase of two kets that are proportional to one another is well defined.

Since all physical results must be gauge-invariant, we should examine all the quantities of our theory to see how they transform under gauge transformations. For example, the eigenvalues E_n and the dynamical phases ϕ_n are clearly gauge-invariant, whereas the eigenkets $|n\rangle$ are not [they transform according to Eq. (29)]. Nor is the gauge potential gauge-invariant, for by substituting Eq. (29) into Eq. (28), we find

$$\mathbf{A}(\mathbf{R}) \to \tilde{\mathbf{A}}(\mathbf{R}) = \mathbf{A}(\mathbf{R}) - \nabla g(\mathbf{R}), \tag{30}$$

which is just as in electromagnetic theory. From this we find that the geometric phase is not gauge-invariant, either, but rather transforms according to

$$\gamma(t) \to \tilde{\gamma}(t) = \gamma(t) - g\big(\mathbf{R}(\tau)\big) + g\big(\mathbf{R}(0)\big), \tag{31}$$

as follows from Eq. (27). But the solution $|\psi(t)\rangle$ to the Schrödinger equation for given initial conditions $|\psi(0)\rangle$ should be gauge-invariant simply because the solution to the Schrödinger equation is unique. Indeed, if we combine Eqs. (13), (21), (29) and (31), we see that the gauge-dependence of $\gamma(t)$ is required to cancel that of $|n(\mathbf{R})\rangle$ and to make $|\psi(t)\rangle$ gauge-invariant. This gives us another reason for carrying out the ϵ expansion through the geometric phase, since without it the results do not transform properly under gauge transformations.

A special case of interest is one in which the system is carried through a cycle, so that the Hamiltonian returns to its original self after time t. In this case the curve in parameter space $\mathbf{R}(\tau)$ is closed, and we can apply Stokes' theorem to the integral (27) for the geometric phase. This gives

$$\gamma = \oint \mathbf{A}(\mathbf{R}) \cdot d\mathbf{R} = \int_S \mathbf{V} \cdot d\mathbf{S}, \tag{32}$$

where the surface S is a surface in parameter space bounded by the closed curve, and where

$$\mathbf{V}(\mathbf{R}) = \nabla \times \mathbf{A}(\mathbf{R}). \tag{33}$$

Here we are assuming that parameter space is 3-dimensional, so we can use the standard notation of vector calculus for Stokes' theorem. In parameter spaces of other dimensionalities, it is necessary to use differential forms, but all the important ideas are conveyed by the 3-dimensional case. Obviously, $\mathbf{V}$ is analogous to the magnetic field in electromagnetic theory, and satisfies the field equation,

$$\nabla \cdot \mathbf{V} = 0. \tag{34}$$

The vector field $\mathbf{V}$ is sometimes called the *curvature form* of the gauge theory.

Notice that for closed curves in parameter space, the gauge-dependence of γ seen in Eq. (31) cancels out, and the geometric phase becomes gauge-invariant. The same conclusion follows from the gauge-invariance of the vector field $\mathbf{V}$,

$$\tilde{\mathbf{V}}(\mathbf{R}) = \mathbf{V}(\mathbf{R}), \tag{35}$$

which appears in the surface integral in Eq. (32). In physical terms, the geometric phase must be gauge-invariant for a closed cycle, because if the final state is proportional to the initial state, as it must be according to the adiabatic theorem, then the relative phases of the two states can be determined by interference experiments.

To return to the case of open curves in parameter space, we might be tempted to say that this geometrical phase is just a nuisance that is due to a poor choice of phase conventions. After all, γ is gauge-dependent, so perhaps it is completely nonphysical. Suppose, for example, we return to Eq. (21), which was derived under the assumption that some phase conventions had been chosen for $|n(\tau)\rangle$ as a function of time along the history of the Hamiltonian, and use it to write the solution (8) in the form,

$$|\psi(t)\rangle = \sum_n e^{i[\phi_n(t) + \gamma_n(\tau)]} c_n(0) |n(\tau)\rangle. \tag{36}$$

Now we say, the kets $|n(\tau)\rangle$ involved a poor choice of phase, because they gave us a nonvanishing γ; suppose we choose a different phase convention,

$$|\tilde{n}(\tau)\rangle = e^{i\gamma_n(\tau)} |n(\tau)\rangle. \tag{37}$$

Then the geometrical phase disappears from the solution,

$$|\psi(t)\rangle = \sum_n e^{i\phi_n(t)} |\tilde{n}(\tau)\rangle, \tag{38}$$

and it seems that we do not have to worry about gauge transformations or gauge dependence. And yet, if the history of the Hamiltonian should take us back to the initial Hamiltonian, so that the curve in parameter space becomes closed, then the phase convention (37) no longer works, because we have one Hamiltonian $H_0 = H_1$ with two different phase conventions. In other words, the phase convention (37) introduces a discontinuity in the definition of the eigenkets when we go around a closed loop in parameter space. This discontinuity is nothing but the geometrical phase associated with the loop; this phase is gauge-invariant and cannot be banished by some clever choice of phase convention.

Therefore it is better to say that the phases of the energy eigenstates, while they do contain a nonphysical element that can be established only by convention, also contain a physical element that is manifest on carrying an adiabatic process around a closed cycle. This is just as in electromagnetic theory, where the integral $\oint \mathbf{A} \cdot d\mathbf{r}$ is the gauge-invariant magnetic flux associated with a closed loop.

At this point you might want to think through a specific problem involving the geometric phase, for example a spin in a slowly varying magnetic field as in Eq. (24).

We now turn to some comments about the Born-Oppenheimer approximation, which is the basic approximation scheme in the theory of molecules, and which relies crucially on notions of adiabaticity. A precise and careful treatment of Born-Oppenheimer theory is a fascinating subject that must lie outside the scope of this course; the following discussion merely aims to present some of the simplest ideas.

Suppose we have a molecule consisting of some number of electrons and some number of nuclei. We will label the electrons with indices i, j, etc., and the nuclei with indices α, β, etc. Let the positions, momenta, and mass of the electrons be $\mathbf{r}_i$, $\mathbf{p}_i$, and m, and let the positions, momenta, masses, and charges of the nuclei be $\mathbf{R}_\alpha$, $\mathbf{P}_\alpha$, M_α, and Z_α. The notation attempts to use capital letters for nuclei, and lower case letters for electrons. Then the basic Coulomb Hamiltonian for the molecule is

$$
\begin{aligned}
H_{\text{molec}} = {} & T_n + T_e + V_{ee} + V_{en} + V_{nn} \\
= {} & \sum_\alpha \frac{\mathbf{P}_\alpha^2}{2M_\alpha} + \sum_i \frac{\mathbf{p}_i^2}{2m} + \sum_{i<j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i\alpha} \frac{Z_\alpha e^2}{|\mathbf{r}_i - \mathbf{R}_\alpha|} \\
& + \sum_{\alpha<\beta} \frac{Z_\alpha Z_\beta e^2}{|\mathbf{R}_\alpha - \mathbf{R}_\beta|},
\end{aligned}
\tag{39}
$$

where T_n and T_e are the nuclear and electronic kinetic energies, and where V_{ee}, V_{en}, and V_{nn} are the potential energies of interactions of electrons with electrons, electrons with nuclei, and nuclei with nuclei, respectively. This Hamiltonian is very complicated, and it must be handled by some approximation scheme.

The Born-Oppenheimer approximation relies on the large disparity in the masses of the electrons and the nuclei, which have a ratio of the order of 10^{-4} for most nuclei. This means that if the kinetic energies of the electrons and nuclei are not too dissimilar, then the velocities of the electrons will

be much higher than those of the nuclei, and we can picture the electron motion as being rapid in comparison to the nuclear motion. Thus, while the heavy nuclei move a short distance, the electrons will whiz around the nuclei or back and forth between them many times, effectively filling out an electron cloud. (This picture is a metaphor of constructed of classical and quantum language, but it does convey many ideas correctly.)

In a rigorous and precise treatment of the Born-Oppenheimer approximation, one expands the molecular Hamiltonian in powers of the small parameter $\kappa \sim (m/M)^{1/4}$, which is of the order of 10^{-1} for most molecules. This expansion is nontrivial because the scale lengths of the wave functions themselves with respect to the nuclear and electronic coordinates are different, and in particular, involve different powers of κ. Thus, the κ ordering is not merely in the Hamiltonian, but also in the spatial dependence of the wave functions.

As a result of this expansion, one finds that the spectrum of the molecular Hamiltonian possesses three distinct energy scales. The largest is that of electronic excitations, involving energies E_{elec} which are of the order of e^2/a_0, where a_0 is the Bohr radius. The next smaller energy scale is that of the vibrational motions of the molecule, in which the separation between energy levels is of the order of $E_{\text{vib}} \sim \kappa^2 E_{\text{elec}}$. Finally, the smallest energy scale is that of the rotations of the molecule, in which the separation between energy levels (for small angular momenta) is of the order of $E_{\text{rot}} \sim \kappa^2 E_{\text{vib}} \sim \kappa^4 E_{\text{elec}}$. This ordering of energy scales is easy to understand from a qualitative standpoint, as was discussed in class last semester.

In the present discussion we will take a simpler approach, but one that captures many of the ideas of the more sophisticated theory. Let us argue that since the nuclei are heavy and slow moving, we can treat them by classical mechanics, while the electrons, which are light, fast, and not easily localized due to the constraints of the uncertainty principle, must be treated by quantum mechanics. Thus, we will treat $\mathbf{R}_\alpha$ and $\mathbf{P}_\alpha$ in Eq. (39) as c-numbers, which presumably obey some classical equations of motion that endow them with some time dependence, and we will treat $\mathbf{r}_i$ and $\mathbf{p}_i$ as operators acting on wave functions for the electrons. This is a hybrid classical-quantum approach.

We now define the electronic Hamiltonian H_{elec} as the last four terms in Eq. (39), which depends on the operators $\mathbf{r}_i$ and $\mathbf{p}_i$ and on the nuclear positions $\mathbf{R}_\alpha$, which are c-numbers. We adopt an abbreviated notation, in which $\mathbf{r}$ stands for all electron coordinates $\mathbf{r}_i$, $\mathbf{R}$ stands for all nuclear coordinates $\mathbf{R}_\alpha$, etc.:

$$H_{\text{elec}}(\mathbf{r}, \mathbf{p}; \mathbf{R}) = \sum \frac{\mathbf{p}^2}{2m} + V_{ee}(\mathbf{r}) + V_{en}(\mathbf{r}, \mathbf{R}) + V_{nn}(\mathbf{R}). \tag{40}$$

Since $\mathbf{R} = \mathbf{R}(t)$ is supposedly a slow function of time, H_{elec} is an example of a Hamiltonian with a slow time dependence, in which the nuclear coordinates $\mathbf{R}$ are the slow parameters. The parameter space is now the nuclear configuration space. Notice that the internuclear potential energy $V_{nn}(\mathbf{R})$ is a c-number, and is just a constant insofar as the electronic operators are concerned; but we include it anyway in the electronic Hamiltonian.

The adiabatic theorem tells us to look at the eigenfunctions of H_{elec} for fixed values of the

parameters $\mathbf{R}$. Physically, this means that we fix or freeze the locations of the nuclei, and then solve for the electronic eigenfunctions and eigenvalues at the given nuclear locations. This problem is similar to atomic structure calculations, except that that the electrons are attracted to two or more fixed centers, instead of just one.

Let the eigenfunctions and eigenvalues of H_{elec} for fixed values of $\mathbf{R}$ be $\phi_n(\mathbf{r}; \mathbf{R})$ and $E_n(\mathbf{R})$, where the dependence on the parameters $\mathbf{R}$ is just as in Eq. (25), so that

$$H_{\text{elec}}(\mathbf{r}, \mathbf{p}; \mathbf{R})\phi_n(\mathbf{r}; \mathbf{R}) = E_n(\mathbf{R})\phi_n(\mathbf{r}, \mathbf{R}). \tag{41}$$

As indicated by the $\mathbf{R}$-dependence, these eigenfunctions and eigenvalues change as the nuclei move around. However, in accordance with the adiabatic theorem, we expect that if the system is prepared in a given electronic eigenstate at some initial time, it will remain in the eigenstate with the same quantum number as $\mathbf{R}$ slowly changes, even if the corresponding eigenfunction and energy should change by a substantial amount. For example, at large distances where two atoms do not interact, the electronic wave functions will be those of two independent atoms; but when the atoms approach one another, the electron clouds will distort and overlap and electrons will begin to be exchanged between the two atoms. But as long as the electronic energy levels do not come too close to one another, we expect the probability of being in a given electronic eigenstate to be independent of time.

[In fact, in important cases, the electronic eigenvalues do come close to one another, and one must consider the transitions due to the breakdown of adiabaticity. Such transitions are called *Landau-Zener* transitions, and are responsible for changes in electronic states in slow atomic or molecular collisions.]

Since the electronic energies change as $\mathbf{R}$ changes, the energy difference must come from somewhere, and it is logical that it comes from the nuclear kinetic energy. In other words, the electronic eigenenergy $E_n(\mathbf{R})$ plays the role of a potential energy for the nuclear motion. Thus, we can take the classical Hamiltonian for the nuclear motion to be

$$H_{\text{cl}}(\mathbf{R}, \mathbf{P}) = \sum \frac{\mathbf{P}^2}{2M} + E_n(\mathbf{R}) = \langle \phi_n | H_{\text{molec}} | \phi_n \rangle, \tag{42}$$

which is otherwise just the expectation value of the total molecular Hamiltonian H_{molec} with respect to the n-th electronic energy eigenstate, as indicated. We see that there is a different nuclear classical Hamiltonian for each electronic energy level; as one says, there is a different "potential energy surface" $E_n(\mathbf{R})$ for each electronic state.

In a more careful treatment of the Born-Oppenheimer approximation, in which the nuclear degrees of freedom are treated by quantum mechanics, one finds exactly the Hamiltonian (42) at lowest order in the perturbation parameter κ, except that it is now a quantum Hamiltonian. At the next order, one finds that the gauge potential $\mathbf{A}(\mathbf{R})$ associated with adiabatic transport of the electronic eigenfunctions is incorporated into the nuclear Hamiltonian, giving pseudo-magnetic effects. These include phase shifts similar to those in the Aharonov-Bohm effect that have observable effects on scattering cross sections and molecular energy levels.

The notion that the electronic eigenvalue is the potential energy for the nuclear motion leads to an immediate understanding of the nature of the chemical bond. The example of the H_2 molecule will be discussed in class.