3.10 Water diffusing through an elastomer O-ring seal from atmosphere at a rate of 0.01 sccm is the principal background gas in a particular vacuum chamber being pumped at 1000 *l*/s. What is the partial pressure, in Pa, of water in the chamber?

3.11 An ion gauge which indicates pressures p_i is being calibrated against a capacitance diaphragm gauge which indicates pressures p_c and which has a zero offset of p_e. The following three data points (p_i, p_c) are taken, in units of Pa: (0.12, 0.04), (0.10, 0.03), and (0.06, 0.01). (a) What is p_e? (b) What is the ion gauge calibration factor $f = p_c/p_i$?

3.12 List the factors affecting the level of water background in a vacuum chamber.

3.8 References

1. Ishimaru, H. 1990. "Developments and Applications for All-Aluminum Alloy Vacuum Systems." *Mater. Research Soc. Bull.* July:23.
2. Strasser, G., H.P. Bader, and M.E. Bader. 1990. "Reduction of Particle Contamination by Controlled Venting and Pumping of Vacuum Loadlocks." *J. Vac. Sci. Technol.* A8:4092.
3. Siefering, K., and H. Berger. 1992. "Improved APIMS Methods." *J. Electrochem. Soc.* 139:1442.
4. Poulter, K.F., M.-J. Rodgers, P.J. Nash, T.J. Thompson, and M.P. Perkin. 1983. "Thermal Transpiration Correction in Capacitance Manometers." *Vacuum* 33:311.

3.9 Recommended Readings

Fraser, D.B. (ed.). 1980. *Pumping Hazardous Gases*. New York: American Vacuum Society.

Hablanian, M.H. 1990. *High Vacuum Technology, a Practical Guide*. New York: Marcel Dekker.

O'Hanlon, J.F. 1989. *User's Guide to Vacuum Technology*, 2nd ed. New York: John Wiley & Sons.

Santeler, D.J. 1990. Topics in Vacuum System Gas Flow Applications. *J. Vac. Sci. Technol.* A8:2782.

Tompkins, H.G. 1991. *Pumps Used in Vacuum Technology*. New York: American Vacuum Society.

Chapter

4

Evaporation

In the previous two chapters, we examined the foundations of vapor deposition in gas kinetics and vacuum technology. Now we can begin to examine specific deposition processes. In this chapter, we will address the thermal evaporation of source material and its transport to the substrate within a high-vacuum environment. This is the basic physical vapor deposition (PVD) process. In the next two chapters, we will look at the deposition behavior of the vapor once it arrives at the substrate. In subsequent chapters, we will discuss other ways of supplying vapor, including the use of chemical vapors (CVD) and nonthermal methods of vaporization: that is, energy beams and sputtering.

4.1 Thermodynamics of Evaporation

We begin the discussion of evaporation by reviewing some basic principles of thermodynamics and chemical kinetics which will be useful in many aspects of thin-film work. Evaporation is a classic illustration of these principles, so it provides a convenient context within which to develop them. Our present objective is to predict the evaporation rate of a material from available data such as the boiling point and heat of evaporation. Readers familiar with the concept of entropy as disorder, with the significance of the Gibbs free energy and the chemical potential in determining equilibrium, and with the origins of the Clausius-Clapeyron equation may wish to skip this section.

The vapor-liquid or vapor-solid equilibrium situation is shown in Fig. 4.1*a*. Here, we are considering either a single element or a pure compound which does not dissociate upon evaporation. Alloys, mix-

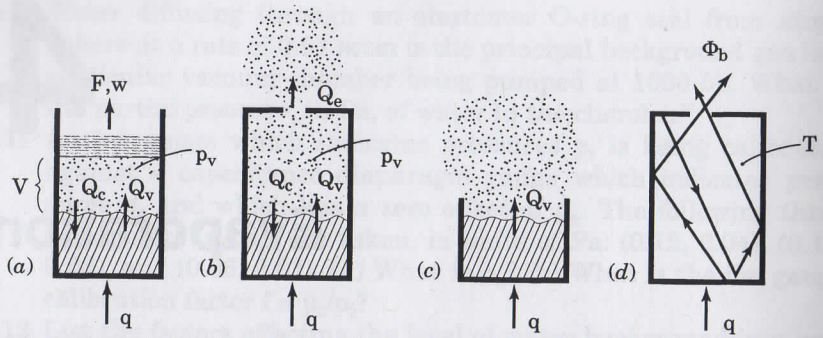

Figure 4.1 Evaporation situations: (*a*) vapor-liquid (or solid) equilibrium in a closed system, (*b*) Knudsen-cell effusion, (*c*) vacuum evaporation, and (*d*) radiation-blackbody analogy.

tures, and dissociatively evaporating compounds will be dealt with in Secs. 4.3 and 4.4. The condensed phase of our material, liquid or solid, is in equilibrium with its pure vapor in a closed isothermal container which will be the system under study. At equilibrium, the pressure in the vapor phase is the (saturation) vapor pressure, p_v, as discussed in Sec. 2.1. It is important to note that this equilibrium is a *dynamic* situation in which the molar condensation rate, Q_c, and the molar evaporation rate, Q_v, are balanced.

The system of Fig. 4.1*a* is closed; that is, no mass crosses its boundaries. However, energy in the form of heat, q, may be added, causing evaporation, and then energy in the form of mechanical work, $w = p_v\Delta V$, is removed from the resulting vapor as it pushes back the surrounding atmosphere or pushes on the piston shown, exerting force F. This is the principle of the steam engine, from which Josiah Willard Gibbs gained much of the inspiration for his development of the theory of thermodynamics. Not all of the heat input can be transformed into work, however. Most of it is consumed in increasing the energy of the molecules *within* the system, which is called the "internal" energy, U. In our earlier discussion of heat capacity leading to Eq. (2.14), we were concerned only with the molecular *kinetic* energy components of U. Now, we must include the much larger molecular *potential* energy component, ε_p, which accompanies the removal of a molecule from the condensed phase into the vapor phase. This removal process is illustrated in the potential-energy diagram of Fig. 4.2. The molecule in the condensed phase sits near the bottom of a potential well created by its bonds to neighboring molecules situated at the material's surface ($z = 0$). Evaporation involves removing the molecule to $z = \infty$ by breaking these bonds and raising ε_p. In the above interchange of heat and work accompanying evaporation and piston-pushing, total energy must be

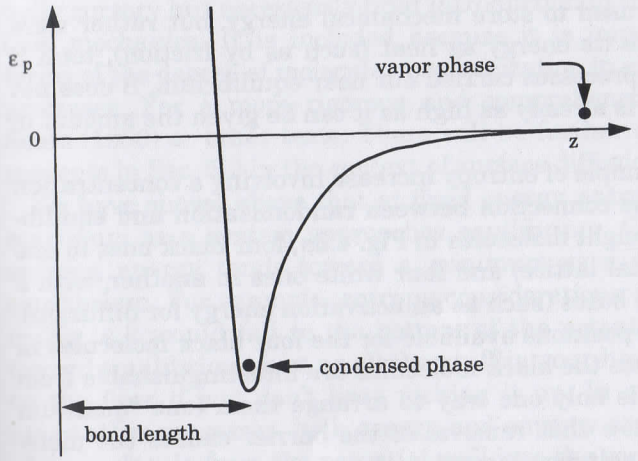

Figure 4.2 Molecular potential-energy diagram for evaporation and condensation.

conserved in accordance with the first law of thermodynamics, which is expressed in differential form as

$$dU + dw = dq \tag{4.1}$$

That is, an increment of heat input must show up either as internal energy stored or as work output.

To further discuss equilibrium, we need to invoke the more subtle concept of entropy. Consider a process involving changes in temperature and pressure (T and p) to our system. Let these changes be carried out slowly—close to equilibrium—and let the system be brought back to its original state at the conclusion. This is called a reversible process. The classic example used in thermodynamics texts is the Carnot engine cycle. Analysis of such processes shows that they obey the second law of thermodynamics, namely,

$$\oint \frac{dq}{T} = \oint dS = 0 \tag{4.2}$$

where the circle indicates integration around a closed loop, and T is in K. This equation defines the entropy, S; that is, $dS = dq/T$, where dq is the heat flow in a process being carried out reversibly. For reversible processes, total S of the system plus the surroundings is constant. For irreversible processes, total S increases. Entropy increase is a measure of the degree of randomization of energy that was initially in a form out of which work could be extracted, such as a difference in p, T or concentration. For example, if the force exerted by the piston in

Fig. 4.1*a* were not used to store mechanical energy, but rather were allowed to dissipate its energy as heat (such as by friction), total S would increase. In processes carried out near equilibrium, S does not increase because it is already as high as it can be given the amount of energy available.

The following example of entropy increase involving a concentration difference shows the connection between randomization and equilibrium. Consider the eight molecules in Fig. 4.3*a*, four black ones in one box (such as a crystal lattice) and four white ones in another, with a barrier between the boxes (such as an activation energy for diffusion). There are only four positions available for the four black molecules in the top box, and since the black molecules are indistinguishable from one another, there is only one way to arrange them (one "quantum state"). Now, we know that removal of the barrier causes the molecules to mix, just as salt dissolves when dropped into water. After mixing, there are most likely to be two black and two white molecules in the top box, and now there are *six* ways to arrange them, as shown in Fig. 4.3*b*. It may be shown by statistical arguments that

$$S = k_B \ln \Omega \tag{4.3}$$

where k_B = Boltzmann's constant
Ω = number of quantum states

Thus, irreversible mixing has increased the entropy of the molecules in the top box from zero to 1.79 k_B. By comparison, recall from the discussion of Sec. 2.2 that the Maxwell-Boltzmann distribution is the equilibrium distribution of molecular speeds because it is the distribution that provides the maximum number of ways of arranging a system of molecules both along the ladder of quantized translational energy levels (the Boltzmann factor) and among the quantum states available at each energy level (the degeneracy), given a fixed total energy for the system. The reason for the appearance of the natural logarithm in Eq. (4.3) is that S appears in an energy term, as we will see below, and energy terms are additive, whereas probability terms such as Ω are multiplicative. The above example has been an admit-

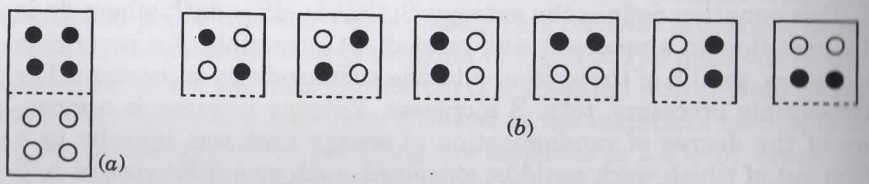

Figure 4.3 Ways of arranging four molecules in a box: (*a*) before mixing, (*b*) after mixing.

tedly cursory but necessarily brief introduction to the subject of statistical mechanics. It is included because it is important to think in terms of the degree of molecular randomization in examining chemical processes. For a more rigorous and comprehensive discussion, see Sears (1950) or other texts. There will be further discussion of these concepts in Sec. 5.2 in the context of surface diffusion.

We have shown above that at fixed energy, entropy tends toward a maximum as a system approaches equilibrium. Conversely, at fixed entropy, energy tends toward a minimum as a system approaches equilibrium. For example, entropy considerations aside, the molecule in Fig. 4.2 would fall to the bottom of the potential well in moving toward equilibrium, just as all the stuff in your house tends to end up on the floor if you don't keep picking it up. In evaporation, and in many other processes, both energy and entropy are varying. Removal of the molecule from the potential well into the vapor phase increases its potential and kinetic energy. Removal also increases its entropy, since clearly a molecule in the vapor phase has more quantum states available to it than in the condensed phase, both in position ("configurational" entropy) because it is in free space, and in energy level ("thermal" entropy) because the translational-energy quantum states are so closely spaced.

The equilibrium relationship between the competing factors of energy and entropy is seen by making the following substitutions in Eq. (4.1): dw = pdV and dq = TdS [from Eq. (4.2)]. Rearranging the result, we have

$$dU + pdV - TdS = dG = 0 \tag{4.4}$$

Here, we have also introduced G, the Gibbs free energy, often simply called the free energy. G is defined as

$$G = (U + pV) - TS = H - TS \tag{4.5}$$

Here, H is the enthalpy, which appeared earlier in connection with heat capacity at constant p [Eq. (2.15)] and which will be useful below. H is the energy term to use for constant-p processes, where pdV work is being done on the surroundings, while U would be the term to use for constant-V processes. Differentiating Eq. (4.5), we have

$$dG = dU + pdV + Vdp - TdS - SdT \tag{4.6}$$

which reduces to Eq. (4.4) for processes carried out at constant T and p, such as evaporation and many other processes in thin-film work. Thus, we see that G provides a concise definition of equilibrium; namely, a system held at constant T and p is at equilibrium when dG =

0 for any disturbance, such as the evaporation of dN_m moles of condensate. This implies that G for the system is at a minimum. Mathematically, G also could be at a maximum (unstable equilibrium), but we know that moving toward equilibrium involves decreasing H and increasing S, which means that G is *decreasing* toward a minimum by Eq. (4.5). Note that T appears in the entropy term in the above equations and gives it the units of energy by the definition of entropy in Eq. (4.2). This means that the entropy term becomes more important in proportion to T. That is, increasing the thermal motion promotes randomization, as one would expect, just as you know that it is more difficult to put your house in order when there is a party going on. During evaporation, H is raised, but this is compensated by an increase in S so that G remains at the minimum. At higher T, more H can be compensated because of the larger TS increase, so p_v increases with T, as we well know. Also note that evaporation can proceed by *absorption* of heat from the surroundings (evaporative cooling), which creates a T difference. This process is driven by the increase in S accompanying the evaporation, and it is an example of an "endothermic" (heat absorbing) reaction.

At this point, it is useful to introduce another function which expresses the incremental change in G for addition of material to a phase at constant T and p. This function also may be applied to multicomponent mixtures by defining it as

$$\mu_i = \left(\frac{\partial G}{\partial N_{mi}}\right)_{T,\,p,\,N_{mj}} \tag{4.7}$$

where μ_i is called the chemical potential of component i, and N_{mi} is the number of moles of that component. Different components consist of different kinds of molecules. The partial differential symbol, ∂, is used above because G also is a function of T, p, and N_{mj}, where j denotes the other components; but here, the amounts of the other components are being held constant. For a single-component system such as our evaporating pure condensate, μ is just the free energy per mole of condensate, G_{mc}. We have shown that, at vapor-liquid equilibrium, G for a closed system (vapor plus liquid) does not change as evaporation proceeds at constant T and p. In practice, the system must be *slightly* removed from equilibrium so that the heat of evaporation can flow into it along a T gradient. In any case, moving dN_m moles of material from the condensate into the vapor involves no change in the G of the total system, and thus,

$$\mu_c = \mu_v \tag{4.8}$$

where c and v denote the condensed and vapor phases, respectively. This is another way of stating the equilibrium condition, and it applies whenever the number of moles of each component is constant (that is, no reactions). The difference in μ from one phase to another when the two are *not* at equilibrium may be thought of as the driving force for motion toward equilibrium; hence the term "chemical potential" for μ. Material will move from phases of high μ to those of low μ until all the μ values are equal.

We now have all the tools we need to quantify the dependence of p_v on T. We know that at any point along the vapor-liquid (or solid) equilibrium curve of Fig. 2.2, Eq. (4.8) holds. As we move up the curve, the μ values of both phases increase, and they must both increase at the same rate in order for Eq. (4.8) to continue to hold; that is, $d\mu_c = d\mu_v$, or for our pure material, $dG_{mc} = dG_{mv}$. Since we are at equilibrium, Eq. (4.4) holds; that is, $dU = TdS - PdV$. Substituting this expression for dU into Eq. (4.6), we have

$$dG = Vdp - SdT \quad \text{or} \quad dG_{mi} = V_{mi}dp - S_{mi}dT \tag{4.9}$$

where i denotes condensate (c) or vapor (v), and m denotes "per mole." Thus, at equilibrium between the two phases,

$$V_{mc}dp - S_{mc}dT = V_{mv}dp - S_{mv}dT$$

or

$$\frac{dp_v}{dT} = \frac{S_{mv} - S_{mc}}{V_{mv} - V_{mc}} = \frac{\Delta S_m}{\Delta V_m} \tag{4.10}$$

Since $\Delta G_m = 0$ for evaporation, we have from Eq. (4.5) that $\Delta_v H = T\Delta S_m$, where $\Delta_v H$ is the "latent heat" (enthalpy change) of vaporization per mole. Note that the heat of vaporization is an H term rather than a U term because it is measured at constant p, not at constant V, and thus it includes the $p\Delta V$ work of expansion. Substituting $\Delta_v H/T$ for ΔS_m in Eq. (4.10) leads to the Clausius-Clapeyron equation,

$$\frac{dp_v}{dT} = \frac{\Delta_v H}{T\Delta V_m} \tag{4.11}$$

The volume term may be eliminated as follows. Consider V_{mc} to be negligible, since it is typically 1/1000 of V_{mv} at 1 atm. For ideal gases, $V_{mv} = RT/p$ from Eq. (2.10). Thus,

$$\frac{dp_v}{p_v} = \left(\frac{\Delta_v H}{RT^2}\right) dT \tag{4.12}$$

and this integrates to

$$\boxed{\ln p_v = 2.3 \log_{10} p_v = \left(\frac{-\Delta_v H}{R}\right)\left(\frac{1}{T}\right) + B} \tag{4.13}$$

where B is the constant of integration. Here, we have assumed that $\Delta_v H$ is independent of T, which is not exactly so but nevertheless is reasonable in regions far from the critical point (Fig. 2.2). Alternatively, we may express Eq. (4.13) in exponential form:

$$p_v = B' \exp(-\Delta_v H / RT) \tag{4.14}$$

The exponential factor here is the familiar one which appears in all thermally activated chemical processes, evaporation being an example of such a process. This factor will be derived in Sec. 5.2 Most of the energy in $\Delta_v H$ goes into raising the condensate molecule out of the potential well of Fig. 4.2, and the remainder goes into the resulting vapor molecule as random translational kinetic energy (ε_t) and as the $p\Delta V$ work which it does on the surroundings (see Exercise 4.1).

Note that a plot of $\ln p_v$ or $\log_{10} p_v$ versus 1/T (T in K) is a straight line according to Eq. (4.13). This is a useful way to estimate the p_v of materials for which p_v versus T data are not directly available (see Exercise 4.2). Almost always, one has available [1, 2] the boiling point of a liquid, or equivalently, the sublimation T of a solid, and the $\Delta_v H$ at that T. For compounds, often just the molar enthalpies of formation from the elements, $\Delta_f H$, are listed for the two phases, from which one can calculate $\Delta_v H = \Delta_f H_v - \Delta_f H_c$. Figure 4.4 shows the Eq. (4.13) plot for water as a dashed line extrapolated from the boiling point downward using the slope calculated from $\Delta_v H$ at 100° C (40.6 kJ/mole). The solid curve shows the actual p_v data. At a p_v level of 1 Pa, which is typically encountered in source-material evaporation for thin-film processes, the linear extrapolation has overestimated p_v by ×4. This is not so bad for a $\times 10^5$ extrapolation in p_v, but clearly it is not good enough for setting source-material flux in a thin-film process. For this and other reasons which we will examine in the next two sections, source flux must always be measured directly. Nevertheless, even an order-of-magnitude estimate of p_v is useful for several reasons. It can predict the source T that will be required fairly accurately; in the case of water, the predicted T for 1 Pa is low by only 15 K. It can predict whether a material is going to have to be cooled or capped off to prevent its evaporation at room T or at vacuum-chamber bakeout T; for example, the Group II (Zn, etc.) and VI (S, etc.) elements are problematic in this respect. It can predict how cold a cryogenic trap has to be to pump a material. For example, over the liquid nitrogen-cooled trap discussed in Sec. 3.4.2, the extrapolated p_v of water is about 10^{-17} Pa, which is totally negligible. Finally, it can predict how hot the walls of a CVD chamber have to be to prevent vapor condensation.

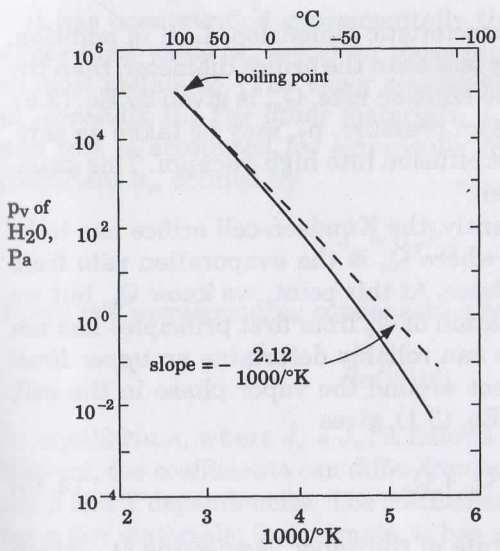

Figure 4.4 Vapor-pressure behavior of water. (Solid line = measured data.)

The p_v is the first piece of information needed to determine evaporation rate, Q_v, which we will discuss next. The p_v data for elements are given in Appendix B. For compounds, p_v sometimes can be found in the literature [3]; otherwise, it can be estimated as discussed above.

4.2 Evaporation Rate

When evaporation is taking place close to vapor-liquid (or solid) equilibrium, the effusion rate, Q_e, from a vapor source can be found from the vapor pressure, p_v, alone. This is the case for the vapor source shown in Fig. 4.1*b*. Here, the closed isothermal system of Fig. 4.1*a* has been modified by opening an orifice in the top which is small enough so that the resulting effusion does not significantly reduce the pressure, p, in the vapor phase; that is, $p \approx p_v$. Suppose that the orifice is small enough so that Kn > 1, and the orifice is therefore operating in the molecular-flow regime. Here, Kn is calculated from Eq. (2.25) using

the orifice diameter as the characteristic dimension, L. If in addition, the orifice length is appreciably less than the orifice diameter, then the orifice is termed "ideal," and the effusion rate, Q_e, is given by Eq. (3.5). In that equation, the downstream pressure, p_1, may be taken as zero, since we are talking here about effusion into high vacuum. This situation is known as a "Knudsen cell."

To avoid reducing p significantly, the Knudsen-cell orifice has to be small enough that $Q_e << Q_v$, where Q_v is the evaporation rate from the surface of the condensed phase. At this point, we know Q_e, but we do not know Q_v. The determination of Q_v from first principles has not yet been accomplished, but we can reliably determine an *upper limit* to Q_v as follows. A mass balance around the vapor phase in the cell, using the steady-state form of Eq. (3.1), gives

$$Q_v = Q_c + Q_e \tag{4.15}$$

where Q_c is the condensation rate of the vapor. Neglecting Q_e, which we have specified as relatively small, we have remaining the expression for vapor-liquid equilibrium in terms of a balance of condensation and evaporation rates. We can write the same balance in terms of fluxes per unit area of condensate surface ($J = Q/A$). We know the *impinging* flux, J_i, from kinetic theory [Eq. (2.18)]. Upon impingement, there is a range of degrees of interaction with the condensate surface which will be analyzed more thoroughly in Sec. 5.1. At the lower limit of this range, the vapor molecule bounces off after a single collision. At the upper limit, it becomes bonded strongly enough to the surface to become incorporated into the condensed phase. The portion of J_i that condenses is denoted as J_c. *If* all of the impinging flux does condense ($J_i = J_c$), then the vapor-liquid equilibrium ($J_c = J_v$) may be written as

$$J_i = J_{vo} \tag{4.16}$$

where subscript o denotes the upper limit to J_v that corresponds to all of the impinging flux condensing. Moreover, we would expect J_{vo} to be unchanged if we were to remove the vapor phase and were to consider evaporation from an open crucible or surface as shown in Fig. 4.1c. We would expect this because the evaporation of an individual molecule of condensate from the surface is not retarded by the impingement of vapor molecules elsewhere on the surface. Molecular potential interactions can extend only over a few atomic distances, and the probability of a vapor molecule landing that close to a surface molecule in the course of its evaporation is very small even at p_v = 1 atm (see Exercise 4.3).

It has been verified experimentally that evaporation occurs at this upper-limit value ($J_v = J_{vo}$) for those metals that have atomic vapors, as most metals do. (The vapor pressure compositions of the elements are given in Appendix B.) For other materials, J_v generally is lower than J_{vo}, and this is accounted for empirically by introducing an evaporation coefficient α_v, defined by

$$J_v = \alpha_v J_{vo} \tag{4.17}$$

There is a corresponding condensation coefficient α_c defined by

$$J_c = \alpha_c J_i \tag{4.18}$$

At equilibrium, where $J_c = J_v$, it follows that $\alpha_c = \alpha_v$. Away from equilibrium, the coefficients can differ from each other and can have differing p and T dependencies. The coefficient α_v has been determined only for a few materials; for example, it has a very low value of about 10^{-4} for As [4]. However, α_v generally is not known, and the factors that influence it are not well understood. This means that unless evaporation is being carried out from a Knudsen cell ($Q_e << Q_v$), effusion rate cannot be predicted accurately and must instead be measured directly in thin-film deposition processes. For the Knudsen cell, it must be verified that the orifice is small enough to avoid depleting the vapor phase (see Exercise 4.4). Even for metals having atomic vapors, where $\alpha_v \approx 1$, a melt surface that is partially obstructed from evaporating freely, such as by an oxide skin, will have an evaporation rate lower than that calculated using the macroscopic melt area.

It is useful to observe that α_c and α_v are analogous to the absorptivity, α, and emissivity, ε, of a surface for electromagnetic radiation (heat and light). Moreover, the behavior of heat radiation will be important later on for the analysis of T control in vacuum, so we digress to discuss it now. For radiation impinging on a surface, the fraction absorbed is given by α; the remainder is reflected (or transmitted). Absorptivity is a function of wavelength, λ, and of incident angle, just as α_c is expected to be a function of molecular kinetic energy and incident angle. The value of α at a given λ is called the monochromatic absorptivity, α_λ. There is a maximum radiation flux that can be emitted from a surface, just as there is a maximum evaporation flux, and both are found from the second law of thermodynamics. The maximum radiation flux is that from a so-called *blackbody* and is given by the Stefan-Boltzmann blackbody radiation law:

$$\boxed{\Phi_b = \tilde{n}^2 \sigma T^4} \tag{4.19}$$

where Φ_b = blackbody radiation flux, W/cm^2
σ = Stefan-Boltzmann constant = 5.67×10^{-12} W/cm^2·K^4
$\tilde{n}$ = index of refraction of the medium over the surface (≈ 1 for vapors at <1 atm)

This radiation is spectrally distributed according to Planck's law; likewise, evaporating molecules are distributed in energy. The radiation flux per unit wavelength interval is $\Phi_{b\lambda}$. The fraction of this spectral blackbody flux that is emitted by surfaces not completely black in a given λ interval is accounted for by the emissivity at that λ, ε_λ; thus

$$\Phi_\lambda = \varepsilon_\lambda \Phi_{b\lambda} \tag{4.20}$$

Moreover, it can be proven thermodynamically that $\alpha_\lambda = \varepsilon_\lambda$ (Kirchhoff's law), just as $\alpha_c = \alpha_v$ for molecules. The conventional way of constructing an ideal blackbody emitter from materials for which ε_λ is not unity is to use a cavity such as that shown in Fig. 4.1*d*. Although ε_λ for the internal surfaces may be <1, the radiation flux from the orifice still is given by Eq. (4.19). This is because, looking into the orifice, one sees the sum of the reflected and the emitted radiation. In effect, the radiation field within the cavity builds up to the blackbody value, just as the pressure within a Knudsen cell builds up to p_v and results in a maximum value for the effusion flux.

4.3 Alloys

So far, we have considered the evaporation only of single-component materials; that is, the elements and those compounds that do not dissociate upon evaporation. With multicomponent materials, we must deal with an additional complication, namely that the composition of the vapor phase generally differs from that of the condensed phase. We will address alloys in this section and dissociatively evaporating compounds in the next section. The distinction between the two is important. The term alloy is used to designate either a solid solution or a mixture of solid phases, and its composition is variable over a wide range. A familiar example of the solid-solution type is the solder alloy Pb-Sn or Pb_xSn_{1-x}, where x is the mole fraction of Pb. Conversely, a compound has a specific ratio of elements, such as the semiconductor GaAs or the dielectric SiO_2. That is, compounds have a specific "stoichiometry." It is also possible to have an alloy of compounds, such as the laser-diode alloy $(AlAs)_x(GaAs)_{1-x}$, more often written as $Al_xGa_{1-x}As$ or (AlGa)As. Finally, there are other three-element solids that are not alloys of binary compounds but are ternary

compounds; that is, they have a specific ratio of all three elements, such as the solar cell material $CuInSe_2$. Each of these categories of material behaves differently in evaporation, and it is important to know which behavior is operative to control composition during thin-film deposition.

Consider a generalized binary metal alloy B_xC_{1-x} whose component elements B and C are completely miscible at the evaporation T; that is, the atomic fraction x can vary from 0 to 1 without precipitating a second solid phase. We will consider evaporation from a well mixed liquid phase to avoid the complication of composition gradients developing within a solid condensate as evaporation proceeds. The total equilibrium vapor pressure over the melt will be the sum of the component p_v values: p_B and p_C. These p_v values will be lower than those of the pure elements because of the dilution of one element by the other; thus,

$$p_B = a_B x p_{vB} \quad \text{and} \quad p_C = a_C(1-x)p_{vC} \tag{4.21}$$

where p_{vB} and p_{vB} are the p_v values of the pure elements and the a values are the "activity coefficients." For simplicity, we will assume "Raoult's law" behavior, where the a values are unity, even though they generally deviate somewhat from unity due to differences between B-C versus B-B and C-C bond strengths. If the evaporation coefficients are unity, as is common for metals, the evaporation flux of each element is given by the Knudsen equation, Eq. (2.18), as discussed in the previous section. Then the ratio of these two fluxes is

$$\frac{J_{vB}}{J_{vC}} = \frac{x}{1-x}\frac{p_{vB}}{p_{vC}}\sqrt{\frac{M_C}{M_B}} \tag{4.22}$$

Thus, the vapor flux will be richer than the melt in the more volatile element for any composition x, so the melt will continue to deplete in that element as evaporation proceeds and will never reach a steady-state flux ratio. This effect is used to advantage in purging a melt of volatile contaminants during a "soak" period prior to deposition, but it is not suitable for the deposition of an alloy film.

There are two ways to achieve a steady flux ratio in such a situation. One is to use separate sources of the two pure elements, operating at different T levels. Here, it is important either to monitor each flux separately and continuously as discussed in Sec. 4.7, or to use isothermal Knudsen-cell or crucible sources whose effusion can be reliably set by T control. In either case, flux ratio will vary over the deposition area due to the separate locations of the two sources, which is undesirable

unless one is specifically studying film properties versus composition. This variation can be quantified by the information given in Sec. 4.6.

The second way to achieve a steady flux ratio is to feed an alloy wire or rod steadily into the melt during evaporation, as shown in Fig. 4.5 for a feed alloy B_yC_{1-y}. If the volume feed rate, W, is adjusted to hold the melt volume V constant, mass conservation dictates that a steady state must eventually be reached in which the evaporating atomic flux ratio of the two elements in Eq. (4.22) is equal to their atomic feed ratio, $y/(1-y)$. During the approach to steady state, the melt composition becomes depleted in the more volatile element until the two ratios become equal at a steady-state melt composition x_s, where

$$x_s = \frac{1}{1 + \left(\frac{1-y}{y}\right)\left(\frac{p_{vB}}{p_{vC}}\right)\sqrt{\frac{M_C}{M_B}}} \tag{4.23}$$

An estimate of the time required to reach steady state can be made by applying a transient mass balance [Eq. (3.1)] to one of the elements. Doing so for B in units of atoms/s, we have:

$$\underbrace{Wyn}_{\text{(input)}} = \underbrace{Wyn\frac{x}{x_s}}_{\text{(output)}} + \underbrace{Vn\frac{dx}{dt}}_{\text{(accumulation)}} \tag{4.24}$$

where n is the atomic density (atoms/cm^3). Below, we make the reasonable assumption that n does not vary much with composition, so it may be cancelled out. Note that the output (evaporation rate) has been expressed as a fraction of the steady-state input value, Wyn, rather than directly as $J_{vC}A$ (A = melt surface area), to simplify the expression. Note also the further simplification of treating V as a constant, whereas in fact V will decrease during the transient period if W is held constant. Integration of Eq. (4.24) from an initial condition of $x = y$ at $t = 0$ gives

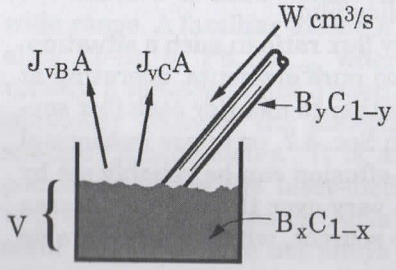

Figure 4.5 Alloy evaporation with continuous feed.

$$\xi = \frac{x_s - x}{x_s - y} = \exp\left(\frac{-t}{Vx_s/Wy}\right) = e^{-t/\tau} \tag{4.25}$$

where ξ is the fraction remaining from the initial state to the steady state at $x = x_s$. The denominator in the exponential term has the units of t and is in fact the time constant, τ, for the exponential approach to steady state. This solution is not exact because of the constant-V assumption, but the important point is that relaxation time scales as V/W. For exponential relaxation problems of this sort in which V is constant, $\tau = V/W$ exactly. This is the case for another common situation in which there is a step change in the concentration of a reactant flowing into a well mixed CVD reactor. The problem is also similar to the chamber-evacuation situation of Eq. (3.11).

For typical values of V and W in alloy evaporation, the approach to steady state can take many hours and consume much source material (see Exercise 4.7). This transient can be avoided by starting with a melt having the steady-state composition. This composition can be estimated from Eq. (4.23), but it may need to be adjusted based on actual film composition measurements if the value of $a_{B,C}$ [activity coefficients of Eq. (4.21)], α_v [evaporation coefficient of Eq. (4.17)], or S_c (sticking coefficient of Fig. 5.1) is not the same for each element. In special cases where the elemental vapor pressures are similar, x_s will be so close to y that the drift of x from y to x_s may not matter. This is the case for Al-2% Cu, which is used to prevent electromigration in integrated-circuit metal lines.

4.4 Compounds

Compounds behave very differently from alloys during evaporation. Some compounds evaporate as molecules that have the composition of the condensed phase. These compounds may be regarded as single-component materials, since they have only one p_v to consider. Ionically bonded compounds, including alkali halides such as the infrared-optics material KI and alkaline-earth halides such as the antireflection coating MgF_2, fall into this category. Oxides vary from one another in behavior, from no dissociation to partial to complete dissociation. There are limited data and some discrepancies about the degree of oxide dissociation, so it is wise to anticipate an oxygen-deficient deposited oxide film. Sometimes, reduction from a higher to a lower oxide occurs during evaporation. For example, SiO_2 evaporates as the monoxide SiO in the presence of a reducing agent M, where M is Si, C, or H_2. Since SiO has a much higher p_v than does SiO_2, these reactions

have been used to thermally evaporate native oxide from Si substrates prior to film deposition. Also falling into the category of partial dissociation are the Group IV-VI compounds (the "four-sixes"), which are used for infrared detectors and diode lasers. Thus, PbTe evaporates mostly as PbTe, but partly as Pb + (1/2) Te_2. Some compounds dissociate completely upon evaporation, and we will examine this behavior below. Among these materials are the Group III-V semiconductors (the "three-fives") such as GaAs, which are used for light-emitting diodes (LEDs), lasers, and microwave devices. Also completely dissociating are the II-VI semiconductors such as ZnSe, which have similar but less well developed applications. Both the III-Vs and the II-VIs sublimate, and their vapors usually consist of atoms of the metallic element, which we will call M, and dimers of the nonmetallic element, which we will call Y_2. The III-Vs and II-VIs differ from each other in that the III metals are much less volatile than the II metals, which has consequences to be seen below.

The behavior of compounds in equilibrium with their constituent elements is described by the "phase diagram" of the system [5]. Fig. 4.6*a* shows a relatively simple phase diagram for a binary mixture of M and Y, which happens to form only one compound, MY. The coordinate x is the atomic fraction of Y in the mixture. The existence of MY is conventionally indicated by a vertical line at x = 0.5, but in fact there is always a narrow *range* of x over which the compound can exist

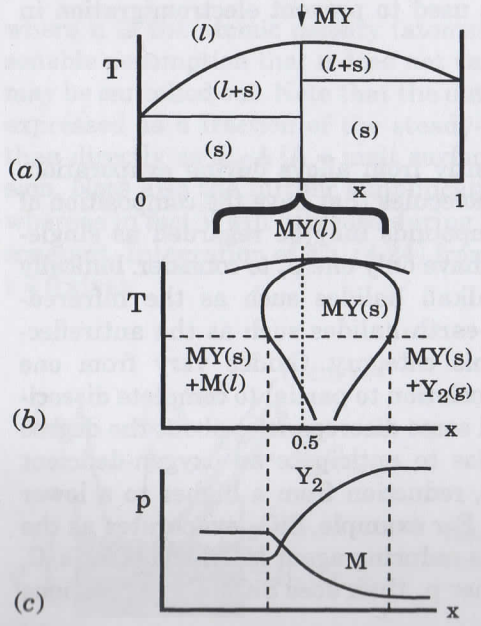

Figure 4.6 General form of phase equilibria for nonstoichiometric binary semiconductors $M_{1-x}Y_x$: (*a*) T-x projection, (*b*) T-x projection near x = 0.5, and (*c*) p-x projection. (Source: Reprinted from Ref. 6 with permission. Copyright © 1979, Pergamon Press.)

without precipitating a second phase of M or Y. This range is known as the single-phase field. A typical situation is shown in Fig. 4.6*b*, which is greatly expanded along the x axis. The distance in x between the M-rich and Y-rich boundaries shown here for the compound is usually less than 10^{-4}. Because it is so difficult to chemically determine such small changes in stoichiometry (see Sec. 10.2), the positions of these boundaries are not known for most compounds; but that is not important for the present discussion. The compound accommodates nonstoichiometry within the single-phase field by generating native point defects within its crystal lattice. The three types of such defects and their effect on stoichiometry are listed in Table 4.1. Usually, not all types are present in a given compound. The term "native" distinguishes these defects from point defects caused by impurity atoms. The nonstoichiometric single-phase compound may be viewed as a solid solution of native point defects in the lattice. Precipitation of M or Y occurs when the solubility range of the defects is exceeded at the M-rich or Y-rich boundary. Note that the two boundaries converge at the melting point in Fig. 4.6*b*; this is a manifestation of Gibbs' phase rule. Note also that the range of x does not always include the stoichiometric composition.

The important consequence of this range of x for compound evaporation and deposition is that the p_v values of M and Y_2 vary by orders of magnitude from one boundary to the other. This behavior is shown in Fig. 4.6*c* for a fixed T, which is indicated by the horizontal dashed line on Fig. 4.6*b*. The p_v of each element increases with the atomic fraction of that element, as one would expect. In each case, upon reaching the phase boundary, the p_v of the enriched element levels out at the value for the pure element, because that is when the pure element begins to precipitate out. Meanwhile, the p_v of the other element over the compound has become much less than that over the pure element. These

TABLE 4.1 Native Point Defects in Binary Compounds MY

Defect type	Symbol*	Stoichiometry
Vacancy (V)	V_M	Y-rich
	V_Y	M-rich
Interstitial	M_i	M-rich
	Y_i	Y-rich
Antisite	M_Y	M-rich
	Y_M	Y-rich

*Subscript denotes lattice location; i = interstitial site.

curves will both shift up with T, and for each element, the p_v behavior may be visualized as a sheet in p_v-T-x space. At the T shown in Fig. 4.6*b*, the p_v levels happen to cross each other within the single-phase region. If the crossover is large enough so that at some point before reaching the M-rich boundary, the evaporation flux of M becomes twice as large as that of Y_2, the surface composition of the compound will adjust itself by evaporation until that point is reached. From then on, evaporation will proceed "congruently;" that is, in the stoichiometric proportion. This is always the case for the II-VI compounds, and therefore they may be evaporated to completion congruently without forming precipitates. This is a good way to maintain a constant and known II/VI flux ratio at the substrate. If this ratio does not happen to be the one which yields the desired film properties, it can be adjusted by adding a second source of the pure II or VI element.

The III-V compounds present a different situation, because the p_v of M is often lower than that of Y_2 even at the M-rich boundary. In that case, M enrichment continues at the boundary by accumulation of a second condensed phase of pure M precipitate (which may also have a small amount of Y dissolved in it). Figure 4.7 shows the p_v versus T data for Ga and As_2 along each phase boundary. The upper curve for each element is the one corresponding to the phase boundary rich in that element. The point where the two curves converge at the left-hand edge is the melting T of GaAs, 1,238° C. Note that for T levels above about 680° C, $p_v(As_2) > p_v(Ga)$, even at the Ga-rich boundary, so that at high T, congruent evaporation is not possible. The maximum T for congruent evaporation depends on the evaporation coefficients, α_v, as well as on the p_v crossover point, and for GaAs that T is just over 600° C. However, at 600° C the evaporation fluxes, J_v, of Ga and As_2 are insufficient for film deposition at reasonable rates, so GaAs is deposited instead from separate Ga and As elemental sources. This is the case for the other III-V compounds as well. Elemental As evaporates as As_4, which has a lower S_c than As_2 in the deposition of GaAs, so the As_4 vapor often is superheated in a small furnace over the source crucible to crack it into As_2.

Clearly, considerable care must be taken to obtain constant and appropriate evaporation rates of the various species when dealing with dissociatively evaporating compounds. When these data are not available, one needs to do flux monitoring of each species (Sec. 4.7) and compositional analysis of both the depleted source material and the deposited film (Sec. 10.2) to define the situation. The dependence of solid-phase existence and composition on elemental p_v levels is also an important factor in determining the composition of the deposited film, and this relationship will be discussed more fully in Sec. 6.5.1. Basically, there are two major effects. One is that the native-point-

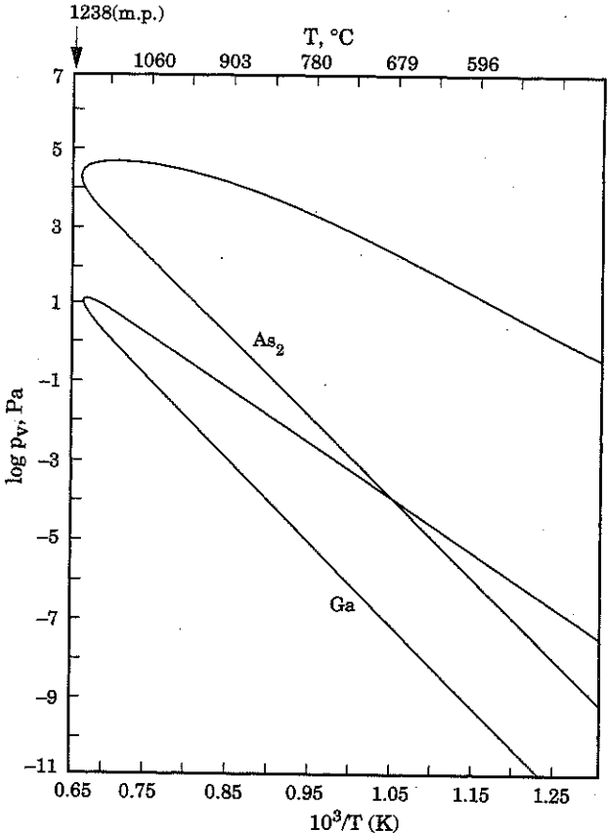

Figure 4.7 Equilibrium vapor pressures of Ga and As_2 along the GaAs liquidus, as functions of T. (Source: Reprinted from Ref. 7 with permission. Copyright © 1979, Pergamon Press. As_1 and As_4 curves deleted.)

defect distribution of the compound film shifts with the impingement-flux ratio of the constituent elements. Another is that if the impingement flux of Y_2 is too low and M is not volatile enough at the deposition T, the film will contain M precipitate as well as MY.

4.5 Sources

4.5.1 Basic designs

The common types of thermal evaporation sources are shown in Fig. 4.8. The simplest are the twisted-wire coil and the dimpled sheet-

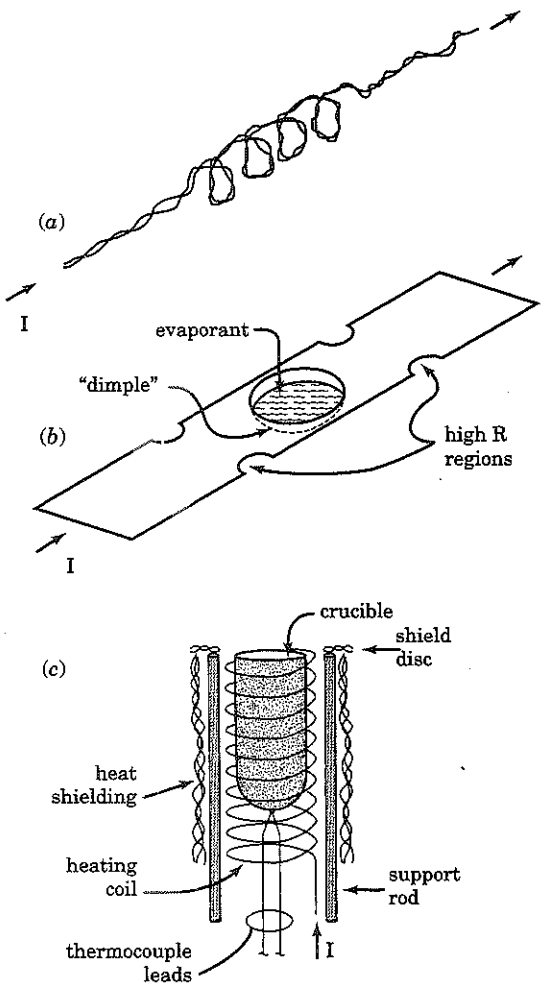

Figure 4.8 Evaporation sources: (*a*) twisted-wire coil, (*b*) dimpled boat, and (*c*) heat-shielded crucible.

metal "boat," both made from one of the refractory metals, W, Ta, and Mo, which were discussed in Sec. 3.4.2. These are heated by passing current of the order of I = 100 A through them (joule heating). The amount of heat generated is I^2R, where R is the parallel resistance of the source/evaporant combination at the evaporation temperature, T. T is nonuniform because of heat conduction down the current contacts. Therefore, if evaporation rate control is important, continuous flux monitoring is necessary, preferably with feedback control of the source I. The wire can be used only for metals that wet the wire. In the special case of Cr, which sublimes, sources are available in the form of a thick electroplated coating on W rod. Although boats can also be used for evaporants which do not wet them, thermal coupling to the evaporant is poor because it is by radiation only, so the boat will have to be much hotter than the evaporant. The thermal coupling problem in vacuum is discussed more thoroughly in Sec. 5.8.1. To prevent wetted evaporant from spreading away from the hot zone of the boat, a narrowed region is often provided on each end of this zone as shown in Fig. 4.8*b*. The resulting higher R increases heating enough in those regions so that evaporation rate exceeds spreading rate.

Alloying of the evaporant with the wire or boat metal may result in emrittlement or melting. If there are no metals suitable for contact with the desired evaporant, ceramic-coated boats or ceramic crucibles can be used. Information on boat materials and configurations suitable for various evaporants is available from their manufacturers. With evaporants which are very reactive at the evaporation T, such as Si, even ceramic sources will dissolve, thus contaminating the melt and eventually disintegrating. In such cases, evaporation must be carried out by focussing energy into the center of the melt, as discussed in Chap. 8.

In applications where simultaneous flux control of more than one evaporant is required, one must provide either separate flux monitoring of each evaporant or closely T-controlled sources which can maintain and reproduce the desired fluxes. Since the former method is relatively awkward and expensive, the latter is employed more often. Because of the exponential dependence of p_v on T [Eq. (4.14)], T control needs to be very good. Typically, p_v increases about 20 percent per 1 percent rise in evaporant T, or 2 percent per 0.1 percent rise; that is, $dp/p \sim 20\ dT/T$ in Eq. (4.12). Crucible sources provide the best T uniformity and control, especially if configured as shown in Fig. 4.8*c*, a design that has evolved from the stringent requirements of molecular-beam epitaxy (MBE; see Chap. 6). Key design features are as follows.

The Ta heater coil is minimally supported within notched ceramic support rods to reduce heat sinking. Alternatively, the coil can be hermetically sealed in a metal can to reduce O_2 outgassing from metal reaction with the ceramic at the high T of the coil. The coil extends well below the crucible bottom so that the thermocouple touching the crucible bottom is immersed within the same heating environment as the crucible and thus stabilizes at the same T. The thermocouple leads are made of fine wire to reduce heat sinking and are made of refractory alloys such as W-Re. Proper thermocouple readout wiring is discussed in conjunction with T measurement in general in Sec. 5.8.3. The heater coil is of finer pitch (denser) near the crucible mouth, so that the T in that location is higher than the T of the evaporant melt.

This prevents condensation of evaporant droplets at the mouth. Such droplets are undesirable because, when they fall back down into the melt, they cause the ejection ("spitting") of macroparticles that end up as defects in the film. The higher mouth T also prevents evaporants which wet the crucible from migrating out of it, as did the notches on the boat source. The radiation heat shielding consists of several wraps of corrugated Ta foil. Ta is used because it is the most ductile of the refractory metals. The shielding reflects radiation from the coil and crucible, thereby improving source-T uniformity and reducing heating and outgassing of nearby hardware. The corrugations minimize thermal contact between successive layers. (Principles of radiative heat transfer which are relevant to the behavior of heat shielding are discussed in Sec. 5.8.1.) The shield disc at the top provides additional heat shielding and also blocks the emission of outgassing contaminants in the direction of the substrate.

The crucible source can be converted to a Knudsen cell by adding to the crucible a closely fitting cover having a thin-lipped (ideal) orifice operating at Kn > 1. The crucible sources used for MBE are often referred to as Knudsen cells, but usually they do not satisfy the Kn > 1 criterion. If, in addition, one wants the Knudsen cell to operate at the saturation p_v of the evaporant, it must satisfy the criterion that $Q_e << Q_v$ in Eq. (4.15).

4.5.2 Contamination

Source structural materials and evaporants evolve contaminant vapors both from their surfaces and from the bulk. The vapor evolution problem was discussed in connection with vacuum materials in Sec. 3.4.2. The same remedies apply here, namely: degrease all materials, minimize surface area, avoid volatile constituents, and employ baking.

Considering source structural materials, surface area is minimized by using nonporous grades of ceramic for crucibles. In the case of graphite or BN, this means the "pyrolytic" grade. Pyrolytic crucibles are formed on the surface of a core or "mandrel" by the thermal decomposition (pyrolysis) of the gases CH_4 or $B_2H_6 + NH_3$. (These are examples of thin-film CVD processes in which the deposit—the crucible—is actually very thick.) Much of the volatile contaminant can be removed from source crucibles and associated components by "firing" the source at a higher T than the desired evaporation T, prior to filling the crucible with evaporant. Crucible materials should be chosen whose residual contaminants are the least harmful to the film being deposited. For instance, BN is a good choice for III-V semiconductor film deposition even though it is not one of the most refractory ceramics. It is a good choice because BN is itself a III-V compound, and therefore its elements do not form charge-carrier traps within the band gap of a III-V semiconductor. The most refractory ceramics are alumina and zirconia, but both evolve some oxygen.

Considering evaporant materials, chemical analysis should be obtained to determine if there are contaminants present that will be harmful to the film being deposited. Be aware that generally, only the metallic impurities are reported by the manufacturer. That is, a 99.99 percent pure ("four-nines" pure) material will have <0.01 percent (100 ppm) total *metallic* impurity content, but it also may contain a great deal more of O, C, Cl, and so forth. The evaporant should be obtained in large pieces rather than in powdered form to minimize surface area. A powder of 10 μm particle size has a very large surface area of about 0.3 m^2/cm^3. The amount of outgassing of adsorbed water from such large areas is considerable and was estimated in Sec. 3.4.2 and Exercise 3.8.

Much of the volatile impurity content in the evaporant can be removed before film deposition. This includes adsorbed gases and also dissolved elements of higher p_v than the evaporant. Removal is accomplished by "soaking" the source at a T somewhat below the T for film deposition, both before substrate loading and again afterwards while blocking the source from the substrate with a shutter; the second soak removes water adsorbed during the loading process. The dissolved impurities having high p_v will progressively deplete relative to the evaporant as described by Eq. (4.25). Conversely, impurities with p_v significantly *less* than that of the evaporant will selectively remain behind in the crucible, as described by the same equation. This phase separation by p_v is essentially the process of purification by distillation, as is done commercially for many materials. However, as the evaporant charge becomes depleted, the fraction of low-p_v impurities in it increases, and therefore so does the fraction of impurity in the evaporant flux. Therefore, it is unwise to evaporate the source material to completion when film purity is important. The only advantage of evaporating to completion is that a film of known thickness can be obtained simply by preweighing the evaporant charge, without having to measure the evaporation flux or time. The third group of impurities, those with p_v *similar* to that of the evaporant, cannot be selectively removed and will remain part of the evaporation flux.

4.5.3 Temperature control

The close control of T in evaporation sources, substrates, and CVD reactors is crucial to the reliable operation of film deposition processes. Here, we discuss the principles of T control within the context of evaporant heating.

To maintain constant crucible temperature, T, in the Fig. 4.8*c* source, the heater coil's power supply is driven by a feedback control loop which compares the thermocouple voltage with a set-point voltage. The simplest kind of T control is "on-off," such as the thermostat in your house. There, the heater power is fully on when T is below set point and fully off when it is above set point, so that T continues to oscillate about the set point. For most thin-film applications, this type of control is inadequate, and more sophisticated controllers are employed instead. The common "PID" type have three modes of control: *p*roportional, *i*ntegral or "reset," and *d*erivative or "rate." If the thermocouple T is lower than the set point by ΔT, the proportional control signal driving the power supply is a fraction of "full on." The fraction is given by $\Delta T/T_b$, where T_b is called the proportional bandwidth. T_b is inversely proportional to the "gain" of the feedback loop. On-off control is essentially proportional control with infinite gain. With proportional control alone, the power will remain "on" enough to maintain steady T only if there remains a finite offset ΔT from set point. The integral control mode gradually removes this offset by adding a control signal proportional to $\int_t \Delta T dt$. Finally, the derivative mode reduces T overshoot past the set point on initial heat-up by adding a third control signal proportional to minus the rate of approach of T to set point, $d(\Delta T)/dt$. The need for derivative control depends on the degree of overshoot and whether this would result in excessive evaporant loss or other problems.

The maximum speed of response to a step change in set point depends on the thermal time constant of the evaporation source or other thermal system. Response speed is important in applications involving multilayer films, where it is desired to quickly change the T of a source or of the substrate between layers or to deposit a layer of graded composition. A shutter may be closed between source and substrate to interrupt the deposition during this transient period. However, this allows background contaminants to build up on the substrate surface, as pointed out in Sec. 2.6. To avoid this problem in multilayer film deposition, separate sources of the same evaporant operating at the fluxes desired for each layer are often employed in conjunction with synchronized shutters. This technique is used, for example, in the deposition of quantum-well structures by MBE, where fluxes must be changed within one monomolecular layer of deposition.

The thermal time constant, τ, for T-control response may be determined by considering that heat must obey the continuity equation [Eq. (3.1)]: input – output = accumulation. To make the solution tractable, we will make various simplifying assumptions which are quite acceptable given that we are only seeking a rough estimate of τ. Thus, in the evaporation-source geometry of Fig. 4.9, we are looking down into a

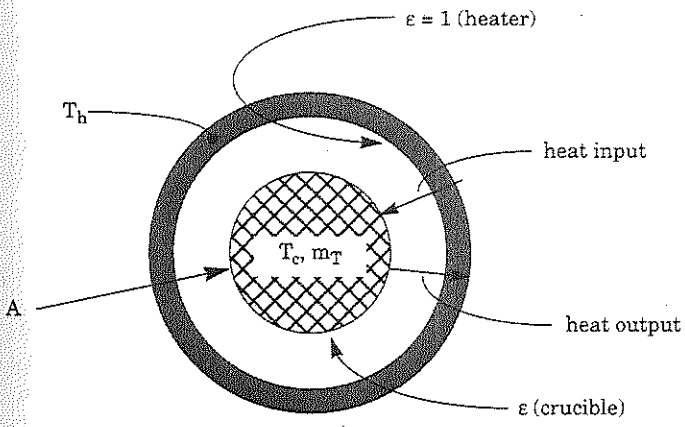

Figure 4.9 Geometry for crucible heating calculation; view looking into crucible mouth.

cylindrical crucible of surface area A and at T_c. It is completely surrounded by a cylindrical heater at T_h. Under vacuum, heat transfer is by radiation and follows Eqs. (4.19) and (4.20). For the crucible, we will make the so-called "gray-body" assumption that ε_λ and α_λ are independent of λ over the λ range of concern here; thus, $\varepsilon = \alpha$ by Kirchhoff's law. For the heater, we will assume that $\varepsilon = \alpha = 1$. For the system consisting of the crucible and its contents, the input term in the continuity equation is the heat radiated from the heater and absorbed by the crucible, and the output term is the heat radiated by the crucible. The accumulation term is the rate of heat buildup in the crucible and its contents. We will assume that the crucible and the evaporant are well coupled thermally. (This assumption is true for a melt; however, the T of a sublimating evaporant would lag that of the crucible considerably.) We may now write

$$\underbrace{A\alpha\sigma T_h^4}_{\text{(input)}} - \underbrace{A\varepsilon\sigma T_c^4}_{\text{(output)}} = \underbrace{(c_{gc}m_c + c_{ge}m_e)\frac{dT_c}{dt}}_{\text{(accumulation)}} = m_T\frac{dT_c}{dt} \tag{4.26}$$

where the c_g and m values are the heat capacities (J/g·°C or J/g·K) and masses of the crucible and evaporant, m_T is the resulting overall "thermal load" or "thermal mass," and the other symbols are as defined previously. We may now substitute ε for α, combine the two terms on the left, and factor the polynomial to give

$$A\varepsilon\sigma\left(T_h^4 - T_c^4\right) = A\varepsilon\sigma\left(T_h^3 + T_h^2T_c + T_hT_c^2 + T_c^3\right)(T_h - T_c)$$

$$\approx A\left(4\varepsilon\sigma\bar{T}^3\right)(T_h - T_c) = Ah_r(T_h - T_c) \qquad (4.27)$$

where $\bar{T} = (T_h + T_c)/2$

$h_r = 4\varepsilon\sigma\bar{T}^3$ = radiative heat transfer coefficient

The $\bar{T}^3$ approximation has changed an equation in T_c^4 into an equation linear in T_c and having an h_r of the same units as the h_c for gas conduction in Eq. (2.32). For typical situations in thin-film deposition, the error introduced by this approximation is negligible, as seen in Exercise 4.9. We now have in Eqs. (4.26) and (4.27) a differential equation which we can solve readily to give

$$\frac{T_h - T_c(t)}{T_h - T_c(t=0)} = \exp\left(\frac{-t}{m_T/Ah_r}\right) = \exp\left(\frac{-t}{\tau}\right) \qquad (4.28)$$

where τ is the time constant for heat-up or cool-down. The expression for τ in the denominator of the exponential is completely analogous to that in Eq. (3.11) for vacuum pumpdown and to that in Eq. (4.25) for alloy composition relaxation. In the electrical analogy, all three of these τ values consist of a capacitance (C) term divided by a conductance (1/R) term, which gives an "RC" time constant. For crucible heat-up, low thermal mass and high thermal conductance or coupling result in rapid response (small τ). Note that according to Eq. (4.28), $T_c \to T_h$ as $t \to \infty$, whereas in practice $T_h > T_c$ at steady state. This discrepancy is due to the simplifying assumption that the heater completely surrounds the crucible. To estimate $\bar{T}$ and h_r, consider that during heat-up, we have $T_h > T_c$, and during cool-down, we have $T_h < T_c$. Thus, we will take $\bar{T} \approx T_c$ to give an overall average h_r for the system.

Let us now calculate τ for a typical situation consisting of a 500° C crucible with A = 4 cm^2, containing 5 g of Sb (c_g = 0.23 J/g·K). If the mass of the crucible can be neglected, then m_T = 1.15 J/K. If ε = 1, then h_r = 0.011 J/s·cm^2·K. Thus, τ = 26 s, which would be a fairly long time to have to interrupt deposition waiting for T_c to reach steady state. The T_c^3 factor in h_r means that a more volatile evaporant will have much slower control response, since it will be operating at a lower T_c. It also means that controllers need to be retuned if T_c is changed significantly, because the control-feedback gain must be matched to τ. The same considerations apply to substrate T control. It does not help to increase the heater power to give a higher $\bar{T}^3$ for heat-up, because the $\bar{T}^3$ for cool-down will still be determined by T_c. Instead, it is best to set "full-on" heater power so that the heat-up rate is the same as the cool-down rate in the neighborhood of the desired T_c. Heater power can be set higher for faster heat-up response, but derivative control may be needed to prevent excess overshoot.

Once heater power has been set, each of the three control modes needs to be tuned. If there is too much gain in any mode, T_c will oscillate at an angular frequency of roughly $1/\tau$. If there is too little proportional gain (too much proportional bandwidth), T_c will drift, and response will be sluggish. These behaviors are shown in Fig. 4.10. The proper tuning procedure is to turn off reset and rate action and then tune the proportional band to just short of the setting which causes oscillation when a step change in set point is applied. Then, reset is added to just short of oscillation. Finally, enough rate action is added to minimize overshoot without making control response too sluggish. The thermal load and the heat supply are analogous to the familiar mass-and-spring system, whose resonant frequency increases with spring stiffness and decreases with increasing mass. Critical damping in either system means having just enough resistance to response so that oscillation is avoided, as with the shock absorbers on your car. With proper tuning, 0.1° C stability of thermocouple T usually can be attained in thin-film applications. Keep in mind, however, that this stability only applies to the steady state. Any disturbance in heat flow conditions, such as the opening of a shutter over the crucible and the accompanying increase in radiative heat loss, will cause a control response transient before re-equilibration.

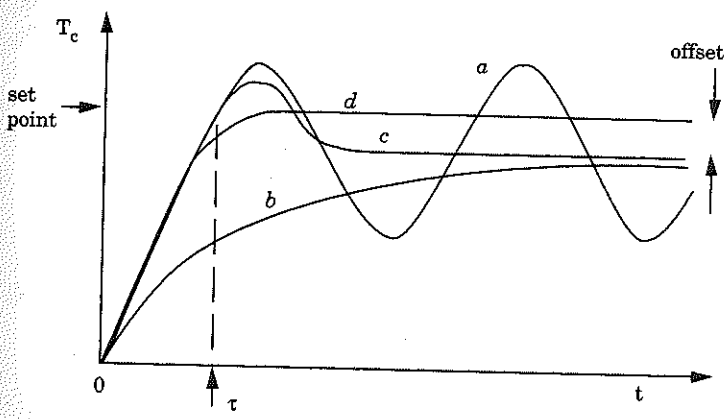

Figure 4.10 Control response to a step increase in set point at t = 0: (*a*) too much gain, (*b*) not enough gain, (*c*) critical damping, and (*d*) critical damping with reset and rate action added.

4.5.4 Energy enhancement

In many thin-film deposition processes, energy input to the film surface is required in order to increase the reaction rate of depositing species with each other or with the substrate, or to assist the depositing species in moving about on the surface. The effects of this energy will be discussed frequently throughout the book, since its control is a major factor in obtaining good films. In thermal evaporation, the energy content of the evaporant is much smaller than typical chemical bond strengths, as can be seen from Table 4.2, so it does not have much effect on film growth. In this table, the physics energy units of electron volts (eV) are given as well as the chemistry units of kilojoules because they relate more easily to the use of electrical means of energy enhancement: that is, electron bombardment and plasmas. One eV is the energy gain of a particle having one electronic charge upon passing through a potential drop of one volt.

There are many ways to enhance evaporant energy so as to modify film growth. Some involve nonthermal methods of source-material volatilization, such as laser ablation and sputtering (Chap. 8). Others use a glow-discharge plasma in the transport step, such as activated reactive evaporation (Sec. 9.3.2). The term "energetic condensation" is applied to any film-deposition process in which the kinetic energy of the depositing species is well above thermal but still below the onset of crystallographic damage and implantation (~30 eV; see Sec. 8.5.2.1). The present discussion will be limited to nonplasma ways of enhancing the energy of the basic thermal-evaporation source of Fig. 4.1*b*.

We first examine the supersonic nozzle. Consider a source having an orifice which is effusing vapor into vacuum, so that the pressure, p, is negligible downstream. We have seen that for such an orifice in molecular flow (Kn > 1), the throughput Q_e (mc/s) is proportional to upstream p [Eq. (3.5)]. Conversely, in the fluid-flow regime (Kn << 1), $Q_e \propto p^2$. One factor of p comes from the vapor concentration increase with p via the ideal-gas law, as in molecular flow. The second p comes from the fact that *fluid* flow velocity through an orifice increases with the Δp driving force. From the molecular viewpoint, collisions between molecules in the approach to the orifice serve to convert random thermal translational energy into directed kinetic energy moving downstream. This conversion occurs because molecules in the orifice are being hit more often from the high-p upstream side than from the low-p downstream side, and thus they accelerate. The limit to this acceleration occurs when the flow velocity in the orifice throat reaches the speed of sound, c* (Fig. 2.4). This situation is known alternately as sonic or choked or critical flow. However, downstream of the orifice, further acceleration into the supersonic regime can occur, and this effect can be enhanced by incorporating a conical nozzle as shown in Fig. 4.11. Such nozzles are commonly employed in rocket engines to maximize thrust. The cone angle is made as large as possible to minimize friction, but not so large as to cause separation of the laminar-flow boundary layer from the cone wall. Separation would lead to randomization of the kinetic energy in eddy currents generated adjacent to the wall. Typical conditions for present applications would be a p of 3 atm, an orifice diameter of 0.3 mm, and a cone included angle of 15°. Note that in the subsonic regime upstream of the orifice, flow velocity, u, is increasing with decreasing cross-sectional area for flow, A; in the throat, u = c* and is thus independent of A; and in the expansion nozzle, u is increasing with *increasing* A.

Two characteristics of supersonic nozzles are important for film-deposition applications. One is that considerable vapor cooling takes place due to the transfer of kinetic energy from thermal to directed. Cooling

TABLE 4.2 Characteristic Energies in Thin-Film Deposition

	eV/mc	kJ/mol
ε_t^* at : 298 K	0.038	3.71
: 2200 K	0.28	27.4
Xe in H_2, 2200 K†	14.3	1380
100 mc, 1 keV cluster ion	10	965
Laser-ablated atoms	40	
Sputtered neutrals	5	
Bond energy: Si-Si	3.29	318
: N≡N	9.83	949

* Mean thermal translational energy of gas molecules, per Eq. (2.11)
†Supersonic nozzle beam [8]

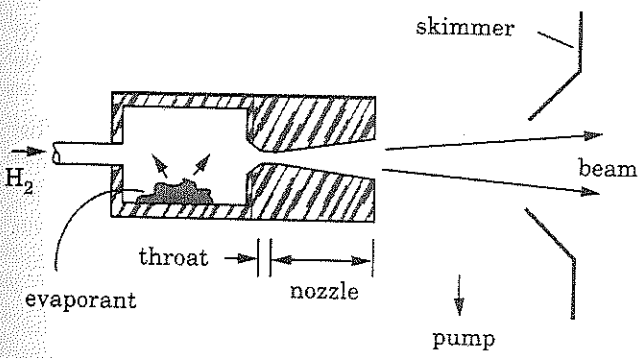

Figure 4.11 Supersonic-nozzle evaporation source.

assists in cluster formation, as we will see later. Cooling is efficient for two reasons. First, because u is so high, little heat is received from the orifice wall; that is, the expansion is "adiabatic." In addition, for an optimized cone angle, there are negligible frictional losses, which would transfer kinetic energy back to thermal energy; that is, the process is "isentropic" (constant entropy). The second important characteristic of these nozzles arises from the fact that all components of a gas mixture get accelerated to the same u in any fluid flow stream. Since $\varepsilon_t = (1/2)mv^2$, the evaporation of a small fraction of high-mass vapor into a low-mass carrier gas such as H_2 results in an ε_t for this "seeded" vapor which is increased by the ratio of masses over what it would be for expansion of the pure vapor. For example, measurements of Xe in H_2 gave ε_t well above chemical bond energies, as seen in Table 4.2. The mass ratio for this combination is 66. This technique requires a large flow of carrier gas which must be pumped away. Most of the pumping is done between the nozzle and "skimmers" (Fig. 4.11) which serve both to isolate the deposition chamber from the higher-p region and to deflect carrier gas out of the supersonic stream to minimize collisional losses. A supersonic-nozzle beam of Ge_2H_6 gas in He has been used to deposit single-crystal Ge film onto GaAs at 50 nm/s, much higher than the rate using thermal effusion [9]. The rate enhancement was attributed to activation of Ge_2H_6 dissociation at the growth surface by the high molecular kinetic energy.

Ionization and acceleration constitute a widely-used method of enhancing the energy of vapor molecules. Ionization of molecules evaporating from a crucible can be done by electron impact from a filament source, as was used in the ion gauge (Sec. 3.5). The ions are then accelerated using an electrical bias between source and substrate. For a typical design, see Ref. [10]. Low ionization rate limits the ion flux of such sources, so they are mostly useful for enhancing the incorporation of minor constituents (dopants) into a depositing film. Much larger ion fluxes are obtainable by using electron-beam or cathodic-arc evaporation (Chap. 8) or by incorporating a glow-discharge plasma into the transport space (Chap. 9).

About two decades ago, Takagi of Kyoto University made the intriguing proposal that film deposition be carried out using accelerated *cluster* ions. With sufficient collisions and cooling, vapor from an orifice is known to condense into clusters of tens or hundreds of atoms or molecules. It was postulated that, upon colliding with the film surface, such a cluster would disintegrate and transfer its perpendicular kinetic energy into energy directed laterally along the surface and distributed among the scattered molecules. Thus, a 100 mc cluster accelerated to 1 keV would result in 10 eV molecules skittering along the film surface (see Table 4.2). This is just the direction and level of energy that is desired for the modification of film deposition behavior. However, it has not yet been demonstrated that this sort of energy transfer from clusters actually takes place. Both the production and verification of clusters are nontrivial, and most of the work issuing from the cluster-ion proposal has been done with uncharacterized cluster masses and with collision rates within the nozzle which were much too low to lead to substantial cluster formation. Much of this work does show evidence of film improvement when the vapor flux is ionized and accelerated versus when it is not, and it is likely that these results are due to bombardment of the surface by single-molecule vapor ions along with a larger flux of neutral thermal molecules. Ion bombardment of a depositing material is widely used to improve film properties, as will be discussed in Sec. 8.5.3.

In the meantime, there has been considerable progress in the understanding of and equipment for cluster formation. Expanded interest in clusters has arisen from the discovery of "magic" numbers of atoms (corresponding to completed shells of conduction electrons) and from the potential of nanoparticles as new materials, such as the 60-atom polyhedral carbon clusters known fondly as Buckminster-Fuller-ene or "buckyballs." A good sampling of recent cluster work can be found in the particle-symposium proceedings listed at the end of the chapter. The basic requirements for cluster formation are that there be enough vapor-vapor molecular collisions to grow the clusters and that there be a sink for the considerable heat of condensation, which would otherwise prevent the cluster from nucleating in the first place. The sink is provided by contact with a cold "third body" during nucleation, this contact being either with the orifice wall (heterogeneous nucleation) or with other molecules of the vapor or the carrier gas (homogeneous nucleation). Under optimized conditions in a supersonic nozzle, there is sufficient cooling and collision rate for substantial cluster generation; for example, Zn beams with 2200-atom average cluster size have been obtained [11]. Alternatively, in the gas-condensation technique, the vapor flows through a chilled tube in an inert carrier gas which can be held at much lower p than in the supersonic nozzle. This technique reduces gas load on the pumps but would be expected to result in substantial cluster loss to the walls. A third mechanism has been proposed to explain observations of clusters in some of the cluster-ion film-deposition work done under conditions unlikely to generate them homogeneously. This mechanism [12] invokes heterogeneous nucleation on a nonwetting orifice wall, followed by entrainment of the clusters in the vapor stream. While a likely explanation, it is unlikely that this mechanism would be controllable or would result in clusterization of a substantial fraction of the vapor.

It may be concluded with regard to cluster-ion deposition that the original proposal remains intriguing but untested. With the present improved state of cluster-generation technology, it is now possible to study the use of cluster ions in film deposition more rigorously.

4.6 Transport

At this point in our development of the principles of thin-film deposition processes, we have completed the first of the four process steps shown in Fig. 1.1, namely the supply of source material. The main issues addressed have been supply rate and contamination. Thus far, these issues have been addressed only for thermal evaporation as a supply method. They will be revisited in Chaps. 7 and 8 for other supply methods. The second of the four process steps is transport of the source material to the substrate. Here, the main issues are arrival-rate uniformity at the substrate and, again, contamination. Contamination entering the process during the transport step has been addressed in Sec. 3.4. Here, we will examine uniformity. The factors determining arrival-rate uniformity at the substrate are very different, depending on whether the medium is a high vacuum or a gaseous fluid, as pointed out in Sec. 1.2. Transport in the fluid medium depends on flow patterns and diffusion, and these factors will be examined in Chap. 7 in the context of CVD. Here, we will examine transport in the high-vacuum medium. For present purposes, high vacuum means that the background pressure is low enough that the Knudsen number, Kn = l/L [Eq. (2.25)], is >1 for L taken as the distance from source to substrate. Under these conditions, transport is along the line of sight from source to substrate, because the probability is small for an evaporant molecule to collide with a background-gas molecule along the way. Consequently, arrival-rate uniformity across the substrate is determined solely by geometrical factors, and we will proceed to examine these below.

Consider the evaporation geometry of Fig. 4.12, where material is being evaporated from a source centered at point B onto a flat substrate situated at a perpendicular distance r_o away. Several source shapes are shown. The circular disc, which is emitting material from the top surface only, well represents the boat, the Knudsen-cell orifice, and the filled crucible. The sphere roughly approximates the wire-coil source. The collimated source represents a partly filled crucible or a nonideal orifice—an orifice whose length is significant compared to its diameter. We first want to know the evaporant flux J_θ at radius r_o from the source and at some angle θ from the perpendicular (say at point R), compared to the flux J_o at the perpendicular (point P). If you think

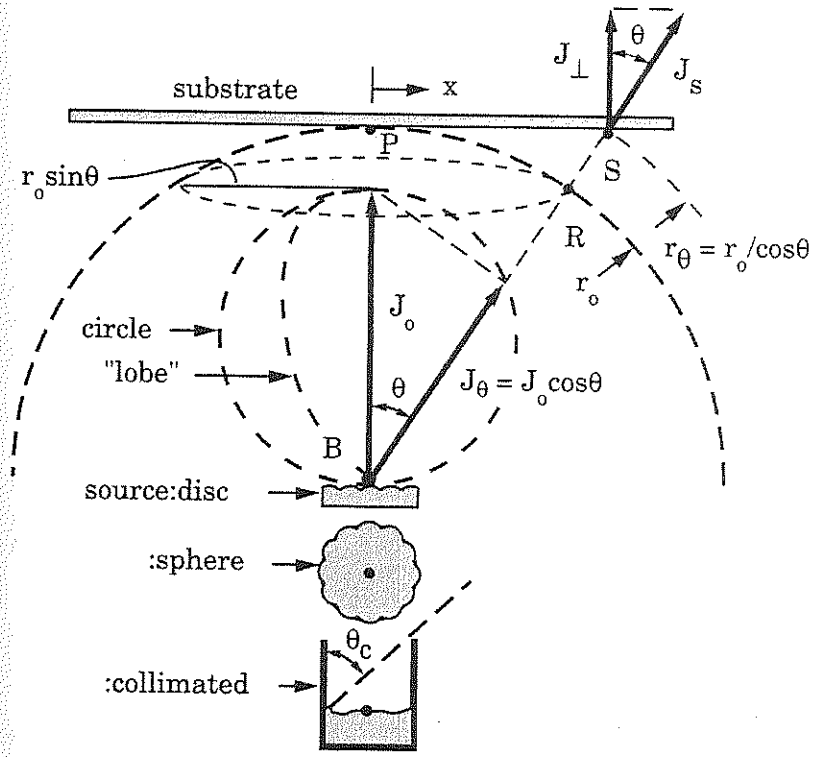

Figure 4.12 Geometry of vacuum evaporation. The flux ratio J_θ/J_o shown is for the "cosine distribution" characteristic of the disc-shaped source centered at point B; alternate source shapes shown below would have different flux distributions when positioned at B.

of yourself as an observer looking into the source from radius r_o, it is clear that J_θ will be proportional to the projected area of the source in that direction. For the sphere, this area is independent of θ, so for a total evaporation rate of Q mc/s, we have simply

$$J_\theta = J_o = Q/4\pi r_o^2 \tag{4.29}$$

where Q (mc/s) = $J_v A$ = total evaporation rate from a source of area A emitting flux J_v from its surface.

More commonly, the source is a disc. Looking into this source from r_o, the projected area changes from a circle to an ellipse to a line as one moves away from the perpendicular; in fact,

$$J_\theta = J_o \cos\theta \tag{4.30}$$

This is known as a cosine flux distribution. In Fig. 4.12, the flux vectors J_o and J_θ have been displaced from their proper locations at points P and R, and to point B, to illustrate that the vector J_θ traces a circle over θ. This is because J_o, J_θ, and θ always form a right triangle, as shown. To calculate the J values, consider that J_θ will pass through the r_o sphere in an annulus having radius $r_o\sin\theta$, circumference $2\pi r_o\sin\theta$, and differential width $r_o d\theta$. Knowing this, we can integrate J_θ over the hemisphere of evaporation, and by mass conservation, this integral must equal the total evaporation rate; that is,

$$Q = \int_o^\pi J_o \cos\theta \cdot 2\pi r_o \sin\theta \cdot r_o d\theta = \pi r_o^2 J_o \tag{4.31}$$

or

$$\boxed{J_o = \frac{Q}{\pi r_o^2}} \tag{4.32}$$

Note that Eq. (4.31) is the same integral as Eq. (2.6) for the average perpendicular component, $\bar{v}_x$, of molecular velocity, v, over the hemisphere. There, we found that

$$\bar{v}_x = \frac{1}{2}v$$

Here, the average flux over the hemisphere would be $Q/2\pi r_o^2$, which is $(1/2)J_o$ by Eq. (4.32).

Looking now into the collimated source from r_o, we see that the projected area drops off much more rapidly with increasing θ than for the disc-shaped source, due to shadowing from the lip, and this leads to the "lobe"-shaped flux distribution shown in Fig. 4.12. Although there is some angle, θ_c, above which the evaporant surface is not visible at all, there will still be some flux at higher angles because of evaporant scattering from the collimating sidewalls. Indeed, if all of this evaporant were to reflect at the specular angle, θ would never change upon such reflection, and therefore the cosine distribution would be preserved. However, molecules do not do this, but instead they scatter randomly, as discussed in Sec. 2.8.2 on viscosity. Some of these scattered molecules return to the crucible rather than escape from it, and this results in the lobed distribution. Calculations of such distributions have been reported [13]. The lobe can be broadened by using conical crucibles. Lobed distributions also are seen from *disc*-shaped sources when they are heated in one spot by a focused beam of electrons or laser light. These energy-beam techniques are discussed in Chap. 8. Supersonic nozzles produce sharply lobed distributions. Deposition uniformity over large substrates is a major concern with lobed flux distributions.

Let us consider uniformity across a flat substrate for the simpler case of a cosine distribution. There are two factors that attenuate the flux with increasing θ besides the projected-source-area factor. One is that the substrate point S in Fig. 4.12 is at radius r_θ rather than at r_o (point R) from the source, and $r_\theta = r_o/\cos\theta$. Since J is inversely proportional to r^2 by Eqs. (4.29) and (4.31), we have for the flux through the r_θ sphere at substrate point S,

$$J_s = J_\theta \cos^2\theta \tag{4.33}$$

Now, the flux that determines deposition rate is the flux perpendicular to the substrate at point S, $J_\perp$, and that flux is reduced from J_s by another $\cos\theta$ factor, as shown in Fig. 4.12. Combining all of these factors, we have, finally,

$$J_\perp = J_o \cos^4\theta = \frac{Q\cos^4\theta}{\pi r_o^2} \tag{4.34}$$

where Q (mc/s) is the total evaporation rate into the hemisphere from a disc-shaped source. The $\cos^4$ factor causes, for example, a 10 percent reduction in $J_\perp$ from center to edge for a substrate whose radius is one-fourth of the source-to-substrate distance r_o. Consequently, there is always a trade-off between nonuniformity at short distances and evaporant waste at large distances. By the way, the flux $J_\perp$ is sometimes expressed in terms of a "beam-equivalent pressure" by using the Knudsen equation, Eq. (2.18).

Various refinements of the Fig. 4.12 geometry can be incorporated to improve flux uniformity. Rotation of a substrate about its own axis, coupled with off-axis orientation of a single source, improves uniformity considerably [14, 15] compared to Eq. (4.34). Numerous small substrates such as lenses can be coated uniformly from a single source with low material waste by placing the substrates on a "planetary" fixture. There, the substrates rotate simultaneously about two different axes so that over the course of the deposition, their complex orbits cover a large fraction of the hemisphere. Note that rotating substrates always incur an oscillatory deposition rate that can affect film properties, especially when the process involves co-evaporating different materials from different sources, since the oscillations of the different material fluxes will not be in phase. For large-area substrates, multi-

ple, distributed sources of the same material and rate can be used, but rate balancing can be difficult. Alternatively, several sources are distributed in a line, and the substrate is passed by perpendicular to the line as shown in Fig. 4.13. This geometry is also used in the sputter-deposition process to be discussed in Sec. 9.3.4, in which case one uses a single source of long rectangular shape, and uniformity is better than 2 percent.

The example of Fig. 4.13 also illustrates the "web" coating process, in which a flexible substrate is supplied from one roll, taken up by a second, and suspended as a web between the two rolls during coating. This process is widely used to apply magnetic metal to videotape, Al or SiO_2 moisture-barrier layers to plastic film for food packaging, and metal electrodes to Mylar film for capacitors. The rolls may be located either in the vacuum, or outside with the web being passed through multiple, differentially pumped sliding seals.

Uniform coating of large *rigid* substrates is also achieved by transporting the substrate past a distributed vapor source as in the web process. For high throughput, entrance and exit load-locks (Fig. 3.1) are used to supply and remove the substrates. Multilayer coatings can be applied by placing sources of different materials in succession along the direction of substrate motion. With these techniques, coatings can be applied uniformly to substrates of many m^2 area, as with heat-control coatings on window glass.

4.7 Vapor Flux Monitoring

Often, it is not possible to predict accurately the flux arriving at the substrate, due to uncertainty in the evaporation rates of the various species or due to the presence of lobed flux distributions, as discussed above. When the arriving flux at least is reproducible for successive depositions, a few measurements of film thickness are sufficient to set the source operating conditions. This can be done for sources having good T control, unity evaporation coefficients, and cosine flux distributions. In other cases, it is difficult to reproduce arriving flux, and therefore flux monitoring is needed. Flux can be monitored either as the concentration of source vapor near the substrate or as actual deposition rate. Monitoring both is especially useful, because then one can calculate the fraction of arriving vapor which becomes incorporated into the deposit—the sticking coefficient, S_c. Deposition monitoring will be discussed in the next section. Here, we will address vapor monitoring.

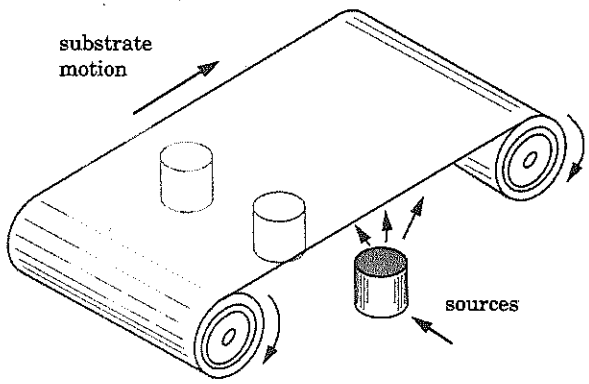

Figure 4.13 Web coating process for flexible substrates.

The basic technique for vapor flux monitoring is to excite the vapor molecules and detect their response. Excitation with a beam of electrons or light raises electrons in the molecules into excited states. This is followed by ionization if the excited electron leaves the molecule, or by photon emission (or Auger-electron emission) if it relaxes back toward its ground state. The simplest type of flux monitor is the ion gauge, which was discussed in Sec. 3.5. Recall that the ionization rate and resulting collected ion current are proportional to the concentration of vapor molecules within the gauge, $n\,(mc/cm^3)$. The flux of any stream of particles is related to its concentration by

$$\boxed{J = nv} \tag{4.35}$$

where v is the particle velocity in the direction of the flux. For flux from thermal evaporation, one can take v as roughly the mean speed, $\bar{c}$, of a Maxwellian distribution at the T of the source, as given by Eq. (2.3). For particles vaporized by laser ablation or ion sputtering, v is much higher.

The ion-gauge flux monitor must be shielded as shown in Fig. 4.14 so that evaporant does not react with the filament. Electrical connectors also must be shielded from accumulating conducting deposits which would short them out, and the grid must be apertured to avoid accumulation of insulating deposits which would charge up. Simultaneous monitoring of multiple evaporants can be achieved by shielding each monitor from the other sources as shown in Fig. 4.14. However, it is more convenient to use a single species-specific monitor. One such device is the mass spectrometer mentioned in Sec. 3.5, which is basically an ion gauge followed by a mass-selective filter. Another is the electron-impact emission spectrometer [16], in which the light emission from the relaxation of electron-impact-excited molecules is spectrally analyzed to separate the emission lines characteristic of each

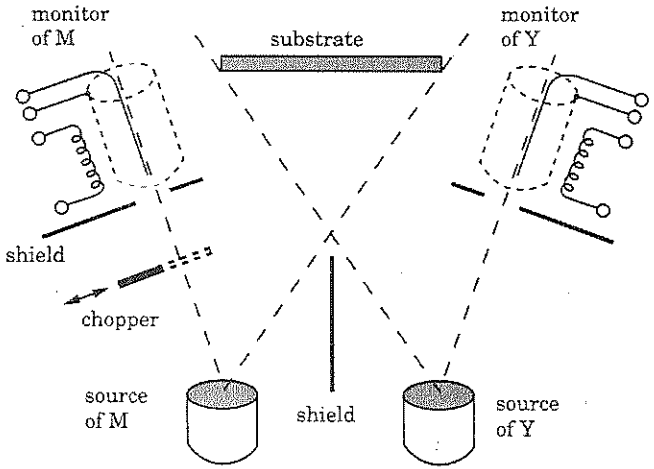

Figure 4.14 Separate monitoring of vapor fluxes from two sources using ion gauges.

element. With any of the above devices, the mechanical beam chopper shown in Fig. 4.14 is useful for distinguishing beam from background species when the background n is a significant fraction of the beam n, as in the case of volatile evaporants such as As.

The above electron-beam techniques are usable only when the electron mean free path is larger than the monitor diameter, which means below 1 Pa or so [Eq. (2.22)]. They also are not usable in a plasma environment because of interference from plasma electrons and ions, although light emission from molecules excited by *plasma* electrons is an important analytical tool in plasmas. For the shorter mean free paths of the fluid-flow processes, or in plasmas, optical excitation and detection techniques must be used to monitor vapor concentration.

Closed-loop control can be set up for flux just as it was for T in Sec. 4.5.3. The flux-monitoring signal is compared to a set point, and the difference is used to drive the source power through a proportional controller. As with T control, the thermal time constant of the source determines maximum response speed. The same control loop can be used with the quartz-crystal deposition monitor which is to be discussed next.

4.8 Deposition Monitoring

We now discuss direct monitoring of deposition flux, which is the vapor flux arriving at the substrate times the sticking coefficient, S_c. Though deposition monitoring is really part of the deposition step of the thin-film process, its close relationship to vapor flux monitoring makes it appropriate to discuss here in the context of the transport step. There are several approaches to deposition flux monitoring, principally:

1. mass deposition on a vibrating quartz wafer adjacent to the substrate
2. change in optical reflection intensity or phase from the film itself with increasing thickness
3. periodicity in the reflection high-energy electron diffraction (RHEED) signal from epitaxial films

Since RHEED is specific to epitaxy, it will be discussed in Chap. 6. The mass and optical techniques are discussed below. All three techniques are extremely sensitive, with submonolayer resolution being achievable routinely. It is important to keep in mind that none of them measures film thickness directly. Thickness must by calculated by knowing or assuming a film property, namely, for the cases listed above, (1) density, (2) index of refraction (except in ellipsometry), and (3) monomolecular layer (monolayer) thickness.

4.8.1 Mass deposition

The vibrating quartz crystal mass-deposition monitor, or quartz crystal microbalance, is one of the most powerful and widely used diagnostic instruments in thin-film technology. It uses the resonant crystalline quartz wafers that were developed for frequency control in radios and are used also for timing in computers and watches. Crystalline quartz is piezoelectric, so a quartz wafer generates an oscillating voltage across itself when vibrating at its resonant frequency, and this voltage can be amplified and fed back to drive the crystal at this frequency. Electrical coupling is done with thin-film metal electrodes deposited on opposite faces of a thin quartz wafer having the proper crystallographic orientation, as shown in Fig. 4.15. For deposition monitoring, one electrode is exposed to the vapor flux and proceeds to accumulate a mass of deposit. This mass loading reduces the crystal's resonant frequency, ν_r. Comparison of the loaded ν_r with the ν_r of a reference crystal located in the instrument's control unit is used to calculate the mass of deposit. Measurement of any quantity relative to a reference value of similar magnitude is always much more accurate than making an absolute measurement, and here it accounts for the submonolayer resolution of the instrument. The ν_r of about 5 MHz, when loaded, generates a beat frequency with the reference crystal which is equal to their ν_r difference. The resonant frequencies are very stable, so one can easily measure a beat frequency of a few hertz,

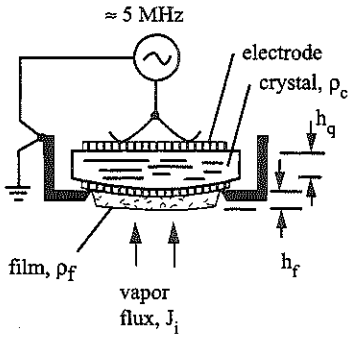

Figure 4.15 Quartz-crystal deposition monitor in cross section.

which corresponds to a sensitivity of <1 ppm in mass, or one monolayer on a typical 350 μm-thick crystal.

The resonant crystal may be crudely modeled as a familiar mass-spring oscillator, for which

$$\nu_r = \frac{1}{T_v} = \frac{1}{2\pi}\sqrt{\frac{k}{m}} \tag{4.36}$$

where ν_r = resonant frequency, Hz
T_v = vibrational period, s
k = spring constant (stiffness), N/m (= kg/s^2)
m = mass, kg

For a mass-loaded crystal, we may then write

$$T_v = T_{vo} + \Delta T_v \propto \sqrt{m_q + m_f} = \sqrt{m_q\left(1 + \frac{m_f}{m_q}\right)} = \sqrt{m_q}\left(1 + \frac{1}{2}\frac{m_f}{m_q}\right) \tag{4.37}$$

which means that

$$\Delta T_v \propto m_f \propto \rho_f h_f \tag{4.38}$$

where ΔT_v = change in T_v due to the loading
m_q = mass of the unloaded crystal
m_f = mass of the deposit
ρ_f = mass density of the deposit, g/cm^3
h_f = thickness of the deposit

The last equality on the right of Eq. (4.37) is valid for $m_f << m_q$ and is sufficiently accurate in practice for $m_f/m_q < 0.1$. Thus, for h_q = 350 μm, a deposit thickness of about $35(\rho_q/\rho_f)$ μm can be tolerated. To further increase the tolerable loading and thereby increase the useful life of a crystal, a correction must be applied to account for the fact that k as well as m is changing in Eq. (4.36). These factors are accounted for together by the acoustic impedance, z, of the crystal-deposit combination, since the resonance is essentially an acoustic standing wave which pervades both the crystal and the deposit. Most instruments allow the programming in of a z-correction factor, the value of which will depend on the film being deposited. This correction extends the tolerable loading to $m_f \approx 0.7\ m_q$. An extensive analysis of the z correction and tabulation of its value for various materials has been reported [17]. Note, however, that the z of a material in thin-film form may be different from the z in bulk because of differences in structure, bonding, ρ_f, and composition. For accurate work, therefore, it is necessary to determine z for the film being deposited by comparing actual film thickness with crystal readings. Deviation of ρ_f from the bulk value also changes the film thickness, which is inferred from the measured mass using Eq. (4.38).

The quartz crystal monitor gives an absolute reading of m_f, and therefore an absolute measurement of arriving vapor flux, if S_c is unity. From this measurement, one can calculate the vapor pressure, p_v, of the source material using the cosine law [Eq. (4.34)] and the Knudsen equation [Eq. (2.19)], provided that the cosine law is valid at the location of the crystal and that a Knudsen-cell source (Fig. 4.1*b*) is employed. This is a convenient way to determine the p_v values of evaporants.

Since the crystal and substrate positions relative to the source are different, the ratio of the vapor fluxes at the two positions must either be determined empirically or estimated using the information in Sec. 4.6. If the cosine-effusion law is being obeyed, the calculation of this geometrical correction factor is straightforward; however, in the case of a sharply lobed or fluctuating flux distribution, large errors can result. Locating the crystal as close as possible to the substrate minimizes geometrical errors. But in that case, too much radiation from a hot substrate can cause drift in the crystal reading, for the reason explained below.

Crystal T control is important. Figure 4.16 shows how ν_r varies with T for the standard "AT-35°" cut of crystallographic orientation. For this cut, ν_r is independent of T near room T, but not at other temperatures. Thus, if the crystal heats up during a deposition measurement due to radiation from the substrate or the source, ν_r will increase and will appear as a *decrease* in mass. If the radiation is coming from the source, transient errors due to quartz T restabilization will occur upon source heat-up or shutter opening at the start of deposition, and again

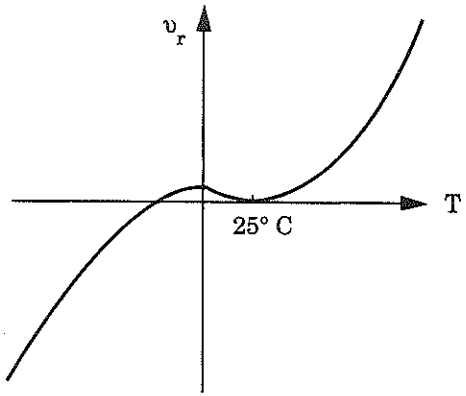

Figure 4.16 Thermal drift in quartz resonant frequency, ν_r, for AT-35° cut of crystal.

upon cool-down at the end, as shown in Fig. 4.17. These transients can be recognized most easily by recording h_f versus time. The actual deposition rate is the slope in the linear region shown in Fig. 4.17. Also illustrated is an abrupt jump in reading due to the crystal spontaneously hopping from one vibrational mode to another. The convex shape of the crystal shown in Fig. 4.15 and the placement of electrode spring contacts off center help stabilize the fundamental vibrational mode, but uneven deposits or high mass loading can cause "mode hopping." Frequency jumps will also occur if the deposit begins to flake off. An additional increase in mass may be seen upon admitting air to the deposition chamber, or admitting any gas which reacts with the film. This increase is an indication of film porosity: the internal surface area along the pores is adsorbing water or becoming oxidized. In such

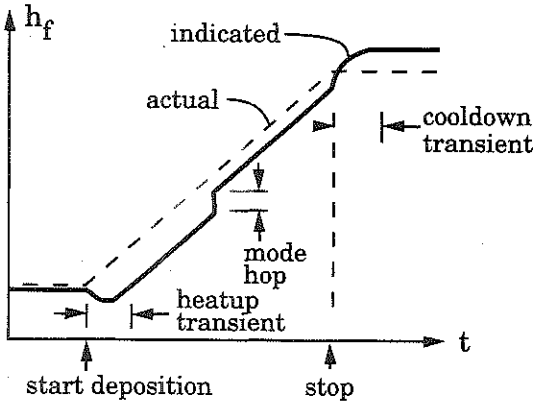

Figure 4.17 Quartz monitor film thickness, h_f, versus time, t (indicated versus actual).

cases, the quartz crystal is a very sensitive microbalance (±0.01 μg) for studying film porosity and reactivity.

Surface T also affects the sticking coefficient, S_c, of many materials. If one wants to monitor the flux actually depositing at the substrate T rather than that depositing at room T, it is possible to operate the crystal at the elevated T despite the steeper dependence of ν_r on T which is shown in Fig. 4.16. In my lab, I have found that satisfactory stability of ν_r can be achieved by applying the feedback T-control techniques discussed in Sec. 4.5.3 to the crystal holder. Above 350° C or so, the piezoelectric effect weakens, and it is difficult to maintain oscillation. If instead of duplicating substrate T, one wants to increase the S_c of a vapor flux toward unity, the crystal can be cooled, or a reactive material can be codeposited. For example, in the monitoring of (AlGa)As deposition, the S_c of As_4 vapor can be increased by codepositing a stoichiometrically excess flux of Al. In the case of the Group II-VI semiconductors, an excess flux of either one of the elemental vapors drives the S_c of the other one toward unity [18].

Loss of oscillation will occur if a liquid is deposited, such as Ga or Hg, or if the film reacts with or dissolves the electrode material, thus destroying the electrical contact. Pt is the most robust of the commonly available electrodes. Quartz monitors can even be used in plasma environments, provided that the holder is well grounded and that the 5 MHz wire is well shielded all the way to the vacuum feedthrough. Materials exposed to plasma or to ion beams can *lose* mass due to ion-sputter removal of surface atoms, as discussed in Sec. 8.5.4. In such cases, the quartz crystal, predeposited with the material of interest, can be used as a sensitive measure of sputtering rate.

4.8.2 Optical techniques

The optical techniques of deposition flux monitoring have several advantages: they are applied directly to the depositing film, yet they involve too low an energy density to perturb the deposition process, and they can be used in high-pressure or plasma environments. On the other hand, their correct interpretation sometimes requires considerable knowledge of the optical properties of film and substrate, and the film must have some transparency. Here, we will briefly review the physics of light interaction with matter and apply it to the two basic rate-monitoring techniques: interference oscillations and ellipsometry. Light *scattering* can be used to monitor film surface roughness during deposition, and this will be discussed in Sec. 5.4.1.

Light (and electromagnetic radiation of other frequencies) consists of sinusoidally oscillating electric and magnetic fields which are oriented transverse to the direction of ray travel and at right angles

(orthogonal) to one another. Interaction with matter is almost entirely due to the electric field, which causes charges in the matter to oscillate. In Chap. 9, we will see how radio-frequency and microwave-frequency electromagnetic fields can energize plasmas by causing their free electrons to oscillate. Here, we begin by considering in-phase light rays a and b of the plane wavefront in Fig. 4.18 as they approach the surface of a transparent solid (or liquid) medium from the gas phase. The incident, reflected, and refracted rays all lie in the "plane of incidence" (the plane of the figure), which is always perpendicular to the solid surface. The electric-field vector, $\mathbf{E}$, may be oriented parallel to the plane of incidence, as shown by $\mathbf{E}_p$, or perpendicular to it, which would be denoted by $\mathbf{E}_s$, since perpendicular is *senkrecht* in German. Other orientations of $\mathbf{E}$ may always be described in terms of its $\mathbf{E}_p$ and $\mathbf{E}_s$ components. The $\mathbf{E}_p$ orientation is known as TM (transverse magnetic) polarization, and the $\mathbf{E}_s$ as TE (transverse electric).

The speed of light in vacuum or gas, c_0, slows down to speed c_1 upon refraction into medium 1 due to the electric polarization which $\mathbf{E}$ induces in the medium, and the speed ratio, c_1/c_0, is called the index of refraction, $\tilde{n}$. Since the frequency of the light, ν, must remain constant, and since

$$\nu = \frac{c}{\tilde{\lambda}} \tag{4.39}$$

the wavelength is proportionately shortened upon refraction; that is,

$$\lambda_1 = \lambda_0 \tilde{n}_0/\tilde{n}_1 \tag{4.40}$$

where λ_0 and λ_1 are the wavelengths in gas and in medium 1, respectively. It is therefore convenient to think of an "optical path length," L,

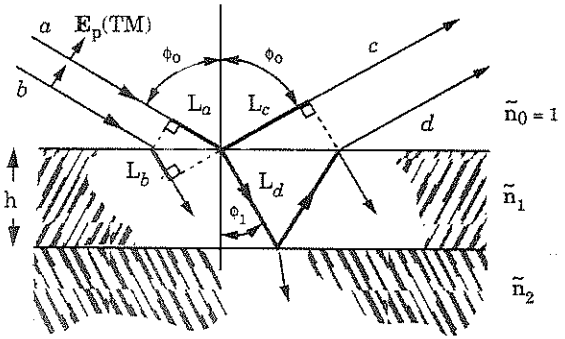

Figure 4.18 Geometry of refraction and interference.

measured in wavelengths, for light traveling in a refractive medium; thus, $L \propto 1/\tilde{n}$. For the rays a and b to continue along in phase as a plane wavefront after being refracted into medium 1, the optical path lengths L_a and L_b, shown by the heavy lines in Fig. 4.18, must be equal. This leads directly to Snell's law:

$$\tilde{n}_0 \sin \phi_0 = \tilde{n}_1 \sin \phi_1 \tag{4.41}$$

where the angles of incidence, ϕ_0, and of refraction, ϕ_1, are measured from the perpendicular to the surface. Another part of ray a is shown being reflected from the surface as ray c. Energy imparted to ray c comes from charge oscillations in medium 1 that are induced by the refracted part of ray a, so these oscillations will be perpendicular to the refracted ray. At a particular value of ϕ_0, ray c will *also* be perpendicular to the refracted ray, and under this condition there will therefore be no component of charge oscillation perpendicular to ray c for TM-polarized light. This means that the energy delivered to ray c will be zero, so no light can be reflected. This particular angle of incidence is known as Brewster's angle, ϕ_B, and it can be shown by trigonometry and Snell's law that

$$\phi_B = \tan^{-1}(\tilde{n}_1/\tilde{n}_0) \tag{4.42}$$

Note that reflection does *not* go to zero for TE polarization, and thus ϕ_B is known also as the "polarizing angle." We will see that the value of ϕ_0 relative to ϕ_B is very important in optical measurements on thin films.

If medium 1 is a thin film of thickness h on a substrate of index $\tilde{n}_2$, light refracted into it will continue to bounce back and forth across h with some reflection and refraction at each interface as shown for ray a in Fig. 4.18. The refracted light leaving medium 1 as ray d after one bounce traverses an optical path length of $2L_d$ within this medium during the time that the reflected ray c traverses a shorter path L_c in the gas before meeting up with ray d in the outgoing beam. The particular film thickness, h_o, for which the path length difference is exactly one wavelength is called the "thickness period," and it can be shown by trigonometry and Eqs. (4.40) and (4.41) that

$$h_o = \frac{\frac{1}{2}\lambda_0}{\sqrt{\tilde{n}_1^2 - \sin^2 \phi_0}} \tag{4.43}$$

The same value of h_o is obtained for rays leaving the medium after any two successive bounces. If all of these rays are in phase, the

intensity of the reflected beam is a maximum; whereas if successive rays are 180° (π radians) out of phase, the rays interfere destructively, and the reflected intensity is a minimum. A phase shift of 2π is induced by the path through a film of thickness h_o, and there can be additional phase shifts upon reflection at each interface depending on conditions, as specified in Table 4.3. There is never a phase shift upon refraction. Note that the reflection situation is less complicated when $\phi_0 < \phi_B$. In that instance, the phase shift is the same at both of the interfaces of Fig. 4.18 when $\tilde{n}_2 > \tilde{n}_1 > \tilde{n}_0$. Then, maxima are obtained when $h = jh_o$, where j is any integer, and minima are obtained when $h = (j + 1/2)h_o$. The same formulae apply to both polarizations. For perpendicular ("normal") incidence ($\phi_0 = 0°$), the first minimum ($j = 0$) occurs at $h = (1/4)\lambda_0/\tilde{n}_1$, and this is the situation for the familiar quarter-wave antireflection (AR) coating used on lenses. If in addition, $\tilde{n}_1 = \sqrt{\tilde{n}_2}$, the reflection from the AR coating is exactly zero. In thin-film monitoring, one can also have $\tilde{n}_2 < \tilde{n}_1$. In that case, the phase shifts at the 0-1 and 1-2 interfaces differ from each other by π for both polarizations, as seen in Table 4.3. Then, the conditions for maxima and minima as described above are reversed.

Deposition rate monitoring by recording the periodic intensity oscillations of monochromatic light is particularly appropriate for optical thin films such as AR coatings, since the optical thickness is the main property of interest. The He-Ne laser ($\lambda_0 = 632.8$ nm) is an ideal light source for films transparent at that wavelength. The technique can be applied to any film which has some transparency. If the film has some absorptivity, the oscillation amplitude will attenuate with increasing thickness as the contribution from the refracted ray becomes weaker. In addition, the phase relationships become more complicated with absorption. In such cases or when the film thickness to be measured is much less than a quarter wave, the more sophisticated optical technique of ellipsometry is preferred.

TABLE 4.3 Phase Shifts upon Reflection of Light from an Interface

Polarization	ϕ_0	$\tilde{n}_b > \tilde{n}_a$*	$\tilde{n}_b < \tilde{n}_a$*
$\mathbf{E}_s$ (TE)	$< \phi_B$	π	0
	$> \phi_B$	π	$0 \rightarrow \pi$ as $\phi_0 \rightarrow \pi/2$
$\mathbf{E}_p$ (TM)	$< \phi_B$	0	π
	$> \phi_B$	π	$0 \rightarrow \pi$ as $\phi_0 \rightarrow \pi/2$

*Light is in medium of refractive index $\tilde{n}_a$; interface is with medium $\tilde{n}_b$; both media are nonabsorbing.

In ellipsometry, one extracts much more information from the reflected light beam by measuring the amplitude ratio and phase relationship of the $\mathbf{E}_p$ and $\mathbf{E}_s$ components; that is, by measuring the polarization ellipse of the reflected light. This concept is illustrated in Fig. 4.19, which shows the **E**-vector oscillations of light waves of both polarizations as they travel in the +r direction. The oscillation of each is described by its variation with time, phase shift, and position:

$$\mathbf{E}_k = \mathbf{E}_{ok} \cos(\omega t + \delta_k - 2\pi r\tilde{n}/\lambda_0) \tag{4.44}$$

where subscript k = p or s polarization
$\mathbf{E}_{ok}$ = amplitude (maximum value) of $\mathbf{E}_k$
ω = angular frequency = $2\pi\nu$
δ = phase angle

The last term may also be written as $\mathbf{kr}$, where the "wave vector" $\mathbf{k} = 2\pi/\lambda = 2\pi\tilde{n}/\lambda_0$. For the two waves shown in Fig. 4.19, inspection shows that $\delta_s = 0$ and $\delta_p = \pi/2$. At any particular t and r, the resultant **E** vector obtained by the superposition of these two waves will have an amplitude and direction equal to their vector sum, $\mathbf{E} = \mathbf{E}_p + \mathbf{E}_s$. Now visualize the evolution of this sum in amplitude and direction as the two waves pass to the right through the r = 0 plane, and visualize the projection of **E** on the p-s plane at r = 0. This projection is shown to the left of Fig. 4.19 for t = 0 and looking into the beam. As t increases and

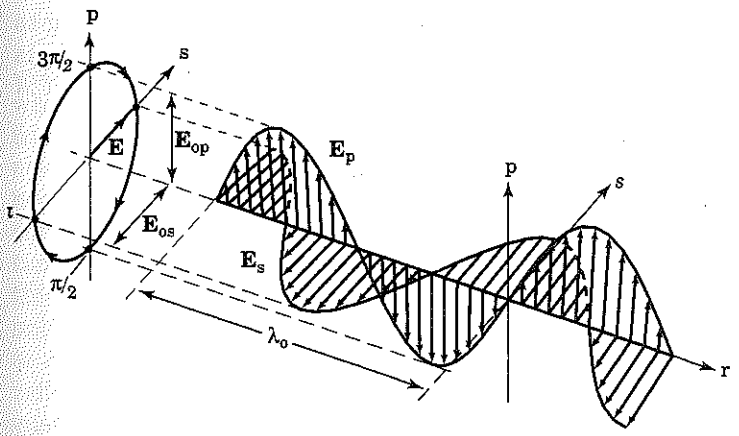

Figure 4.19 Superposition of two light waves to form the polarization ellipse. Here, the p-polarized wave leads the s-polarized wave by $\pi/2$, resulting in right-hand elliptical polarization.

the waves pass to the right, $\mathbf{E}_s$ decreases from its maximum value toward zero, while $\mathbf{E}_p$, starting from zero, becomes negative. For equal amplitudes, $\mathbf{E}_{ok}$, this progression results in $\mathbf{E}$ tracing a circular path in a clockwise direction as shown, and this is called right-hand circular polarization. Had δ_s been $-\pi/2$, the path would have been counterclockwise and the polarization left-hand. If the waves are instead *in* phase ($\delta_p = \delta_s$), the path becomes a line bisecting the top-right and bottom-left quadrants; if they are out of phase ($\delta_p = \delta_s \pm \pi$), the path is a line bisecting the other two quadrants. Also, if one of the amplitudes is zero, a line is obtained. These three latter cases are examples of *linear* polarization. Finally, for other phase and amplitude relationships, the path is an *ellipse* whose major axis is inclined at some azimuthal angle θ as shown in Fig. 4.20. The maximum excursions of the ellipse in the p and s directions are the amplitudes of the electric fields in those directions. The significance of the angle ψ will be clear later. Note that monochromatic light is required to obtain a time-invariant phase relationship between the two waves. The smaller the bandwidth, $\Delta\lambda$, of the light, the longer is the distance of travel in r before the phase relationship is lost. This distance is the coherence length, Δr, and

$$\Delta r = \lambda\left(\frac{\lambda}{\Delta\lambda}\right) \tag{4.45}$$

A good monochromator or a laser is required for Δr to be long enough for ellipsometry.

In ellipsometry, we are concerned with the phase of the reflected light but not with the time dependence, which is simply sinusoidal. The time dependence can be factored out of Eq. (4.44) by switching to exponentials using Euler's formula: $\cos x + i\sin x = e^{ix}$ (which is obtained by comparing the series expansions of these functions). If $\Re$ denotes the real part of these complex numbers, then $\cos x = \Re(e^{ix})$, and

$$\mathbf{E}_k = \Re\left(e^{i\omega t}\cdot e^{i\delta_k}\cdot e^{-i\cdot 2\pi r\hat{n}/\lambda_0}\right) \tag{4.46}$$

It turns out that operating on the complex exponential terms is equivalent to operating on the real part only, at least for present purposes, so the $\Re$ is dropped henceforth. The substitution of $\tilde{n}$ by $\hat{n}$ indicates that the refractive index is also in general a complex number,

$$\hat{n} = \tilde{n} + i\kappa \tag{4.47}$$

Here, κ is the extinction coefficient, which accounts for light absorption and scattering in the film. The reason that the extinction term is imaginary becomes clear upon factoring the $\hat{n}$ exponential term in Eq. (4.46):

$$e^{-i\cdot 2\pi r\hat{n}/\lambda_0} = e^{-i\cdot 2\pi r\tilde{n}/\lambda_0}\cdot e^{-(2\pi\kappa/\lambda_0)r} \tag{4.48}$$

The second term on the right is a *real* number, and it describes the exponential attenuation of light passing through an absorbing medium. Photodetectors actually measure light *intensity*, $I = |\mathbf{E}_k|^2$, and applying this relationship to the attenuation term leads to the familiar Beer's law:

$$\boxed{\frac{I(r)}{I(r=0)} = e^{-(4\pi\kappa/\lambda_0)r} = e^{-\alpha_r r}} \tag{4.49}$$

where α_r is the light absorption coefficient. Note that α_r is different from the radiation *absorptivity* discussed in Sec. 4.2. The absorptivity equals the fraction of incident light that is absorbed in the *total* thickness of the medium rather than being reflected from its surface or transmitted through to the other side, whereas α_r expresses attenuation per unit of thickness. In ellipsometry, the effect of extinction in a thin film shows up as a reduced amplitude for light reflected from the substrate interface. Additional loss of reflected amplitude occurs due to refraction into the substrate. There are also phase changes due to reflection from interfaces and due to path length differences, as discussed above.

The amplitude and phase changes upon reflection of each polarization component (k = p or s) are accounted for together by the complex reflection coefficients

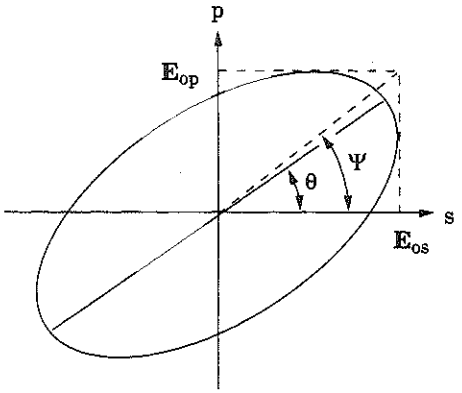

Figure 4.20 Geometry of the polarization ellipse of the reflected light. The ellipsometric angle, ψ, is oriented as shown only when the *incident* amplitudes are equal.

$$r_k = |r_k| e^{i\delta_k} \tag{4.50}$$

where $|r_k|$ is the reflected/incident amplitude ratio, and where it has been assumed that the phases of both *incident* components are zero. Here, we have dropped both the t and the r factors of Eq. (4.46) and are considering only the δ factor at a fixed t and r. This expression is called the "phasor" representation of a periodic function, which should not be confused with the energy-beam sidearm carried by Enterprise crew members. It is common in ellipsometry to direct at the surface linearly polarized light which is inclined at an azimuthal angle of 45° to the plane of incidence, so that the amplitudes and phases of each component are equal. The light is polarized by passing it through a "dichroic" material such as a Polaroid sheet. Dichroic materials cause polarization by having a much higher extinction coefficient for one polarization than for the orthogonal one. Using such light, the ratio of the complex reflection coefficients is obtained directly by measuring the polarization ellipse of the reflected light. These are the basic data which are obtained from ellipsometry, and the ratio is expressed as:

$$\rho = \frac{r_p}{r_s} = \frac{|r_p|}{|r_s|} e^{i(\delta_p - \delta_s)} = \tan\Psi e^{i\Delta} \tag{4.51}$$

In the case of equal amplitudes for the incident light, ψ is also given by the ratio of the reflected *amplitudes* as shown in Fig. 4.20, which also shows the azimuthal orientation angle of the ellipse, θ. Since ψ can vary over 90° and Δ over 360°, the data space forms a hemisphere in polar coordinates.

In practice, the reflection ellipse is mapped by passing the reflected light through a spinning polarizer into a photodetector. From the ellipse shape and azimuth, the ellipsometric angles, ψ and Δ, can be calculated. These angles uniquely determine the $\hat{n}$ of a bare substrate. Upon film deposition, the new ψ and Δ plus the previously measured $\hat{n}$ of the substrate uniquely determine both the $\hat{n}$ and the thickness, h, of the film. The complex dielectric constant of the film, $\varepsilon_d = \varepsilon_1 + i\varepsilon_2$, can also be calculated, since it is related to $\hat{n}$ by $\varepsilon_d = \hat{n}^2$. The imaginary part of ε_d is associated with optical attenuation in the film, as was the case with $\hat{n}$. The mathematics of transforming the raw ellipsometric data into ψ and Δ and thence into film properties is very complicated; but fortunately, good software is available. It is possible, for example, to plot ψ and Δ in real time during film deposition. In this mode, if the incident angle, ϕ_0, is set near to the ϕ_B [Eq. (4.42)] *of the film* which is being deposited, Δ can detect less than one monolayer of deposition! This seems surprising, since the light wavelength amounts to thousands of monolayers, but it comes about as follows. Recall that there is no reflection of the p component at ϕ_B. Therefore, only the s component is reflected from the surface of the film, while the p component is completely refracted through the film and thus becomes retarded in phase by the added optical path length. We know that one thickness period of film, h_0 in Eq. (4.43), corresponds to a 360° phase retardation, so for a typical h_0 of 100 nm, the h sensitivity of this technique is 0.3 nm per degree of Δ. The choice of ϕ_0 is important, because if it is too close to ϕ_B, too much sensitivity results, meaning that small errors in measuring ϕ_0 will result in large errors in calculated $\hat{n}$ and h. Another characteristic of ellipsometry to be wary of is that the sensitivity of the ellipsometric angles to changes in the optical properties of the film varies widely with conditions. For example, sensitivity to the $\hat{n}$ and h of a film are low when h is near the thickness period of Eq. (4.43) (Tompkins, 1993).

Figure 4.21 shows results obtained [19] for the ψ, Δ trajectory in depositing $Al_{0.25}Ga_{0.75}As$ onto a GaAs substrate by CVD at 600° C. Point 1 corresponds to the ψ and Δ for the GaAs substrate before deposition. With increasing deposition, the trajectory follows a logarithmic

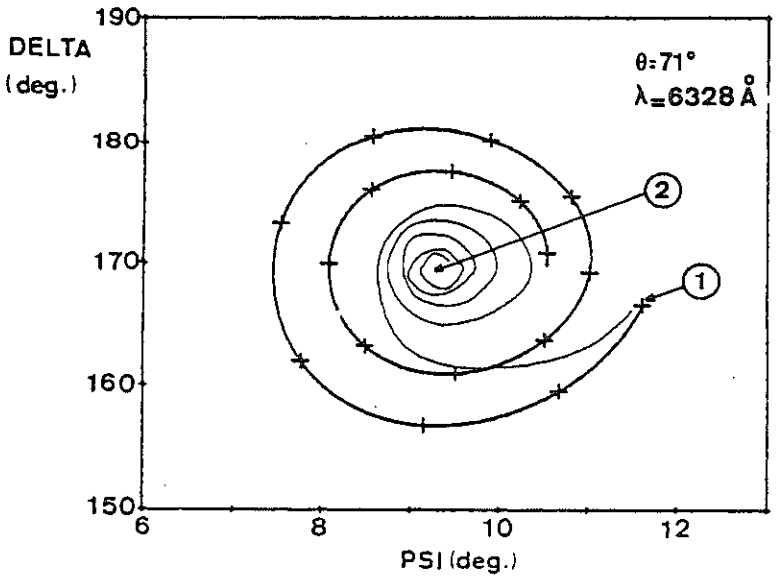

Figure 4.21 Experimental Ψ–Δ trajectory during CVD of $Al_{0.25}Ga_{0.75}As$ onto a GaAs substrate (heavy line with + marks) and calculated trajectory using abrupt-interface model (light line). (Source: Reprinted from Ref. 19 by permission. © 1981 by Annual Reviews, Inc.)

spiral as indicated by the heavy line, which has been marked with crosses at 10 nm increments of h. The spiral would eventually converge to point 2, which corresponds to the ψ and Δ for bulk $Al_{0.25}Ga_{0.75}As$. If the film were completely transparent to the 632.8 nm He-Ne light used, the spiral would not converge, but would repeatedly circulate back on itself, passing through point 1 after the deposition of each thickness period, h_0. The observed convergence results from the attenuation, or extinction, of the refracted ray as h increases. Comparison of these data with a model calculation provides additional information about the depositing film. The narrow-line trajectory plotted in Fig. 4.21 is the theoretical trajectory for the deposition of a smooth film having a uniform $\hat{n}$ equal to that of $Al_{0.25}Ga_{0.75}As$. The considerable deviation from the data indicates that the $\hat{n}$ of the film is *not* initially that of $Al_{0.25}Ga_{0.75}As$, but instead it gradually changes from that of GaAs to that of $Al_{0.25}Ga_{0.75}As$ with increasing h. Intermixing of interfaces is a common problem in thin-film deposition, especially in CVD, and its measurement in real time is a powerful tool in pursuing its solution.

The above results at the same time point up the principal shortcoming of ellipsometry, which is that the calculated film properties must be based on an assumed model for the optical structure. Thus, the more that is known about the film(s), the more likely it is that a correct and unambiguous interpretation of the data can be made. For example, the development of roughness in the depositing film can be misinterpreted as a reduced $\hat{n}$. This is because roughness on a scale much smaller than λ_0 will not cause detectable scattering, but rather will appear optically as an "effective medium" composed of film material plus gas inclusions. This too can be modeled, but it needs to be known whether roughness is indeed the cause of the reduced $\hat{n}$. In addition, effective-medium theory is complicated, and most of the effective-medium models, including the popular Bruggeman approximation, assume isotropic inclusions. Unfortunately, rough or porous thin films often have a columnar structure rather than an isotropic one. These ambiguities, like all ambiguities, are best resolved by collecting more data, and the easiest way to do this in ellipsometry is by varying λ_0; that is, by employing *spectroscopic* ellipsometry. Clearly, fitting a model to the ellipsometric data over a wide λ range is a much more rigorous test of it than fitting it at only one λ. Moreover, it usually happens that a certain λ is particularly responsive to the film layers under study, in which case that λ can be selected for real-time monitoring and feedback control of deposition rate and composition [20]. Spectroscopic ellipsometry is also useful for determining the h and $\hat{n}$ values of multilayer film structures *after* deposition, as will be discussed in Sec. 10.1.1.

4.9 Conclusion

In this chapter, we have examined the supply of source material by thermal evaporation and the transport of the resulting vapor to the substrate within a high-vacuum environment. Although a material's p_v versus T behavior can be estimated by thermodynamics knowing only the boiling point and heat of evaporation, its net evaporation *rate* can be calculated from p_v only in the cases of true Knudsen-cell sources or atomic vapors of metals. In other cases, p_v only gives an *upper limit* to net evaporation rate. This points up a general fact of chemistry, namely that the kinetics of a process are always more difficult to estimate than the equilibrium situation. The evaporation of alloys and compounds introduces the additional complication of vapor component ratios drifting with time due to changing source composition. Materials knowledge, source design, and T stability are all important in controlling supply rate. Once the vapor has been generated, it must be transported uniformly to the substrate. In that step, key factors are the geometry between source and substrate and the monitoring of vapor flux or of deposition rate. We have seen that there are some very sensitive monitoring techniques available, which can also be used for feedback control of supply rate. Having now dealt with evaporation and transport, we are ready to examine the deposition itself.

4.10 Exercises

4.1 For Al evaporating at 1100° C, $\Delta_v H$ = 318 kJ/mole. Assuming atomic vapor, what fraction of this energy goes into $p\Delta V$, what fraction into kinetic energy, and what fraction into potential energy?

4.2 The sublimation T of TiF_4 is 284° C. At that T, $\Delta_f H_c$ = –1639 kJ/mole and $\Delta_f H_v$ = –1551 kJ/mole. (a) Write an equation for p_v (Pa) with T (in K) as the only unknown. (b) What is p_v at room T (25° C)? (c) What is the TiF_4 effusion rate at room T from an ideal Knudsen cell having a 1 mm-diameter orifice?

4.3 For water in equilibrium with its vapor at 100° C: (a) What is average time between arrivals of vapor molecules within a radius of three atomic distances (about 1 nm) from a particular surface site? (b) How much time does it take the average arriving molecule to traverse the last three atomic distances above the surface, neglecting acceleration due to potential interaction? (c) What is the probability that there will be an impinging molecule within three atomic distances of a particular surface site?

4.4 The p_v of arsenic vapor over the elemental solid is 1 Pa at the operating T of a particular Knudsen cell. If the evaporation coef-