

# 1

---

## *Introduction*

This book is intended to introduce plasma processing and technology, so that the reader can readily understand the issues involved in processing and can immediately access the state-of-the-art literature. The reader is not assumed to have any knowledge of plasmas, but only some undergraduate background in basic electromagnetism. In fact, wherever possible the treatment even avoids quoting results from one part of the book for use in another part. In many cases it develops new ideas more than once, each time they are needed, to make the discussion easier to follow.

Plasma reactors and the physical and chemical processes that take place in them are discussed in considerable detail, the main emphasis being on capacitive, inductive, and electron cyclotron resonance (ECR) reactors. However, this is an area in which there are few simple (or perhaps even right) answers. It is not usually possible to give a complete description of a plasma reactor, so we are largely concerned, here, with showing how one should go about thinking about what is happening in the reactor.

The chapters of the book are organized in the order in which we need to consider material to understand the reactor. This can sometimes mean that we do not treat a single topic in one chapter separately from other topics. Instead, subject matter will often be introduced where it is needed as we build up the picture of how systems work. The discussion will move from an overview of processing systems based on simple arguments and considerations to more complex analyses as we progress.

Chapter 2 introduces the processes taking place in a plasma reactor, including the materials processing done by the plasma. This chapter describes the different levels at which the plasma may be understood; it touches on electrostatics, transport of particles (including random walks and “continuity” or “conservation” equations), and the role of feedback in the plasma.

The next two chapters are concerned (in Chapter 3) with the simplest plasma/reactor models – circuit models – and (in Chapter 4) more detailed (mostly analytic) plasma models. Inductive, capacitive, and ECR reactors are all introduced and discussed in these chapters.

Chapter 5 begins to handle the specific issues that make processing plasmas particularly difficult to describe, with the topic of plasma chemistry. The chemistry, like some other aspects of these systems, is not well understood – much of the needed data is not available. The strategies used to deal with this uncertainty, and the “recipes” used in choosing plasma chemistries, are outlined.

The next two chapters extend the discussion of chemistry, relating the chemistry and long mean free path transport (since the pressure in many of these reactors is low), in Chapter 6, and the effects of the chemicals and their transport on the features being etched, in Chapter 7.

Chapter 8 returns to the hierarchy of means to describe the plasma, going into more detail than previous chapters. The rest of this introduction, and the final chapter, Chapter 9, are concerned with overall strategies for describing such a complex system and the results that can be expected from our attempts to do so.

Finally, a glossary of technical terms is provided, since much of the difficulty of learning a new field is caused by the unfamiliar expressions used to describe what are sometimes familiar ideas.

The remainder of the introduction is devoted to describing the approach that one might employ to study a physical system that is genuinely difficult to describe completely. It is an assumption that we make, in much of this book, that these plasmas are not capable of being described by means of simple analytic expressions. This may turn out to be untrue. Our approach is suitable for dealing with such a system – but it is also appropriate for finding an underlying simple pattern of behavior, within the complex behavior. In other words, whether the complexity of the system is real or apparent, the overall strategy we are advocating should be useful.

### *Strategies for a Complex System*

This book is concerned with the process by means of which we can build up a useful understanding of a complex physical system. By “complex physical system” we mean a system that is sufficiently difficult to describe that it is not possible in practice to describe the system both completely and accurately. In such a case, there are few right answers to problems. The description of a complex system requires that the system be examined from many directions. Perhaps by starting with simple estimates, to show which effects are important, we can build enough understanding to construct more detailed calculations.

With many calculations, one can win; with few, one cannot. How much less chance of victory has one who makes none at all! By this means I examine the situation and the outcome will be clearly apparent. [1]

Examples related to plasmas, used in fabrication of semiconductors, will illustrate a process where understanding of the plasma is developed, using whichever approaches are appropriate in a given instance.

These are the strategist’s keys to victory. It is not possible to discuss them beforehand. When confronted by the enemy respond to changing circumstances and devise expedients. How can these be discussed beforehand?

More important than the details of the way the estimates are done is the recognition that, by developing physical insight into the particular situation, a set of useful estimates and calculations can be arrived at, which will vary from one example to the next.

Of the five elements, none is always predominant; of the four seasons, none lasts forever.

The purpose of these calculations is, ultimately, to be able to design a near-optimal system or process.

The overall strategy, used in designing the optimum process, must be developed with the widest possible view of the purpose of the design.

Know the enemy and know yourself; in a hundred battles you will never be in peril.

He whose generals are able and not interfered with by the sovereign will be victorious. The general should act expediently in accordance with what is advantageous and so control the balance.

Because the system we are focusing on is too complex to allow a complete description, there will be few neat, simple conclusions in what follows. We will not give the impression that we can divide the problem into separate topics that can be handled completely and separately, each in its own chapter. Instead, the chapters represent stages in the process of going from very simple estimates to complex calculations.

Before we can set up many different types of estimates of the behavior of our physical system, we need to have some idea of how to go about doing approximate calculations using the equations of applied physics. This is the topic of the next two sections. The last section of this chapter addresses the way in which such models are tested against experiments, in order to develop better models.

### *Problem Solving in Applied Physics*

It is important to realize that problem solving is a skill in itself. Most people working in applied physics and engineering have more than enough knowledge to understand the material in this book. Using this knowledge to describe experiments needs something more than this.

The hardest part of solving many problems is the identification of the problem. Experiments raise many questions, and attempting to answer them is one of the main ways in which progress in science is made. Typically, the explanation of experimental observations in plasma processing of materials requires knowledge of the relevant physics, chemistry, and so on, and insight into which of the potentially relevant effects are actually important. For now, we will assume that these important effects have been identified. The task is then to describe them and their interaction.

The description of physical phenomena is usually done in one of two ways: by analogy (for instance, when we think of ions being like raindrops falling out of the plasma – see Section 2.1) or by means of mathematical equations. The challenge with analogies is to know when they break down. Equations could probably be said to embody a sort of analogy as well – and again, we need to know when they break down. The ability to formulate a description of either kind is very useful, and coming up with a good analogy can be most helpful in setting up an appropriate equation. The word “equation” means “setting equal,” so when we write an equation we are describing two or more quantities, which are combined in the equation so that there is some kind of balance between them. Often, our task is to figure out which effects are “in balance,” so we can write the equation that describes their balance. Once we have set up the equation(s), we will almost always be forced to solve them approximately. Exact solutions are rarely possible in practice.

One of the most frequently used equations in what follows is the continuity (or conservation) equation for the particle number density  $n$ :

$$\frac{\partial n}{\partial t} + \nabla \cdot \mathbf{\Gamma} = S. \quad (1.1)$$

This equation provides an excellent example of some issues which arise in the approximate solution of equations in applied physics. In steady state,  $\nabla \cdot \mathbf{\Gamma} = S$  implies that  $S$ , which is the rate of production of particles per unit volume per second, is equal to  $\nabla \cdot \mathbf{\Gamma}$ , which is defined to be the rate they are lost from the surface of a unit volume per second. ( $\mathbf{\Gamma}$  is the flux per unit area per second.) In the time-dependent version, given above, the rate of production  $S$  is balanced by the sum of two effects: the loss from the surface and the rate of increase in the density in the (unit) volume,  $\frac{\partial n}{\partial t}$ .

It is necessary to know when to use this equation, and what to use it to calculate. The short answers are: The equation is quite general (unless you can think of a term in the particle balance that has been left out) but it is only really useful if we know (expressions or values for)  $S$  and  $\mathbf{\Gamma}$ . It is normally used to find a particle number density,  $n$ .

These answers raise more questions: Why do we find  $n$  from it, not, say,  $S$ , and when do we know (expressions or profiles for)  $\mathbf{\Gamma}$  and  $S$ ?

The first part of the answer is that  $S$  is a “source” in the equation, which describes the production of density  $n$ . It is usual in such a case to assume we shall find  $n$ , from the equation and from knowledge of  $S$ . Other examples of equations with a source are Maxwell’s equations, two of which read:

$$\nabla \cdot \mathbf{D} = \rho, \quad (1.2)$$

that is, the free charge density  $\rho$  “produces” the field  $\mathbf{D}$ , and

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}, \quad (1.3)$$

that is, the current density  $\mathbf{J}$  “produces” the “rotation” of the field  $\mathbf{H}$ . In some practical examples it seems reasonable that we specify the source first. In an electromagnet, we “impose” a current density  $\mathbf{J}$ , and we calculate the magnetic field  $\mathbf{H}$  it produces. If we create particles by some mechanism, in doing so we impose a rate of production  $S$ , and from this the density  $n$  must be deduced.

Reality is more involved than this, because the sources we impose are affected by the response of the system. Although this complicates the calculation, it does not change the fundamental relationship between the source and the response. (See Ref. [2].)

The conservation equation for the density is expressed in terms of the flux  $\mathbf{\Gamma}$  and the source rate  $S$ . The quantities  $\mathbf{\Gamma}$  and  $S$  can be found from the distribution function  $f$ , or from an infinite set of equations in terms of other “moments” of the distribution [3]. Moments of the distribution are integrals over velocity of  $f$  multiplied by integer powers of the velocity  $\mathbf{v}$ .

When the mean free path  $\lambda$  is small compared to other dimensions, we may be able to use approximate expressions for  $\mathbf{\Gamma}$  and  $S$ , but these tend to be problematic, even

then. If the mean free path is not small, we need to do a more accurate calculation based on a kinetic equation. The kinetic equation is a conservation equation, in many ways very like the equation for conservation of density. It uses a density per unit of “phase space” instead of a density in physical space. Phase space is a set of independent variables that combines physical space and velocity variables. Just like the more usual conservation equation, it has a source term, which is the rate of production of particles per second per unit volume of phase space. The density in phase space is the quantity we usually attempt to calculate when we use the kinetic equation, which is also what we should expect by analogy with the standard particle conservation equation.

### *Sensitivity to the Order of Calculation*

The fact that the calculations we perform are approximate has an important consequence for the order in which we do calculations. If we know one quantity  $A$ , which appears in an equation, approximately, we may be able to find another quantity  $B$  in the equation, from our knowledge of  $A$ , with reasonably good accuracy. However, if we know  $B$  approximately we may not be able to find  $A$  from it at all accurately. The diffusion equation (1.1) illustrates this point well. In steady state, and if  $\Gamma = -D\nabla n$ , the equation is  $-D\nabla^2 n = S$ .

The particle number density  $n$  does not depend very sensitively on the ionization rate  $S$ . This means that if we know  $S$  approximately, we can find  $n$  from it with some confidence. However, an approximate knowledge of  $n$  would not be sufficient to allow us to find  $S$  with any confidence. One way to see this is to examine the analytic solutions of the diffusion equation given elsewhere in the book. The density is seen to tend to “fill-in” the center of a region, even when the ionization peaks far away. The density seems somewhat insensitive to the shape of  $S$ , although (for instance) doubling  $S$  in a linear problem will clearly also double  $n$ . Another way to look at this is to suppose that  $S$  had been measured and had errors of (say) 10%. This means that  $\nabla^2 n$ , the curvature of  $n$  has errors of 10%, but (in a plasma where  $S > 0$ ) it is probably still negative everywhere. The density  $n$  itself when calculated may not be badly in error, because of a 10% error in its curvature. However, a 10% error in  $n$ , which could change from being +10% to -10% from one point in space to its neighbor, could make both  $\nabla n$  and  $\nabla^2 n$  completely wrong – and could for instance easily make them both have the wrong sign. If we find  $S$  from  $n$ , by setting  $S = -D\nabla^2 n$ , then  $S$  will similarly be quite wrong.

As a consequence of this, we should set up calculations where we iterate to find the profiles of  $n$  and  $S$ , so that we use the diffusion equation to find  $n$  from  $S$ . We must use another, direct calculation to find  $S$  – that is, we must use a method directly related to the way in which the ionization takes place. This might involve calculating the heating rate in the plasma and then determining how much of the power deposited is used for ionization, for instance.

In summary, the (partial differential) equations we use often have a “source” and a “response.” We usually should specify the source, and calculate the response from it. This is physically reasonable; it is also the case that the response is less sensitive to the exact details of the source, than the source would be to the details of the response, if we tried to use the equation the “wrong way round.” This is

important, given that our knowledge of any variable is usually approximate at best. Mathematically, the source usually appears directly in the equation, whereas the response is usually differentiated one or more times with respect to space and/or time. The differentiation of the response makes the equation very sensitive to the details of how the response varies in space or time.

### *Design of Calculations*

In a text on statistical design of experiments [4], some guidelines are given that, with very slight modifications, are extremely useful in designing the calculations we need to do.

We expect to perform a series of calculations, where we speculate about the nature of the process we are studying in order to describe the process. We compare the results of our calculations to experiments, to see whether we successfully included the important aspects of the process. The results of the comparison should lead to improvements in our speculations. It may also suggest new experiments. The cycle of comparison and improvement is expected to be iterated multiple times.

Understanding the nature of the problem and the variables involved requires careful consideration and the use of all the available information about the process. This aspect of the procedure resembles Sun Tzu's approach to a military campaign, where the importance of: 1) recognizing what is really important, 2) seeking out information, and 3) making estimates of various kinds, are stressed.

Montgomery [4] summarizes the stages of designing a statistical experiment, in a fashion we shall rewrite to make it apply to our situation, as:

1. Use your knowledge of the problem to determine which effects are likely to be important.
2. Keep your analysis as simple as possible. Including the important effects is what is important, rather than using an enormously complex model.
3. Recognize what is actually significant in practice. Do not get bogged down in minor details.
4. Iterate. We usually learn what is most important as we go along. Often one should not invest more than 25% of one's effort in the initial calculations, which are frequently just learning experiences.

These points should be illustrated with examples. Taking them in order, we have:

1. Many people put a great deal of effort into the wrong aspect of a model, and they fail to recognize that other aspects of the problem are more critical. Most plasmas are controlled to a great extent by surface processes. There is little point having an exact plasma model, if we fail to properly consider how the surface processes control the plasma.
2. Complex plasma models are often set up, with a "module" to describe each aspect, but only some of the modules are critical. It is more important to get the critical aspects right, than to have a module for every conceivable process.
3. One of the common mistakes, made most frequently, but by no means only by inexperienced researchers, is to find a very small error in an existing calculation

and put a great deal of effort into correcting it. Suppose we can do calculations, or make measurements, to an accuracy of 5%. Suppose further that we find errors in our calculations, which can be corrected with a certain amount of effort. Suppose there are three errors, as follows:

	Size of Error	Time to Correct Error
Error 1	50%	1 month
Error 2	5%	1 hour
Error 3	5%	6 months

A surprising number of people will simply decide to fix all three errors, and do them in no particular order, when in fact error number three should certainly be tackled last, if at all.

- Frequently, we do not know which effects to include. In one situation, it took many years of estimating the magnitude of different heating processes to discover that Coulomb collisions were effective in heating “cool” electrons in a dc discharge. For this reason, we need to iterate – we could not have set a correct calculation up at the outset. It took many iterations to rule out some processes and hence to deduce which process was missing.

### Exercise

- What are the major considerations in choosing a plasma reactor for etching silicon-based wafers? Try to be realistic, for a) a new industrial plant and b) a small-scale industrial or university project. What typically determines the diagnostics and modeling that are employed? Analyze the cost of adding various diagnostics, and employing models, to optimize the process. Discuss whether it is economically justified to increase the use of diagnostics and modeling.

## 2

---

### *Plasma Processes for Semiconductors*

The plasmas used in processing semiconductors are usually partially ionized gases, the neutral gas pressure being from about 1 mTorr to about 100 mTorr and the plasma (that is, the charged particle) number density  $n$  being in the range  $10^{10} \text{ cm}^{-3} \lesssim n \lesssim 10^{12} \text{ cm}^{-3}$ . To describe these plasmas precisely and in detail is difficult for a number of reasons. Plasma physics textbooks are largely devoted to detailed analysis of the special cases that can be described by analytic theories. Instead, it is often more useful to try to obtain approximate descriptions of the real plasma using simple physical reasoning.

In trying to understand a complex physical system it is usually useful to try several different approaches, such as studying different physical pictures, and to attempt to reconcile them with available data. In this book we will attempt to describe each situation we consider starting from “first principles” in each case. A major theme throughout will be the development of approximate quantitative models to explain experiments, based on simple reasoning. Statistical design and analysis of experiments, for example [5], provides a valuable tool for building up physical understanding.

It is tempting to say that theoretical models of processing plasmas are like houses made of cards. After a few layers (of assumptions) the whole edifice is in danger of collapsing. A slightly better metaphor might be a house of straw, where each extra layer (assumption) we add bends the structure so that after a very few layers it is hopelessly distorted.

For this reason it is important to avoid taking a simple analytic model and pursuing it as far as we possibly can. Instead, we need to be aware of the analytic models that might be relevant and that might explain experimental data, and we need to constantly check them for realism when we try to use them. Space will not permit this testing process to be illustrated fully in what follows, but it is important to keep in mind the necessity for vigilance in this regard. Typical assumptions we will try to avoid, but which are commonplace in some relevant literature, are that inelastic rates of reaction can be found assuming a Maxwellian electron distribution or that the transport of particles (neutral or charged) at long mean free path can be described by “fluid” equations.



## 2.1 Plasmas

The plasmas we are interested in are a particular subset of all the possible types of plasma: Processing plasmas are partially ionized gases. In general, the definition of a plasma could be given as: A plasma is a collection of charged particles where the long-range electromagnetic fields set up collectively by the charged particles have an important effect on the particles' behavior. In our case, the fields the plasma sets up will be mostly electric fields. Further, we are concerned with plasmas confined in "tanks," and the electric fields are usually strongest where the plasma is in contact with the surface of the tank. This electric field is created because electrons in the plasma tend to move much faster than ions. The fast-moving electrons hit the wall before the ions do and some stick to it, so the wall charges up negatively and this negative charge pushes other electrons away at the same time as attracting positive ions (see Fig. 2.1). In steady state the wall no longer charges up, and thus electrons and positive ions have to arrive at the same rate. The field near the wall holds the electrons away from the wall, allowing only the most energetic electrons to get there. The field also accelerates the positive ions toward the wall, and in this way the rates of arrival of electrons and positive ions are made equal.

The electrons on the wall push the other electrons back from the wall, exposing the positive charge in a region close to the wall. This positively charged region is called a "sheath." The positive charges in the sheath are the starting points for electric field lines. These field lines end on electrons on the wall. The electric fields in the positively charged sheath region are very strong. The interior of the plasma is usually nearly neutral and the electric fields in the plasma interior are much weaker. The strong electric field in the sheath accelerates the ions in a direction straight toward the surface. The high-energy ions moving nearly normal to the surface are a vital aspect of the plasma processing of materials.

The positive ions are produced somewhere inside the plasma (as are most of the electrons) and fall outward (see Fig 2.2). Like raindrops steadily falling from a cloud, the number of ions or raindrops falling through an area per second (in steady state) is the total number created each second above that area. The speed of the ions, or drops, depends on how far they have fallen, and on whether they hit anything while they were falling. The density of ions, or drops, will be greater if the number passing per second is high and if they move slowly.

In this analogy the cloud corresponds to the source of ionization. In the plasma, gravity is not usually important. It is the electric field that tends to pull ions out. The field usually points outward from the "center" of the plasma where the field is zero. The ions passing through a plane are those produced between the center where the field passes through zero and that plane. The ion density and the number crossing an area per second depend on where the ionization is, on where the electric field goes to zero, and on how the field varies in space.

Another metaphor for the ion motion, in contrast with the electron motion, is that the ions are as slow as molasses running downhill. The plasma potential appears to the positive ions like a hill, with the peak of the hill being somewhere in the middle of the plasma volume. The molasses (ions) is being poured onto the hill around the peak and moves very slowly outward.

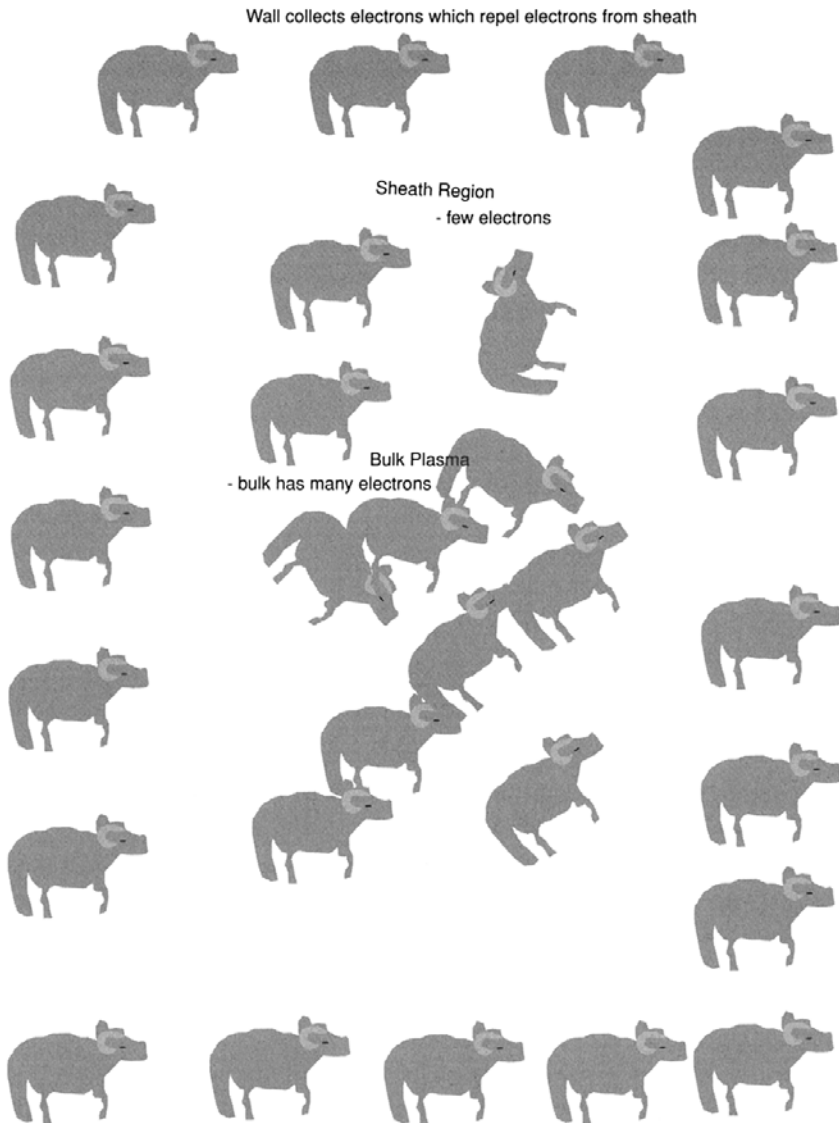


Fig. 2.1. A flock of sheep as a metaphor for charging of the walls by electrons. The fast-moving electrons hit the wall and stick to it, providing a negative wall-charge. The negative charge initially builds up faster than positive ions can get to the wall to neutralize it. In steady state the negative wall-charge repels other electrons from the region close to the wall (called the sheath) and attracts ions, so the fluxes of electrons and ions can be equal.

The electrons behave somewhat like sheep in a croft. See Figs. 2.1 and 2.3. (A croft is a small field; it might be confusing to talk about the sheep being in a field.) The croft slopes a little upwards from the point of view of the sheep (electrons) as they move out from the middle. There is a very high fence (the sheath) around the croft. Very few sheep can leap over the fence. Some sheep are so weak that they are trapped in the center of the croft by the slope, but many of them are energetic enough