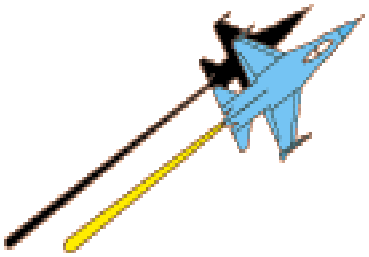
A Solid State Physics Primer

ONE: The Photon



The study of the structure of matter on the atomic level is called **quantum mechanics.** Just as astronomy progressed little beyond natural philosophy until glass lenses were fashioned into telescopes, that allowed a more exact observation of the planets, so too in physics. The tool that lets scientists view the atom is light, which reacts with electrons because it has a particle as well as a wave property. This important quality was discovered by Albert Einstein (1879 - 1955), and is called the **photoelectric effect**. If a light of a known energy content is focused upon a metal surface, an electric current is generated; that is, electrons are freed from their atomic bindings. There is a direct correlation between the energy of the light and the energy of the electrons that are generated.

In order to explain this phenomenon, they needed a model of the atom. It could not be an actual observed system as in celestial mechanics. It had to be created from properties implied by the interaction of the system with light. It had to have certain properties and characteristics that when investigated analytically served to explain physical phenomena. Much as there were many early models for the closest objects to earth, most of which were centered on earth rather than on the sun, there were a number of early models of the atom. As more was learned about the atom, the model was modified accordingly. Unlike in astronomy, which completely discarded early models in favor of the sun-centered system, the early atomic models are still applicable.

The light spectrum

Now that light was proven a valuable way to investigate the nature of the atom, many experiments were conducted. A simple particle of matter was bombarded with light and the resulting electromagnetic emissions from the matter were analyzed. The wavelength of the light returning from hydrogen, the simplest atom, was found to be different from the energy of the light focused on it. Just as Kepler needed volumes of accurate data to reason out orbital dynamics, physicists gathered much experimental data on the emission spectra of the hydrogen atom.

The wavelength of the light emitted from the hydrogen atom was found to follow a repetitive pattern of a similar spacing or lines called the spectrum See Figure 5.1). An interesting relationship was discovered among the lines of each of these series. Three such series were proven.

Lyman Series: $v = cR (1^2 - n^2)$ n = 2,3,4, ... (5.1)

Balmer Series: $v = cR (2^2 - n^2)$ n = 3,4,5, ... (5.2)

Lyman Series: $v = cR (3^2 - n^2)$ n = 4,5,6, ... (5.3)

Where: v = frequency

c = speed of light

R = Rydberg constant

You will notice that these functions are quite similar to the equations for the conic sections, most notably for the hyperbola.

These spectra energies were graphed as a function of energy according to

$E = h\nu$ (5.4)

Where: E = energy

ν = frequency

h = Planck's constant

The result is called an **energy band diagram** and is shown in Figure 5.2. The vertical arrows show the difference in energy between the incoming light and the light leaving the hydrogen atom. It was theorized that the incoming light is absorbed by an electron, causing it to be temporarily excited to a higher energy state. The electron is not able to remain in this unstable state, so it drops to the next stable state, releasing the excess energy in doing so. This energy is released in the form of a photon of light.

Suddenly the phenomenon of the emission spectrum is modeled very nicely in an elegant math and a simple theory. Given the format of the equations and the existence of discrete energy levels it was only natural to turn to celestial mechanics for the underlying physical model of a system capable of producing such affects.

The Bohr atom

Niels Bohr (1885 - 1962) postulated a planetary system model for the atom. The nucleus was at the center, and the single electron had a series of stable orbits available. The system had three important rules:

1. The orbit of each electron is a circle with the nucleus at the focus
2. The electron may shift to an orbit of higher or lower energy. It gains or loses energy equal to the difference in energy levels by the absorption or emission of a photon of energy
3. The angular momentum of the electron in orbit is always equal to an integral multiple of Planck's constant divided by 2pi.

These postulates are consistent with the most fundamental precepts of celestial mechanics. Circular orbits are a special case of elliptical orbits, the angular momentum of an object in stable orbit is constant, and objects can change orbits by the addition or loss of energy. The only difference is that in classical mechanics there is a continuum of allowable energy states, not discrete stable orbits. However, detailed analysis of the three-body problem has recently revealed certain preferred, optimal orbits and behaviors that otherwise might be considered as a continuum of energy states.

The major divergence from accepted theory by the **Bohr model** is the assumption that the orbiting electron does not emit any radiation. Classical electromagnetic theory requires a charge experiencing angular acceleration to generate radiation. This assumption, however, is consistent with celestial mechanics, which does not recognize the long-range affects of electromagnetic charge, only mass. Consequently, it is possible to perform a stability analysis on the electron, similar to that done in celestial mechanics.

For an electron in stable orbit, as indicated in Figure 5.3 there must be a balance between the centripetal force and the electrostatic force between the charges. The latter is governed by an equation exactly analogous to the universal law of gravitation, known as Coulomb's Law.

$$F = k \frac{q_1 q_2}{r^2} \quad (5.5)$$

Where: q's = charge of the two objects

r = distance between the charges

k = Boltsmann constant

This central force is balanced by the centripetal force due to the velocity of the electron in a circular orbit. Equating these two forces,

$q_1 q_2$ mv^2

k r^2 = r (5.6)

These two functions can be used to find the potential and kinetic energy of the electron in two orbits, and solve for the difference in total energy. The result is

v = cR (m^2 - n^2) (5.7)

This is a general expression for the series in Equations 5.1, 5.2, and 5.3. Consequently, the Bohr model is a good model for the experimental data for the hydrogen atom.

The applicability of the Bohr model to the hydrogen atom was good for the overall conceptual perspective, but upon further inspection, it was flawed. First, there was experimental evidence that there were more stable levels that the Bohr model predicted. Second, the model could not be extended to include models of more complicated atoms than hydrogen. *(We shall soon see that even the current model experiences a similar breakdown in the realm of semiconductor physics.)*

From the perspective of orbital mechanics, it is obvious that the Bohr model could be easily extended to cover many additional contingencies. For example, allowing elliptical electron orbits will immediately allow additional energy levels; a continuum of them, in fact. Further investigation into these possibilities was not carried out. Instead, quantum mechanics veered off onto a completely different mathematical route: probability and statistics.

The infinite well

The deterministic course in the investigation of quantum mechanics was due almost exclusively to a simple but very powerful statement, known as the **Heisenberg Uncertainty Principle**. It states that the exact position (x) and momentum of a particle (p_x) can never be known exactly. Any quantitative measurement of these values is bounded by an specific uncertainty such that

(delta-x) (delta-p_x) = h/2pi (5.8)

where h is Planck's constant. This mathematical statement says that if the position is known quite well, the momentum can only be known with great uncertainty. The uncertainty of each entity multiplied together always is a constant value.

With this development, quantum mechanics was forced to accept that the atomic system could never be known with any amount of certainty. So, the leading physicists of the day built a model to take this into account. Erwin Schrodinger (1887 - 1961) devised the model, which is illustrated eloquently in the infinite well problem. It is notable that Schrodinger was very disconcerted at this indeterminate model, and devoted the last forty years of his life to an unsuccessful search for an alternative scheme.

The basis for the potential well problem is the **Schrodinger Wave Equation**.

Y" + cY = 0 with c = (2 pi)/ h^2 (5.9)

Which is a second order differential equation for a simple harmonic oscillator, similar to that used to model the two-body problem. This is the system that, in two dimensions, was shown in part one to exhibit elliptical motion. With one degree of freedom, it is simple harmonic motion.

Applying the boundary conditions for the infinite well problem, the solution to Equation 5.9 is the function

Y = A sin kt with k = sqrt (2 m E) / h (5.10)

The solution requires that k is a multiple of pi/L, where L is the width of the infinite well.

k = sqrt (n pi) / L with n = 1, 2, 3, ... (5.11)

Equating the values for k in equations 5.10 and 5.11 it is possible to derive an expression for the energy of the system.

$n^2\ pi^2\ h^2$

$E_n = 2mL^2$ (5.12)

Notice that all the values are constants except for n, which takes on values of 1, 2, 3, etc... Consequently, the energy E takes on discrete values of energy. In so doing, *energy is quantized.* The integer n is called the **quantum number**.

(NOTE: the body of quantum mechanics is thereafter derived based on the assumption that n takes on only the values of positive integers. The negative integers also satisfy these functions - what happened to the other half of the theory?)

The solution to the Schrodinger Wave Equation requires a solution in three dimensions. The solution is derived in spherical coordinates, and the allowable energy states are quantized, or represented by discrete integer values of the quantum numbers:

N = 1, 2, 3 .... (5.13)

L = 0, 1, 2, ... (N-1)

M = -L, ... -2, -1, 0, 1, 2, ... L

These are the familiar quantum numbers that determine the probability distribution of the different electron orbital shapes (e.g. circular, dumbbell, doughnut, etc.. In addition, there is a spin state, which can be +/- ½. This is used to allow to directions of spin by the electron. The orbital itself is configured using the three quantum numbers N, M, L.

A Continuum

Quantum theory very nicely explains the allowable energy states of atoms, as well as the specific configurations and energy content of the electrons in orbit around the nucleus. These electron states are represented by electron "clouds" or particle distributions, because the exact path of the electron cannot be known. The familiar **Periodic Table** of elements is constructed using these quantum numbers, based on a law that states that no two electrons in a system can have the exact same quantum numbers. This is known as the **Pauli Exclusion Principle**. This system worked very nicely for many years. It led to the discovery of many of the elements, complex compounds, and great advances in physics as well as chemistry. Complications occurred in semiconductor physics, however.

It will be shown that when materials in the class of compounds known as semiconductors combine in a *coherent matrix*, the outermost electrons in all the atoms enter what is called the **conduction band**. A single allowable energy level subdivides into many smaller, but still discrete, energy levels. Macroscopically, all of the electrons occupy the same discrete energy level, as has thus far been described - i.e. the same orbital as characterized by a single set of quantum numbers. On the atomic level, however, they must comply with the Pauli Exclusion Principle: every single electron must occupy a separate, *distinguishable,* **energy state**.

This is not hard to imagine, for dozen or even a hundred atoms in a compound. However, the "coherent matrix" term must apply to blocks of semiconductor compounds on the order of 10^{25} atoms. That is, one atomic energy level must be divided into 10^{25} different, discrete levels of energy. Now, these are electrons we are talking about here. They already are limited to quantum increments of energy equivalent to Planck's Constant, or 6.63 (10^{-34}) Joule-seconds. This is a physical limitation on the smallest allowable level of energy, contingent upon the entire body of quantum mechanics by the conclusions of Einstein's photoelectric effect experiment. Semiconductor physics requires this value to be reduced to 6.63 (10^{-60}) or less, which is an unimaginably small value - even if it were theoretically possible. Yet, this quantity is not only possible, it is necessary.

The current atomic model has no way to configure the atomic orbitals to allow for this practical continuum of variations. It allows one discrete atomic orbit, not 10^{-25} possible variations on this specific orbit. Needless to say, this is a problem. Evidently, the probabilistic representation has been extended to such a level that the uncertainty

constraints have become discrete energy values and exact orbital paths. In other words, semiconductor physics has experimentally found that the Heisenberg uncertainty principle is does not apply to a single, "coherent matrix" of conductive material after all.

Orbital elements

Celestial mechanics had a very similar, unsettling stage in its development. In the very beginning, even the two-body problem was not solvable. The position and velocity vectors had six unknowns, and it was not until **regularization** produced the eccentric and mean anomalies that a solution was developed that included time. The solution included a system of equations with six variables, which are known as the **orbital elements**. They translate the x-, y-, and z-coordinates of the position and velocity vectors into six characteristics of the ellipse and a satellite's motion on that ellipse.

The same situation exists at this stage, in quantum mechanics. Time is not a function of the equations. For example, the capture by an electron of a photon of light energy and its subsequent escape in the photoelectric effect was assumed to occur instantaneously. It was not possible to measure the time it took for this complex event to occur, thus it was not possible to propose and test a model for what occurred in this subatomic transfer of energy. At the other end of the spectrum, there is no time in the Schrodinger probability "cloud" plots. There is only the idea that at any given time, the probability that the electron exists at a certain point is known - with a certain probability.

Now, the ideal solution to the quantum dilemma is a deterministic system: a model which can say where a particle is at any given time, just like we can do in celestial mechanics. That is a very big jump from where we are now. Why not try to get halfway there? Why not develop a system that explains clearly the path of the electron, and not just its probability distribution? If you consider that solutions to the restricted three-body problem have many stable solutions that are shaped just like the Schrodinger distributions, plus the fact that the original quantum theory is based upon the two-body dynamics, it is logical to look in that direction for a deterministic resolution to this situation.

Let us look at the atomic model of the atom. There are three quantum numbers. In the context of orbital mechanics, if you know three variables in three-dimensional space, you know half of the solution; i.e. the entire solution has six variables, for the position and velocity vectors. Therefore, with just these three knowns, you construct a probability distribution of possible locations. To arrive at a complete analytical solution three more unknowns must be solved. Then you will have the six orbital elements needed for the classical dynamic solution.

There is an extra known in quantum mechanics that is not there in celestial dynamics, because of the uncertainty principle. The extra three unknowns lie in equation 5.12. The entire body of theory to this point has been developed based on n being only positive integers. Yet, negative integers are just as good a solution to the function. Presumably, these additional values will lead to a set of three more quantum numbers analogous to those in equation 5.13. Therefore, you will have six orbital elements.

Actually, the total solution will have two more quantities: for the spin direction of the two vectors. One of these must account for the direction of motion around the orbit. The other may actually indicate spin rate of the orbital element.

The Heisenberg uncertainty principle provides yet another unknown, in the relationship between angular momentum and position. This is a very important value to have because it would allow the solution to the subatomic equivalent of the three-body problem with one more degree of freedom. In other words, if the corresponding relationship can be derived in the context of celestial mechanics, then this will allow a solution of the three-body problem in the general case, rather than just the restricted case. Conversely, it may stipulate that only certain orbits within the context of the restricted three-body problem are allowable, or stable over the long term.

TWO: Probabilities

The last chapter paralleled the early progress in celestial mechanics to the development of the fundamental theory of quantum mechanics. The latter body of theory progressed swiftly in the early stages, and generated models that offered an exact explanation for experimental results. Then, just as celestial mechanics made rapid progress in formalizing the two-body problem, only to be stymied by the three-body problem, quantum mechanics hit the wall. They made of this wall the infinite well problem, made of the stuff of probability, and advanced the science from there.

Then semiconductor physics happened. Unusual phenomena were measured in the laboratory, and a model was required. The electrical engineers overwhelmed the applicability of quantum theory because of the multiplicity of forces: practical electrical phenomena transpire in a continuum of forces, not in discrete energy states. They were faced with what for all practical purposes is the n-body problem, without having a clue as to the solution even of the three-body problem. Confronted with very much the same situation, astronomers and aerospace engineers can turn to brute force numerical methods for a reasonable solution within acceptable limits. Electrical engineers had no such luxury, for the simple reason that their theory encompassed no time steps to be iterated!

Negative mass

The fundamental assumption of **solid state physics** is that electrons in a conducting metal are no longer bound in the quantum mechanical way, one electron orbiting its nucleus. At least the outermost, most loosely bound, electrons are not restricted to such a condition. Instead, many free electrons exist in a supra-atomic state, bound to no atom in particular but collectively limited in their movement by the host matrix or lattice.

The three known orbital characteristics from quantum mechanics were condensed into a single entity, a **wave vector**. When an electron is in a stable orbit around a nucleus, its wave vector is balanced by an equivalent wave vector emanating from the nucleus. This is Coulomb's law in terms of vector calculus. It endows the particles with forces and directions, per the attraction of positive and negative charges. This same process was traced in celestial mechanics, whereas the magnitude of the force of gravity was given a direction, toward the opposite mass.

This conglomerate of outermost electrons in the conducting metal was called the **valence band**. No current flowed (i.e. the movement of charged particles did not result in the net displacement of energy) because the sum total of the magnitudes and directions of all the wave vectors in the valence band summed to zero. The total material was therefore in overall equilibrium.

The application of an electric field to this equilibrium state has a net affect upon the wave vectors. Just like a magnetic field will cause little slivers of steel to align in the direction with the field lines, an electric field causes a general realignment of the wave vectors so that the net affect is current flow. Overall, the electrons have more energy because they are in motion, so they are in a different - though still common - energy band, called the **conduction band**.

Both band phenomena are governed by the wave equation from quantum mechanics

$\underline{h^2}$.

$E = 2m^* \times k^2 + E_g$ (6.1)

Where: m^* = effective mass of the electron

k = wave vector

Equation 6.1 is the equation of a parabola. It relates the energy of a particle to the magnitude of its wave vector. Figure 6.1 shows the relationship graphically. The valence band function is shown on the bottom. The maximum of the parabola is the most energy that an electron in the valence band can have. As k increases in magnitude, the energy decreases. There are only two allowable directions for k, as shown, and they always balance each other. Consequently, there is no net direction of motion or force. The conduction band is represented by a parabola pointing in the positive energy direction.

This presentation, which is consistent with the quantum theory, indicates that electrons are able to adopt a continuous range of energy values. However, to be analytically consistent, the electrons in the valence band must have a negative energy mass, according to Equation 6.1. This is a problem, but it does not surface until the theory is successfully applied to a significant range of applications. At this point, it is tantamount to assuming the analysis is of a restricted three-body problem, the k-vector serving as a regularizing substitution. This same tactic was employed successfully in celestial mechanics.

Energy Bands

Thus far, the semiconductor physics adheres closely to quantum mechanics. However, at this point it turns to a different statistical interpretation, one that allows a continuous energy distribution rather than a discrete one. This is done by assuming a characteristic energy distribution of electrons as a function of temperature.

$\underline{. 1 .}$

$f(E) = 1 + \exp [(E - E_F)/kT]$ (6.2)

Where k = Boltsmann's constant

T = absolute temperature

E_F = Fermi level

This function describes the **Fermi-Dirac distribution function**. It gives the probability that a given energy state will be occupied by an electron at the given temperature, T. Substituting E for E_F in equation 6.2, you can see that f(E) will be 1/2. In other words, the **Fermi level** is means that this level has a probability of 50% of being occupied by an electron.

A similar function to equation 6.2 was developed in celestial mechanics. Replacing cosine with the exponential of imaginary numbers in equation 2.15 results in a similar form. This is consistent with the idea that the solid state physics is developing in the frequency domain, versus celestial mechanics; more on this later.

The usefulness of the Fermi distribution is evident when considering materials that have been "doped" with special impurities. The most common substrate material for semiconductors is silicon. It has four valence electrons, i.e. electrons in the outermost shell. If some of the silicon atoms are replaced in the material lattice with atoms having five valence electrons, four of them bind comfortably in the matrix, but the fifth electron is free to move through the material. This is called an **n-type material**. If silicon were doped with atoms having three valence electrons, there is a hole in the matrix - or the equivalent of a negative mass. This is a **p-type material**.

These special types of material can be represented on an energy band diagram, in Figure 6.2. The upper solid line represents the energy of electrons in the conduction band, which are the free fifth electrons of an n-type semiconductor. The lower solid line is the energy level of the holes in the valence band, or the absent electrons in a p-type semiconductor. The

dashed line right in the middle is for an intrinsic material, or one without any dopants.

The range between these lines is occupied by materials having a range of doping concentrations. Silicon with a light concentration of p-type dopants has a fermi level slightly below the intrinsic level (6.3a). Silicon with a very heavy concentration of n-type dopants is very close to the conduction band energy level (6.3b).

Referring to Figure 6.1, the energy of electrons in the valence band is negative. In terms or orbital mechanics, these electrons are in stable elliptical orbits (which have negative energy). On the other hand, the energy of the electrons in the conduction band is positive. These are the equivalent of electrons in hyperbolic orbits (with positive energy), having sufficient energy to escape the host nucleus. The range between these extremes is occupied by electrons in parabolic orbits. They have sufficient energy to escape the host nucleus, but not altogether. This is consistent with the dynamics of central force theory in celestial mechanics.

The actual concentration of dopants is governed by the relationship

$n_o p_o = n_i^2$ (6.3)

where: n_o = concentration of n-type dopants

p_o = concentration of p-type dopants

n_i = intrinsic value for the material

= 1.5 x $(10^{10})/cm^3$ for silicon at room temperature, a constant

These concentrations are the variables in the Fermi-Dirac statistical distribution formula. The fermi level is a function of the dopant and intrinsic concentrations.

$p_o = n_i \times \exp [(E_i - E_F)/kT]$ (6.4)

$n_o = n_i \times \exp [(E_F - E_i)/kT]$ (6.5)

These quantities are called **carrier concentrations**, as they represent the number of carriers - electrons or holes - in a specific type of material. In particular, p_o is the concentration of free holes and n_o is the concentration of free electrons in the semiconductor matrix.

Galactic atom

The next step in the development of semiconductor physics is to model a pn junction, when a p-type material and an n-type material are created together. Some intricate processes occur at the junction of these two very different materials. Before exploring these processes, let us be sure the current level of solid state physics is consistent with our understanding of celestial mechanics.

The major departure from orbital mechanics is the idea of a continuum of possible energies for a parabolic orbit. Parabolic orbits are usually ignored completely, as they represent a very specific discrete energy state, E = 0. Most practical applications of celestial mechanics involve either elliptical orbits, or hyperbolic ones. The only exception are the trajectories of comets. They are all parabolic in shape.

A very similar situation exists in the solid state model. There seem to be many more possible energy states in the range where the energy value is zero, when in the valence or conduction bands. In this sense, the charged carriers in the middle energy bands are all on parabolic trajectories. They are the glue that holds the matrix together, so to speak. There are also many, many more possible trajectories they can follow - between any two nuclei in the entire crystal lattice - versus one nucleus per trajectory for electrons in either the valence or conduction bands.

The analogy of the semiconductor physics to celestial mechanics is now on a much different scale. The equations governing phenomena in the semiconductor would logically apply to a much large scale than even the three-body problem, or even the n-body problem. Now the

solar system is just another atom, with the sun for a nucleus, electrons for planets, comets for electrons in the conduction band, and moons for electrons in the valence band. Let us return to the solid state physics, to see if this analogy holds.

The p-n junction

The juxtaposition of a p-type and an n-type material is governed by the need for the resulting matrix to be in equilibrium. Figure 6.4 shows the two fermi level of the two materials before, and after combination. The fermi level is constant across the junction. The result is a **contact potential** or an energy barrier. The physical significance is that an electron must have extra energy in order to pass from the p-type material into the n-type material.

This potential barrier is not instantaneous. It takes place across a definite region, called the **transition region**. This region is characterized by the absence of any charged carriers - electrons or holes. The electric potential generates enough of a field strength to sweep the region clean. The width of this region is directly proportional to the concentration of dopants on either side of the junction.

There are two ways to perceive this phenomena. When looking at individual objects, or a group of objects. In the discrete analogy, a satellite might be given enough extra velocity to escape earth, only to find that a great deal more energy is required to escape from the solar system itself. This phenomena occurs as the satellite moves through outside of earth's sphere of influence, in a transition to the heliocentric system in which the sun is the most significant influence. The continuum analogy must look at many objects, such as comets, with a wide range of energies and trajectories, to resolve the dynamics of the system. At this scale, it was noted that the affect of Jupiter's orbit is to sweep the solar system free of debris, to keep the transition region free of "carriers."

Solid state physics is based upon a probability distribution, so you expect the model to be a continuum. The mathematics are based upon a diffusion equation. This is a second order differential equation that describes the carrier concentration as a function of distance from the junction, x.

$p(x) = p_o + \text{delta-p} \exp(-x/L_p)$ (6.6)

Where: L_p = diffusion length

This is a simple exponential function, illustrated in Figure 6.5. The process is the diffusion of holes across the p-n junction, into the n-type material. Once the holes enter the n-type material, they are quickly recombined into the material. The average distance they are able to move into the n-type material is called the **diffusion length**. This length is a function of the diffusivity of the material and the carrier lifetime.

The diffusion equation is based on the premise of carrier concentrations, as a continuous process. Applied to a single carrier, the discrete situation must follow the same profile. That is, the more energy a single carrier has the farther it will pass into the n-type material, as an exponential function of the relative distance traveled.

A somewhat analogous situation was seen in orbital mechanics. Once an object achieves escape velocity, it enters a hyperbolic trajectory. If there were no other forces extant upon this isolated two-body system, the trajectory would indeed be exactly hyperbolic - the greater its energy, the greater its divergence from the parabolic minimum. Even looking at only the half of the trajectory succeeding the delta-v, it would be stretching matters to model this transition as an exponential curve. However, before the object reaches the full extent of the hyperbola, it enters a transition and the influence of the sun modulates the trajectory. Overall, then, it is not a bad assumption to assume an exponential trajectory as the initial leg of the patched conic interplanetary transfer.

The BJT Transfer

The next system modeled in solid state physics is the p-n-p junction, which is two p-n junctions back to back. All the same relations hold, except that now there are opposing

"bookends" for exponential distributions, as illustrated in Figure 6.6.

This particular configuration is called a Bipolar Junction Transistor (BJT). The working model of this device has two voltages applied across the two junctions to overcome the contact potentials. The first "emitter" junction is forward biased, and the second "collector" junction is reverse biased (i.e. the second junction is a n-p junction, so its contact potential goes in the other direction). Once these voltages are applied in the proper proportion, a very remarkable phenomena is noted.

Notice in Figure 6.7 that there is a connection to the "base" region of the BJT. A very small current here will correspond to a current leaving the collector junction that is amplified on the order of one hundred times greater. Thus the ability of the transistor to amplify signals.

Returning to the celestial mechanics analogy, compare this transistor junction to the interplanetary transfer. The objective is the most efficient transfer of charged carriers from the emitter to the collector junction. The properties of the materials are, in fact, coordinated in such a way that the transfer is not only efficient but the final product is hundreds of time greater. This is analogous to the phenomenon known as **gravity assist** in celestial mechanics: using the gravitational energy of a planet to boost the speed of a satellite. The Cassini deep space probe, for example, was able to use a close pass by earth (after first entering an orbit around Venus) to increase its energy by a factor of seventy. The same process happens in the transistor.

The applicability of the BJT as an electronic model for the earth-mars interplanetary transfer will be shown by virtue of several tables. First, Table 6.1 shows the reference points between the two systems. In the case of celestial mechanics they are the sun at the focus of the Hohmann transfer, plus earth and mars at the foci of the hyperbolic trajectories. The corresponding reference points are the emitter, base, and collector junctions on the transistor.

Next there are the external forces needed to make the transfer. These are the two delta-v's for the patched conic method, one to escape earth's orbit to enter the heliocentric ellipse, and the other to enter a hyperbolic capture orbit around mars. These correspond to the bias voltages applied to the two junctions of the BJT. Consistent with the patched conic model, there is a forward bias applied at the beginning of the transfer, and a negative one at the end.

Finally, these processes can be modeled by actual electronic components, as shown in Figure 6.8. These items correspond to the internal forces given in Table 4.3.

Now the analogy is complete. Two practical applications of systems designed to optimize a process are matched, both in theory and in practice. The modeling of the interplanetary system using the semiconductor physics is not a method in which the various components are "patched" together to simulate the actual process. Since all of the components are integral to the electronic system, there is a seamless, continuous integration of each discrete portion of the trajectory.

THREE: Field Effects

One of the main themes in the development so far is the idea of a central force. This is a force proportional to the distance squared, and occurs in celestial mechanics as The Universal Law of Gravitation and in quantum mechanics as Coulomb's Law. No such relationship was evidenced in the solid state analogy, however.

The BJT model very nicely models the affect on the satellite of the two planets at either end of the trajectory. The heliocentric, central force is lacking. There is only a base junction, as indicated in Figure 6.8. This significant discrepancy is resolved in a transistor with an integral p-n-p junction, plus an assembly used to take advantage of the "field effect," or what can be thought of as the central force.

The MOSFET Transfer

There is one final problem with the transistor analogy. It is supposed to represent the most efficient transfer of an electron from one orbital to another. So far, the model has been quite close to the patched conic model of celestial mechanics. However, you may recall from Figure 4.5 that the patched conic trajectory was not the best solution. Two additional delta-v's were required in transit, while on the Hohmann Transfer portion of the path. These are very significant course corrections which are not in evidence in the transistor model.

It takes an even more sophisticated transistor to realize these in-flight corrections: the metal oxide field-effect transistor, or MOSFET (Figure 7.1). This device is a modified BJT, with three distinct phases. Each of these phases is described in Figure 7.2. These three phases are related to the actual earth-mars transfer as follows:

Accumulation. Notice how close the satellite remains to earth during just after it escapes. The satellite is being pulled back so that "charge" is accumulated. A delta-v is needed to overcome this prolonged transition phase in the actual orbit. The equivalent in the MOSFET, as in the BJT device, is a bias. In this case it is called the **drain voltage**.

Depletion. This is the heliocentric, two body part of the Hohmann Transfer as the satellite crosses the sun's sphere of influence, W. The affect of earth and mars upon the satellite are relatively balanced. In fact, since both earth and mars are moving in the same direction as the satellite, the affect is more than balanced: they expedite the path. In the convention of the transistor model, the "charge" is depleted.

Inversion. Notice how close the satellite is to mars during the final fourth of the trajectory. The satellite is being pulled toward mars so that "charge" is inverted (the opposite of depleted), and the satellite rushes toward mars more than the patched conic transfer permits. Thus a negative delta-v is required to reduce the speed of the satellite, thus counteracting the sling shot affect. The equivalent in the MOSFET, again, is a reverse bias at this junction, called the **source voltage**.

You will notice that the addition of these two new junctions to the system allow the simulation of the parking orbits around earth and mars. In addition, the diffusion aspect of each junction can be used to model atmospheric drag.

Frequency domain

Before continuing with the analysis, it is interesting to compare the BJT and the MOSFET devices. Superficially, they seem to be a nice model for the interplanetary transfer process in celestial mechanics. The MOSFET differs only in that it includes a central force that projects a strong common influence upon the balance of the processes. Actually, it is a lot more sophisticated than that.

You will notice that *the BJT optomizes the transition process across the base junction*. The junction bias voltages are adjusted properly, and the most efficient transit is achieved. The MOSFET, on the other hand, has these same adjustments but also a gate voltage. The magnitude of this voltage does not optimize the process - it controls it. It is not until the threshold voltage is achieved that the BJT type of continuum is achieved. Below the threshold voltage the process is represented incrementally: a low voltage gets only to the accumulation phase, for example; a little higher and the device is in inversion. That is, at a low voltage the interplanetary transfer is fixed in time and space. The more voltage, the further along the path the satellite is situated.

In the language of signals analysis, what has transpired is the system has been changed so that it is no longer represented in the time domain, but in the frequency domain. The field effect portion of the device has performed a Fourier Transform on the system.

Modeling in the other direction

It would be very helpful if it were possible to manage a coordinate transformation that would project the time-bound system we have been studying in celestial mechanics into the frequency domain. Evidently, we only need to do to the orbital mechanics scheme the same kind of field effect transformation that the MOSFET device does to the p-n-p junction.

As long as we are in the frequency domain, it is instructive to study the frequency response of the MOSFET. It is striking that the majority carriers do not participate in the response at low frequencies; only at high frequencies. This is extremely odd behavior in any physical system. It is not logical for what are evidently the most important objects in a system not to have the greatest affect upon that system. That is exactly what happens in the MOSFET, though, once the threshold voltage has been exceeded. Beyond that, what is the significance of the **saturation voltage**?

OK, so what happens in the orbital mechanics system when the threshold voltage is exceeded. Evidently, the satellite gets enough delta-v to enter a hyperbolic escape trajectory. The focus for this trajectory is presumably mars. Logically, as the voltage continues to increase then the satellite will get enough energy to escape the sun, i.e. the saturation voltage.

Now, in all of these processes the majority carriers are mysteriously absent. (On the scale of things it is reasonable to assume that a satellite taking 200 days to reach mars is low frequency in comparison to an electron taking 10^{-6} to make the same transition in the transistor.) Everything is predicated by the sun only. That means our analysis of motion in the solar system must be independent of the masses of the planets if we are to devise a transformation into the frequency domain. This is a very radical departure from all of the most basic principles of celestial mechanics, with the possible exception of special relativity.

The transformation cannot be done without defining the parameters. Most importantly, what are the equivalent of majority and minority carriers (i.e. electrons and holes) in celestial mechanics: planets and moons. Next, what is it that the MOSFET did with respect to the BJT dimensionally: it limited motion to a plane. Recall the Heisenberg uncertainty criteria, p x h = constant; the only equivalent is to project the interplanetary transfer into a plane at an inclination to the sun, until such time that all the parameters have the given degree of uncertainty. At this point, the two intermediate delta-v's will presumably not be required because the force from the planets will not be in the plane of motion. This is how the MOSFET is able to cause a more efficient transfer of charged carriers, if not faster than the BJT.

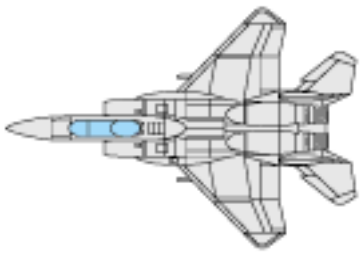
Evidently, it is necessary to enact the same kind of thing with the celestial system: if we

find this plane by projecting the actual motion of the planets, we will have made the transformation into the frequency domain. Fortunately, there is plenty of good data to use to find this frequency plane because the exact motions of all the majority and minority carriers are quite well known, e.g. the planets and their moons. All we must do is neglect the mass of the planets.

A good first estimate is to assume all the moons have negligible mass and to generate a plane about which they all show consistent motion. If such an **invariant plane** can be shown to exist, then the motion of the planets would be as a frequency or energy level. Presumably, each will occupy a specific frequency band and altogether they will form a continuum in the distribution. Thus, the equivalent of discrete energy levels as required by the Pauli Exclusion Principle.

This will in fact been shown to be the case, in the next chapter in the reprint of a paper previously titled, "The Inverse Problem of Celestial Mechanics."

FOUR: Frequency Domain



A least squares fit is used to generate a symmetric coordinate system for the planets in the solar system. The movement of each planet is uniform about this plane, each being in simple harmonic motion. A sinusoidal forcing function is assumed. The resultant linearized system analysis shows two distinct orbital schemes, one for the inner planets and one for the outer. These are interpreted to be the upper and lower portions of a prolate cycloid. Known astronomical phenomena validate the system. The sun-Jupiter Lagrange point, at the Trojan asteroids, links the inner and outer waveforms. The unique earth-moon system is indicated, as are new stable quasi-satellite orbits centered on Earth and Venus, and the location of the "Trojan" orbit of Mars occupied by Eureka. Each body occupies a unique place on the system wave. The entire wavelength is occupied without duplication, indicating that the wavelength selected is ideal. Relative positions are indicative of the axial inclination and orbital ellipticity. Mathematical substantiation is developed by showing the system is a solution of the regularized three body problem. Reasonable assumptions made from the graphical study confirm this mathematical solution.

Regularization

The goal of the analysis is to develop a solution to the regularized three body problem. The first objective in this pursuit is to create a coordinate system in which the data – i.e. the planetary orbits – are presented in a symmetric context. This benefits the analysis later because the function will be regularized with respect to m1 and m2, the large masses, leaving only m3 as the unknown.

The data used to evaluate the positions of the planets in the Solar System were obtained from a published computer program (Ref 4). The program was modified to output data in the coordinates of the celestial sphere. This computer routine used standard values to establish the elliptical and spatial characteristics of the orbits. These were used to generate a single set of data: the positions of each planet at increments of ten degrees of heliocentric longitude, throughout the course of their movements around the sun.

These points for each of the planets were analyzed to determine a least squares fit at each of the ten degree increments of heliocentric longitude, or pi/16 radians. In other words, the least squares fitting process was the generating function to transform the known planetary coordinates into a new coordinate system. The slope and intercept of each line were calculated (see Figures 1 and 2). Both values vary sinusoidalally. A line fixed at one end rotating around 360° while changing in slope sinusoidally defines a flat plane in three-dimensional space.

Further analysis shows that the position of each planet in its revolution around the sun is itself sinusoidal with respect to this new plane, with the movement also being proportional above and below the heliocentric equatorial plane. The plots of the planets are provided in Figures 3 through 11. Notice the inner planets all have the same shape (cosine waves) while the outer planets are all 180 degrees out of phase (with the exception of Pluto). The transition area, at Jupiter, is unstable as evidenced by the irregular curve.

The geometry of the change in the slope line is notable. If a circle of radius 0.2 Au were centered on the sun with a line attached at the outside diameter and the opposite end attached at the Asteroids, the slope and intercept match the calculated values. This indicates a transition in the vicinity of the Asteroids and substantiates the ragged patter on Jupiter, which is near this event horizon.

The objective of the first part of the study was to create a more symmetrical system in which to present the data. This has been done successfully. The plane established will be used in all the subsequent analysis as the main frame of reference.