Chapter 19

Introduction to Quantum Theory

19.1 Introduction

Toward the later part of the 19th century, several new observations caused people to question the basic ideas that were the cornerstone of the physics of the time. Primary among these was the growing acceptance of an atomic theory of matter. Successful predictions in chemistry and the development of a statistical particle basis for thermal phenomena being key. To us today, the atomic basis of matter is so obvious that we do not question it. On the other hand to the physicist of the early 19th century, the continuous nature of matter was obvious. Given the technology of the day, any attempt to measure the size of the atom was impossible. The scale of phenomena at which the discreetness of the atoms could be observed was inconceivably small, see Sec 1.4.2 on “Things That Everyone Should Know.” Even Dalton, the father of Chemistry, had his doubts about the atomic nature of matter. Although his model of atoms described with great success the rules of chemical composition, he could not understand chemical structures like gaseous O_2. If two oxygen atoms were attracted and bound to one another to form a stable O_2, why wouldn't two O_2's be even more attracted to one another and form an O_4? Continuing this line of reasoning, he would believe that oxygen should be a solid and not a gas. Regardless of these conceptual difficulties, by the later part of the 19th century, because of its success in chemistry and statistical mechanics, the atomic theory became dominant and, with it, the idea that the atom had definite properties and a definite size. At this same time, the discover of the electron provided an opportunity for a model of atoms based on bound electrons. The development of a model of atoms became a dominant effort of the period. We now know that all efforts based on a classical model were destined to failure. It would take the development of quantum mechanics to solve this problem and it would be some time after the first successes of quantum mechanics before a satisfactory model of the atom was possible. Today, the success of quantum mechanics in "explaining" chemistry is phenomenal. We can compute complex reactions and propose molecular configurations before they are seen in the laboratory.

Although it was the successful application of quantum mechanics to atomic theory in the early parts of 20th century that lead to its acceptance, the conceptual development of quantum mechanics begins much earlier and deals with a much simpler system – light. We will start with this pre-atom development not only because it is the historically correct approach to the study of quantum mechanics but also because, in its simplicity, it makes the conceptual basis of the theory most clear. For this we will need some understanding of thermal processes. The fact that the application of quantum mechanics to atoms was so successful makes the description of the quantum mechanics of contained massive objects such as electrons.

19.2 Blackbody Radiation

19.2.1 Background on Thermal Processes

Before discussing the phenomena associated with what we call black body radiation, we will have to understand some of the basic issues involved in thermal phenomena. The official name for the field is thermodynamics, the study of processes (dynamics) developed by temperature (thermo) differences. The classic example is the aeolipile of Hero of Alexandria which converted heat energy to rotational energy in the first century, see Figure 19.1. Thus ideas about the use of heat to do useful work began in the earliest recorded times. As is often the case, the understanding of the phenomena required a more careful articulation of the processes and more refined development of the vocabulary of the processes. This also almost always involves a qwerty use of words with implied meanings. The earlier sentence on the aeolipile is typical since we have been rather cavalier about the words heat and energy. Hopefully, we will be able to clarify these usages.

At some time in prehistory, humans became conscious of the sense of hot and cold. With the advent of the scientific era, there was a systematic effort to measure the primary physiological response to what today we call

Figure 19.1: **The Aeolipile of Hero** A small fire boils water and the pressure builds up to ultimately force steam out of angled jets. The flow of steam from the jets rotate the sphere.

temperature difference. Although originally crude, by the 17th century, the development of precise instruments for the measurement of temperature became reasonably widely accepted. It is worthwhile to point out that this idea of temperature is not to be identified with the physiologically phenomena of hot and cold. The deviation of hot and cold from the measured entity called temperature is observed in the statement that the outside temperature is 95^0 F but feels like 100^0; a strong indication that something subtle was going on. This scientific epoch in which the measurement of temperature was established is best called the period of thermometry.

It was because of these careful measurements in thermometry and the development of an instrument called the calorimeter that it became clear that beside temperature there was an important concept called quantity of heat. It with the invention of the calorimeter that the effects of heat flow could be carefully measured. The independence of temperature and this quantity of heat idea was most obvious in cases in which the material in question went through a phase change, i. e. water to steam or ice to water, without a temperature change. The idea being that the same matter, say water stuff, had different heat quantities that was intrinsic to their manifest nature. Ice had less heat in it than liquid water that had less heat content than steam. At first this quantity of heat was thought to be something special created to change the system either by a change in temperature or a change in state, liquid or solid. The next major development in our understanding of the thermal behavior was brought about by the experiments of J. P. Joule, the namesake for our SI unit of energy, which lead to the realization that heat, the unit of which was the calorie, defined to be the amount of heat that is needed to change the temperature of liquid water at 0^0 C by one degree, was, in fact, just another form of process energy such as mechanical or motional energy. In other words, if carefully done, the aeolipile's rotational energy would increase in direct proportion to the rate of heat flow from the flame once the conversion of units from calories per second to Joules per second. Of course, this is in an idealized world. In the realistic situation, there is heat flow to the surrounding environment and to mechanical friction. If the friction forces at the rotational axis of the aeolipile are a constant, the work done against friction will increase as the flame continues to provide heat energy. In fact, the aeolipile's rotational energy does increase when first exposed to the flame. This increase in the rotation rate continues until the energy loss to friction and the surrounding environment consume all the heat input from the flame and, thus, the aeolipile's equilibrium state is a constant rotational energy established at this balance point of dissipation. There are two important ideas here. The first is that heat has meaning only during process. It is not in the state of the system. Note that this is contrary to the ideas expressed in the previous paragraph in which heat content or latent heat was identified. The other important point is the idea that heat flow was the same as an energy flow as long as the correct units were used. This is the basis for the entry in 'Things Everyone Should Know' titled the 'energy content of food'. These two ideas are combined and are expressed as the first law of Thermodynamics: The energy processes manifests themselves in many forms, mechanical, thermal or heat, material flow, etc. that can morph into each other but in all these processes once suitable units are identified the total energy is always conserved. For our purpose of establishing the origins of quantum physics, it would be sufficient to understand thermodynamics at the level of the first law but thermal processes are so ubiquitous and important and they are so subtle that it is worth a full articulation of all of the laws that we will continue to develop the history and introduce the other three laws of thermodynamics.

The development of the steam engine in the late 18th and early 19th centuries provided a strong stimulus to articulating a complete theory of heat and heat processes and their relationship to mechanical processes. The development of a this theoretical framework started its full articulation in this later period when Sadi Carnot, a French military engineer, published his book, *Reflections on the Motive Power of Fire and on Machines Fitted to Develop that Power* in 1824. Unfortunately at the time, Carnot's ideas were generally ignored until picked up later in the 19^{th} century by R. J. E. Clausius and Lord Kelvin who articulated what is now called the second law of thermodynamics. The primary goal of Carnot was an understanding of the efficiency of heat engines such as the steam engine. To do so, he developed a theory of heat flow as energy falling from a higher to a lower temperature. Since he was working on the engines, he also showed the importance of cyclic, return to their original state, machines such as steam engines. The final piece of his contribution was showing the universal importance for thermodynamics of a specific cyclic process now called the Carnot engine or cycle. This observation of Carnot provide the basis for the 2^{nd} Law of Thermodynamics, The statements of Clausius and Kelvin as modified by Planck of the second law is

> **Kelvin-Planck:** No process is possible whose sole result is the extraction of energy from a heat bath, and the conversion of all that energy into work.
>
> **Clausius:** No process is possible whose sole result is the transfer of heat from a cooler to a hotter body.

In more careful language than that used by Clausius at that time, we would use for the word cooler the phrase lower temperature and for hotter the phrase higher temperature. It is important to note that the Kelvin-Planck implies an intrinsic inefficiency in processes that are driven by heat. We will see in the 1^{st} Law that energy is conserved but regardless all the heat energy is not available to do useful work. This intrinsic inefficiency is often quite large and as implied above cannot be reduced or eliminated.

Modern treatments of thermodynamics are generally based on the realization that matter is composed of small entities called atoms. This is historically backwards. Thermodynamics was fully developed and later Maxwell and others showed that the atomic hypothesis could be used to "explain" the thermal properties of gases; the atoms act in simple mechanical ways as articulated by Newton but macroscopic collections of these microscopic constituents have emergent behaviors such as temperature that emerge from the complications of underlying system [Atkins 2007]. This realization is correct but misses the point that the laws of thermodynamics were established prior to the acceptance of the atomic theory of matter and these laws constitute a complete, coherent and correct description of all processes using or generating heat flow which constitutes almost all evolutions of macroscopic systems. **The behavior of these macroscopic systems is simply summarized in four laws and these laws like all laws in physics are found to hold in all cases.**

19.2.2 Basic Thermodynamics

Thermodynamics emerged as a formal system from early studies of the use of heat energy to produce useful work, basically the effort to catalogue the operations of the steam engine. What emerged was a beautiful complete set of theoretical constructions that could then be used to describe operations of huge classes of phenomena including biology and chemistry. In fact, the ubiquitous nature of thermal processes is one of the great steps forward in our understanding of natural phenomena. For that reason and because of the subtlety of the subject, we will spend more time on this than is needed for our establishment of basis for modern physics. This is just something that all people should know. Also the formal constructs of the theory developed after some of the vocabulary of the phenomena had become a part of the vernacular and, as often happens, the words of thermodynamics have a vernacular connotation that can at times be misleading. Also as stated above, thermodynamics was developed classically and is a consistent construction that matured before the establishment of the atomic description of matter. The atomic theory of matter provides a valuable example of a basis for matter that provides a picture of the "why" for thermal phenomena but is actually only an example of a possible system that behaves thermally. This leads to an interesting anomaly in the use of thermal descriptions that the underlying theoretical constructions are "explained" by the picture of how the atoms behave. This is misleading since the concepts of thermal systems are actually very general and stand on their own. This leads to the interesting speculation that the success of classical thermodynamics was among the first examples of verification of the atomic nature of matter before this theory was articulated. The interesting point though is that the constructions of thermodynamics are general enough that the radical transformation of our picture matter after the discovery of the atom and the subsequent development of quantum mechanics does not weaken the edifice that thermodynamics gave us about heat transfer processes. In the following, we will deal with the ideas of thermodynamics classically and without the crutch of the idea of the disorder of systems of "atoms".

19.2.3 -1^{st} Law: Equilibrium & Variables

The usual exposition of Thermodynamics is the articulation of the four basic laws, 0 through 3. but it is very important to understand the time scale and meaning of processes in Thermodynamics. Although the greek word thermodynamics is translated as heat processes, thermodynamic processes take place on an long time scale compared to the usual processes of mechanical and electrical phenomena. This is due to the fact that thermodynamic descriptors such as volume, temperature, pressure... have to be articulated consistently for system or subsystems under consideration. For example, a macroscopic volume of gas confined to a space with a piston is said to be at temperature T or pressure P. The piston is moved to change the volume. if this change is done too rapidly there will not be enough time for the system, the volume of gas, to have the same temperature or pressure throughout. But thermodynamic analysis requires the state variables to be defined for the entire system under consideration.[1] This need to have the state variables of the system defined requires time on some usually rather long time scale. This is generally summarized as allowing all the fast things to have happened before the slow processes of the thermodynamic changes take place. The processes of thermodynamics thus has two time scales. When the system is manipulated on time scales that are too short to allow 'thermalization' the process is irreversible. When the time scale is long enough for the 'thermalization' to be obtained at each infinitesimal step the process is said to be reversible; so slow that at any step in the process the change can be reversed back to the original configuration. In this introduction to thermodynamics, we will deal with reversible processes.

It is also important to have some understanding of thermodynamic variables. These are the organized in two different ways, state and process. For specificity, you should consider the simplest system, a simple gas in a containing volume. For Thermodynamic purposes the situation is described by state labels; pressure, P, volume, V, amount of gas[2], N, and the tempera-

---

[1] In cases like this, there are two ways around the problem. One is to subdivide the system into smaller parts so that in each part equilibrium is meaningful. The other is to allow irreversible processes but to determine the final state of the process by means of conservation laws. The 1^{st} Law discussed later is a good example of these.

[2] Since for gases in reasonable circumstances, the most useful measure of how much gas is in the volume is measured by the number of atoms in the volume. In the case of gases, the mass unit is the gram-mole which using Avogadro's number, $N = 6 \times 10^{23} \frac{\text{atoms}}{\text{gm·mole}}$, sets the number of atoms in the volume. The gram mole is the mass in grams divided by the atomic mass number of that atomic specie, i.e. for helium whose atomic number, the number of nucleons in a helium nucleus, is 4, the mass of N grams is $\frac{N}{4}$ in gm moles.

ture T. These variables are constrained by an equation of state which for the simple gas such as helium at modest pressures is PV = NRT where R is a universal constant called the ideal gas constant[3]. A system with these state variables is called a PVT system. Many students in their first exposure to thermodynamics have the impression that the only systems that are dealt with are these ideal gases. This is not the case. We could be analyzing the energetics of electrical storage batteries which in the simplest case is a Voltage, V, charge Z, and temperature T system. We could be dealing with elastic extensive systems more commonly called rubber bands. Here we would have a tension, t, length, l, temperature, T, system. For all these, we would identify a constraining arrangement of the variables called the equation of state.

The other important aspect of the thermodynamical variables is how the variables scale with the system size. Simply stated, the variables come in two categories: extensive and intensive. A thermodynamic variable that scales with the system size is an extensive variable. If you make an exact copy of the system with which you are dealing, remove the separator between them and combine them into one system, extensive variables will double and intensive variable will remain the same. In the ideal gas case, the volume and the number of atoms are extensive and the pressure and temperature are intensive. In the battery case, the charge is extensive and the voltage and temperature are intensive. In the rubber band, the length is extensive and the tension and temperature are intensive. It is important to note that since the classifications, extensive and intensive, are based on a scaling law of size of system that the ratio of two extensive variables is an intensive variable. Specifically, in the ideal gas case, the density of gram moles is $\rho \equiv \frac{N}{V}$ and thus ρ is an intensive variable.

Remember that the basic role of thermodynamics is to investigate processes involving changes in the variables of the system to produce things like work and heat. We started our discussion of thermodynamics with a statement that pressure, an intensive variable, differences between two systems is what leads to changes volumes, an extensive variable. The thermodynamical variables all come in pairs one extensive and one intensive; in gases pressure and volume, in elastic systems tension and length ... and processes come about by the relationship because of changes in the value of a member of a pair. The usual example is work done by changing the volume of a gas that is contained in a volume that is variable by having a movable wall element, a piston between two systems. The relevant phrase is that the pressure brings

---

[3] R is basically the energy content of one gram mole of a gas at a temperature T.

about the change in volume and in doing so work was done or energy in the form work was transferred to or from a member of the system. Shinola Shinola

19.2.4 0^{th} Law: Temperature Exists

Depending on the macroscopic system under study, there are small number of properties of the system that can be measured, see Chapter 2 on page 33, and the values of these attributes define the state of the system. These are things like the volume, pressure, concentration of species, magnetic field strength and so forth. Among these is the temperature, T. The 0^{th} Law of Thermodynamics is basically the statement that temperature exists; it is a measured quantity whose values can be put on an objective scale. Like all measurement processes, there is an instrument or set of instruments that are standards and a protocol for use. The temperature is measured through the process of contact with a system whose response has been calibrated. The home mercury thermometer is a simple and classic example of a calibrated system that serves as the basis for measurement of temperature; a quantity of mercury is contained in a volume under zero pressure and with elongated space for expansion and the volume measured and calibrated to predetermined situations that allow us to define a temperature scale.

The protocol for measuring the temperature follows from the 0^{th} Law of Thermodynamics which states that **two or more bodies held in thermal contact for a long enough time will come to the same temperature**. There are several things that need clarification. What do you mean by contact and what constitutes a long enough time? In many cases, contact means literally what it says – touching. A more general and more appropriate definition is best stated as that somehow it is reasonable to talk about the two systems separated by a permeable membrane that allows an exchange between the systems. Of particular interest here is the exchange of energy as driven by a temperature difference. By the way, we will see that this sentence also provides a definition of heat, the energy that is exchanged between two systems in contact with the separating wall that allows temperature driven energy exchange. Note also that the systems under discussion have no other contact; they are isolated from everything else. The exchange can carried out in different ways but for now we restrict our discussion to contact by touching, better said enveloping, but not allowing any other transfers. In fact, this particular form of exchange is called conduction. Shortly, we will expand our contact regimens beyond touching but for now let's leave it at that. The other issue with the statement of the 0^{th} Law is how long is long enough. The -1^{st} Law discussion above clarifies this issue. If the objects have very different temperatures when brought into contact, they will begin to show changes in other measures of their state such as density, length, color, etc. These or, at least, one of these other measures will start to change. Long enough is when these changes stop. We say that the systems have achieved thermal equilibrium. In equilibrium, all state variables are unchanging including the temperature. This is exactly how the home thermometer is used to measure temperature in a person. In this case, when the expansion of the mercury stops, the thermometer and the person or more precisely the part of the person in contact with the thermometer bulb are at the same temperature; the temperature is indicated on the thermometer by the current volume of the mercury. We would then establish a temperature scale based on these volumes.

I should expand on some of the subtleties of the previous discussion. As stated above temperature is one among several measures that are required to specify the state of the system. Note that I have not been precise about the idea of "a system." Often in introductory discussions of thermodynamics, the system is a finite volume of gas but this is much too restrictive to contain all the physical cases. Some examples and their state variables are: the box of gas case described by the pressure, temperature, volume, species content; a rubber band suspended between two fixed posts with temperature, tension, and density specifying the state; the mercury thermometer discussed above with its vacuum volume and liquid mercury bounded by a glass envelope. These constitute systems. A physical entity that is described by an appropriately complete set of state variables and with specified boundaries. In some cases, the system is made of subsystems. The mercury in class case is a simple example. At a higher level of analysis it can be considered three connected subsystems, the volume of mercury, the volume of the vacuum and the volume of the glass enclosure all with their state variables. This is because we have idealized the parts. The mercury is assumed to be incompressible and if so there is no need for a discussion of the pressure. We also assumed consistently that the vacuum does not have pressure. In fact the mercury is slightly compressible and the vacuum is actually filled with mercury vapor and the spatial separation of the mercury and the mercury vapor allows mercury atoms to move between the two. For now and since each case is unique and needs its own considerations lets leave it that there are system and systems are specified by contents at a certain state labeled by measurable variables and in some sense isolated by proscribed boundaries. The boundaries may or may not allow passage of something between the system. For example in our thermometer, the boundary between the mercury vapor and the liquid mercury can or cannot allow the passage of mercury atoms and thus allow for changes in the amount of mercury in either subsystem. When there is passage of matter or energy between connected systems this is a process.

It is very important to realize that Thermodynamics was developed to describe processes utilizing temperature and heat flow to produce useful work. We also have to return to our earlier observations about measurement and ‘what is’ in Section 2.2 on page 34 in which we state that a thing is what is measured. The example used there of distance and length is important to our understanding of temperature. In the example of length, we articulated that the identification of a standard and a protocol defined the meaning of distance. Similarly here we define temperature through the establishment of standard and protocol. For our purposes it is adequate to consider our earlier discussion of the mercury example as our standard and protocol, the extension of the mercury in the thin tube in contact with the system whose temperature is being assessed. We must keep in mind that our goal is to describe thermal processes which do mechanical work or some other useful activity. This brings about a reinterpretation of our state labels. The comparison of temperature with another more familiar state variable, pressure for example, is worth pursuing. Pressure was an established concept and it role in energy transfer clearly indicated by Newton. If two isolated systems, labeled 1 and 2, are enclosed by a movable partition and the pressure in 1 is greater than the pressure in 2, The volumes of both systems are changed and, if we do this process slowly, and wait a long time to allow the pressures to equilibrate, the volumes will be changed and, although this will sound like a strange way to say it, both volumes will change. Said another way, pressure differences drive volume flow. Actually from simple mechanics we see that the pressure driven volume changes exchanges energy between the systems. The higher pressure does work on the lower pressure system; the work done during a volume change is $P\delta V$. Temperature is a state variable of a system like pressure. These variables come in pairs. Pressure is paired with volume and volume changing processes are work flow. Pressure is a state variable that is intrinsic. It is the same throughout the system and two identical systems have the same pressure but two identical systems have twice the volume. Volume is an extrinsic variable. Similarly two identical systems have the same temperature; temperature is intrinsic. We have not defined the partner extrinsic variable which is related to the heat flow. That is the problem of the 2^{nd} Law. We will get to that after we deal with the 1^{st} Law.

The example above using a thermometer to determine of body temperature measurement is also an illustration of another important property of thermodynamic systems. There are systems which for some reason or another do not change important thermal variables in processes. These are called baths. In the example above of measuring body temperature, we assume that the body temperature does not change as the process of thermal equilibrium established within the mercury in glass thermometer. Heat flows from the body to the thermometer but the body's temperature does not change. For this process, the body is said to be a temperature bath. In fact, this experience is an indicator that physiologically we sense heat flow not temperature since as the thermal equilibrium is established the thermometer goes from 'feeling' cool to okay. The human body achieves its thermal bath status by two means. Firstly, the body is so large compared to the thermometer that the heat flow needed to thermalize the thermometer is insufficient to change the body temperature. Also even if we place the human body in contact with another large system, our metabolism will function to maintain our internal body temperature. In a room at some nominal temperature, the body does thermalize in the sense that our surface temperature becomes almost the same as the room temperature which is acting as the temperature bath. Since this is different from our internal body temperature maintained by our metabolism, the average person transfers about 70 Watts or 70 Joules per second of heat energy into the room. In this sense, we never achieve thermal equilibrium with the room but this amount of heat flow is sensed as normal. More than this is a cold room and less is hot. Another useful example of a thermal bath is a pot of boiling water at standard atmospheric pressure. If you increase the heat flow from the stove all that happens is that the water boils more vigorously but its temperature does not change. In fact, this thermal bath is used to fix one of the standard temperature designations. In this case, $212\,^0$ F. The other end of the Farenheit temperature scale is defined from a temperature bath of ice, water, and salt defined to be at $0\,^0$ F. With these two points, the temperature scale can be set.[4]

This brief discussion of the definition of temperature is characteristic of all thermodynamic discussions. Although we characterize systems with labels such as pressure, temperature and volume, the important issues deal with the processes of change. In this discussion of the 0^{th} law, we defined

---

[4] The current temperature standard is not the simple mercury in glass simple thermometer using only two temperature baths. The National Institute of Science and Technology, NIST, follows the International Temperature Scale of 1990, ITS-90. ITS-90 sets the temperature scale using several temperature baths in the range from 83.4^0K to 962^0C to calibrate a platinum resistance thermometer.

temperature as a measurable quantity but used the idea of process occurring manifest by change to the each of the systems labels due to an initial temperature difference between systems in thermal contact. The basic issue is change brought about by differences in state variables and the ability of all systems to have the changes ultimately stop. This condition is called equilibrium. The existence of a state of equilibrium between systems is what allows us to define the temperature scale but temperature is not the only state variable that is measured by establishing equilibrium. Two cylinders of gas connected by a movable wall but at different pressures will change by moving the wall toward the smaller pressured cylinder. Change stops occurring when the pressures equilibrate. We say that the pressures are equal. Note that this discussion of pressure does not use the concept of force or area. We can define a pressure scale using pressure baths and equilibrium just as we did with temperature. As in the case of temperature, there are systems that can be considered pressure baths. A huge volume of gas such that the movement of the piston of a small system connected to it does not change the volume significantly enough to modify the pressure or any of the other thermal variables of the large system. The atmosphere is a pressure bath to all intents and purposes for man-made processes on the earth. The connection between this purely thermal discussion of pressure and its usual force and area description will be clarified by the other laws of thermodynamics.

There is also an important point to note and that is that the process under consideration can take place in two ways. We can bring into contact two systems with very different temperatures. As required these systems will ultimately come to equilibrium as indicated by the stability of their state variables. Another approach is to bring about the temperature change by means of a system of thermal baths that makes the process a continuous set of very small changes. This later case is called reversible. The reversible process has the advantage that the system is in a certain state as manifest by its state labels at all steps in the process. The first case is labeled an irreversible process. Both can be analyzed by thermodynamics but the the outcomes of the processes will be different if the process is irreversible or reversible. How there can be a difference in the final states of reversible and irreversible processes will be clarified in the 2^{nd} law. In our original discussion of temperature equilibrium, we just discussed making the contact membrane between the systems permeable to the flow of heat. In some sense, all partitions are permeable to the flow of heat. What really matter is the time scale over which the heat flow operates. Any material wall with a finite temperature difference will pass some amount of heat on some time scale, this is the based on the thermal conductivity, $\tau \equiv \frac{\delta Q}{\Delta \mathrm{time} \cdot \Delta \mathrm{T}}$ of the material. The rule for reversible processes is that the rate of heat transfer must be small for any time scale for thermalization of each of the systems. Another way of saying that the time scale for heat flow through the wall is infinitely long.

This discussion of thermal equilibrium above is based on two systems in thermal contact. The idea is that the systems are both contained in some sense; enclosed so that there is no communication between them. In our discussion of the process of thermal equilibrium above the two systems initially with different temperatures are brought into contact by joining them at a surface of each one and that the surface allows heat to flow between them and it will flow until the two system temperature are the same. This simple set up is the epitome of the operation of processes in thermodynamics. Two or more systems are connected by a membrane which is porous to some particular one or more important thermodynamic labels. The contact wall allows flow of something and that flow is driven by the difference in the relevant state variable. In the case of the 0^{th} Law, the contact wall allows heat to pass if there is a difference in temperature between the systems.

Before closing this section on temperature equilibrium processes, we should look at the situation of two like systems connected by a movable wall but well insulated to heat flow between them. We know that volume will flow to the low pressure system from the high pressure system and that equilibrium will be reached. Using our ideas of mechanical energy we would say that there was energy flow between the systems with the high pressure system losing mechanical energy to the low pressure system. In this way of looking at this, there is in complete analogy with the temperature difference system. Is this volume exchange the same as the energy concept in mechanical systems? This is the subject of the next section.

1^{st} Law: Conservation of Energy

The separation membrane between two systems can effect each other in different forms than conductive heat flow driven by temperature differences. For example, consider two volumes of gas with differing pressures and differing temperatures. The barrier between them can be movable, a piston, and heat conducting. In language that follows our discussion above about heat flow, we can describe the process as pressure equalization by movement of the piston and the movement continues until the pressure difference vanishes. In the process of pressure equilibration the mechanical measures, volumes, of the component systems change. Again, as in the case of two systems in contact and having a temperature difference other differences emerge as the equilibration takes place. In our simple glass thermometer as the equilibration comes about the volume of the mercury changes and we conclude that the equilibration process is taking place until all measures on the system stop changing.

In this case, the energy transfer is called work and the idea that it is done by one system on the other, work flows from the higher pressure system to the lower pressure system. In the case of temperature equilibration, the heat flows from the higher temperature body to the lower temperature body.

Another example a thermodynamic process would have the two systems separated by barrier that allow matter to pass but nothing else, no movement and not heat flow and so forth.

In our thermometer example the mercury expands as heat is added but because the other space in the thermometer cavity has no pressure, $P_{vac} = 0$, another system state variable, In this case, there is no work done since work done for a volume change is the $P\Delta V$. P is called an intensive state variable since its defining statement that in a process which changes the volume, the work done is the energy exchange driven by pressure differences. The First Law of Thermodynamics relates the various forms of energy change in a thermodynamic process. It is simply the identification that, in any process, the total energy is conserved. As applied to the example of the mercury in the thermometer, the heat flow into the mercury that changed the temperature did not $P\Delta V$ work but went into energy content of the mercury which manifested as the temperature increase. If there was some pressure in the 'empty' part of the thermometer, there would be work done by the expansion and more heat would have been needed to raise the temperature the required amount.

2$^{\text{nd}}$ Law: Increase of Entropy

The next two laws, two and three, are much less well known and thus a bit more subtle to discuss. The 2$^{\text{nd}}$ Law clarifies the nature of heat. It is important to realize that heat is defined only in a process.

3$^{\text{rd}}$ Law: Absolute Zero

19.2.5 Radiation in a Cavity

Now that we have a vocabulary, let's examine an especially simple thermal system. Consider a massive block of stuff, say aluminum, that is in thermal equilibrium at some temperature T_1. We can change the temperature by

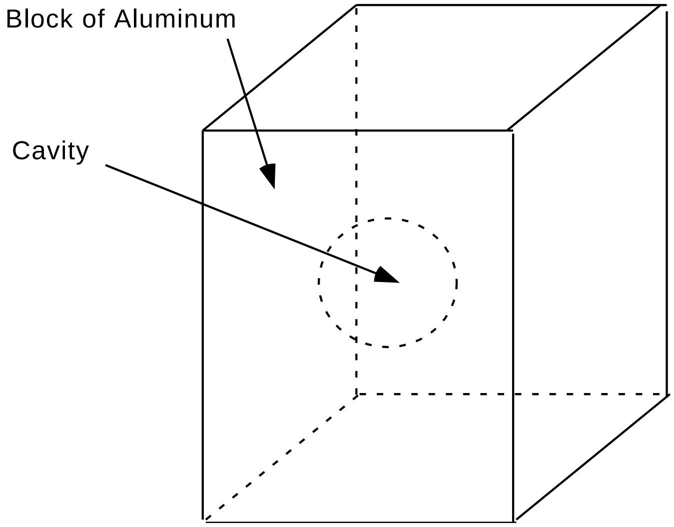

Figure 19.2: **Black Body Cavity** Inside a block of material an empty cavity absorbs heat. The amount of heat needed to raise the temperature of the cavity scales as the volume of the cavity.

placing the block of aluminum into contact with a succession of temperature baths to study the thermal properties of the aluminum. These temperature baths have very closely spaced temperatures so that we can consider the heating to be a slow or equivalently reversible process. We can measure the heat flow into the aluminum when we incrementally change the temperature of the aluminum. The heat that is required to raise the temperature a small amount scales with the mass of the aluminum block in use. We construct the quantity, $C \equiv \frac{\delta Q}{\Delta m \; \Delta T}$ called the specific heat.[5] Also note that there are now two different symbols for the changes brought about by the thermal processes. Lower case δ is a change in something that is not a state variable. Heat is an energy but it is not a state variable, see the discussion in Section 19.2.4 on page 417. It is only defined during a change; there is no such thing a Q. Make a hole in the center of the stuff. The hole is empty, a perfect vacuum or at least as near as we can get, see Figure 19.2. We have studied all the thermal properties of the stuff, the aluminum in this

---

[5] There is a technical issue here. Adding heat in the open air is different than adding the heat at constant volume; the block of aluminum expands against atmospheric pressure, whereas at fixed volume there is no mechanical work done on the atmosphere. Therefore, we always specify the process in which the the heating is done. In this case, we are using C_V, the constant volume specific heat.

case, and completely understand it. In order to raise the temperature of the aluminum, we have to add heat or energy to it including the hole. If we take into account the heat to raise the temperature of the stuff, we find that it takes energy to raise the temperature of the nothing that is in the hole. In the case of the empty cavity, the heat to increase the temperature scales as the volume of the hole.

This has to be a surprising result. Even though the hole is empty, a vacuum, it takes heat input to raise its temperature. If you make a bigger hole, you need more heat input for the same temperature change. The amount of heat required for a given temperature change scales as the volume of the cavity; for a spherical cavity of radius R as R^3. If the energy being put in the cavity were a surface phenomena, the heat input to raise the temperature would scale as R^2.

Put a hole in the side and look at what comes out. If the temperature is high enough, light that you can see comes out. As you raise the temperature, the light gets whiter. We now know what is going on. We are filling the cavity with light. The heat energy added to raise the temperature goes into making the light. There are several simple observations that confirm this. The nature of the light as indicated by the spectrum is independent of the material bounding the cavity On the other hand, the spectrum of the light is universal because there is nothing in the cavity, the vacuum. The spectrum depends only on the temperature. We can verify that the energy delivered to the cavity scales as the volume and not the surface area of the cavity. Again, this makes sense since there is nothing in the hole but the light that is produced. Hypothesizing a particulate nature for that light, what we now call the photon, Planck developed the following formula for the the light intensity at any wavelength in a wavelength interval $\Delta\lambda$:

$$\rho(\lambda, T)\Delta\lambda = 8\pi c\frac{h}{\lambda^5}\frac{1}{e^{\frac{hc}{\lambda kT}} - 1}\Delta\lambda \tag{19.1}$$

where λ is the wavelength, c is the speed of light, k is Boltzman's constant, T is the temperature and h is a constant that he used to fit the spectrum. That constant, h, is now known as Planck's constant and considered to be one of the fundamental dimensional constants of nature[6]

---

[6] The form of the function $\rho(\lambda, T)\Delta\lambda$, 19.1 on page 419, is a shorthand representation for $\int_{\Delta\lambda,\Omega} I(\lambda, T) d\lambda d^2\Omega$ where $I(\lambda, T)$ is the light intensity at a wavelength λ. Ω is the directions of the beam of light being analyzed. If the light is going in all directions $\Omega = 8\pi$. This is an angle and is dimensionless. The integration range is $\Delta\lambda$ is taken to be infinitesimally small so the λ is the wavelength at that value.

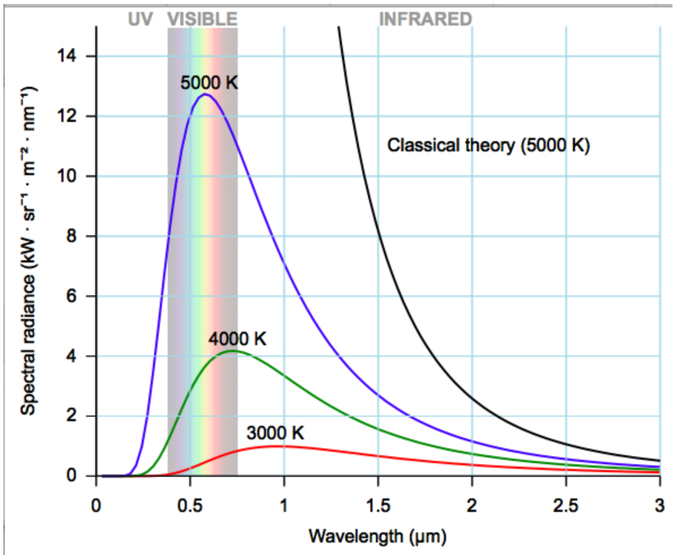

Figure 19.3: **Intensity of Light from a Cavity as a Function of Wavelength.** Plots of the intensity of the light as a function of the wavelength for three temperatures, 3000 °K, 4000 °K and 5000 °K. These curves are plots of Planck's Law, Equation 19.1 on Page 419. The higher the temperature the lower the peak and the greater the area under the curve. Also shown is the location of the visible window on the spectrum and the classical Rayleigh Jean's predicted spectrum for T=5000 °K

The primary features of the observed energy density spectrum that Planck found, see Figure 19.3 on page 420, had sharp peaks at a characteristic wavelength that increase and shift to lower wavelength, λ, as the temperature increases. In addition, the area under the intensity curve grows larger as the temperature increased. Equation 19.1 on page 419 is the function that Planck found that fit the data and it is in excellent agreement with the observed data. He managed to get this formula through a process that is common in physics research labeled phenomenological. He picked a rather simple formula that had the pattern of development that looked like the then known observed data. His reasoning was functional simplicity. As more data was becoming available, he was impressed with how closely the new data matched his simple function. He immediately realized that he had to develop a physical picture of the thermally developed light to 'explain' the data.

It is worthwhile to point out at this point that the dimensional content of Planck's fitting parameter, h, is action $h \overset{\text{dim}}{=} \frac{ML^2}{T}$.[7]

19.2.6 Attempts to describe the spectrum

The first effort to predict the spectrum of the light in the black body cavity from first principles was by Lord Rayleigh and Sir James Jeans. They followed what would be natural to us. We treat light as a field phenomena. The earlier descriptions of light as particulate based phenomena developed by Newton were discredited even in his own time.

The only trick in developing the prediction of Rayeigh-Jeans is to get the form of the intensity of the thermal light from dimensional analysis. Intensity is energy per unit area per unit time in a wavelength interval or $I \overset{\text{dim}}{=} \frac{E}{T \cdot L^2} L \overset{\text{dim}}{=} \frac{ML^2}{T^2 \cdot L^2 \cdot T}$. Using only temperature and wavelength which is all you have for a cavity, dimensionally the spectral intensity must take the form

$$\rho_{RJ}(\lambda, T)\Delta\lambda = 8\pi \frac{kT}{\lambda^4}\Delta\lambda, \tag{19.2}$$

where k is the Boltzman's constant and the 8π is a geometric factor.

From Figure 19.3 on page 420, the Rayleigh-Jeans law looks awful. It does well only at very long wavelength, the far infrared. More difficult was the fact that energy at small wavelengths diverged. This was referred to as the ultraviolet catastrophe and was a clear indication that this pure field approach to explain the spectrum was inadequate.

Another early effort to derive the spectrum was by Wilhelm Wien. Studying the thermodynamics of a field filled cavity, Wien found that, if instead of plotting the spectrum against the wavelength for fixed temperature you scaled the wavelength with the temperature $\lambda_T \equiv \lambda T$, the curves other than a scaling with temperature were universal. This was called the Wien displacement law. He then modeled the spectrum following the pattern of the well known earlier success of Maxwell who had developed a velocity spectrum for gas atoms confined to a fixed temperature volume. It is worthwhile to review Maxwell's descriptions of the thermodynamics of atomic system to understand both the limitations and the success of the treatment that Planck ultimately applied to the thermal cavity. Wien attempted to describe the

---

[7] The dimensional content of h is determined either by the overall scale of the spectral intensity, $\rho(\lambda, T)\Delta\lambda = 8\pi \frac{ch}{\lambda^5}\Delta\lambda$ or the requirement that the power of the exponential, $\frac{ch}{\lambda kT}$, be dimensionless. Also note that in these sections on thermodynamics the symbol T is used for temperature but in all the discussions of dimension the symbol T is reserved for the dimension of time. The reader will have to distinguish the usages by the context.

light in the cavity as Maxwellian like but satisfying his displacement rule[8]. Wien's formula has two fitting paramaters, c_1 and c_2 but still could not match the data in the infrared.

$$\rho_{Wien}(\lambda, T)\Delta\lambda = \frac{c_1}{\lambda^5}\frac{1}{e^{\frac{c_2}{\lambda T}}}\Delta\lambda \tag{19.3}$$

This was not a rigorous derivation but it did fit the small wavelength part, the peaks and the motion of the peaks with changes in temperature.

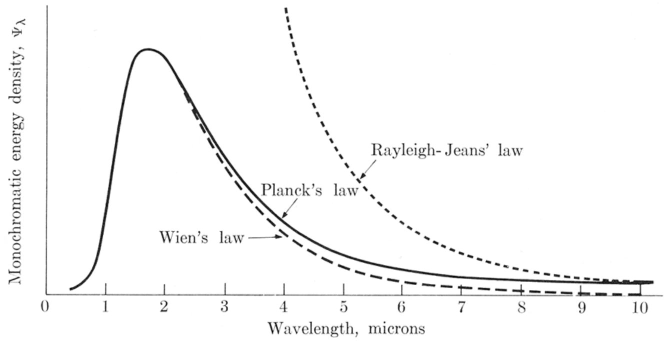

Figure 19.4: **Classic Attempts to Predict Spectrum** Plots of the intensity of the light as a function of the wavelength for a fixed temperature for the case of the Wien Law and Rayleigh-Jeans Law. It is no surprise that Planck's formula took the form that it did. In a sense all that it had to do was interpolate between the two classical laws.

19.2.7 Planck's Explanation of the Spectrum

The issue for Planck was not to get the relationship, equation 19.1 on page 419, but to justify it from some fundamental understanding of the nature of light. To derive his formula, Planck had to assume that thermal energy was distributed in a system of discrete entities whose energy was proportional to the frequency. His reasoning followed from Wien's displacement condition and the Maxwell form. The $\lambda_T \overset{\text{dim}}{\approx} E \cdot L \overset{\text{dim}}{\approx} \frac{E}{c \cdot T}$ was a term

---

[8] Maxwell's rule for speed distribution for a system of atoms of mass m and energy $m\frac{v^2}{2}$ where v is the speed is

$$\rho_{Maxwell}(v, T)\Delta v = 4\pi v^2 \left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} e^{-\frac{mv^2}{2kT}}\Delta v$$

that naturally could fit into a exponential, but that the statistics, the way that multiple copies of the same thing are counted, were not the normal ones. He had to count the physically independent identical states in a very unusual and inexplicable ways. The proportionality constant between the energy and the frequency that he had to use was the fitting constant h that he used in Equation 19.1 on page 419 is now called Planck's constant.

$$\epsilon = h\nu = \hbar\omega, \tag{19.4}$$

where ν is the cycle frequency, ω is the radian frequency and $\hbar = \frac{h}{2\pi}$.

The second aspect of Plank's attempt to completely describe the phenomena of thermal radiation by articulating how a collection fundamental structural entity dealt with the issue of identity of these elements. In order to get his function, he had to count the possible configurations in an unusual way. Too get some understanding of the counting issue, consider a collection n coins of with no other distinguishing features except a two labeled degree of freedom such as H and T for heads and tails. For example consider 5 coins. The configuration HHHHT and the configuration HTHHH would be considered distinct. For his counting, he had to consider these the same configuration. For a collection of n of his entities, he had to require that be only n unique identifiable configurations. The basis for this strange result was clarified by S. N. Bose and Albert Einstein some time later. The resulting counting scheme is called Bose-Einstein statistics and is one of the cornerstones of our understanding of modern physics [Wali 2009].

19.3 Photo-Electric Effect

We are all well aware of the use of light to control electronic devices. Photoelectric devices are used to open doors, send optical signals, and operate computers. Today, we usually see these devices as small solid state elements. This was not always the case. The early photoelectric devices were large vacuum tubes. Light hitting the clean metal surface inside the tube would cause an electric current to flow. This phenomena was just being discovered shortly after the turn of the 19^{th} century, see Figure 19.5.

Einstein clarified the conflict with the observational data available at the time by saying that Planck was right and that light is composed of countable entities that have energy in proportion to the frequency of the light, the color. He suggested that the process of the photo cell was that the electrons in the metal of the anode absorb the light to have the energy

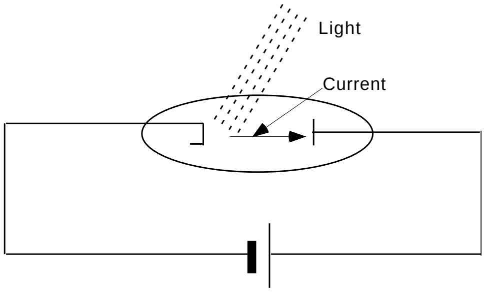

Figure 19.5: **Photo Electric Effect** Light shining on a metal plate in a vacuum enclosure releases electron into the evacuated space which form a current.

they need to move out of the metal into the space between the cathode and anode and move ac across the gap to complete the current.

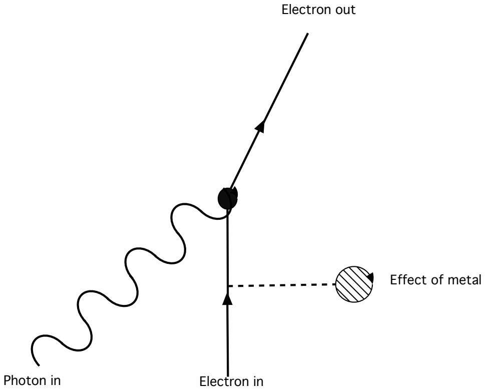

Figure 19.6: **Einstein Photo Electric Effect** A single photon of light interacts with a single electron. The energy of the photon removes the electron from the metal and gives it kinetic energy, see Equation 19.5.

The energy of the photon, $h\nu$, is transferred to an electron which gives it both kinetic energy and the energy to overcome the weak binding of the metal, see Figure 19.6 on page 424.

$$h\nu = m\frac{v^2}{2} + \phi \tag{19.5}$$

Where ϕ is the energy required to move the electron out of the metal.

This picture predicts that the number of electrons, the current through the tube, is equal to the number of photons. Thus, for a given frequency, color, the current is proportional to the intensity of the light. He also noted that you could measure the velocity of the electron by back-biasing the tube and seeing what voltage just stops the current[9]. This phenomena is like the ability to measure the velocity of a vertically thrown ball by measuring how high it rises against the earth's gravitational force; $\frac{mv^2}{2} = mgh$.

$$\frac{mv^2}{2} = eV_{stop} \tag{19.6}$$

In other words, by measuring V_{stop} and the frequency you can measure h as the slope of the straight line.

$$eV_{stop} = h\nu - \phi. \tag{19.7}$$

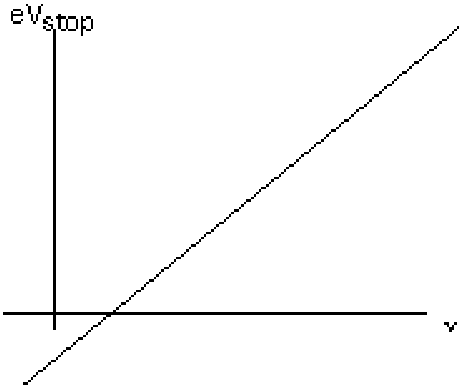

Figure 19.7: **Plot of Stopping Potential Versus Frequency** The stopping potential versus frequency of the light curve as predicted by Einstein, Equation 19.7 on Page 425. Thus the relationship between the frequency and stopping potential was linear with slope h. The experiment was successfully executed by Millikan and confirmed the prediction of Einstein.

Thus Einstein, by extending Planck's analysis of the black body experiments, makes several predictions. When you plot the stopping potential against the frequency of the light, you get a straight line and that the slope of that line is the value of $\frac{h}{e}$ where e is the charge of the electron and that the h is the same as Planck's value, see Figure 19.7. The value of the

---

[9] Einstein was awarded the Nobel Prize in 1922 for this work. Interestingly to modern physicists, he was never awarded the Prize for his development of the General Theory of Relativity.

charge of the electron was already measured in an earlier experiment by J. J. Thompson.

Another prediction from Einstein's construction is that the current in the phototube illuminated by monochromatic light is proportional to the intensity of the light but that no matter the intensity of the monochromatic light there was a cutoff for a phototube independent of the intensity and depends only on the frequency.

All of these predictions were satisfied by a series of very difficult experiments carried out by the great american physicist, Robert Millikan. These energy packets of light were subsequently called photons and are now universally recognized as the fundamental constituents of light.

Subsequent experiments importantly also showed that you could run photo-dectors at such a low intensity of light that the current that you could count time separated photon events.

It is the combination of the well understood properties of light as manifested by the Young's double slit experiment and the result of the Einstein analysis of the photo-electric effect that is the foundation of quantum mechanics. In addition, when these are combined with the understanding of the black body experiment that was articulated by Planck, the emergence of particles in the vacuum and the counting scheme for identical particles, we have the basis for a modern theory of matter that is manifest in the quantum theory of fields, see Chapter 22 on page 483.

19.4 Young's Double Slit Experiment Revisited

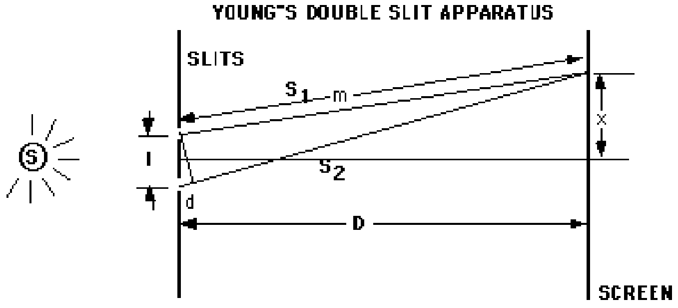

Figure 19.8: **Double Slit Revisited** Light from a single source illuminates two slits. When both slits are open, there is an interference pattern at the screen.

In Section 4.4 on page 88 while discussing light, we introduced the double slit experiment of Thomas Young. Monochromatic light passing through two narrow slits and illuminating a distant screen produced a pattern of illumination that at some places produced a brightness that was four times what would be present if only only one slit was exposed. There were intervening places that had no illumination. The spatial average of the brightness of the illumination was twice that of one slit being open. The only working description that was possible was that the light's causal agent was not the brightness but an underlying descriptor, the amplitude, whose square was the brightness. Not only that but using Newton's observation that light had an underlying structure that manifested itself as color and that the label for the color was associated with a label that was periodic. Fresnel, Section 4.6 on page 106, extended this analysis to diffraction phenomena which lead to our discovery that light is a phenomena the travels over all paths in all of space when going between two places. The only construction that would describe the pattern of bright spots and dark places on the illuminated screen exposed to monochromatic light was that light was an amplitude definable at every point in space, a field which was not directly measurable but whose square was the brightness at that place. Figure 19.9 on page 427 shows the

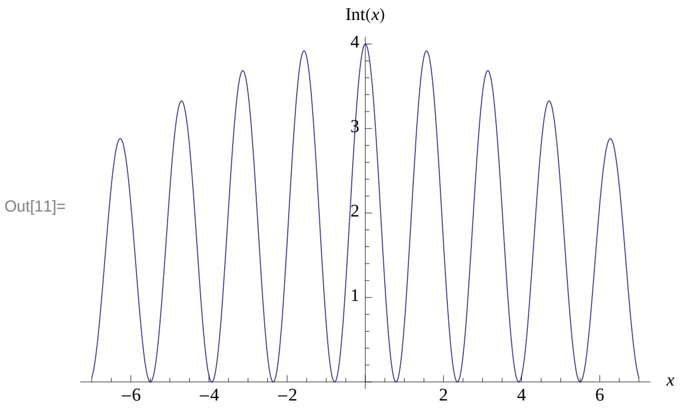

Figure 19.9: **Two Slits** When monochromatic light from two narrow slits illuminates a screen, a pattern of bright and dark lines are produced, see Figure 19.8. The figure above shows the brightness of the light, Int(x), for the positions x measured up the screen of the light as you move up the screen. There is small decrease in the brightness of the peaks as you move away from the cental maximum due to diffraction.

intensity pattern for positions varying as you move up the screen in Figure 19.8 on page 426. These developments in our understanding of light was further extended by Maxwell, see Section 5.4 on page 140, when he unified the electric and magnetic force system to include the observation that disturbances of the electric or magnetic field would propagate as a wavelike disturbance in the fields traveling at the speed of light; light had become the electric and magnetic field effects operating at very high frequencies.

Now let us replace the remote screen on which we displayed the illumination pattern with an array of of very small Einsteinian photo detectors[10]. What happens now? Instead of intensity, you count the photons at each location. There are clicks all over the surface of the array. You can run the intensity down so low that, for that color of light, you get only one count a week. Actually, this is an exaggeration. You can run the experiment at such a low rate that you are confident that there cannot be two photons in the system at the same time[11]. When you do this, you do not see two half hits at the same time. You also do not see hits in places where the Fresnel/Young analysis indicated that there was no brightness. Ultimately, as the counts accumulate, the number of counts at a place x on the screen, $n(x) \equiv \frac{\text{counts in patch of area A}}{A}$ per unit time follow the Young/Fresnel pattern seen in Figure 19.9 on page 427.

$$< n(x,t) >_T \hbar\omega = \mathrm{I}(x) \equiv < A^2_{\mathrm{Tot}}(x,t) >_T = < \{A_1(x,t) + A_2(x,t)\}^2 >_T, \tag{19.8}$$

where I(x) is the intensity or more correctly the energy per unit area per unit time of the light at a place on the screen and $<>_T$ indicates that a time average over times T large compared to the period of the light and also large with respect to the inverse photon arrival rate.

Earlier in our discussion of the double slit, we used a Fresnel construction that the light was a traveling wave phenomena to describe the brightness pattern. Now we need to incorporate the knowledge from the photoelectric effect that the detection of a photon by interaction is a local probabilistic effect. In fact, any attempt to correlate the time and place on the screen

---

[10] Actually we do this all the time. The CDD plane that is at the heart of current digital cameras is an array of Einsteinian photo detectors. Since the back plane of the camera is a couple of inches square, an array capable of producing Mega pixel images would use photo detectors that have an exposure face of about 10^{-3} inches in width and height.

[11] Just to be specific, you can have the distance from the source of the monochromatic light to the screen be tens of meters, say D. If this is the case and you associate the constituent that makes the clicks to be traveling with the speed of light, it is reasonable to say the photon that made the click is in the apparatus for $\frac{D}{c} \approx 10^{-8}$ seconds and, if the interval between clicks is about a second, there is no time when any two photons are in the apparatus at the same time.

of the event in relation to the previous event fails. There is no correlation on an event by event basis. **At the single event level the relationships between events are purely probabilistic, stochastic in both space and time**[12]. Only the long time accumulation of many events show a pattern; in this case, the classical double slit pattern. The photons like classical field phenomena travel over all paths although individual photons are always sensed locally. The pattern developed by $n(x,t)$ requires the photon locations detected at the array of Einsteinian photo detectors must be determined by a field that satisfies a Fresnel construction. In a very real sense, it is impossible to talk about individual photons. Although they may be widely separated in space ant time, they are effectively entangled.

This was a difficult problem for Einstein to come to terms with. In fact, he never really did. He believed that our descriptors of elementary objects, causal agents in my vocabulary, had to possess 'local reality'. On the other hand, this phenomena is real and has become very important in applications to what is today called quantum information systems, see Chapter 21 on page 471.

It is also worthwhile reminding ourselves that the pattern on the screen or array of photo detectors is determined by the wavelength of the light. The Planck condition on the energy of the photons is determined by the frequency. It is also important to point out that, when the Maxwell field system or any pure wavelike field system is developed from an action perspective and has both time translation symmetry implying a conserved quantity called energy and space translation symmetry implying through Noether's Theorem a field energy and momentum. This construction for the Maxwell field requires that the energy and momentum are related by $\frac{\mathrm{Pc}}{\mathrm{E}} = 1$ where P is the field momentum density, E is the field energy density and c is the velocity of the disturbances or in this case the speed of light[13] Using this condition for the photon requires that

$$p_{\text{photon}} = \frac{E_{\text{photon}}}{c} = \frac{h\nu}{c} = \frac{h}{\lambda}. \tag{19.9}$$

It should be noted that since c is huge compared to our usual velocities and that the photon momentum is very small compared to its energy. Thus

---

[12] Stochastic is the simple idea that the system has no memory. In a sequence each event is independent of its predecessors. After tossing five heads in a row, the probability that the sixth toss is a head is still one half; fair coin tosses are stochastic.

[13] For the free particle of mass m with the usual non-relativistic free particle action, $L = \frac{mv^2}{2}$ the momentum is $P = mv$ and the energy is $E = \frac{mv^2}{2}$ this ratio $\frac{Pv}{E} = 2$ clearly a constant of the motion.

an energetic disturbance on a stretched rope or water surface produces a noticeable momentum whereas an intense beam of light heats the surface but produces little pressure on it. A beam of n photons of light transfers an energy $nh\nu$ and a momentum transfer of $n\frac{h}{\lambda}$ or the ratio of the momentum transferred to the energy transferred is $\frac{1}{c}$ which in usual MKS units is $\approx 3 \times 10^{-9}\ \frac{\mathrm{sec}}{\mathrm{m}}$. This is why, although you sense the warming of sunlight, you do not sense the pressure of the sunlight. The sunlight does not blow on your face.

19.5 Electrons and Young

We now make one more change to our Young's double slit apparatus. We replace the light source s in Figure 19.8 on 426 with a cathode to provide a current of electrons[14] and we replace the screen with a plane of electron counters. The pattern on the screen follows the same general rule as it does for the light; the electrons come to the detectors in the same way as the photons of the light. The electrons are detected discretely. There are never half counts. As the counts are accumulated a pattern emerges with definite minimum and maximums. In the case of light, we used a field property of light to describe the pattern. The immediate implication is that electrons as elementary countable entities have field properties also. Some how we have to identify a measure of the electron beam the describes the pattern.

The wavelength of the light which determines its pattern is given by $\mathrm{e}_{\mathrm{photon}} = \mathrm{h}\nu_{photon} = \frac{\mathrm{h}c}{\lambda_{\mathrm{photon}}} = \mathrm{p}_{\mathrm{photon}}c$. Similarly, setting the wavelength on the momentum or a non-relativistic free electron is $p_e = m_e\, v = \frac{\mathrm{h}}{\lambda_e}$. Thus the energy[15] of the electron will be determined by its wavelength through

$$E_e = \frac{p_e^2}{2m_e} = \frac{\mathrm{h}^2}{2m_e\lambda_e^2}. \tag{19.10}$$

Even when the beam of electrons is very sparse, for example one electron per second, the pattern predicted by the electron wave length above develops over a long time at a rate of a click per second.

---

[14] Actually this is not what is actually done. It is impossible realize the usual Young set up for electrons. The slit separations would have to be on the scale of atom sizes. Instead, we use X-ray diffraction from a crystal to measure the distance between atoms in the crystal. We then send a beam of electrons onto the same crystal and from the resulting diffraction pattern determine the wavelength of the electrons in the beam.

[15] The appropriate form for the energy in this case is the non-relativistic kinetic energy, $\frac{m_e v^2}{2} = \frac{p_e^2}{2m_e} \approx \frac{m_e^2 c^2}{\sqrt{1-\frac{v^2}{c^2}}} - m^e c^2$, for $\frac{v^2}{c^2} \ll c$.

The conclusion is that all things elementary, in this case electrons and photons, have both constituent and field properties. Although we have only dealt with electrons and photons, since they are primary constituents of matter, modern physics assumes that all matter and energy is quantum in nature and that the mechanics of the constituents is determined by descriptors that were capable of manifesting this. It was also important to develop these from an approach based on the principal of least action and follows the usual rules mechanics in which Noether's Theorem implies conservation laws based on symmetries. It will turn out that this formulation will lead to a single descriptor set known today as a quantum field degree of freedom that is the basis for the theory of Quantum ElectroDynamics, QED, see Chapter 22 on page 483.

19.6 Action and Quantum Mechanics

To understand the two results, the photo-electric effect and the double slit experiment, we need to have light be different than our earlier constructions. To Maxwell, light was purely a field phenomena. There was nothing discrete to be counted and the non-zero field values traveled following the construction of Fresnel. With Planck and Einstein the picture has changed drastically – all interactions are discrete transitions that take place instantly and locally and are stochastic in both space and time. nature and that light travels over all paths and generates an interference pattern from a superposed amplitude whose square is the probability that you will find energy and momentum deposition at that place at that time, see Equation 19.8 on page 428. The implication being that light is composed of discrete entities that are at the place of the transition at that time. if you make a position measurement at that place you will find the constituent there. These points lead to all the quandaries that are associated with quantum mechanics. It is the combination of superposition and localization of interaction that is at the heart of such conundrums such as Schrödinger's cat.

In the case of the double slit, we have a superposition of the two sources, light from slit 1 and light from slit 2, and the local transition that says the the light hits a point on the screen.

We will also follow a parallel development that we saw in the case of Fermat's Least time and the Fresnel/Huygens's construction. We will find that the simple principle of Least Action as the process of selection of the natural trajectory from all possible trajectories is replaced by a process of adding an amplitude to each trajectory and calculating the phase advance over that trajectory and then at the final event adding the phases from all the trajectories to get the amplitude at the final event.

In order to describe quantum phenomena, we will require that the light travels over all paths and as it travels over all paths it generates a field that is composed of the sum of the phases from each path. In the language of the double slit, it comes from both the slits. The square of the resulting amplitude is the probability that there will be a photon at that point. In contrast to the Fresnel/Huygens construction, the phase that is advanced over the trajectory is the action in units of $\hbar$.

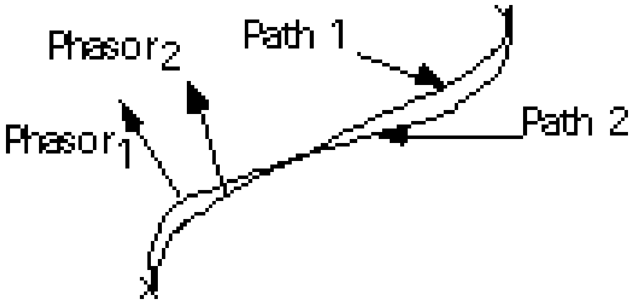

Figure 19.10: **All Paths Formulation of Quantum Mechanics** Light and, in fact, all things travel over all possible paths. Along the path, a phasor is advanced by computing the action over the path in units of $\hbar$, Equation 19.11. The resultant amplitude is the superposition of all the path phasors at that point.

$$\Delta\phi = \frac{\Delta S_{path}}{\hbar} \tag{19.11}$$

The situation here is the same as that which was obtained from the analysis of the Fresnel/Young construction and the Fermat Least Time goes; the paths around the least action path reinforce $\Rightarrow$ that the particles are found around the classical path. There is a wavelength in the field that the path forms just like in Huygens/Fresnel case. Once there is a natural trajectory, a connected region that has the phasors reinforcing, you have all the usual ideas of a particle mechanics. For a path made of phasors, you have a wavelength. But also from Neother's theorem, $\frac{\delta S_{nat}}{\delta x_{nat}}$, is the momentum. Momentum is not just a particle concept. It is related to any thing with a dynamic.

$$\lambda = \Delta x_{nat} \tag{19.12}$$

for which $\Delta\phi = 2\pi$ or

$$\hbar \Delta\phi_{path} = \frac{\delta S_{nat}}{\delta x_{nat}} \delta x_{nat} = p\lambda \tag{19.13}$$

or

$$p = \frac{h}{\lambda} = \hbar \frac{2\pi}{\lambda} = \hbar k \tag{19.14}$$

where k is called the wave number; the number of complete waves per unit length. For light this is consistent with our identification of $\epsilon = \hbar\omega$ since for light even in the classical wave theory, a light beam with energy density E has a momentum density $p = \frac{E}{c}$.

19.7 Constructing the Amplitude.

Our problem is to find a closed form for the amplitude. Our technique will be to follow the procedures of the Fresnel construction using phasor clocks that advance as the action accumulates on the trajectory. The way that this is described is to say the field or particle propagates from (x_0, t_0) to (x_f, t_f). As it propagates the phase and the magnitude of the amplitude changes. There are a few essential differences between this and our earlier algorithms. In this case, we are dealing with trajectories, connected events in space-time, not paths in space. An additional complication is that the phasor clock advances as the action not just the time advances. A third difference is that, since actions are time sliced, we will not rectify the segments of the trajectory, see Figure 19.11.

Divide the time interval into small segments each of size ϵ. There are $\frac{(t_f - t_0)}{\epsilon}$ of these time slices. At each time on a time slice, you let x take on all values. Between the ends of intervals let the path be a straight line. In this sense, we are saying that the set $\{x_i\}$, of positions at each time slice is the designation of the trajectory. Between each time slice, given the $\{x_i\}$, or each little leg of the trajectory, we can calculate the the position and velocity and thus the action and thus know how much to advance the phasor. We then do this for all sets $\{x_i\}$ to obtain all possible trajectories. You could say that there is propagation between each time slice and that the final propagation is the effect of the total of all these.

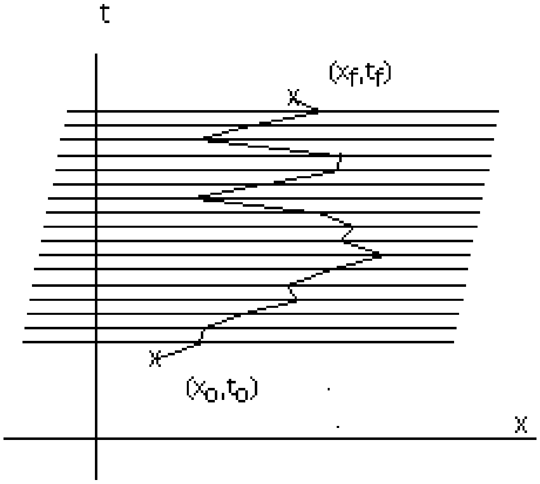

Figure 19.11: **Time Slicing** As light or any particle travels from event (x_0, t_0) to event (x_f, t_f), it travels over a path that is time sliced in small segments of size ϵ. The path is given by the set of values of x at each time slice.

$$\begin{aligned} Prop(x_0, t_0; x_f, t_f) &= \\ &\quad Prop(x_0, t_0; x_1, t_1) Prop(x_1, t_1; x_2, t_2) Prop(x_2, t_2; x_3, t_3) \ldots \\ &\quad \ldots Prop(x_{n-1}, t_{n-1}; x_f, t_f) \end{aligned} \tag{19.15}$$

In each time slice the phase advances by the change in the action in that time slice in units of $\hbar$. Note that by time slicing in this way guarantees that we only deal with trajectories that are always advancing in time. There are not trajectories that have segments that run backwards in time.

Using Equation 19.15 on page 434, we have a different picture of the Fresnel construction. At each time slice multiply the propagators. Before the problem was to add the hands of the little clocks for each possible path. Now we seem to be multiplying. From Equation 19.15, we see that we need something that multiplies and yet adds as you advance through the time slices. This can be reconciled if we understand the exponential function a little better. Remember that $x^a \times x^b = x^{a+b}$. This gives me an excuse to make an excursion into some useful information about the exponential function that everyone should know. It is also a great source of Fermi problems.

19.7.1 A Mathematical Aside – The Population Equation The Exponential Function

Any system that develops or decreases at a rate proportional to the size of the system at any time is an example of a population system. Compound interest is an example. If you have a bank account that pays back 5% interest per year and you just leave the money there but do not add other money, the change in value of the account. in any year, is $\Delta P\ |_{year} = 0.05 P_{year_0}$, and the value at the end of the year is P_{year_0} at the beginning of the year plus $\Delta P\ |_{year}$ so that the principle for the next year is $P_{year_0} + \Delta P\ |_{year} = P_{year_0}(1+0.05)$. After n years, the value of the account is $P_n = P_0(1+0.05)^n$. Now instead of compounding it once in a given year, you compound it α times per year. Then the value after n years is $P_n = P_0(1+\frac{0.05}{\alpha})^{\alpha n} = P_0((1+\frac{0.05}{\alpha})^{\alpha})^n$. The function, $\left(1+\frac{0.05}{\alpha}\right)^{\alpha}$ can be plotted.

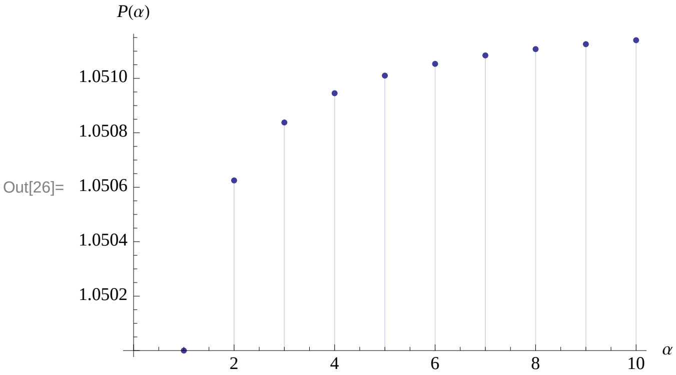

Figure 19.12: **Compound Interest** When interest of 5% per year is compounded frequently, α times per year, the value of an initial investment P_0 after n years is $P_0((1+\frac{0.05}{\alpha})^{\alpha})^n$. As α increases, the function $P(\alpha) \equiv (1+\frac{0.05}{\alpha})^{\alpha}$ quickly rises to the value $e^{0.05} = 1.05127$ and thus for α large the principle is $P_0 e^{0.05\,n}$ after n years.

There appears to be a finite limit at large α. In fact, this is the definition of the exponential function

$$\exp(x) \equiv \lim_{\alpha\to\inf} (1+\frac{x}{\alpha})^{\alpha}. \tag{19.16}$$

You can simply show that $\exp(\beta x) = (\exp(x))^{\beta}$ and $\exp(x+y) = \exp(x) exp(y)$, so that it is convenient to write $\exp(x)$ as e^x. The value of $\exp(x)$ at $x = 1$, e, is the number that is the base of the natural logarithms. Putting all this

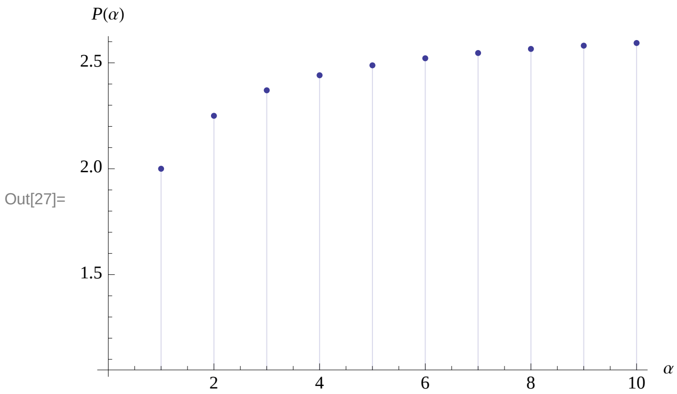

Figure 19.13: **Definition of e** A more dramatic example occurs for the case $P(\alpha) \equiv (1+\frac{1}{\alpha})^{\alpha}$ which is the definition of e.

together, we can say that if we have a population that grows or decays at a rate that is proportional to its current population,

$$\frac{\Delta P}{\Delta t} = kP, \tag{19.17}$$

then the population at some time t is

$$P(t) = P_0 e^{kt}. \tag{19.18}$$

In the language of derivatives, the population equation is

$$\frac{dP}{dt} = kP \tag{19.19}$$

and Equation 19.18 is the solution to that differential equation[16].

For any $k > 0$, the exponential function grows very rapidly, faster than any power of t. This implies that there is no such thing as a small rate of growth for a population that has any positive rate of growth no matter how small. There are an incredible number of applications of the population equation. We will look at some interesting examples.

---

[16] It is worthwhile making this result and Equation 19.18 more mathematical, i. e. dimensionless. P is a number and therefore already dimensionless. t is a time and thus $k \overset{\text{dim}}{=} \frac{1}{T}$. Rescaling t to $\beta \equiv kt$, Equation 19.19 is reduced to $\frac{dP}{d\beta} = P$ and the solution to $P(t) = P_0 e^{\beta}$. It is in this sense that physicists say that the times scale of this system is $\frac{1}{k}$.

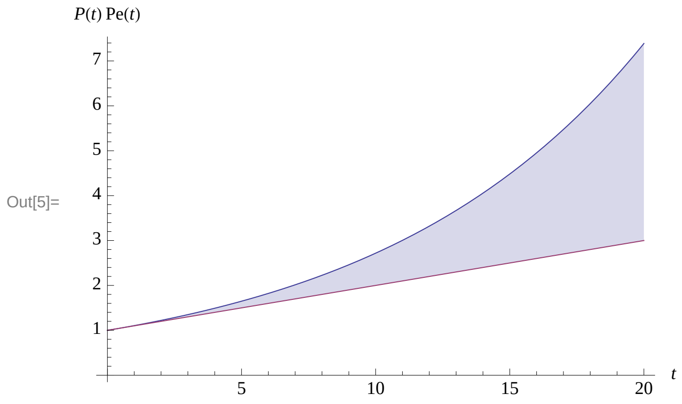

Figure 19.14: **Comparison of Compound Interest and Uniform Growth** A comparison of the rate of growth of a compounded interest and uniform growth. The lower curve is $P(t) = (1{+}0.1{\times}t)P_0$, the uniform or linear growth curve, at 10% per year and the upper curve is the $P_e(t) = P_0 e^{0.1t}$ a growth of 10% compounded instantaneously. You can see here the genesis of the statement that when a growth is very large it is called "exponential." In the figure $P_0 = 1$ for both cases.

In figure 19.14, there is a comparison of a linear growth ar 10% per year, $P(t) = \left(1 + \frac{1}{10}t\right) P_0$ with the exponential growth resulting from instantaneous compounding with the same growth rate of 10% per year, $P(t) = P_0 e^{0.1t}$. The effects of compounding are significant and dramatic.

The best mnemonic for the use of the exponential growth is with the idea of doubling time. There is really nothing special about the value of the natural logarithms and any base can be used. A very convenient base is 2. At a given rate of growth, how long is it before you double the population.

$$P(t_2) \equiv 2P_0 = P_0 e^{kt_2} \Rightarrow t_2 = \frac{\ln(2)}{k} = \frac{0.69}{k} \tag{19.20}$$

If the growth rate is k per year, in $t_2 \equiv \frac{0.69}{k}$ years, the amount will double. For convenience, most people do two things. They express k as a percent instead of a fraction and they round out the 0.69 to 0.70, see 1.4.2. Combining these two things, you get the **Rule of Seventy** or

$$t_2 = \frac{70}{k_{\text{per cent}}}. \tag{19.21}$$

The population equation becomes

$$P(t) = P_0 2^{\frac{t}{t_2}} \tag{19.22}$$

A population example: Some strains of bacteria, if given adequate food, will divide every minute. This is a doubling time of one minute or a rate of 0.69 $\text{minutes}^{-1} \simeq 0.7\ \text{minutes}^{-1}$. If you start the population with one cell at 11 how many cells do you have at noon? $e^{\frac{60}{0.7}} \simeq 6 \times 10^{37}$

If this number of cells fills a bottle, at what time is there half a bottle? – one minute before noon. You also have two bottles at 12:01 and so forth.

A second example: Electrical power has a growth rate of 7% per year. This is a doubling time of 10 years. In 100 years you have 10 doubling times or 2^{10} times the number of power plants in the country if we do not change things. Estimate the number of power plants now and how many there will be in 100 years. Estimate when there will be a power plant for every person. Consider the SUV phenomena. By ratcheting up the gasoline consumption per mile and combined with a population increase, the rate of gas consumption is increasing at about 5% per year. If the oil reserves are 100 years worth of consumption at current consumption rates, how long will that supply last in the face of the SUV phenomena and population growth. Suppose we discovered new reserves to extend that supply to 1000 years. How long will it last in the face of the SUV phenomena and population growth.

A third example: Your parents are giving up about \$10,000 per year to send you to college. For 5 years that is \$50,000. If they had not spent that money, they could have put it into their retirement account. That account earns 7%. If they are in their late 40's and retire in the late 60's, this allows two doubling times. Thus at retirement this is worth \$200,000. If they continue to get 7% per year and expect to live another 15 years this is worth about \$25,000 per year in their golden years.

On the other hand if you have negative growth, $k < 0$, the population disappears asymptotically and fairly rapidly, $P = P_0 e^{-|k|t}$. Instead of doubling times, you now have halfing times, often called the half life of the sample.

$$t_{1/2} = \frac{0.69}{\mid k \mid} \tag{19.23}$$

Using this you can write

$$P(t) = P_0 e^{-\frac{0.69}{t_{1/2}} t} = P_0 2^{-\frac{t}{t_{1/2}}}, \tag{19.24}$$

similar to the relationship for positive rates of growth, $P(t) = P_0 2^{\frac{t}{t_2}}$.

Another important feature of the population equation follows from the defining equation:

$$\frac{dP}{dt} = kP \rightarrow \frac{d(e^x)}{dx} = e^x \tag{19.25}$$

The slope of the exponential function is equal to the exponential function. This is the special case of the more general case,

$$\frac{d(e^u)}{dx} = e^u \frac{du}{dx} \tag{19.26}$$

19.8 New Clocks or Even More on Phasors*

Manipulation of the clocks in doing the analysis of Fresnel's construction is often difficult and unintuitive. Now that we have the exponential function there is a simple way to handle these clocks or better called phasors. Let's summarize the use of the Feynman clocks. At each place on the initial place on a path, we identified a clock as designated by a two dimensional arrow. The clock was then carried over the path of interest to the end place on the trajectory. On each segment of the path, the clock hands were rotated by the time of travel over that segment. At the end of the path, the carried clock was added with the clock hands of all the other clocks from the other contributing paths to the signal at the place. The addition was a 'tip to tail' addition of the two dimensional vectors the clock hands represent.

We can accomplish the same end by representing Feynman's clocks as $A_0 e^{i\theta}$ where A_0 is called the magnitude and is an ordinary number, θ is called the phase, hence the name phasor, and i is the famous, infamous?, number whose square is $i^2 = -1$ usually denoted the unit imaginary number. The analysis of phasors as clock hands is the same as that of the complex numbers in the form $A_0 e^{i\theta}$ except that the conventions are different. For the clock hands, the angle was measured from the vertical and advances in a clockwise direction. For the complex numbers the angle is measured from the horizontal and is positive in the counterclockwise direction.

In the quantum mechanics case, the angle θ is set by the action

$$\Delta\theta = \frac{\Delta S}{\hbar} \tag{19.27}$$

The best way to get A, the amplitude, is $A_0 e^{-i\theta} A_0 e^{i\theta} = A^2$ where the process of replacing i by $-i$ is called conjugating so that this process is to

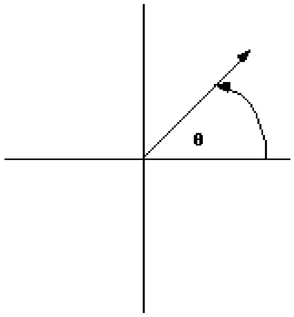

Figure 19.15: **Phasor as a Complex Number** A phasor can be interpreted as a complex number. In this figure the horizontal represents the real part and the vertical the imaginary part of the complex number $A_0e^{i\theta}$ where A_0 is the amplitude or length of the vector and θ is the phase.

take the amplitude, conjugate it, and multiply the conjugate and the original amplitude. That result is the amplitude squared.

$$\psi(x_f, t_f; x_0, t_0) = \sum_{\text{traj.}} e^{\frac{iS}{\hbar}} = \sum_{\text{Path}} e^{\frac{i\sum_{\text{traj.}(x_0,t_0)}^{(x_f,t_f)} L(V,x)\,\Delta t}{\hbar}} \tag{19.28}$$

where a path is designated by some set $\{x_i\}$ of coordinate values at each time slice. All paths are achieved by now allowing each $\{x_i\}$to take on all values.

To be specific lets do the free particle. In between any time slice, we use the straight line path to evaluate the action. This implies that the velocity is a constant and is the inverse slope between the end points of the segment, $v = \frac{x_f - x_0}{t_f - t_0}$, for that interval. Thus,

$$L(v, x) = \frac{m}{2}v^2 \rightarrow S = \sum_{(x_0,t_0)}^{(x_f,t_f)} L(v, x)\,\Delta t = m\frac{\frac{(x_f - x_0)^2}{(t_f - t_0)}}{2} \tag{19.29}$$

Using this for each time slice and setting the interval as ϵ with $n\epsilon = t_f - t_0$ and using the path designation as the set $\{x_i\}$ with each x_i allowed to take on all values:

$$\psi(x_f, t_f; x_0, t_0) = \prod_{i=1}^{n-1} \int_{-\infty}^{\infty} e^{im\frac{(x_i - x_{i-1})^2}{2\epsilon\hbar}} dx_i \Big(2\frac{\pi i\hbar\epsilon}{m}\Big)^{\frac{-n}{2}} \tag{19.30}$$

This is a series of gaussian integrals. Using the following and a great deal of patience you get

$$\psi(x_f,t_f;x_0,t_0) = [2\pi i\hbar\frac{(t_f-t_0)}{m}]^{-\frac{1}{2}}e^{im\frac{(x_f-x_0)^2}{2\hbar(t_f-t_0)}} \tag{19.31}$$

This object is called the propagator. It is like the Fresnel construction, see Sec. 4, in optics tells you how things get from (x_0,t_0) to (x_f,t_f). The product of it and its conjugate is the probability that you will find the particle at (x_f,t_f).

By direct substitution you can show that

$$-\frac{\hbar}{i}\frac{\partial\psi(x_f,t_f;x_0,t_0)}{\partial t_f} = -\frac{\hbar^2}{2m}\frac{\partial^2\psi(x_f,t_f;x_0,t_0)}{\partial x_f^2}. \tag{19.32}$$

This is the free particle Schrödinger equation. For the more general case,

$$\begin{aligned}\psi(x_f,t_f;x_0,t_0) &= \sum_{\text{traj.}} e^{\frac{iS}{\hbar}}\\ &= \sum_{\text{traj.}} e^{\frac{i\int_{(x_0,t_0)}^{(x_f,t_f)}L(V,x)dt}{\hbar}}\\ &= \sum_{\text{traj.}} e^{\frac{i\int_{(x_0,t_0)}^{(x_f,t_f)}\left(m\frac{v^2}{2}-V(x)\right)dt}{\hbar}} \end{aligned}\tag{19.33}$$

and, using the time slicing, you can show that

$$-\frac{\hbar}{i}\frac{\partial\psi(x_f,t_f;x_0,t_0)}{\partial t_f} = \left(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x_f^2}-V(x_f)\right)\psi(x_f,t_f;x_0,t_0). \tag{19.34}$$

This the full Schrödinger equation. A general state is propagated from an initial configuration and, in general, is not at some point but distributed

$$\psi(x,t) = \int\psi(x,t;x_0,t_0)\psi(x_0,t_0)dx_0 \tag{19.35}$$

The probability that the object will be at the place x is

$$\psi^*(x,t)\psi(x,t)dx = P(x,t)dx \tag{19.36}$$

where $P(x,t)$ is the probability that an experiment at time t will find the particle at the position x. Note that, because of the manner in which is it constructed,this state also satisfies the Schrödinger equation for any starting state.

19.9 The Uncertainty Relations

Consider the problem of finding the location of something in a microscope. Light shines on the specimen and enters the microscope lens and is deposited on a photographic plate at a position x_2.

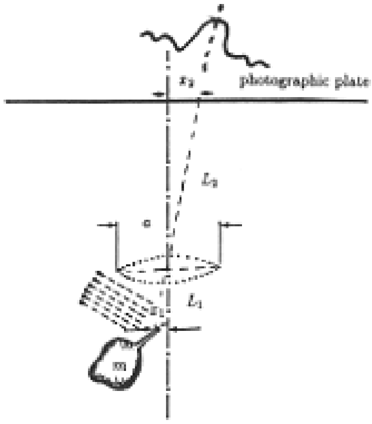

Figure 19.16: **Uncertainty Microscope** The position of a small object is recorded on a photographic plate with the use of a microscope.

Given the wave nature of light, the position of a bright spot behind a lens on the photographic plate is uncertain. There is diffraction. The light is focused by a lens. Since the brightness from the entering lens is spread you can only know the position of the grain of silver by

$$\Delta x_2 \approx \frac{1}{6}\frac{\lambda}{a}L_2 \tag{19.37}$$

Using $\frac{\Delta x_1}{\Delta x_2} \approx \frac{L_1}{L_2}$, we get

$$\Delta x_1 \approx \frac{1}{6}\frac{\lambda}{a}L_1 \tag{19.38}$$

Since the light that hit the electron in the material has a momentum $p = \frac{h}{\lambda}$, and it entered the lens, we know that the electron now has a random momentum

$$\Delta p \approx \frac{h}{\lambda}\frac{a}{2L_1} \tag{19.39}$$

or

$$\Delta x_1 \Delta p \approx \frac{h}{6 \times 2} \approx \frac{\hbar}{2} \tag{19.40}$$

This the famous Heisenberg Uncertainty Relationship.

It a special case of a very general set of relations in quantum systems. Variables are paired in sets that have an incompatibility. You can measure either of them with great precision but if you measure one with a certain dispersion, the other will have a dispersion also and the products of these dispersions are related – you cannot measure both of the variables with precision simultaneously. Two variables such as momentum and position that have this relationship are considered incompatible. Note that momentum and space translation, unimportance of position, are related through Noether's Theorem and symmetry.

19.9.1 The Uncertainty Principle and the Quantum Mechanical Harmonic Oscillator

$$V(x) = \frac{1}{2}\kappa x^2 \tag{19.41}$$

We can use that uncertainty principle, Equation 19.40, to determine the energy of the lowest configuration.

$$\Delta p = \frac{\hbar}{2\Delta x} \tag{19.42}$$

Using $p \approx \Delta p$ and $x \approx \Delta x$ and plugging into the energy relationship for the harmonic oscillator,

$$E = \frac{p^2}{2m} + \frac{1}{2}\kappa x^2 = \frac{1}{2m}(\frac{\hbar}{2\Delta x})^2 + \frac{1}{2}\kappa \Delta x^2 \tag{19.43}$$

To find the Δx that minimizes the energy look at plot of $\frac{1}{\Delta x^2}$, Δx^2, and the sum.

For our problem the minimum occurs at

$$\Delta x = \frac{\sqrt{\hbar}}{\sqrt{2} m^{1/4} \kappa^{1/4}}. \tag{19.44}$$

or defining $\omega \equiv \sqrt{\frac{\kappa}{m}}$,

$$E_0 = \frac{\hbar\omega}{2} \tag{19.45}$$

Even in the lowest energy state, the particle still has energy – no surprise. Thinking in terms of our particle traveling over all paths, even though we are in the lowest energy state, the particle has some spread in position and momentum and thus has some energy.

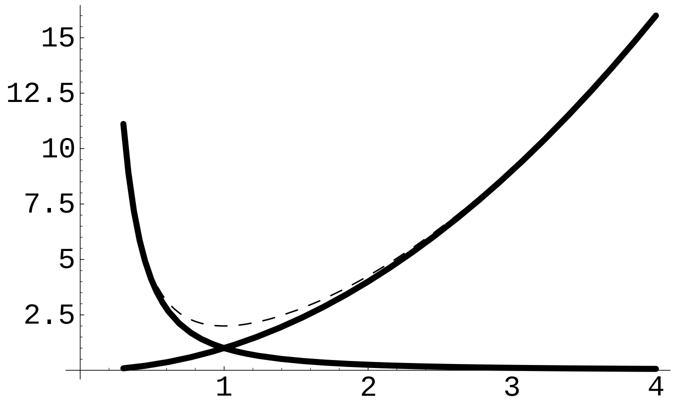

Figure 19.17: **Ground State of Oscillator** A plot of equation 19.43, the energy of the quantum oscillator, in the lowest energy state as a function of the uncertainty in position, Δx. The term that comes from the kinetic energy and goes like $\frac{1}{\Delta x^2}$ is large and dominates for small Δx. Where as, the term from the potential energy that goes as Δx^2 dominates for larger Δx. Thus there is a minimum somewhere between these two domains.

19.9.2 Oscillator Ground State Wave Function

In our interpretation of the Feynman path integration, the field that is associated with the position of the mass at the end of the spring is spread out. Again, this occurs even in the lowest energy state. If we guess that the field, called the wave function, is a gaussian, which it is in fact,

$$\psi(x) = Ne^{-\frac{x^2}{2\sigma}} \tag{19.46}$$

Then $P(x)$ is $\psi^2(x)$ or

$$P(x) = N^2 e^{\frac{-x^2}{\sigma}} \tag{19.47}$$

The condition on N is found from the fact the $\sum_{\Delta x=-\infty}^{\infty} P(x)\Delta x = 1$. This is the statement that the particle must be found somewhere. This satisfies

$$N^2 \int_{-\infty}^{\infty} e^{-\frac{x^2}{\sigma}} dx = N^2\sqrt{\pi\sigma} = 1 \tag{19.48}$$

or the lowest energy state wave function is

$$\psi(x) = \frac{1}{\sqrt[4]{\pi\sigma}} e^{-\frac{x^2}{2\sigma}} \tag{19.49}$$

to get the $\Delta x = \sqrt{\frac{\hbar}{\sqrt{m\kappa}}}$, then $\sigma = \frac{\hbar}{\sqrt{m\kappa}}$.

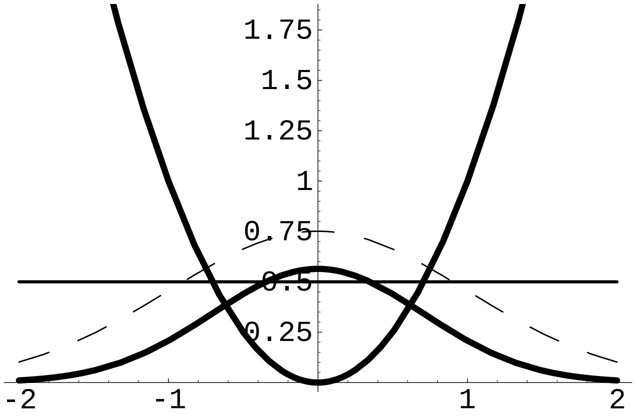

Figure 19.18: **Wave Function and Energy of Ground State** The thick curve that is concave down is a plot of the wave function for the lowest energy state of the quantum oscillator, Equation 19.49. The thick concave up curve is the potential energy of the oscillator. The dashed curve is the probability the the mass in the oscillator will be found at the position x, Equation 19.47. The horizontal line is the energy of the ground state, $\frac{1}{2}\hbar\sqrt{\frac{\kappa}{m}}$.

The thick curve that is concave up is the potential energy, the thin horizontal line is the energy of the lowest energy state on this energy scale. The thick curve that is concave down is the amplitude, ψ, and the dashed curve is the probability distribution.

Notice how the solution 'leaks' into the classically forbidden region. There are places at which the potential is greater than the total energy.

19.10 Localized States

19.10.1 An Aside on the Particle in the Box

Consider a system that has the following potential energy.

$$V(x) = \begin{cases} \infty, & x \leq 0; \\ 0, & 0 < x < L \\ \infty, & L \leq x \end{cases} \tag{19.50}$$

A particle free to move in this potential is the best model of a particle in a box. To be consistent, we require that P(x) be zero everywhere outside the region, $0 < x < L$, the inside of the box. If the particle could be found outside the energy would be infinite. It is in this sense that this is the model for a particle in a box. Since $P(x)$ must be zero outside the box, ψ must be zero and since it is continuous, ψ must be zero at the edges of the box or $\psi(0)$ and $\psi(L)$ must be zero. The simplest function that does that is

$$\psi(x) = \begin{cases} 0, & x \leq 0; \\ N \sin\left(\frac{\pi}{L}x\right), & 0 < x < L \\ 0, & L \leq x \end{cases} \tag{19.51}$$

The probability requirement, $\int_{-\infty}^{\infty} P(x)dx = \int_{-\infty}^{\infty} \psi^2(x)dx = 1$ is that

$$\int_{-\infty}^{\infty} \psi^2(x)dx = \int_0^L N^2 \sin^2\left(\frac{\pi}{L}x\right) dx = 1 \Rightarrow N = \sqrt{2\frac{\pi}{L}} \tag{19.52}$$

In the box, the particle has a wavelength which implies that we know the wavelength, $\lambda 2L$ and thus we know the momentum in the box, $p = \frac{h}{2L}$. This makes the energy of this state

$$E_{low} = \frac{p^2}{2m} = \frac{h^2}{2m4L^2} = \pi^2\frac{\hbar^2}{2mL^2} \tag{19.53}$$

This is the lowest energy state. There are other states that have $\psi(0)$ and $\psi(L)$ equal to zero.

$$\psi_n(x) = N_n \sin{(n\frac{\pi}{L}x)} \tag{19.54}$$

for $n = 1, 2, 3, \cdots$. It turns out that N_n is independent of n and thus is the same as before. In this case, $\lambda = \frac{2L}{n}$ and the $p_n = h\frac{n}{2L}$. The energy is

$$E_n = \pi^2\frac{\hbar^2}{2mL^2}n^2, \;\; n = 1, 2, 3, ... \tag{19.55}$$

Several of the wave functions are shown in Figure 19.19. These states are separable and thus are stationary, the time dependence is trivial canceling out of the probability distribution functions.

The probability distribution functions are the square of the wave function. For these separable or stationary states there is no time dependence. These are shown in Figure 19.20.

The energies are discrete. These are the states that have a fixed energy.

We add energy by adding nodes, places where ψ is equal to zero.

You can make states of almost any energy, none less than the lowest. You can make states where the particle is found at least initially in some region

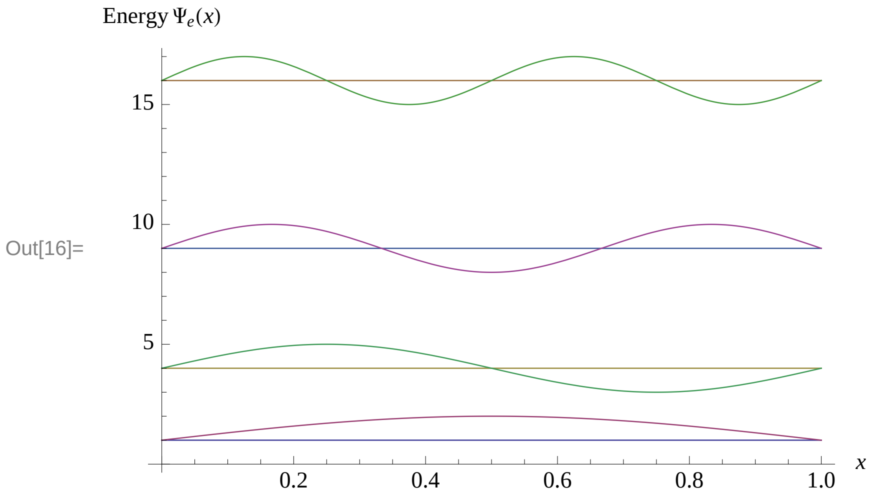

Figure 19.19: **wave Functions for Energy States** A plot of the wave functions for stationary or definite energy states of the particle in the box. The plots are a combination of the fixed energy values, represented by horizontal lines and the wave functions of the states with that energy raised so that each one is plotted on the horizontal energy value associated with that state.

of the box. These are just like the normal modes in the stretched string. You can start the string with an arbitrary pluck. It will be a superposition of the normal modes. Here you superimpose the definite energy states. These definite energy states are called stationary states since they have a given energy they can be interpreted as having a given frequency and they just oscillate like the normal modes did.

19.10.2 Returning to the Oscillator*

Since we now know that the higher energy states are constructed by adding nodes we could guess that there are higher energy definite states and they are finite polynomials times the gaussian that we have discovered above. We also know that at large x, the energy value is negligible and thus all solutions at large x have the same fall off.

$$\psi_n = N_n P_n e^{-\frac{x^2}{2\sigma}} \tag{19.56}$$

The condition that the energy be definite forces the polynomial to take on a very specific form.

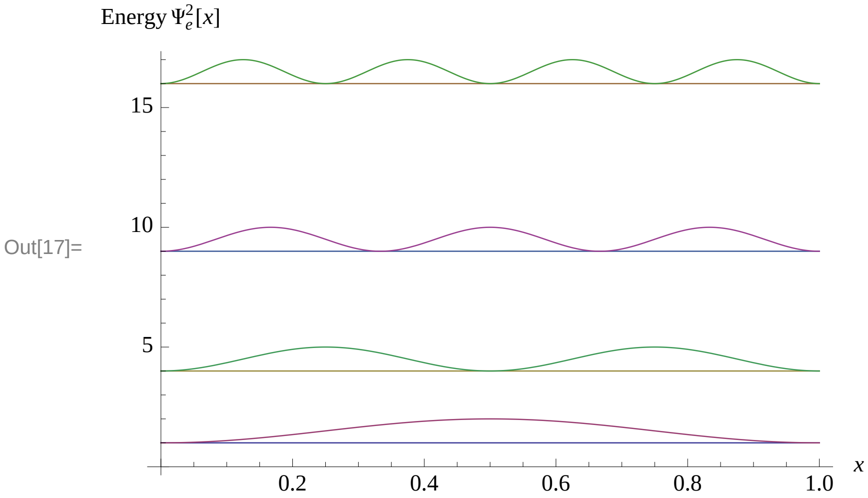

Figure 19.20: **Probability Distributions for Definite Energy States** A plot of the probabilities for finding the particle at the position x for stationary or definite energy states of the particle in the box. The plots are a combination of the fixed energy values, represented by horizontal lines and the probability distributions of the states with that energy raised so that each one is plotted on the horizontal energy value associated with that state. See Figure 19.19 for the wave functions.

We find that the energy is

$$E_n = \frac{1}{2}(2n+1)\hbar\sqrt{\frac{k}{m}} \text{ for } n = 0, 1, 2, 3, \cdots \tag{19.57}$$

We recover our lowest or zero point energy.

We can plot the first few wave functions, ψ_n, for states with definite energy, $E_n = \frac{1}{2}(2n+1)\hbar\sqrt{\frac{k}{m}}$, Figure 19.21.

Here I have raised the height of each ψ so that it is at its energy level and can be seen. Note that all the energies differ by same amount.

Plotting the probabilities, Figure 19.22

The fact that the energy of the state is linear in n cannot be over emphasized. **The energies of the simple harmonic oscillator count.** There is a minimal excitation $\epsilon = \hbar\sqrt{\frac{k}{m}}$ and that the nth state is one with n of these excitations. The unit of energy, $\hbar\sqrt{\frac{k}{m}}$, is the energy of a particle and the state has n particles in it.

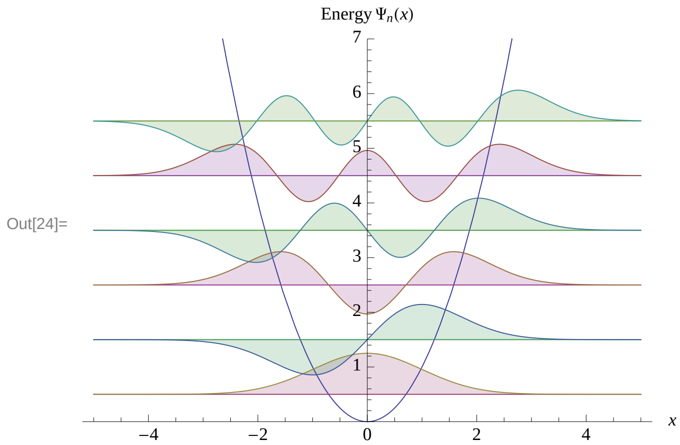

Figure 19.21: **Oscillator Energy States** A plot of the wave function for definite energy states of the quantum oscillator. Also shown is the potential energy and, in order to show them more clearly, the wave functions are raised so that each one's zero is at the corresponding height for that energy.

In addition, there is this magical lowest energy that is not zero energy and is $\frac{\hbar}{2}\sqrt{\frac{\kappa}{m}}$. That the lowest possible energy state has non-zero kinetic and potential energy is a direct reflection of the uncertainty principle, Section 19.9. In order to have zero potential energy, the oscillator mass would have to be located at the origin. But if it localized to just the origin, the uncertainty principle, Equation 19.40, would require that the state have a huge range in momentum and thus a huge kinetic energy. Clearly this is not the lowest energy state. Similarly, the state with no kinetic energy would have a huge position uncertainty and thus a huge potential energy. As we saw in Section 19.9.1, the lowest energy state is achieved with a compromise of spread in position and momentum.

19.11 Time Development of Quantum Systems*

19.11.1 Motion in Quantum Mechanics

For energy states, the time development of the state is especially simple. So far the P(x) that we generate for the energy states are time independent; the ψ are time independent. How do the ψ develop in time? In the stretched string we know that each normal mode just oscillates and that the motion

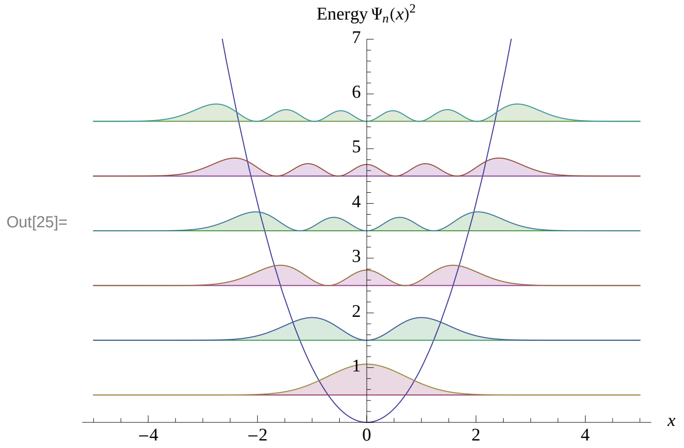

Figure 19.22: **Oscillator Probability Amplitudes** A plot of the probability that you will find the mass at the position x for the quantum oscillator. Also shown is the potential energy and, in order to show them more clearly, the probability curves are raised so that each one's zero is at the corresponding height for that energy.

comes from the fact that the normal modes have different frequencies and that as time advances the amplitudes of the normal modes change.

What we have found above are the states of definite energy. We treat these as normal modes in the sense that, since they have a definite energy, they have a definite frequency, $\epsilon = \hbar\omega$. In the case of quantum systems, we can use the all paths arguments to show that advance in phase by the rule that $\Delta\theta = \frac{\Delta S}{\hbar}$, to show that the states of definite energy develop in time with a phase that advances at the rate of this frequency, $\omega = \frac{\epsilon}{\hbar}$.

For any definite energy state, the $P(x,t)$ is time independent. In order to get any motion, we need to have the system in a superposition of several definite energy states. Then you get $P(x,t)$ in which there is motion.

So that we see that the quantum oscillator changes with time by the interference of the phasors associated with the definite energy states.

$$\psi(x,t) = e^{i\Delta\theta_0(t)}\psi_0(x) + e^{i\Delta\theta_1(t)}\psi_1(x) + \cdots \tag{19.58}$$

where each phasor advances in angle as $\Delta\theta_i = \frac{\epsilon_i}{\hbar}t$.

19.11.2 Relation between the Quantum and the Classical Oscillator

It is interesting to compare the classical oscillator with the quantum mechanical one. In section 19.11.3, we will actually try to find how we can make a quantum oscillator act like a classical oscillator. In this section, we will go the other way. What is the corresponding classical configuration that looks like the quantum case. Remember that in the quantum case, you are dealing with small systems. In fact, it is only in the last few years that experimentalists have been able to manipulate few or single atom systems. Since the quantum mechanics is intrinsically probabilistic, see Chapter **??**, we need to look at a configuration of the classical system that has a probabilistic interpretation. Since in the early days of quantum mechanics, everyone had been trained in only classical systems, there was a tendency to interpret the newly emerging quantum phenomena from a classical perspective, this will allow us to better understand the earlier interpretations of quantum mechanics. Using a classical oscillator with a random start with the same energy as a very excited quantum oscillator, say $\epsilon = \hbar\omega(n + \frac{1}{2})$ with $n = 20$.

The classical probability of seeing the mass at point x is inversely proportional to the time spent in that interval, i. e. $P_{cl}(x) \propto \frac{1}{\text{speed}}$. This should be compared with the quantum case.

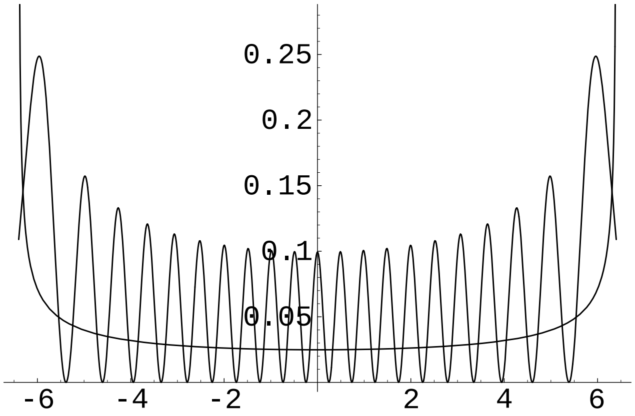

Figure 19.23: **Large Energy Quantum Oscillator** The probability amplitude for quantum oscillator that is in a large n state.

A more interesting question is "What is the state that has the mass pulled to the side and released?". We can construct this state. Like the stretched string, it is a superposition of the definite energy states. Each energy state will evolve in time as its phasor advances and it is the interference of the states that determines how the probability distribution changes.

19.11.3 Classical Motion of the Quantum Oscillator*

How do we recover the classical limit? How do we get something that oscillates back and forth? If we displace the ground state solution a large distance compared to the spread in the ground state wave packet, we should have a solution that moves back and forth like the classical mass and spring. This state should walk, quack, and act like the classical oscillator.

$$\psi(x,0) \equiv \psi_d(x) = \psi_0(x-d) = \frac{1}{\sqrt[4]{\pi\sigma}} e^{-\frac{(x-d)^2}{2\sigma}} \tag{19.59}$$

where $\sigma \equiv \frac{\hbar}{\sqrt{m\kappa}}$.

We want to find the time development of this state. The separable or definite energy states of the oscillator are the states with the simple time development; a separable state with energy E_n has the time evolution $\psi(x,t) = \psi_x(x)\psi_t(t) = \psi_x(x)e^{-iE_nt}$. Expand our stretched ground state as a superposition of the definite energy states.

$$\psi_d(x) = \sum_{n=o}^{\inf} \alpha_n \psi_n(x). \tag{19.60}$$

There is a definite procedure for finding α_n. Because the $\psi(x,t)$ satisfy the Schrödinger equation, you can show that the energy eigenstates satisfy $\int_{-\infty}^{\infty} \psi_{n'}^*(x)\psi_n(x)dx = \delta_{n',n}$ where $\delta_{n',n} \equiv 1$ if $n' = n$ and 0 otherwise. Thus $\alpha_n = \int_{-\infty}^{\infty} \psi_n^*(x)\psi_d(x)dx$. Using equation 19.59 and a table of integrals, we can show that

$$\alpha_n = \frac{(\frac{d}{\sqrt{2\sigma}})}{\sqrt{n!}} e^{\frac{d^2}{4\sigma}} \tag{19.61}$$

The probability of finding the displaced state with n excitations is α_n^2. This is $P_n = \frac{\left(\frac{d^2}{2\sigma}\right)^n}{n!} e^{-\frac{d^2}{2\sigma}}$. This is a well known probability distribution, the Poisson distribution, see Section 19.11.4. The mean of this distribution is $\frac{d^2}{2\sigma}$ and the standard deviation is $\frac{d}{\sqrt{2\sigma}}$.

19.11.4 An Aside on the Binomial & Poisson Distribution

Both the related Binomial and Poisson Distributions are very common distributions and you should know about them independently of their importance in quantum mechanics. These are the distributions that arises when you select a sample from a population. The classic example is the large bag of socks that are half red and half black. What is the chance of getting five red socks in a sample of ten socks? What is the chance of getting four red and six black socks. Although it may not seem to be important when we are dealing with socks and the condition that the population be huge may seem artificial when applied to socks, this distribution is extremely important is many cases. It is a special case of the binomial distribution which applies to sock sampling when you have finite or smaller bag of socks. The Poisson distribution is the limit that you obtain from the binomial distribution when the bag of socks becomes infinitely large. If you have a large population, preferably infinite, and want to draw samples from it, and if you expect to draw N, the probability that you will draw m is

$$P_m(N) = \frac{N^m}{m!} e^{-N}. \tag{19.62}$$

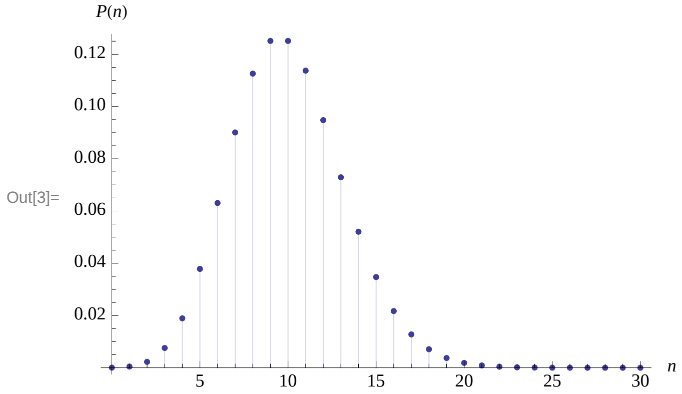

Figure 19.24: **Poisson Distribution with Expected Value of Ten** A plot of the Poisson Distribution for the case when the mean or most likely value is 10. For example in the case of the huge bag containing half red and half black socks, if you sample the bag by removing 20 socks, you expect that most of the time you will draw ten red and ten black socks. This is the distribution of number of red socks that you will get for a sample size that you expect will draw 10 red socks.

A dramatic feature of Figures 19.24, 19.25, and 19.26 is how much the spread of the distribution narrows as the mean gets larger. This is an important property of the Poisson distribution. The mean, N, which is also its

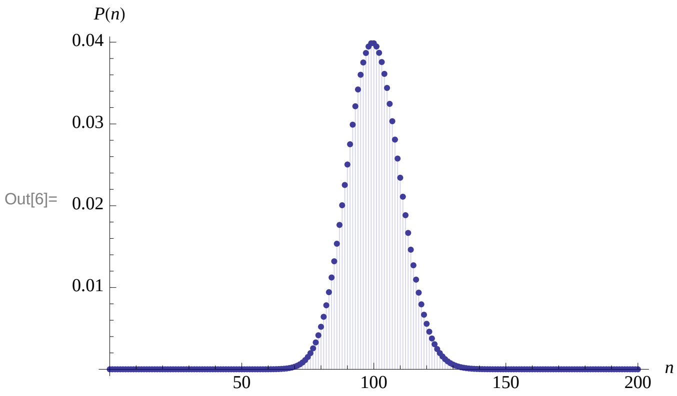

Figure 19.25: **Poisson Distribution with Expected Value of One Hundred** This is the distribution for a sample of a population in which you expect to draw 100 red socks.

peak or most likely value and the width are related. The width of the distribution is the range of values that have a certain likelihood. For instance, σ_{99} is the range of values in a sample of the population that have a 99% likelihood. You pick your comfort range and that sets the size of the σ with which you will live. The range of values set by σ_{99} says that, if your sample value was outside the range $N \pm \frac{1}{2}\sigma_{99}$, your chances of selecting that sample was one less than one in a hundred. σ_{99}, is

$$\sigma_{99} = 2.6 \times \sqrt{\text{mean}}. \tag{19.63}$$

This is often referred to as the square root of N rule. The rule is that in general the width of the distribution is proportional to $\sqrt{N}$ where N is the most likely value or mean for samples selected from that population. You can have weaker criteria for satisfaction than σ_{99}. A useful rule of thumb is to use the usual definition of the width called the standard deviation which holds for about two thirds of the cases. In that case, the rule is simply that

$$\sigma = \sqrt{N}. \tag{19.64}$$

The fact that the distributions, Figures 19.24, 19.25, and 19.26, narrow as N increases is a consequence of the square root of N rule. As N increases the range of values around N that are likely increases but the fraction of values that are likely divided by N gets very small as N gets large.

$$\frac{\sigma}{N} \to \frac{1}{\sqrt{N}} \tag{19.65}$$

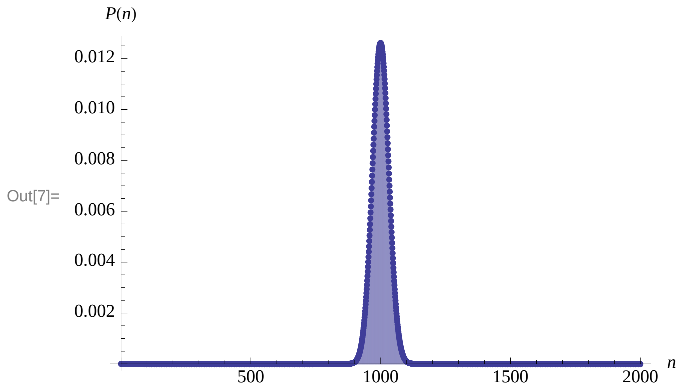

Figure 19.26: **Poisson Distribution with Expected Value of Thousand** This is the distribution for a sample of a population in which you expect to draw 1000 red socks. Note the dramatic narrowing of the width as the expected number of red socks increases. This is result of the $\sqrt{N}$ rule, equation 19.64.

for large N. The fact that there is a range of possible values around the expected value is called statistical fluctuations about the expected value. For these ideal sampling distributions, the square root of N rule means that as the size of the sample grows, the fractional size of the statistical fluctuations shrink. This is the basis of the fact that although there are always sampling errors, as the size of the sample grows, these go to zero. Out of chaos comes certainty.

Let's consider the simple case of political polling. Suppose that there are two candidates and 60,000,000 voters. This is a bag of red and green socks with a million socks, a really big bag. Since the bag is so big, we can use the Poisson distribution and its associated $\sqrt{N}$ rule. Suppose the real preference of the voters is about 60% for candidate A and 40% for candidate B. How we know that before sampling is a interesting problem that will be discussed shortly but let us just blissfully proceed and see what happens. If we take a sample of 100 voters, we would expect around 60 for candidate A and 40 for candidate B. But any set of 100 that we pick is a sample of the real population and we only have a finite chance of getting the most likely mix. Since this is a sampling problem, we realize though that our chances of any result have a distribution whose width is set by the expected value. Suppose we made our 100 calls and we found that there were 57 for candidate A. Do we conclude that 57% of voters like A and 43% like B. That's what the pollsters do. How do they handle the problem of the possibility that they had a non-representative sample? They assume that they did not but then quote a margin of error for the poll. Using the standard deviation as our level of confidence, $\approx \frac{2}{3}$ of the samples will fall within one standard deviation, they have that $\sigma = \sqrt{57} = \sqrt{64-7} \approx 8 - \frac{7}{16} \approx 7.5$ or a % of $\pm\frac{4}{57} \times 100 \approx \pm 8\%$. If they want to use a more certain basis, they will use a two standard deviation error which is a probability of success of 95 %.

Of course, you see the circularity of the argument. They have to assume that their sample value for candidate A divided by the sample size is the correct fraction for candidate A for the population. They then base their error estimate on the sample value because there is nothing else available. This is better than nothing although what they should give is the sample size and the population size. The trouble with this approach is that most people would consider it ridiculous to call a hundred people to determine the preferences in a population of one million. From our square root of N rule, Equation 19.65, we see that they should increase their sample size to reduce the statistical fluctuations. If they went to a thousand people in their sample, they would reduce their fractional uncertainty by $\frac{1}{\sqrt{10}} \approx \frac{1}{\pi}$. If they went to a sample of 10,000, they would reduce the fractional error by a factor of $\frac{1}{10}$. There is a certain point in which it is not worth it to reduce your statistical error below what may be systematic errors in your sampling. In this case of polling, because they are using the phone, they may have a bias in their sample. The people with or at phones may be more likely to prefer one candidate over the other.

The Poisson is the distribution that you get when you look at rare events or background in a large sample. The famous example of the number of deaths due to horse kicks in corps of the German Cavalry recorded in the period 1875 to 1894.

Deaths per year	Number of Corps
0	144
1	91
2	32
3	11
4	2
5 or more	0

There are a total of 280 corps. The average number of deaths in a corps is $\frac{0\times144+1\times91+2\times32+3\times11+4\times2}{280} = 0.7$. Each corps is a sample of the population of all cavalry soldiers and thus the number of deaths should be distributed as

m	$P_m(0.7)$	Observed Fraction
0	0.5	0.51
1	0.35	0.33
2	0.12	0.11
3	0.03	0.04
4	0.005	0.01
5	0.007	0

Three mile island – downwind cancers. down wind population $\approx 25,000 \Rightarrow$ 250 deaths per year $\Rightarrow$ 50 cancer deaths per year. In three years there were 144 deaths. The expected rate was 142. Are there 2 excess deaths per year?

19.11.5 A Return to Classical Motion of the Quantum Oscillator*

Following Section 19.11.1 and particularly Equation 19.58 on Page 450, the time development of this state, 19.59 on Page 452, is

$$\psi_d(x,t) = \sum_{n=o}^{\infty} \frac{(\frac{d}{\sqrt{2\sigma}})^n}{\sqrt{n!}} e^{-\frac{d^2}{4\sigma}} e^{-i\sqrt{\frac{\kappa}{m}}(n+\frac{1}{2})t} \psi_n(x) \tag{19.66}$$

Here the term for the phasor is written out explicitly as $e^{-i\sqrt{\frac{\kappa}{m}}(n+\frac{1}{2})t}$. What is the $\langle x\rangle_d(t)$?

$$\langle x\rangle_d(t) = d\cos\left(\sqrt{\frac{\kappa}{m}}t\right) \tag{19.67}$$

Problem: Show that

On the other hand if you calculate the expected value of x in a state with a definite energy you find that it is zero. **In any state with a definite number of excitations the expected position is 0.**

On the other hand, the expected value of the position squared is related directly to the energy and thus is not zero.

Some examples of superposition

An example of superposition that we will need later on. Up till now we have considered light of only one frequency and superimposed multiple sources and had it interfere. Now consider three sources with almost the same frequency at the same point, say one at some average frequency and one at a slightly lower frequency and one higher by the same amount, $\Delta\omega$. Over time what do you get at the point?

If you look on a short time scale compared to $\frac{1}{\Delta\omega}$:

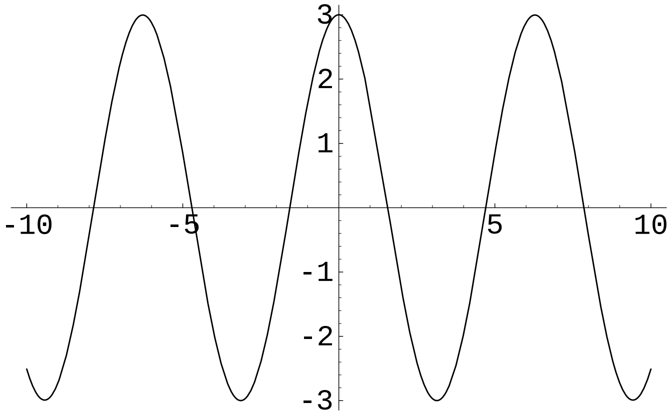

Figure 19.27: **Three Harmonic Amplitudes** The superposition of three harmonic amplitudes with close frequencies looked at on a short time scale. In this case, the three signals have a radian frequency, ω, of $\sec^{-1}$, $1.01\sec^{-1}$, and $0.99\sec^{-1}$. On the time scale of seconds, the resultant is a signal with frequency of about $\sec^{-1}$ and an amplitude that varies on a larger time scale. To view the resultant on a longer time scale, see Figure 19.28

In this case, $\Delta\omega = 10^{-2}$ in inverse units of what ever the time unit is. We are looking over a 20 second time period.

If you look on a large time scale compared to $\frac{1}{\Delta\omega}$.

The signal becomes localized in time. There are periods of large amplitude and periods of small amplitude.

In the double slit that had two spatially separate sources that we superimposed. That lead to a pattern in space for the signal, here we get a pattern in time.

The more pieces you include the better. This case has five.

This seems to be a trivial idea and yet it leads to many of the quandaries of quantum mechanics.

The effect of adding sources is very dramatic. In Figure 19.30, shows the superposition of fifty sources with a total spread in the frequency that is $\frac{\Delta\omega}{\omega} = 10^{-3}$.

Again though, if you look at times short compared to $\frac{1}{\Delta\omega}$ it looks okay.

For times long compared to $\frac{1}{\Delta\omega}$ you get it active and then inactive for a long time. Notice how tightly you can pack the cluster when you have lots

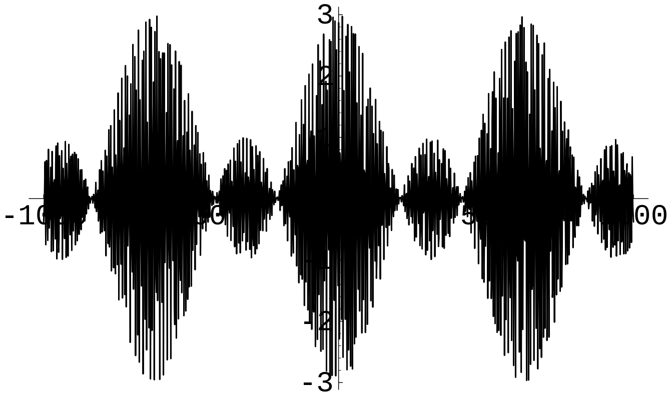

Figure 19.28: **Superposition of Three Harmonic Amplitudes** The superposition of three harmonic amplitudes with close frequencies looked at on a short time scale. The same situation as in Figure 19.27. On the time scale of hundreds of seconds, the resultant is a signal with frequency of about $\sec^{-1}$ and an amplitude that varies on this larger time scale.

of terms.

This is part of a general pattern. If you spread in frequency, you localize in time. You can localize only if you spread in a related variable. In other words we get a tight localization in t if we spread out in ω. If you use a tight value of ω you are spread out in t. The spread in t, Δt, and the spread in frequency, $\Delta\omega$ are related by $\Delta t \Delta\omega \approx 1$.

Since we know that there is a momentum associated with motion in quantum mechanics and that the momentum is proportional to the wavelength, $p = \frac{h}{\lambda}$, we get a similar relation with wavelength and position as with frequency and time.

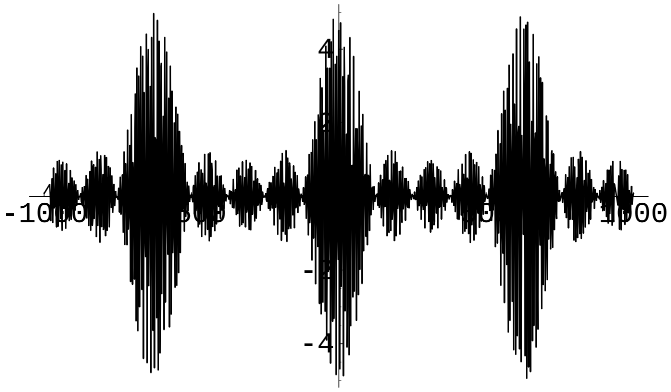

Figure 19.29: **Five Harmonic Amplitudes** The superposition of five harmonic signals with close frequencies on time scales that are long compared with $\frac{1}{\Delta\omega}$.

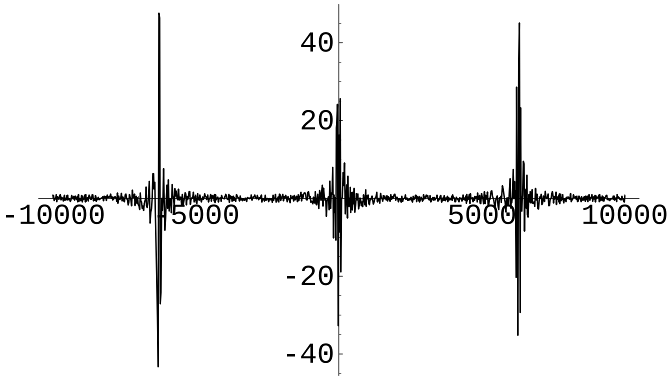

Figure 19.30: **Fifty Harmonic Amplitudes** The superposition of fifty harmonic signals with close frequencies on time scales that are long compared with $\frac{1}{\Delta\omega}$.

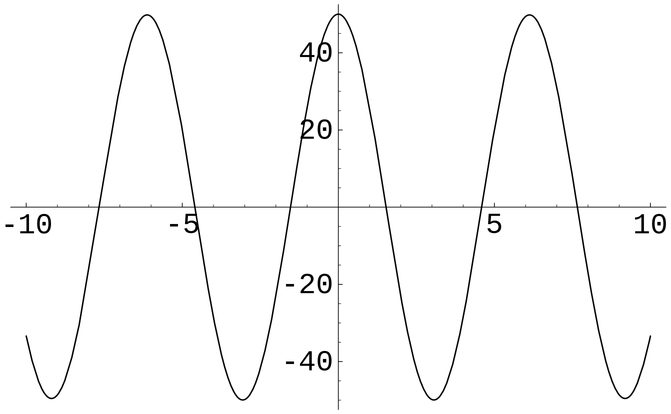

Figure 19.31: **Five Harmonic Amplitudes Revisited** The superposition of five harmonic signals with close frequencies on time scales that are small compared with $\frac{1}{\Delta\omega}$.