4. Properties and Characterization of Thin Films

4. 1. Film Thickness

4.1.1. Introduction

To make sure that coatings which were produced by a given process satisfy the specified technological demands a wide field of characterization, measurement and testing methods is available. The physical properties of a thin film are highly dependent on their thickness. The determination of the film thickness and of the deposition rate therefore is a fundamental task in thin film technology.

In many applications it is necessary to have a good knowledge about the current film thickness even during the deposition process, as e. g. in the case of optical coatings. Therefore one distinguishes between thickness measurement methods which are applied during deposition ("in situ") and methods by which the thickness can be determined after finishing a coating run ("ex situ").

4.1.2. Gravimetric Methods

4.1.2.1. General

These are methods which are based on the determination of a mass. The film thickness d can be calculated from the mass of the coating m if the density ρ and the area A on which the material is deposited are known:

$$d = m/(A\rho) \qquad (4.1)$$

For this method one has to bear in mind that the density of a coating may deviate significantly from that of the bulk (e. g. due to porosity or implanted interstitial atoms). For exact measurements calibration is necessary.

4.1.2.2. Weighing

The simplest method for film thickness determination is most probably the determination of the mass gain of a coated substrate with an exact balance. Although, together with the problem of film density mentioned above, also other obstacles exist (e. g. condensation of water vapor from the ambient) it is possible to determine film thickness with sufficient accuracy for several practical applications.

4.1.2.3. Quartz Oscillator Method

This set-up, which is commonly called "quartz oscillator microbalance (QMB)", is generally used for the in-situ determination and control of the film thickness and deposition rate in the case of PVD methods. In commercially available designs film thicknesses in the range from 0,1 nm - 100 µm and deposition rates in the range from 0,01 - 100 nms^{-1} are permanently displayed (see Fig. 4.1.).

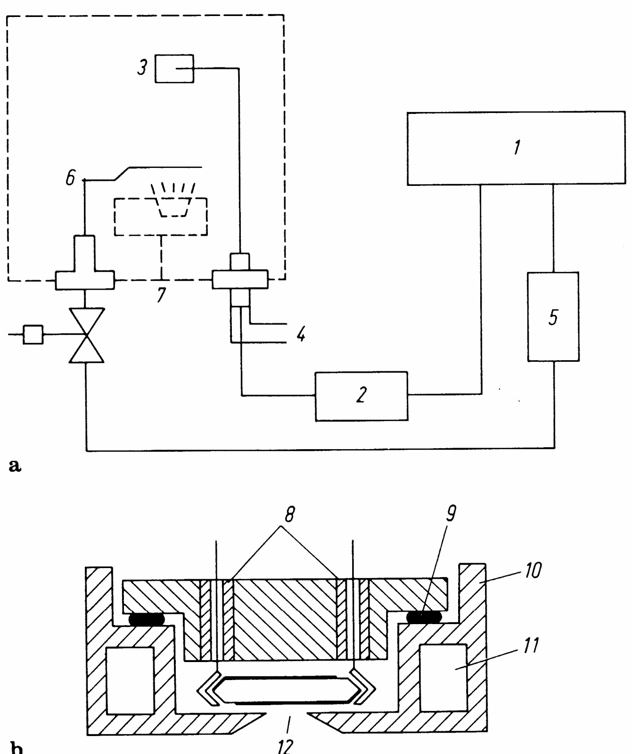


Fig. 4.1.: Quartz oscillator microbalance and control of coating thickness and deposition rate [1, p. 35]:

a schematic of experimental set-up, **b** Oscillator head (schematic)

1 Quartz oscillator microbalance with D/A-converter; **2** oscillator;
3 measurement head with quartz oscillator; **4** water cooling of quartz crystal;
5 shutter control; **6** shutter; **7** vapor source;
8 electrical connects; **9** seal; **10** Cu block;
11 water cooling; **12** aperture

The method of mass determination was developed in 1959 by Sauerbrey and is based on the change of the resonance frequency of an oscillating quartz crystal

$$f = N / d_q \tag{4.2}$$

if the crystal is coated by a film with the mass

$$\Delta m = \rho A d\,. \tag{4.3}$$

d_q is the thickness of the quartz crystal and N is the spring constant of the crystal, which amounts to 1,67mmMHz in the case of the so-called AT cut (see Fig. 4.2., minimum temperature coefficient of the resonance frequency). A is the area, ρ the density and d the film thickness in the coated region of the quartz crystal.

The mass amount Δm acts similar to a thickness change of the quartz crystal by

$$\Delta d_q = \Delta m / (\rho_q A_q) \tag{4.4}$$

where A_q and ρ_q are the area and the density of the quartz plate, respectively. In this case the resonance frequency decreases proportionally to d if $\Delta f << f$ is valid. With

$$\frac{\Delta f}{\Delta d_q} = -\frac{N}{d_q^2} \tag{4.5}$$

one obtains

$$-\Delta f = \frac{N\Delta m}{d_q^2 \rho_q A_q} = \frac{A f^2 \Delta m}{A_q N \rho_q A} = C\frac{\Delta m}{A} = C\rho d \tag{4.6}$$

where the constant $C = \frac{A}{A_q}\frac{f^2}{N\rho_q}$ is a measure for the weighing sensitivity. A more exact calculation yields a function $F(\frac{A}{A_q})$ instead of $\frac{A}{A_q}$.

Example:

For a quartz crystal with f = *6MHz* and d_q = *0,28mm* C = *8MHz/(kgm^{-2})*. A coating thickness d = *0,1nm* and $\rho = 10^4 kgm^{-3-3}$ yields $\Delta m/A = 10^{-6} kgm^{-2}$.

From these values one obtains Δf = *-8Hz,* which can be measured reasonably well. Also for d = 1µm Δf with -80kHz is still small when compared to f, so that the region of measurement extends from approx. 0,1 nm to some µm. In most cases one oscillator platelet allows to monitor 10 - 100 deposition runs.

If the quartz crystal changes its temperature by an amount ΔT because of radiation emitted from the evaporator and/or because of the substrate heating or because of the heat of condensation of the deposited material the resonance frequency changes by $\Delta f = \beta f \Delta T$. The temperature coefficient β of the quartz crystal has a minimum if the single crystal is cut according to the so-called AT cut and amounts to $\beta = -1.10^{-6} K^{-1}$ (see Fig. 4.2.). For f = *6MHz* the temperature dependence $\Delta f/\Delta T$ = *-6Hz/K*. By water cooling the quartz crystal the temperature influence is suppressed so far that film thicknesses of 100 nm can be measured with a resolution of 0,1%.

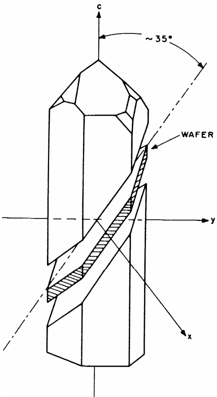


Fig. 4.2.: Quarz AT-cut [5, p. 1-108]

For commercially available equipment good linearity is only ensured for mass gains of 10% relative to the initial mass of the oscillator crystal. Recently developed devices can sustain much higher mass gains because non linearities are mathematically implemented in the algorithm of measurement. Benes (Institute of General Physics, Vienna University of Technology) has developed a method of measuring acoustic impedances (z-Values). By exciting different higher harmonics and by their exact measurement it is possible to determine the z-values which are necessary to experimentally tackle with the determination of high mass additions.

4.1.2.4. Microbalance

This very exact method is, unfortunately, not suitable for practical applications. It is therefore mostly used for calibrating other measurement processes. All microbalances used for film thickness measurement basically work by compensating of the coating weight by a counteracting force. The compensation can be accomplished by optical or electrical (turning coil) systems. It is possible to measure the mass thicknesses m/A as well as the coating rates $\dot{m}/A$ (see Fig. 4.3.).

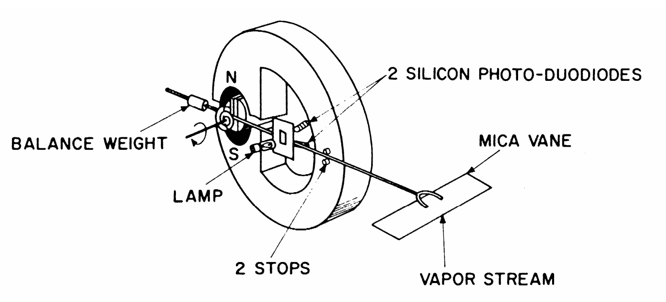


Fig. 4.3.: Microbalance [5, p. 1-104]

4.1.2.5. Dosed Mass Supply

Many deposition processes are executed with dosed mass supply, i. e. by using a mass flow (given in kgs^{-1}) which is held constant due to defined geometry of the plant and due to temporarily unchanging process parameters. This method demands calibration e. g. by weighing or chemical microanalysis to find the connection between the mass thickness *m/A* at the substrate and the mass flow.

After calibration it is sufficient in most cases to keep the relevant process parameters constant to obtain the same thickness values in similar deposition times within certain margins.

If the mass supplied per unit time, *m*, or equally the increase of film thickness per unit time, *d*, is known and the deposition rate *a* is constant: Ist die zugeführte Masse pro Zeiteinheit *m* bzw. der Schichtdickenzuwachs pro Zeiteinheit *d* bekannt und ferner die Depositionsrate *a* konstant:

$$a = \left(\frac{dm}{dt}\right) / A = \rho\left(\frac{dD}{dt}\right) = konst \qquad (4.7)$$

then the film thickness *D* grows as

$$D = \frac{at}{\rho} \tag{4.8}$$

i. e. proportionally to the coating time *t.*

Because of its simplicity the method of dosed mass supply is widely used: for thermal spraying, CVD processes, for galvanic and electroless deposition, for build up welding and in many applications based on sputtering and ion plating.

4.1.3. Optical Methods

4.1.3.1. General

Optical coatings, unlike other applications, require the measurement of the film thickness as exacty as possible during deposition. Therefore film thickness monitors are used which, especially in the case of multi coated optics (interference filters etc) are incorporated into closed loop controls by (in some cases quite complex) software components.

4.1.3.2. Photometer Method

This metod is mostly used in PVD processes for the production of single layer and multiplayer coatings for optical applications. It measures the optical thickeness *nd* so that it allows for a compensation of changes in the refractive index *n* by corresponding changes of the film thickness *d*.

With a photometer (see Fig. 4.4.) the intensity of light reflected on both interfaces of the sample or transmitted through the sample is measured. These intensities, related to the initial intensity, determine the reflectivity R and the transmittivity T of the film. The impinging light passes an interference filter which acts as a monocromator and is modulated by a chopper so that disturbances by stray light from the surrounding are avoided. The light ray is directed towards the substrate or a reference glass which is located in a test glass exchange unit. The light intensities are measured by photomultipliers. Computers allow for the automated control of the coating process which is especially important for multiplayer coatings.

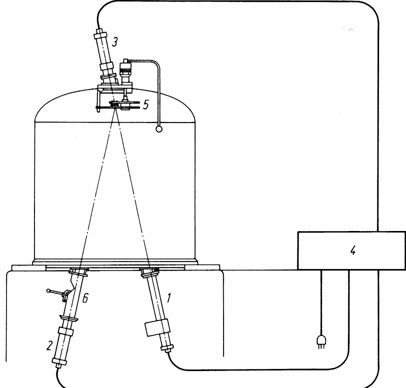


Fig. 4.4.: Photometer-methode for film thickness measurement [1, p. 38]:
1 modulated light source; **2** detector for light reflected at the coating;
3 detector for light transmitted through the coating;
4 display unit; **5** test glass exchanger; **6** beam deflection

Although it is possible to calculate the film thickness d from R and T by using given values of the optical constants n and κ (i. e. real part and imaginary part of the refractive index, respectively) it is in most cases easier to deduce d from measured curves (see Fig. 4.6.). In the case of absorption free or weakly absorbing films R and T change periodically with increasing d due to interference effects if coating and substrate have different refractive indices n (see fig. 4.5.).

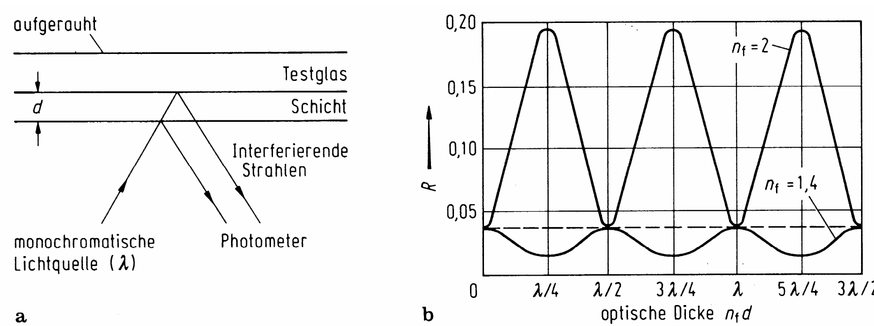


Fig. 4.5.: Two beam interference for a dielectric absorption free coating [1, p. 39]: **a** experimnental situation; **b** reflectivity R as a function of the optical thickness $n_f d$ of the film for different refractive indices n_f. The refractive index of the test glass is n_s = 1,5.

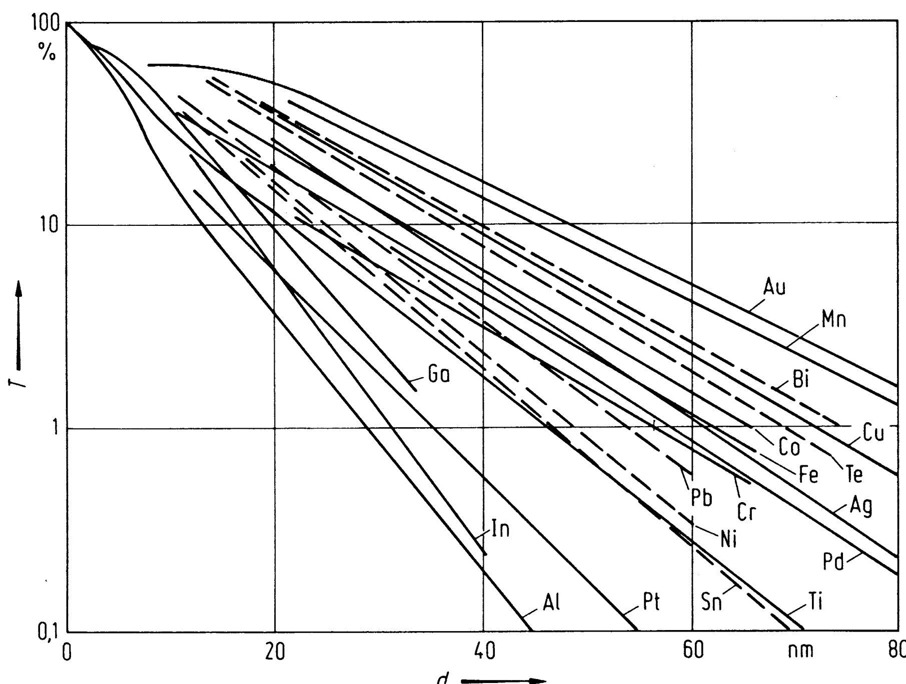


Fig. 4.6.: Transmissivity T in dependence on the film thickness d, measured at 550 nm; substrate: glass. [1, p. 39]

As a practical example Fig. 4.7. shows the development of R and T for the production of a highly reflecting dielectric mirror consisting of highly refracting ZnS and lowly refracting MgF_2 $\lambda/4$ coatings.

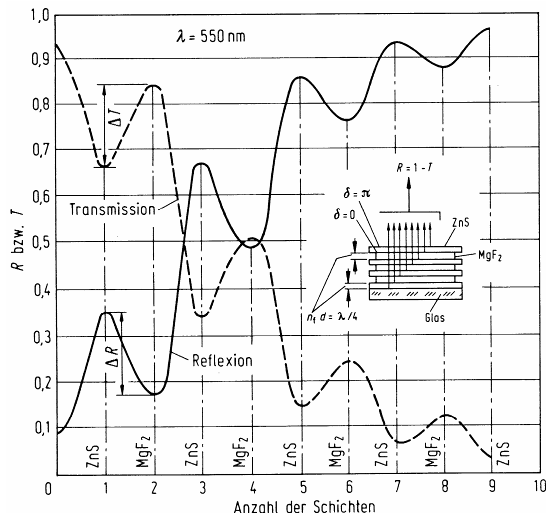


Fig. 4.7.: Reflectivity *R* and Transmittivity *T* of a system of alternatingly evaporated highly refracting ZnS (n = 2,3) and lowly refracting MgF_2 (n = 1,38) $\lambda/4$ coatings [1, p. 40]

4.1.3.3. Tolansky Interferometer

The basic principle of this multiple beam interferometer is as follows: a thin transparent glass slide is put onto a highly reflecting surface. The glass slide is tilted against the surface by an angle α. If this set-up is illuminated by monochromatic light, multiple beam interferences lead to the formation of equidistant interference lines. The more partial beams interfere, the sharper the lines. Their distance depends on α. A film thickness can be determined by this method as follows: First a schratch is made into the film which reaches down to the substrate (it is also possible to mask a part of the substrate during deposition which leads to the formation of a step). Then the sample is coated by a highly reflective layer. Because of the scratch the distance reflector/interference slide is changed, which leads to an offset of the interference lines relative to each other. This situation ids shown in Fig. 4.8. The film thickness *d* is given by

$$d = \Delta N \frac{\lambda}{2} \tag{4.9}$$

where ΔN is the number (or the part) of the lines by which the interference lines are shifted by the scratch or the step.

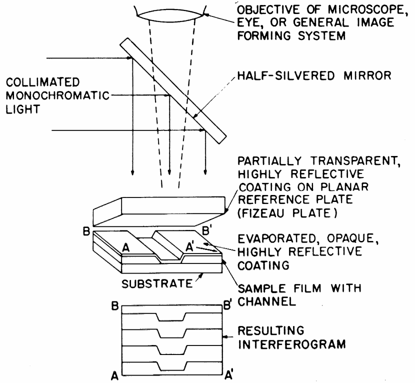


Fig. 4.8.: Schematic of the Tolansky interferometer [5, p. 11-8]

Tolansky interferometer are commercially available as add ons for optical microscopes. The advantage of fast and easy use is counteracted by the disadvantage that a suitable scratch or step has to be present. In addition the step (scratch) has to be covered by a highly reflective coating, mostly Ag. The resolution lies at approx. ± 1 nm if one works with strictly monochromatic light and highly reflactive samples and interference slides.

4.1.3.4. FECO Method

The FECO (Fringes of Equal Chromatic Order) method has an even higher resolution than the Tolansky interferometer. Given a careful measurement a resolution of 0,1 nm can be achieved. The principle is displayed in Fig. 4.9. Parallel white light is illuminating the combination of sample and reference slide. The reflected light is focused into the entrance slit of a spectrograph via a semitransparent mirror. The image of the step has to be normal to the entrance slit. The obtained spectra (interferograms) are displayed in Fig. 4.10. Dark interference lines are observed at the wavelengths

$$\lambda = 2t / N \tag{4.10}$$

where N is the interference order.

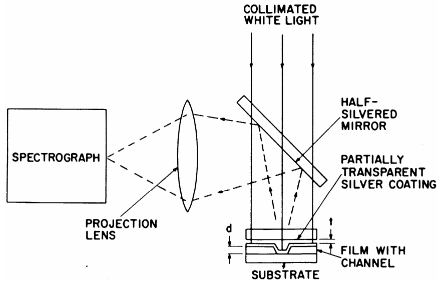


Fig. 4.9.: Schematic of the FECO method [5, p. 11-9]

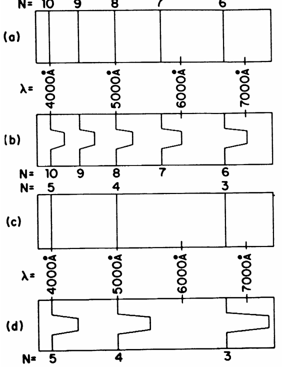


Fig. 4.10.: **a** Interferogram for a slide distance t=2µm
b Interferogram for a, but with a scratch of d=100nm depth
c Interferogram for a slide distance t=1µm
d Interferogram for b, but with a scratch of d=100nm depth
[5, p. 11-9]

Since t is not known the following analysis has to be performed: If the order N_1 is associated with the wavelength λ_1 then the order N_1+1 is associated with a shorter wavelength λ_0. Then

$$N_1\lambda_1 = (N_1 + 1)\lambda_0 = 2t \tag{4.11}$$

is valid and

$$N_1 = \frac{\lambda_0}{\lambda_1 - \lambda_0} \tag{4.12}$$

This allows the determination of the distance *t*:

$$t = \frac{N_1 \lambda_1}{2} = \frac{\lambda_1 \lambda_0}{2(\lambda_1 - \lambda_0)} \tag{4.13}$$

To determine the film thickness *d* the shift of the lines of order N_1 is used. If these lines are located at the wavelength λ_1

$$t + d = \frac{N_1 \lambda_2}{2} \tag{4.14}$$

has to be valid. Then *d* can be calyulated by

$$d = \frac{N_1 \lambda_2}{2} - \frac{N_1 \lambda_1}{2} = \frac{\lambda_0(\lambda_2 - \lambda_1)}{2(\lambda_1 - \lambda_0)} \tag{4.15}$$

4.1.3.5. Other Optical Methods

Apart from the methods described previously the optical determination of film thickness can be done by various other techniques. Some examples are: The Nomarsky-interferometer, wnich is available as add on for optical microscopes; the VAMFO (Variable Angle Monochromatic Fringe Observation) method allows to determine the film thickness and the refractive index in situ; the same is possible with Ellipsometry.

4.1.4. Direct Methods

4.1.4.1. General

Direct methods allow the determination of film thickness either by mechanical profiling or by observation in microscopes.

4.1.4.2. Stylus Method

The film has to exhibit a step on a plane substrate. A diamond stylus (tip curvature approx. 10 µm) is pulled along the surface at constant velocity. The step height is measured by a pick-up system. Prerequisite for an exact measurement are a suitable hardness of the film and a plane substrate. To prevent destruction of the sample the load on the tip can be reduced to approx. 10 µN.

The metering range ranges from some 5 nm to some 10 µm at a resolution of some Å in the most sensitive range. An additional application of these commercially available devices is the imaging of surface profiles and the acquisition of roughness values.

4.1.4.3. Optical and Electron Microscopy for Thickness Measurement

For the determination of the film thickness with the optical microscope e. g. metallographic cross sections are used. In the Transmission Electron Microscope (TEM)

replica of a step in the film or cross sectional preparates can be investigated. In the Scanning Electron Microscope (SEM) the step itself or a fracture surface of the film can be imaged.

The resolution and the metric range depends on the instrumnent and on the magnification. In the optical microscope film thicknesses up to some mm with a resolution of 0,1 µm can be determined. In the SEM a resolution of 5 nm can be achieved and in the High Resolution TEM (HRTEM) 0,1 nm are possible.

4.1.5. Film Thickness Measurement by Electrical or Magnetic Quantities

4.1.5.1. Resistance Method

This metod is used in the case of PVD processes for the determination of the film thickness of metallic coatings. The monitoring element is an insulating plate with two parallel line contacts between which a film is deposited through a mask. The resistance R as a measure of film thickness is controlled via a bridge circuit. The deposition rate is determined by electronic differentiation.

With the aid of a zero point indicator the deposition process is stopped if the setpoint of the film thickness is reached. The metric range lies between 1 nm and 10 µm. Applications are metal films for integrated circuits, resistance films made from NiCr, metallized foils etc.

4.1.5.2. Capacitance Method

In analogy to the previous method the film thickness of insulating coatings can be determined by a monitoring element consisting of comb shaped, interlocking plane electrodes which allow the measurement of the capacity change during deposition.

4.1.5.3. Eddy Current Method

The thickness of insulating coatings on non ferrous metal or of non ferrous coatings on insulating substrates can be measured by this method. The measurable quantity is e. g. the voltage applied to a RF coil which is modified by eddy currents in the non ferrous metal. Since this quantity is also dependent on the conductivity of the non ferrous metal a calibration is necessary. The method is mostly applied in polymer metallization.

4.1.5.4. Magnetic Method

This method is applied to films which are deposited on a substrate of plane ferritic steel. It is based on the measurement of the adhesive force of a magnet put on the coating (non ferrous metal, lacquer, polymer) which depends on the film thickness. Since this force also depends on the permeability of the steel a calibration is necessary. Also Ni as coating material is accessible to thickness measurement after calibration.

4.1.6. Thickness Measurement by Interaction with Particles

4.1.6.1. Evaporation Rate Monitor

Film thickness or deposition rate monitors were developed especially for applications based on evaporation technology. To control the vapor density in the vicinity of the substarate the vapor is ionized at this position by collisions with electrons emitted from a glow filament. The ion current is measured. Some elder designs for these devices are displayed in Fig. 1.11.

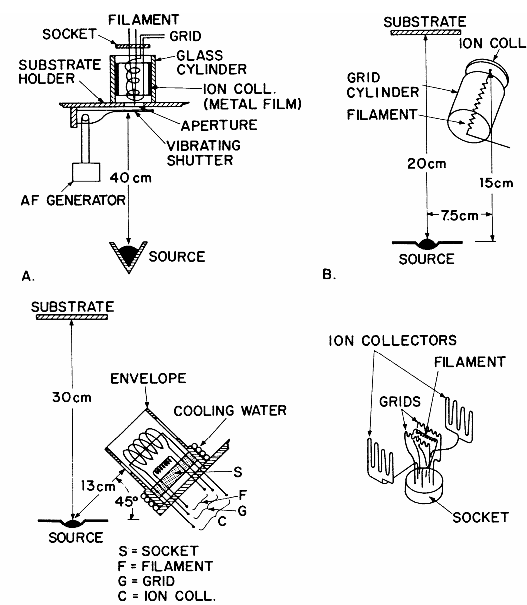


Fig. 4.11.: Thickness and rate monitors based on the ionization principle [5, p. 1-98]

More recent devices analyze the ion current by a quadrupole mass spectrometer. By this method the evaporation rates of simultaneously evaporated materials can be determined.

4.1.6.2. Other Methods

... Beta (electron)-backscattering: especially suited for the film thickness determination of noble metal films of and of metal films on printed circuit boards.

... X-ray fluorescence: also the thickness of multilayer systems can be determined. Because the method is non-destructive and exhibits a high throughput it is used perferrable in quality control e. g. for abrasion resistant coatings.

... Tracer-methods: either the coating or the substrate has to contain radioactive "tracer" atoms.

4.2. Roughness

4.2.1. Introduction

With decreasing thickness the surface structure of a coating gains more and more importance. In the extreme case of ultrathin films the surface roughness may be in the order of the film thickness and can influence all film properties such as mechanical, electrical, magnetical or optical properties. Also film morphology, inner structure, texture and crystallinity are strongly connected to roughness evolution. This section shall briefly discuss basic roughness types, the mechanisms of their origin, roughness measurement and roughness quantification.

4.2.2. Types of Roughness

Generally, the reason for the development of roughness during the deposition process is the finite extension of the film forming particles and their random, temporally and spatially uncorrelated impingement at he growth front. The "building blocks" of the film do not necessarily have to be single atoms as it is the case for PVD coatings. They can also be complex molecules (e. g. for organic coatings) multi-particle aggregates as e. g. for cluster deposition or macroscopic aggregates like the ceramic or metallic droplets in the case of thermal spraying.

4.2.2.1. Stochastic Roughness

The simplest possible model of roughness development is the perpendicular impingement of not nearer specified particles with a finite extension a on random positions of a quadratic lattice at random times on an initially completely flat surface. A particle is added to the material ensemble (from now on called "aggregate") as soon as it has a nearest neighbor beneath itself. Particles which are members of the aggregate but have no nearest neighbor above form the so-called "active surface" and constitute the growth front of the film. Particles can only be incorporated into the aggregate upon attachment to the acitve surface. This situation is diaplayed for a one dimensional profile h(x) in Fig. 4.12.

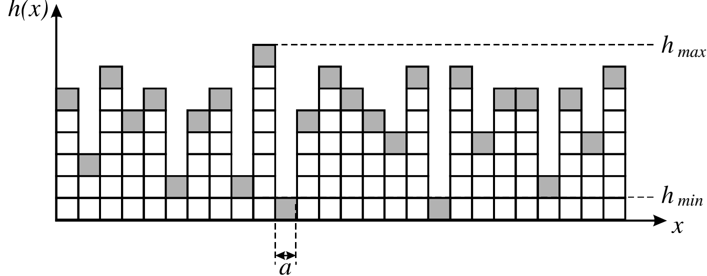


Fig. 4.12.: *Stochastic growth: particles marked gray constitute the active surface.*
h_{max} is the maximal, h_{min} is the minimal heighth value

The simple growth mechanism sketched above leads to the formation of an aggregate which consists of neighboring collumns with completely uncorrelated height values. The terminus "uncorrelated" means that from the height of one column it is impossible to draw conclusions on the height even of the nearest neighbor collumn. The aggregate is dense beneath the active surface, it contains no volume defects as it is visible from Fig. 4.12. Models of roughness evolution in which the active surface is a single valued function of the position on the substrate are called "Solid on Solid" (SOS)-models. Overhangs or closed pores in the bulk do not occur in these models by definition.

The roughness of the active surfaces increases with time. No quantification of "roughness" was given up till noe, but for the time being roughness can be considered as a

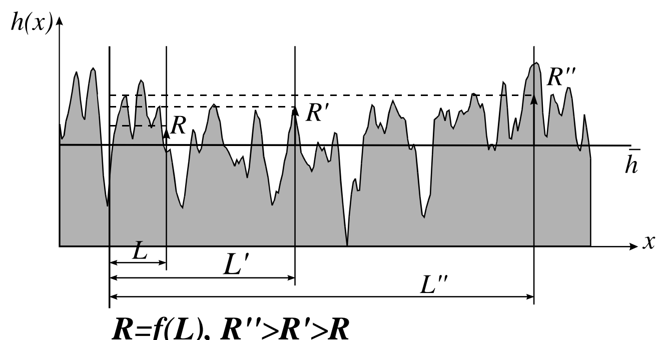
measure for the difference between the maximum and the minimum height value in the aggregate. The temporal evolution of roughness is dependent ofn the specific model used and therefore experimentally determined roughness values can be correlated to certain growth models.

4.2.2.2. Self Affine Surfaces

Even small modifications of the previously described model of stochastic growth have significant consequences for the detailed shape of the active surface. At first, the SOS character of the growth model shall be conserved. A first approximation to realistic conditions can be implemented if one allows a particle to find a site with as many nearest neighbors as possible within a certain distance from the impingement site. This is a model of particle migration along the surface by surface diffusion and particle attachment to energetically favoprable sites with many nearest neighbors, such as steps, kinks or point defects.

These relaxation mechanisms lead to the formation of lateral correlations in the active surface. This means that height values cannot change abruptly within the vicinity of a given point at the surface. Within the zone where a particle may find the energetically most favorable site, which enters the model as a free paremeter one can assume that height values change only gradually.

A big difference to the model of stochastic growth is that the roughness of the surface is not only dependent on the deposition time but also on the length of the interval on which the roughness (by which parameter ever it is defined) is determined. This behavior is shown for a surface after some deposition time in Fig. 4.13.



$R=f(L)$, $R''>R'>R$

Fig. 4.13.: *Dependence of the roughness R on the measurement interval L for a self affine surface. $\overline{h}$ is the mean height of the surface*

Surfaces like these are called "self affine" since their shape at a given time t on a given spatial interval L can be scaled to the shape of the surface at time t' present on L' by a temporal scaling factor T, by a vertical scaling factor Z and a horizontal scaling factor X.

4.2.2.3. Pore Formation and non "Solid-on-Solid"-Surfaces

Even the model of stochastic growth can lose its Solid-on-Solid charakter by a simple modification: If a particle is added to the aggregate when it just has a nearest neighbor, regeardless if this neighbor is below or beside the deposited particle, overhangs and pores can develop (see Fig. 4.14).

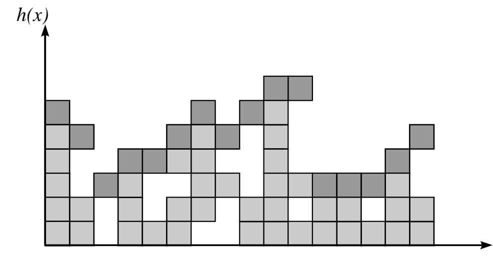
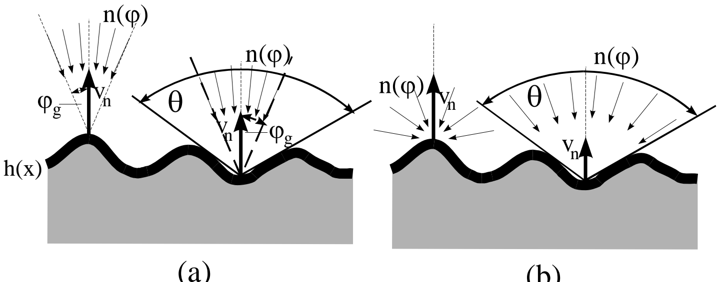


Fig. 4.14.: *Ballistic aggregation: light grey particles are part of the aggregate, particles belonging to the active surface are marked dark grey. One can easily observe the existence of closed pores.*

This growth model is called "ballistic aggregation" and shows lateral correlations and self affinity in the active surface as opposed to stochastic growth. The function which describes the active surface is still single valued as Fig. 4.14. shows. Only if all particles which are conncted to the vacuum above the highest point of the aggregate via a continous path are considered to be path of the surface the surface becomes a multi-valued function and the non SOS character of the model becomes obvious.

Also loosening the constraint of perpendicular particle incidence leads to the formation of long deep pores in the growing film because of the mechanism of shadowing if the particle mobility is sufficiently low. If the film forming particles hit a structured growth front from a wide range of impingement angles the normal growth velocity v_n at a peak is higher than v_n in a valley as Fig 4.15b shows. If, however, the distribution of impingement angles is narrow, this effect does not occur (Fig. 4.15b).



Fif. 4.15.: *Shadowing for different distributions of impingement angles $n(\varphi)$;*
θ is the opening angle of a valley
a *particles only impinge from $\varphi<\varphi_g$*
b *particles impinge from all angles φ*

The shape of the starting profile h(x) can be generated by islanding in the first phases of growth, by stochastic or ballistic roughnening or by an arbitrary profile artificially given to the substrate.

By the faster growth of peaks the shadowing effect is amplified and columnar features are formed which are separated by deep, narrow crevices. Fig. 4.16. images this effect for a sinusoidal starting profile.

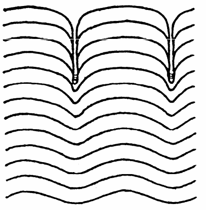


Fig. 4.16.: *Shadowing for a sinusoidal starting profile [7, p. 692]*

Fig. 4.17. shows that shadowing can also lead to the formation of closed pores in the case of self affine surfaces. Therefore also shadowing dominated growth can be counted to the non SOS mechanisms.

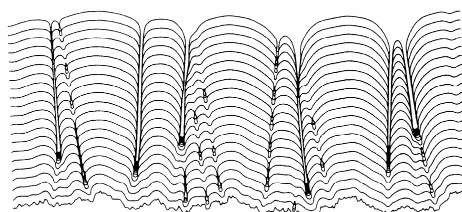


Fig. 4.17.: *Formation of closed pores by shadowing in the case of a self affine starting profile [8, p. 148]*

Broad distributions of particle incidence are observed for sputtering processes or plasma supported PVD processes where particles have to cross rather large regions of elevated gas pressures and therefore are scattered often. Generally one can say that the higher the gas pressure the broader is the distribution of impingement angles at constant distance between source and substrate. Therefore shadowing is present for sputtering at high gas pressures which is also known from the structure zone diagrams of film growth (see section 4.3.2.).

4.2.2.4. Waviness and defined Structures

Material surfaces and coating surfaces can of course exhibit roughness or, in general, structured surfaces, due to a countless number of other processes. Even an ideal single crystalline surface shows corrugations because of the periodic arrangement of atoms. Islanding in the first phases of growth may lead to the formation of more or less regular three dimensional features.

Material removing processes such as turning, milling or honing often generate very complex wavy structures which are hard to characterize quantitatively. Finally, surfaces can be willingly structured with very well defined microscopic structures as e. g. in microelectronics. This can be accomplished by photolithographic methods which mostly consist of selective material removal ("Top-Down-methods"). In recent time, however, defined micro and nanostructures are often manufactured by skilledly exploiting basic fim growth mechanisms like islanding and stress triggered self organization ("Bottom-Up-methods").

4.2.3. Roughness Measurement

The basic principle of each kind of roughness measurement is, in the case of a 2-dimensional surface, the acquisition of each height value $h(x,y)$. Theoretically the assumption of a SOS surface (i. e. $h(x,y)$ single valued) does not have to be made, but practically all methods yield only the uppermost height values. Regions of the surface which are located beneath overhangs are not imaged. In addition the intrinsic roughness (i. e. the roughness resulting from the deposition process only) is always convoluted with the

roughness, tilt or waviness of the substrate. The separation of these two morphologies is not trivial and has to be done by the observer in most cases. Modern measurement systems offer a wide choice of data manipulation features (e. g. correcture of the acquired values for arbitrary tilts or filtering of components with specified wavelengths by fourier algorithms) but these shuld be applied only by skilled persons because otherwise they could lead to massively corrupted results. Basically, data can be acquired in real space and in fourier space. Both methods shall be described briefly in the following.

4.2.3.1. Stylus or Probe Methods

This set of measurement methods allows the acquisition of surface data in real space by the close approach of a probe ("stylus") to the surface. The lateral resolution is limited by the tip curvature. Generally, a surface which is given by a continuous function h(x,y) on a quadratic area with side length L is mapped to a discrdte set of height values $h_i(x_i \pm \Delta x/2, y_i \pm \Delta y/2)$ which are given on areas with the lateral extension Δx and Δy. This fact is displayed in Fig. 4.18.

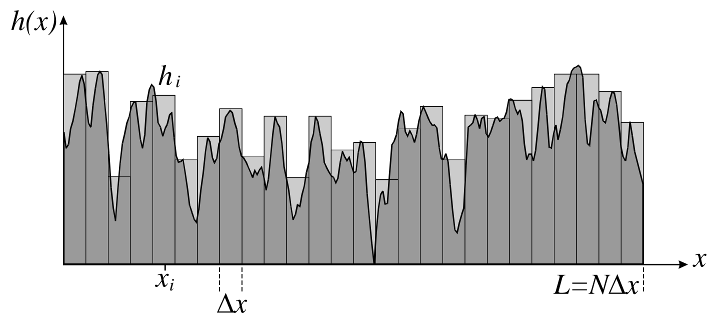


Fig. 4.18.: *Transformation of a continuous profile h(x) into discrete height values h_i, which are defined on an intervall Δx definiert sind. The h_i are produced by convoluting h(x) withn the geometry of the probe. One notices the loss of details in the surface morphology caused by the measurement process.*

The intervals Δx und Δy are called "sampling interval". The scanned area with the side length L contains L/ Δx measurement positions in x-direction and L/ Δy measurement positions in y-direction. If Δx and Δy are equal, the total measurement area contains $N = (L/\Delta x)^2$ measurement positions. L limits the maximum lateral extension of surface features within the measurement region, Δx and Δy determine the minumum lateral extension of surface features. The number N of measurement positions will have an essential role for the quantification of roughnes which will be discussed in section 4.2.4.

The acquisition of surface profiles was limited to one-dimensional profiles until approx. 1985. After this date the rapid increase in speed and sorage capacity of modern computers has allowed the easy storage and manipultion oft wo dimensional surface data with a high N. In the following the most important probe microscopic techniques shall briefly be discussed together with their advantages and disadvantages:

Stylus profilometer: this instrument was briefly discussed in section 4.1.4.2. It is the prototype of instruments dedicated to the acquisition of surface profiles. A diamond or sapphire tip is gliding along a surface. Its displacement is transformed to an electrical signal by an inductive bridge. The vertical resolution of this set-up can reach the sub-nanometer region. The lateral resolution (sampling interval) can be chosen between sub mm and µm. The biggest advantage of the system can be considered to be the high scan length L which can reach some 10 cm. This makes stylus profilometers suitable devices for the semiconductor industry where they are used e. g. for determing the curvature of big wafers. Disadvantages are the relatively high mechanical load which is applied to the surface by the tip and the limitation to one-dimensional profiles.

<u>Scanning Tunneling Microscope:</u> The Scanning Tunneling Microscope (STM) was developed by Binning und Rohrer in 1986 and allowed, for the first time, to image two dimensional surfaces with atomic resolution. The measurement principle is based on the quantum mechanical tunneling effect: if a conducting tip (mostly W) is very closely approached to a conducting surface a tunneling current begins to flow. The magnitude of the current exponentially depends on the distance of the tip to the surface. If the tip is scanned across the surface the surface morphology can be extracted from the variations in the tunneling current. The vertical resolution is within the pm (10^{-12} m) region, the lateral resolution lies well within the Å region. The exact vertical and lateral positioning of the tip is made possible by the use of piezoelectric positioning elements (which were also used in the prototype of the instrument).

The STM also provides a good example for the so-called feedback principle which is widely employed in measurement technique (see Fig. 4.19.).

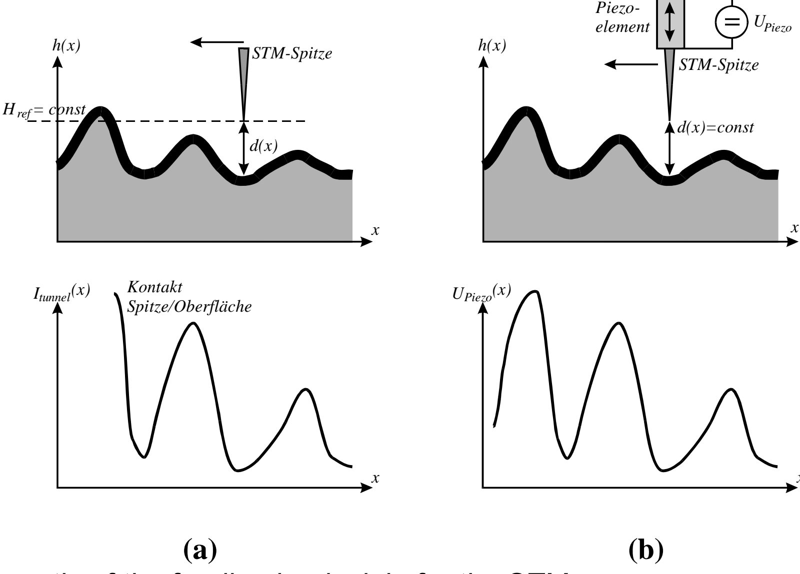


Fig. 4.19.: *Schematic of the feedback principle for the STM*
a *fixed tip, measurement of the tunneling current I_{tunnel}; there may be contact of the tip with the surface*
b *constant distance between tip and surface, d; the piezo voltage U_{Piezo}, which is necessary to keep d constant is measured.*

As previously mentioned the movement of the STM tip over the surface at a height which is kept constant relative to a reference height H_{ref} leads to a variation in the tunneling current I_{tunnel} (Fig. 4.19a). This principle has several disadvantages: if e. g. the height of the surface exceeds H_{ref} the tip will contact the surface which, in most cases, will lead to the distruction or modification of the tip. Also I_{tunnel} is a very small quantity with a low signal to noise ratio. It is therefore more useful to keep I_{tunnel} and therefore the distance tip/surface constant and to measure the piezo voltage U_{Piezo} of the vertical positioning element which is necessary to accomplish this (Fig. 4.19b.). The contact between tip and surface is avoided and the signal gains intensity since typical piezo voltages range between 1 - 1000 V which can easily be measured with high resolution.

The advantages of the STM are its robust set-up and the high imaging accuracy. Disadvantages are the limitation to conductive surfaces and the fact that the tunneling current is not only dependent on the distance tip/surface but also on the tip geometry, on the chemical constitution of the tip which cannot accurately be controlled and on the chemical costitution of the surface. Nonetheless, especially the last point may also be a big advantage since it allows for a chemical distiction between single atoms on a surface.

Atomic Force Microscope: The Atomic Force Microscope (AFM) is the extension of the STM to insulating surfaces. A micromechanically manufactured cantilever (mostly made of Si or Si_3N_4) carries a tip with an opening angle of 20 - 50° on its end. The tip curvature is approximately 20 - 50 nm. If the tip is brought into contact with the surface the cantilever is deflected. The deflection is measured and kept constant by using the feedback principle. From the signal applied to the vertical piezo the surface morphology can be reconstructed.

The most common method to measure the cantilever deflection today is to detect the displacement of a laser beam reflected from the backside of the cantilever by a four quadrant photodiode. This allows the detection of vertical and lateral cantilever deformations. Another method can be the application of a conductive tip onto the backside of the cantilever which transforms the deflection to a tunneling current signal. This principle is common for AFMs used in Ultra High Vacuum (UHV).

One advantage of the AFM is, as previously mentioned, the independence from the electrical conductivity of the sample. In addition the direct contact between tip and sample surface also yields information about mechanical properties of the surface as e. g. the local coefficient of friction and the local hardness. Unfortunately this direct contact also brings disadvantages: atomic resolution cannot be achieved due to the high tip curvature. The surface is loaded mechanically and there is a high probability of tip contamination. All these points can be (at least partially) suppressed if the instrument is operated in "non-contact" mode. In this mode the free cantilever is excited in its resonance frequency. If the tip is approached to the surface the intermittent contact with the surface leads to a reduction in amplitude and to a phase shift compared to the free oscillation. The changes in amplitude or phase are used as a signal which enters the feedback loop and can again be used to reconstruct the surface morphology.

In recent time many probe techniques have evolved from the principles of STM and AFM. They allow the determination of a great variety of surface properties on different materials. As examples Magnetic Force Microscopy (MFM) and Scanning Near field Optical Microscopy (SNOM) can be mentioned. MFM is used for the investigation of magnetic surfaces and SNOM allows to perform optical spectroscopy on a molecular level, thus yielding chemical information on the nm scale.

4.2.3.2. Optical Methods und Scattering

Roughness determination by optical methods is the complementary approach to probe techniques. Here the information about surface morphology is not acquired in real space but in reciprocal space. Generally, if one considers the reflection of an electromagnetic wave (optical or X-ray) on a surface, two components are present: They are displayed in Fig. 4.20. For reflection on a smooth surface the specular component is dominant (Fig. 4.20a), for rough surfaces the diffuse component dominates (Fig. 4.20b).

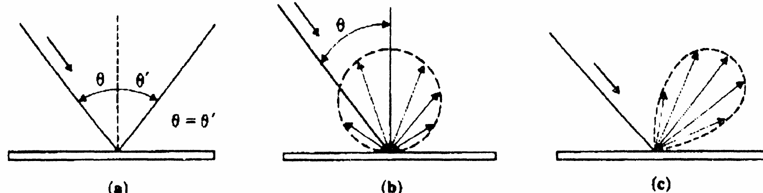


Fig. 4.20.: *Reflection types at surfaces [9, p. 35]:*
a *specular reflection*
b *diffuse reflection*
c *combination of specular and diffuse reflection; real measurement signal*

The change in the intensity of the specular component during film growth can yield important qualitative information about roughness evolution. Quantitative information about surface morphology can be obtained from the diffuse component by determining the fourier coefficients of the surface fourier spectrum. These can be calculated from the spatially resolved intensity of the diffuse component. The shape of a (for simplicities sake) one dimensional surface profile h(x) is given by its fourier series as

$$h(x) = \sum_{k=0}^{N} A_k \sin(kx + \varphi_k) \tag{4.16}$$

Unfortunately h(x) cannot fully be reconstructed because the phase information is lost upon reflection. It is possible to determine the intensities A_k, but not the phases φ_k. What this means is illustrated in Fig. 4.21. for a simple surface profile consisting only of three fourir components.

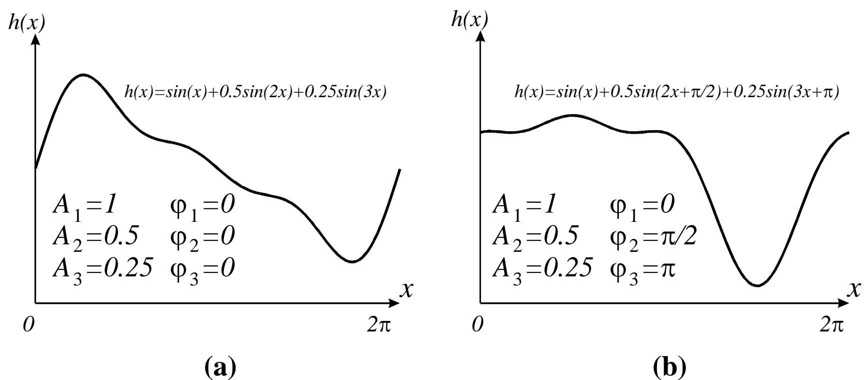


Fig. 4.21.: *Surface approximation with three fourier components in the interval [0,2π]*
a *all phase factors φ_k=0*
b *φ_1=0, φ_2=π/2, φ_3=π*
The knowledge of the A_k alone is not sufficient to reconstruct the surface morphology unambiguously.

As radiation visible light as well as X-rays are used. Especially X-rays allow the determination of the roughness of inner interfaces within a large variety of materials due to their high penetration depth. Also particle beams as e. g. electrons or ions can be used for scattering experiments and yield results similar to electromagnetic radiation due to the wave/particle dualism, but on a different length scale because of their significantly shorter deBroglie wavelength.

Scattering based techniques are non destructive and allow the determination of surface parameters in real time and in situ during a deposition process. Disadvantages are the loss of the phase information and the low reflection coefficients of X-rays which gain reasonable values only in grazing incidence. The low intensity of the diffuse component in the case of Xrays often requires intensive syncrotron light sources which is linked to high apparatve and administrative effort.

4.2.4. Roughness Quantification

As previously mentioned in section 4.2.3., Roughness Measurement, the scanning length L and the sampling interval Δx play an important role for the determination of roughness in real space. In the case of a one dimensional scan line the number of measurement points N is given by $L/\Delta x$, in the case of scanning a quadratic region with the side length L, N amounts to zu $N=(L/\Delta x)^2$ given eaual lateral resolution in x and y direction. This means that with known Δx L can be substituted by N. Therefore the definitions of

roughness given in the following are essentially the same for one and two dimensional data arrays.

4.2.4.1. Global Quantities

Before a roughness value can be determined a reference niveau within the active surface has to be defined. The easiest definition of this niveau is given by the mean height $\overline{h}$ calculated from all measured height values, h_i,

$$\overline{h} = \frac{1}{N}\sum_{i=1}^{N} h_i \tag{4.17}$$

The roughness R can then be defined by deviations from $\overline{h}$ as it is schematically displayed in Fig. 4.22.

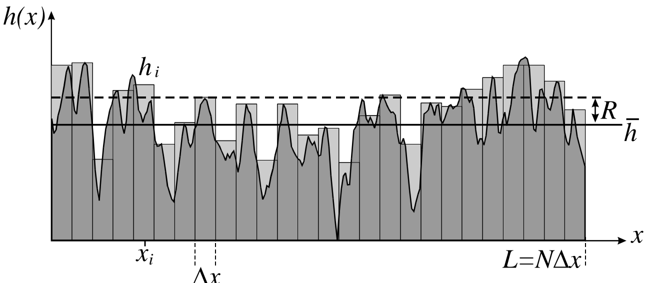


Fig. 4.22.: *Schematic visualization of rhe roughness R as deviation from the mean height $\overline{h}$ of a given surface profile.*

The rule for the calculation of R, however, can differ significantly. Due to historical reasons several definitions of R have emerged. In engineering mainly the so called R_a value is used which is defined as the mean value of the absolute deviations from $\overline{h}$:

$$R_a = \frac{1}{N}\sum_{i=1}^{N} \left|\overline{h} - h_i\right| \tag{4.18}$$

In scientific literature, however, the mathematically more meaningful value called R_q, R_{RMS} oder RMS-value is used. It is the square root of the mean quadratic deviation from $\overline{h}$:

$$R_q = R_{RMS} = RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\overline{h} - h_i)^2} \tag{4.19}$$

R_q is tightly connected to the so-called correlation functions and their fourier transforms which basically correspond to the diffuse component of scattered electromagnetic radiation as dercribed in section 4.4.3.2. This is why R_q is preferredly used for a theoretical description of scattering data. R_a and R_q do not differ significantly in their numerical values. For a simple sinusoidal profile $R_a/R_q = 2^{3/2}\pi \cong 0.9$ is valid. For more general profile shapes $R_a/R_q \cong 0.8$ can be assumed.

Different profile shapes can exhibit the same R_q or R_a values as Fig 4.23. shows for some simple examples.

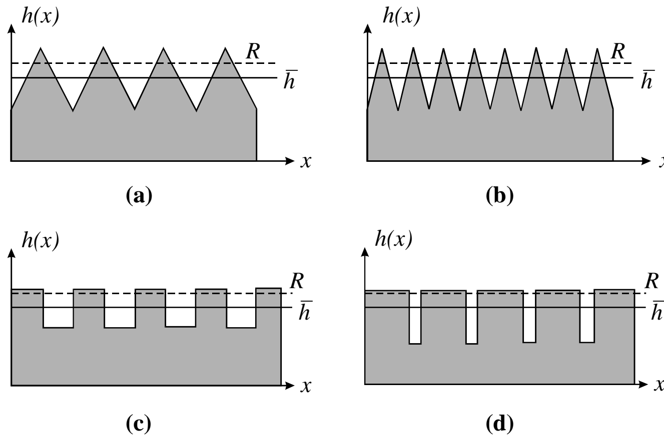


Fig. 4.23.: *Surfaces with different shapes, but equal roughnesses*
a, b *different periodicities*
c, d *different symmetries*

Parameters, which are at least partially shape specific are: the so called "Skewness" Sk,

$$Sk = \frac{1}{NR_q^3}\sum_{i=1}^{N}(h_i - \overline{h})^3 , \qquad (4.20)$$

the sign of which is (because of the odd third power in Eq. 4.20) a measure for the symmetry of a profile in respect to $\overline{h}$*, i. e. a statement is made wether more height valuzes ere located above or below* $\overline{h}$*. The so-called "Kurtosis" K*

$$K = \frac{1}{NR_q^4}\sum_{i=1}^{N}(h_i - \overline{h})^4 , \qquad (4.21)$$

quantifies the mean edge steepness od a profile. Despite the definition of these shape specific parameters further information about the lateral shape of a profile can only be derived with mathematical tools which take into account the horizontal distance between separate areas of the surface. These will be described in the following section.

4.2.4.2. Correlation Functions

In many disciplines of physics spatial and temporal correlations play an important role. Correlation can be considered as the increased amount of specific events (these may be height values, the number of radioactive decays or certain signal levels occurring during a measurement) within a certain interval around a measurement point. This interval is called correlation length. In the case of a surface profile this may e. g. mean that around a large height value the profile cannot decay arbitrarily steep as it was discussed for self affine surfaces in section 4.2.2.2. In contrast to this case it is not possible to draw a conclusion about the height value $h_{i\pm1}$ *upon knowing* h_i *in the case of an uncorrelated profile (see section 4.2.2.1.).*

To quantify correlations several types of functions can be used. The so called autocovariance function R(X) can be written for an interval of N points with a distance Δx *from each other as:*

$$R(X) = R(n \cdot \Delta x) = \frac{1}{N-n}\sum_{i=1}^{N-n}(h_i - \overline{h}) \cdot (h_{i+n} - \overline{h}). \qquad (4.22)$$

Here X is n times the sampling interval Δx. n can assume values from 0 to N-1. The autocovariance function is the mean value of the products of two height values h_i und h_{i+n}, which are separated from each other by a distance X. Of course, if X is increasing at constant profile length only a decreasing number of height pairs can be used for the calculation of the mean value as it is shown schematically in Fig. 4.24.

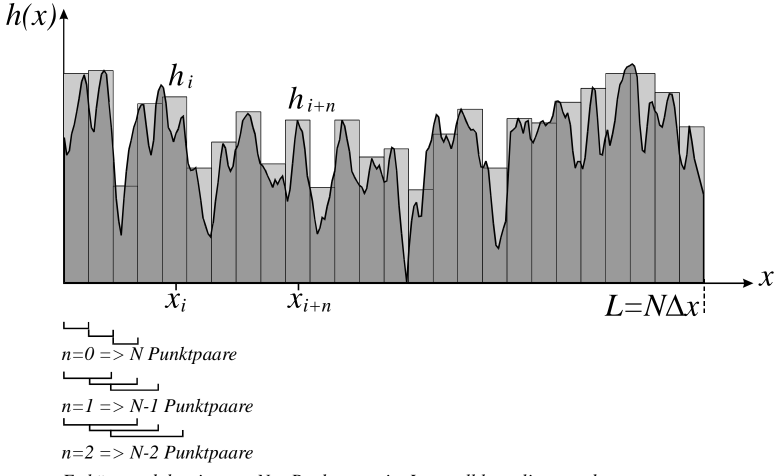


***Fig. 4.24.:** Calculation of the autocovariance function for a discrete profile. The bigger the distance between the points to be correlated the fewer point pairs can be used for averaging.*

From Fig. 4.24. the structure of R(X) given in Eqn (4.22) is resulting. For a continuous profile R(X) is transformed into the integral expression given in Eqn. (4.23):

$$R(\tau) = \frac{1}{L-\tau}\int_0^{L-\tau} (h(x) - \bar{h}) \cdot (h(x+\tau) - \bar{h})dx \qquad (4.23)$$

L is the length of the profile and τ is the continuous distance of the point pairs. From Eqns. (4.22) or (4.23) one can easily see that R(0) corresponds to the RMS value. This connection of R_q with the autocovariance function is an explanation for the higher mathematical significance of the RMS value when compared to R_a. Often also the normalized autocovariance function or autocorrelation function

$$\rho(X) = R(X)/R(0) \text{ oder } \rho(\tau) = R(\tau)/R(0) \qquad (4.24)$$

is used. For profiles which exhibit correlations, but no periodicities, both previously defined expressions decay monotonically. The coeerlation length ξ is given by the steepness of the decay around X (or τ) =0. This behavior is shown schematically in Fig. 4.25.

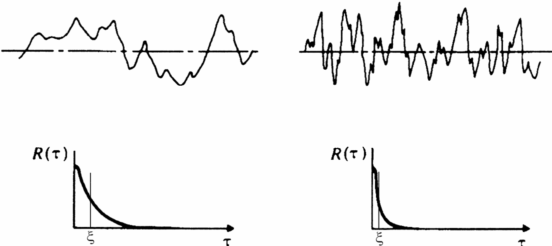


***Fig. 4.25.:** Autocovariance function for surfaces with equal roughness, but different correlation length ξ [9, p.151]*

Often ξ is defined as the value of X or τ at which R(X) (or $R(\tau)$, $\rho(X)$, $\rho(\tau)$) are decayed to 1/e times their value at X, τ = 0. Periodicities in the profile show themselves by the occurance of maxima at values of $X, \tau \neq 0$.

An analogous function to R(X) is the so called structure function S(X),

$$S(X) = S(n \cdot \Delta x) = \frac{1}{N-n} \sum_{i=1}^{N-n} [(h_i - \bar{h}) - (h_{i+n} - \bar{h})]^2 . \qquad (4.25)$$

in its discrete form and

$$S(\tau) = \frac{1}{L-\tau} \int_0^{L-\tau} [(h(x) - \bar{h}) - (h(x+\tau) - \bar{h})]^2 dx \qquad (4.26)$$

in the integral formulation for continuous profiles. $R(\tau)$ and $S(\tau)$ are connected to each other by the expression

$$S(\tau) = 2R_q^2 [1 - \rho(\tau)] . \qquad (4.27)$$

Although both functions contain the same information due to Eqn. (4.27) S(X) (or $S(\tau)$) is easier to calculate and mathemarically more robust. While e. g. high frequency parts strongly influence $R(\tau)$ at small τ $S(\tau)$ is not influenced by this because is approaches zero for small τ. Additionally, $S(\tau)$ is much more sensitive to surface roughness because it approaches $2R_q^2$ for large τ. This is made evident in Fig. 4.26. which shows a comparison of $R(\tau)$ and $S(\tau)$ for unworn and worn surfaces.

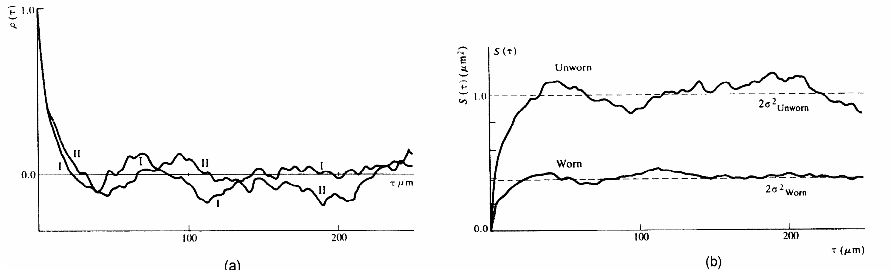


Fig. 4.26.: *Correlation functions for surfaces with different roughness*
a *Autocorrelation function: no significant differences*
b *Structure function: Differences in RMS roughness (here: σ) can easily be recognized. [9, p., 156]*

The decrease in roughness due to surface wear is easily visible for $S(\tau)$ while $R(\tau)$ basically does not allow for a discrimination of both surface types.

In the case of two dimensional surface scans one can decompose them into M one dimensional profiles and average the correlation functions. The one dimensional profiles may be parallel or perpendicular to the scan direction or may be formed by radial lines within a circular area of the surface. In this case one talks about "radially averaged correlation functions".

4.2.4.3. Methods Based on Fourier Analysis

As mentioned in section 4.3.2.2., Optical Methods und Scattering, scattering of electromagnetic radiation on a surface basically yields the fourier transform of the surface profile. From the mathematical point of view the result of the scattering experiment is the so called "Power Spectral Density", P(k). $k=2\pi/\lambda$ is a wave vector which represents the

wavelength λ of an isolated fourier component of the surface profile. P(k) is connected to the surface profile by

$$P(k) = \lim_{l \to \infty} \frac{1}{L \cdot (2\pi)^2} \left| \int_{-\infty}^{\infty} h(r) e^{ikr} dr \right|^2 . \qquad (4.28)$$

Additionally P(k) is the fourier transform of the autocovariance function,

$$P(k) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} R(r) e^{ikr} dr . \qquad (4.29)$$

So, a scattering experiment yields essential global parameters of the surface as e. g. correlation lengths and roughnesses if P(k) is transformed back into real space. The detailed shape of the surface profile, however, cannot be assessed because of the loss of phase information, as it was discussed in section 4.3.2.2.

The mechanisms of roughness evolution, the measurement metods and the mathematical tools for roughness quantification which were discussed in the previous sections are by no means complete. The structural evolution of thin films during growth depends on many other processes which are incorporated into the so called structure zone models These will be treated in the following and correlations with thin film properties will be made.

4.3. Mechanical Properties

4.3.1. Introduction

The structure of thin films is mainly responsible for their mechanical properties. Therefore at first the connection between film growth, structure and morphology and the mechanical properties resulting thereof shall be discussed.

4.3.2. Structure Zone Models

4.3.2.1. General

For the growth of a film and for the development of its structure three factors are important: the roughness of technical surfaces, the activation energies of surface and volume diffusion of the film forming atoms and finally the binding energy of ad-atoms to the substrate. Substrate roughness leads to shadowing which, in turn, triggers a porous structure. Shadowing can be overcome by surface diffusion at elevated temperatures.

The energies mentioned above are proportional to the melting temperature of many pure metals, T_m [K]. Therefore one can guess that one of the three effects, shadowing, surface diffusion and volume diffusion, is dominant in one given region of T/T_m, i.e. the ratio of the substrate temperature T [K] and the melting temperature T_m [K]. Within this region the dominant effect has the primary influence on the microstructure. This is the basis of the so-called structure zone models.

4.3.2.2. Model of Movchan und Demchishin

Movchan and Demchishin, in 1969, investigated the structure and the properties of films evaporated in HV (10^{-4} to 10^{-3}Pa) as a function of T/T_m. The materials deposited were Ti, Ni, W, ZrO_2 und Al_2O_3 and the thickness reached values up to 2 mm. Their results yielded the three zone model which is displayed in Fig. 4.27a.

4.3.2.3. Thornton-Model

The M.-D. model was expanded in 1974 by Thornton by experiments using a hollow cathode discharge at Argon pressures between 0,1 and 4 Pa. Thornton added one additional variable, the Ar pressure, to describe the influence of a gas atmosphere (without ion bombardement) on the film structure. Additionally, a transition zone (zone T) vas introduced between zone 1 and zone 2 (see Fig. 4.27b). This transition zone is not very pronounced for metals and single phase materials, but can well be observed for refractory compounds and multiphase materials which are produced by evaporation in HV or, in the presence of inert or reactive gases, by sputtering or ion plating. The other zones show equal properties in both models.

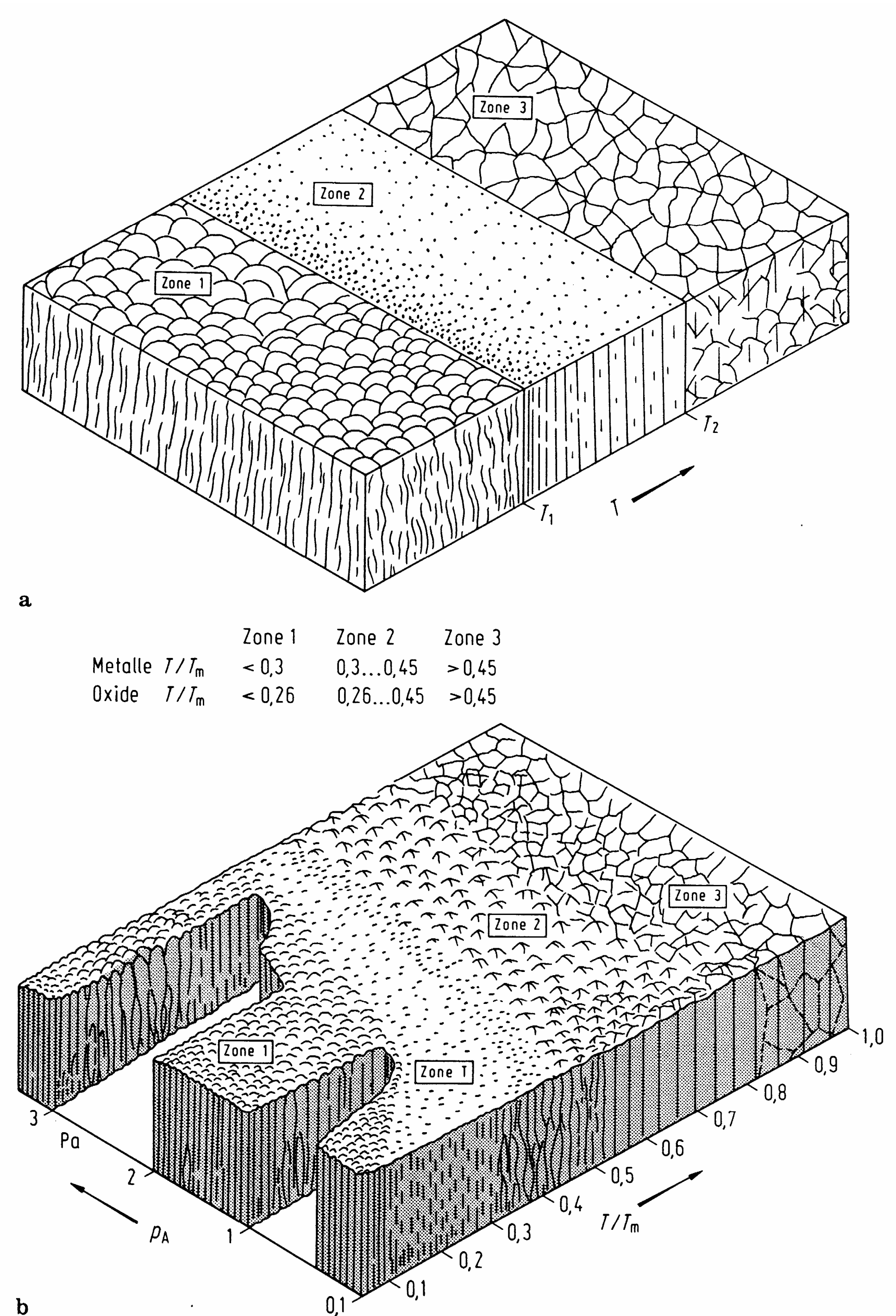


Fig. 4.27.: Structure zone models: **a** Movchan and Demchishin; **b** Thornton;
Zone 1: porous structure consisting of needle or fiber shaped crystallites
Zone T: dense fibrous microstructure
Zone 2: columnar microstructure
Zone 3: recrystallized microstructure
$\boldsymbol{T}$ Substrate temperature [K]; $\boldsymbol{T_m}$ Melting temperature [K]; $\boldsymbol{p_A}$ Argon pressure [1, p. 22]

Zone 1

Zone 1 comprises the microstructure forming at low values of T/T_m. Ad atom diffusion is not sufficient to overcome shadowing. Therefore needle shaped crystallites emerge from relatively few nuclei. The crystallites broaden with increasing height by lateral capture of atoms so that inverted cones are formed. Their tips consist of spherical caps. The film is porous and the single crystallites have a distance in the region of some 10 nm. They exhibit a high dislocation density and show compressive stress inside due to the high defect density. Globally, on the other hand, zone 1 films are subjected to tensile stress because increasing film thickness leads to the coalescence of neighboring crystallites.

Zone T

Zone T is characterized by the fact that ad atoms can compensate the effects of shadowing because of their mobility due to surface diffusion. Additionally, especially at low working gas pressure, a current of energetic particles is present which increases the density of nuclei by the formation of surface defects. A fiber shaped and, compared to zone 1, much denser structure is formed.

Zone 2

Zone 2 is defined by the region of T/T_m in which surface diffusion is the dominant factor of growth. A columnar microstructure is formed. The column diameter increases with the substrate temperature T while porosity decreases.

Zone 3

Zone 3, finally, comprises the region of T/T_m in which growth is dominated by volume diffusion. A recrystallized dense microstructure of three dimensional, equiaxed grains is formed. This temperature region is important for epitactic growth of semiconductors by evaporation, sputtering and CVD.

Influence of the inert gas on the microstructure

Acoording to Thorntons model the transition temperatures T_1 and T_2 are decreasing with falling inert gas pressure p_A. This is mainly due to the fact that, as mentioned before, a permanent stream of energetic particles is present. The reason for this is the low number of collisions of the film forming particles with working gas atoms. This, on the one hand, leads to the formation of surface defects which increase the nucleation density. On the other hand their impulse is transferred to loosely bound adsorbates (e. g. single ad atoms) and increases the transient mobility of the adsorbates. Finally, the additional energy input also heats the substrate. All these effects lead to a temperature reduction for the transition from zone 1 to zone T.

Influence of ion bombardement on the microstructure

Ion bombardement generates point defects at the substrate and therefore increases the density of nuclei. The energy transfer to the ad atoms increases their transient mobility. Therefore, at given T/T_m a denser crystallite structure is generated when compared to the situation without ion bombardement.

This means that ion bombardement influences the film structure in a way that the zone limits, especially the border between zone 1 and zone T are shifted to lower values of T/T_m. This effect, which cannot be explained by additional heating alone, is indeed observed (see Fig. 4.28.).

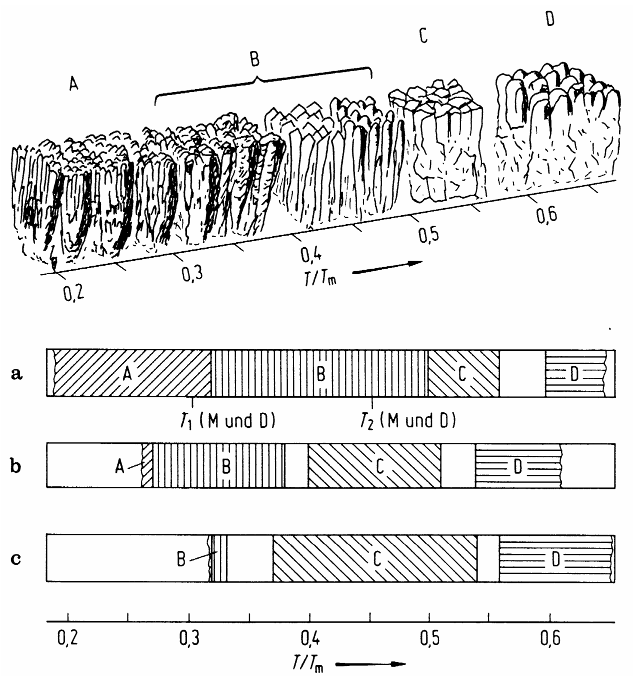


Abb. 4.28.: Influence of the substrate temperature *T* and the bias voltage U_b at the substrate on the film structure of Ti. **a:** U_b = 0; **b:** U_b = -5kV; **c:** U_b = -10kV; Argon-pressure: 2,7Pa; [1, p. 24]

The Ti coatings in Fig. 4.28. were produced by evaporation from an electron gun. The substrate was kept at bias voltages up to -10 kV. A current density of 2 $mAcm^{-2}$ was observed at the substrate (this means about 1 ion per film forming particle). It is visible that porous coatings are produced in a situation corresponding to pure evaporation ($U_B = 0$), while ion plated films ($U_B \neq 0$) exhibit a dense structure.

The dense structure of evaporated films at high substrate temperature differs from the microstructure of ion plated films: in the case of evaporation at high substrate temperature the structure is a result of recrystallization and grain growth triggered by volume diffusion. In the case of ion plating volume diffusion is less important due to the relatively lower substrate temperature. The intensive ion bombardement, on the other hand, continuously forms new nuclei which leads to the fine grained, dense structure.

4.3.3. Incorporation of Foreign Atoms

Foreign atoms which are incorporated into the growing film can form compounds with the film atoms or can assume interstitial or substitutional positions. Resulting from these options they can change film properties as e. g. intrinsic stresses. Impurities located at grain boundaries can lead to embrittlement. Within the grain they can form polymorphic phases. Foreign substances can also prevent columnar film growth due to permanent renucleation so that a fine grained mirostructure is formed.

An example for permanent renucleation are the so-called "brighteners" which are used in galvanic baths to achieve brilliant films. Usually galvanic coatings exhibit a dendritic structure which forms because of the high mobility of the hydratized metal ions and, in analogy to PVD, due to the preferred condensation on special crystal faces. By adding brighteners (i. e. certain organic additives) to the electrolyte these effects are suppressed and a very fine microstructure is formed which provides the desired brilliance.

In all deposition processes foreign atoms are incorporated into the film, either as undesired, but unavoidable impurities or as a desired medium to achieve certain film properties. Undesired impurieties ere present e. g. in galvanic coatings which can contain water or organic and anorganic additives. In the case of electroless deposition catalyzing agents (Phosphorous, Boron) may be incorporated into the coating.

In the case of vacuum based processes the residual gas contains H_2O, O_2, N_2, H_2 and carbohydrates which result from outgassing materials, incoming gas currents, leaks and backstreaming pumping oils. If material is e. g. evaporated these gaseous impurities can be incorporated into the film. In the case of sputtering, especially bias sputtering, ion implantation of the sputtering gas is also possible.

For CVD methods the carrier gas and reaction gas which was not fully transformed can enter the coating. Since even small amounts of impurities may have significant influence on the film properties, the expense to achieve a clean deposition device may sometimes be very high, but justified.

The incorporation of foreign substances, nonetheless, can also be a desired means for the following methods:

... Reactive PVD methods are used to transform an elementary material to one of its chemical compounds (nitride, carbide, oxide, boride) upon deposition in the presence of a reactive gas.

... In the case of CVD deposition of Chronium or Tungsten onto steel or hard metal the carbon contained in the substrate diffuses into the coating and forms carbides which is important for the production of hard material coatings.

... Ion implantation of foreign atoms is used to an increasing amount for manufacturing abrasion resistant tribological coatings.

4.3.4. Stresses

4.3.4.1. General

All coatings are in a state of more or less pronounced internal stress which is composed of two componets, thermal stress σ_T and intrinsic stress, σ_i:

$$\sigma = \sigma_T + \sigma_i \qquad (4.30)$$

The thermal stress σ_T is caused by different Coefficients of Thermal Expansion (CTS) of the coating and the substrate and is given by:

$$\sigma_T = E_S(\alpha_S - \alpha_U)(T_B - T_M) \qquad (4.31)$$

with

E_S = Elastic Modulus of the fiml
α_S = mean CTE of the film
α_U = mean CTE of the substrate
T_B = Substrate temperature during coating
T_M = Substrate temperature during measurement

Therefore the thermal stress σ_T can, in principle, be calculated as opposed to the intrinsic stress σ_i.

The intrinsic stress σ_i depends on the deposition parameters and is caused by structural disorder within the film, i. e. by incorporated foreign atoms and by coating atoms which are located out of a potential minimum. These stresses can be compressive or tensile depending on the deposition parameters, so that the coating tries to contract (tensile stress) or to expand (compressive stress) parallel to the substrate surface.

If low melting point materials are deposited at sufficiently high substrate temperature (T/T_m>0,5) the intrinsic stresses remain low due to the high number of diffusion events while thermal stresses are dominating. Thermal stresses can be reduced by temperature treatment after deposition. The flow of material which is triggered by temperature treatment can lead to the formation of hillocks or voids with sizes in the µm and sub µm region depending on wether the film is set under compression or tension by the temperature change. Hillock formation is observed for Al, Pb or Au coatings which were deposited by PVD methods at elevated substrate temperature.

High melting point materials are generally deposited at rather low T/T_m values (<0,25) so that the intrinsic stresses are dominant relative to the thermal ones. For sufficiently thin coatings (<500nm) the intrinsic stresses can be considered constant throughout the whole film thickness. For evaporated coatings they are mostly tensile, while sputtered coatings mostly show compressive stress. Stresses can approach the yield strength of the bulk material and can reach values of $10^3 Nmm^{-2}$ for refractory metals. In some cases intrinsic stresses even exceed the bulk yield strength which indicates certain solidification processes in thin films.

The bonds within the interface between coating and substrate have to withstand the shear forces which are produced by the intrinsic and thermal stresses. Since the contribution of the intrinsic stress to the shear force grows with film thikness the film cann spall off the substrate after a certain critical thickness is exceeded. Under adverse conditions this may even happen at thickness values as low as 100 nm.

Intrinsic stresses may be influenced by the process parameters like substrate temperature, deposition rate, impingement angle and energy distribution of the film forming atoms, gas incorporation and residual gas composition in regard to their type, compressive or tensile (see Fig. 4.29.). This was demonstrated for evaporated and sputtered coatings in extensive work. Therefore the possibility to produce mostly stress free and, resulting from this, rather thick fimls by PVD methods is given.

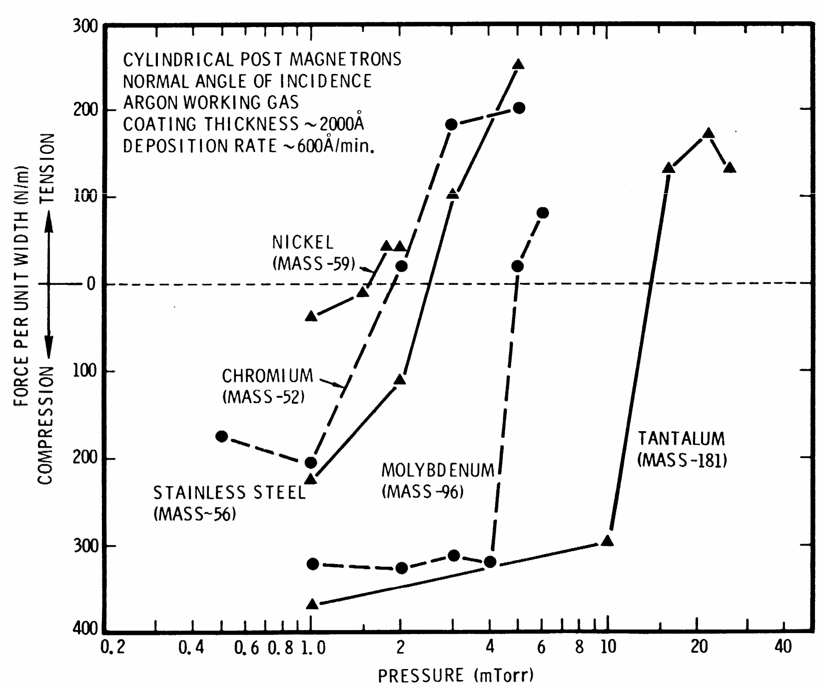


Fig. 4.29.: Investigation of internal stresses for sputtered coatings in dependence on the Ar pressure [2, p. 222]

4.3.4.2. Stress Measurement

General

All macroscopic stress measurement methods are based on the principle that a rather thin substrtate (quartz or glass) is coated and the curvature caused by the internal stress in the film is measured either in situ or to another arbitrary time at a given temperature. Temperature measurement during deposition is mostly done on a spare substrate which is locatied in close vicinity to the actual substrate. The measurement substrates are generally rectangular and are fixed either on one or both ends. In the first case the deflection of the free end is measured, in the second case the deflection in the middle of the sample is observed. Also circular substrates may be used.

σ_T and σ_i may be separated as follows: The stress at room temperature and at deposition temperature is measured. At the deposition temperature $T_B = T_M$ is valid and therefore $\sigma_T = 0$. According to Eqn. (4.30), $\sigma = \sigma_i$. For room temperature $\sigma_T = \sigma - \sigma_i$ if σ is the total stress measured at room temperature. The deformation of the coated substrates can be measured either by electrical or optical methods.

Interference Optical Measurement

The coated and curved substrate is put onto a plane glass slide with the coated side facing the glass surface. In the case of tensile stress it is beared on both ends and in the case of compressive stress it is beared in the middle (see Fig. 4.30.).

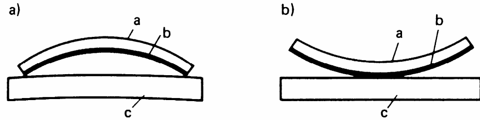


Fig. 4.30.: Substrate deformation: **a)** tensile stress within the coating
b) compressive stress within the coating
a substrate; **b** coating; **c** reference slide
[3, p. 295]

Using the set-up displayed in Fig. 4.31. the radius of curvature, R_S of the substrate can be determined with sufficient accuracy.

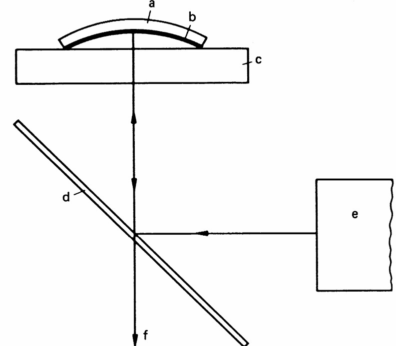


Fig. 4.31.: Schematic for the measurement of internal stress by interference optics:
a substrate, **b** coating; **c** plane glass slide; **d** beam separator; **e** light source;
f acquisition optics [3, p. 295]

For the above set-up the radius of curvature of the sample can be determined by

$$R_s = \frac{D_m^2 - D_n^2}{4\lambda(m-n)} \tag{4.32}$$

with

D_m = diameter of the m-th Newton ring
D_n = diameter of the n-th Newton ring
λ = wavelength of used light

The total stress within the coating can be calculated by:

$$\sigma = \frac{E_s d_s^2}{6(1-\nu_s)d_F}\left(\frac{1}{R_{s1}} - \frac{1}{R_{s2}}\right) \tag{4.33}$$

with

E_S = Elastic modulus of the substrate
ν_S = Poisson number of the substrate
d_S = Thickness of the substrate
d_F = Thickness of the film
R_{S1} and R_{S2} = Radius of curvature before and after coating, respectively

Measurement by Geometrical Optics

While the application of the interferometic method is rather time consuming the measurement of stresses by geometrical optics is quite fast.

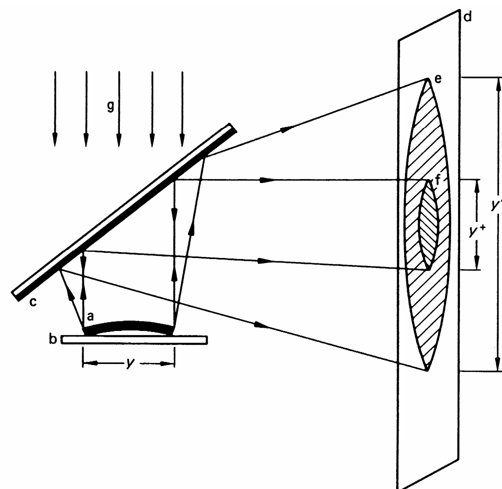


Fig. 4.32.: Schematic for the determination of internal stresses by geometrical optics:
a coated substrate; **b** glass slide with thin silver film;
c semitransparent mirror; **d** screen; **e** image of the coated substrate;
f image of the uncoated substrate; **g** einfallendes Licht;
y Durchmesser des Substrates; $\mathbf{y^+}$ diameter of the image of the uncoated substrate (for parallel light: $y=y^+$); **y`** diameter of the image of the coated substrate
[3, p. 296]

For samples with small extensions the resolution which can be achieved given the set-up in Fig. 4.32 is only small. The following expression is approximately valid:

$$R = \frac{2yD}{y' - y^+} \tag{4.34}$$

with

y = sample diameter
y+ = diameter of the image of the uncoated substrate
y` = diameter of the image of the coated substrate
D = distance sample/screen

The stress in the coating can then be calculated using Eqn. (4.34). Also asymmetries in the stress distribiution can easily be assessed by this method.

Electrical Measurement

For this metod the substrate is coated prior to deposition with a conductive layer (e. g. Ag) on the side which is facing away from the vapor source. The substrate and a metallic plate located in a distance of approximately 1mm form a condenser which is part of a oscillating circuit. If the substrate is deformed due to internal stresses the distance between the condenser plates is changing which shifts the resonance frequency of the oscillating circuit. This frequency shift is very well measurable.

Stress Measurement by X-rays

This method determines the change in lattice constant due to stresses with a X-ray diffractometer.

4.3.5. Adhesion

4.3.5.1. General

For all applications of coated materials sufficient adhesion regarding the respective application is paramount to guarantee a reasonable lifetime of the coated part. Adhesion is a macroscopic property which depends on the inteatomic forces within the interface between substrate and film, on the internal stresses and on the specific load the coating is subjected to. The latter can be mechanical (pulling or shearing), thermal (high or low temperatures or thermal cycling), chemical (corrosion, either chemical or electrochemical) and can also result from other effects.

According to Mattox a "good adhesion" is generally reached if:

1. a strong atom-atom bonding exists within the interface zone,

2. low internal stresses exist within the film,

3. no easy mode of deformation or fracture exists, and

4. no long term dgradation is present in the composite substrate/coating.

The adhesion depends primarily on the choice of the partners present in the composite, on the interface type, on the microstructure (and therefore on the deposition parameters of the coating) and on the pre treatment of the substrate. In the following these dependences shall be discussed with a special focus on PVD coating methods.

4.3.5.2. Interface between Substrate and Coating

Nucleation and Film Growth

The processes of nucleation and growth were thoroughly discussed in chapter 3. It was shown there that the aggregation of ad-atoms leads to the formation of nuclei and to the growth of islands which - dependent on the deposition parameters - coalesce to a more or less continuous film.

The nucleation density and the growth of nuclei determine the effective contact area at the interface or, vice versa, the surface which is exposed to voids. Given a low nucleation density the adhesion is low due to the low contact area and the easy fracture propagation through voids and pores. The nucleation density can be enhanced by ion bombardement, surface defects, impurities, the surrounding gas and therefore in general by the choice of a suitable deposition method.

Mattox distinguishes the five interface types described in the following:

Mechanical Interlocking

The substrate surface is rough and exhibits pores in which the coating material is locked (see Fig. 4.33a). This yields sufficient, purely mechanical adhesion for many applications. One has to take into account that substrate roughness leads to shadowing and therefore to void formation and a porous structure. On the other hand a crack through a coating which was deposited on a sand blasted surface cannot propagate easily because it has to change its direction frequently or it often has to pass more solid material. Examples for mechanical interlocking are thick coatings which were produced by thermal spraying or which were electrochemically deposited on sintered ceramics or etched polymer.

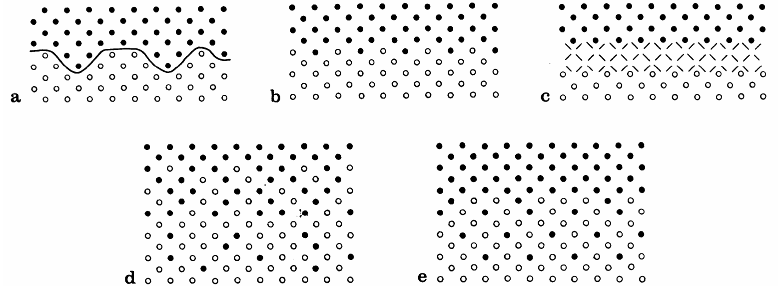


Fig. 4.33.: Shematic of the five possible transition zones between substrate and coating: **a** mechanical interlocking; **b** monolayer/ monolayer; **c** chemical bonding; **d** diffusion; **e** pseudodiffusion [1, p. 20]

Monolayer / Monolayer Transition

Here an abrupt transition from the substrate to the film within a few atomic layers is observed (see Fig. 4.33b). Such a transition is formed if no diffusion or chemical reaction is present between the partners, i. e. if there is only low particle energy available, if there is low solubility or if impurities are present on the substrate. Evaporated coatings which were produced under these conditions exhibit this characteristic feature.

Reaction Transition

This transition, which forms in the case of a chemical reaction between the substrate and a coating, demands the formation of a chemical bond between the involved partners and is promoted by elevated temperatures (see Fig. 4.33c). Within a zone which can comprise many atomic layers an intermetallic compound, an oxide or some other compound may form. Intermetallic compounds are often brittle, prone to crack formation and may trigger fractures within the composite. Oxides may form thermal barriers or good transitions from metal to a ceramic. Examples for the latter case are oxidic interlayers which are formed by CVD or by thermal spraying.

Diffusion Transition

If subatrate and coating material are mutually soluble and if suitable temperature is present a continuous transition of the chemical composition, of the lattice parameters and of the stresses from the substrate to the coating is formed by interdiffusion (see Fig. 4.33d). For different diffusion rates porosities may be formed at the interface due to the Kirkendall effect if the transition zone is sufficiently thick.

The diffusion transition forms everywhere, where two mutually soluble materials with atomically clean surfaces ace in close contact to each other at elevated temperatures. This is the case for many coating processes as e. g. ion plating, CVD, vacuum plasma spraying or melt dipping. It can also be achieved for evaporation and sputtering if the substrate is treated properly before deposition (Ion Etching) and if it is heated during deposition.

Pseudo Diffusion Transition

This transition is formed between materials without mutual solubility in the presence of strong energy input to the substrate or to the forming film. This can be accomplished by ion bombardement or ion implantation (see Fig. 4.33e). Energetic ions or neutrals are penetrating the the lattice of the substrate to a certain depth depending on their energy and stick there without further diffusion.

Also ion bombardement prior to deposition can increase the solubility and, together with that, the diffusivity in the inerface zone due to the formation of a high defect concentration and of internal stresses. Another example is the contact of two molten metals which mix and then are rapidly quenched as it is the case e. g. for plasma build up welding.

Practically the obtained interface zones are often a combination of the different transition types. Regarding adhesion those transitions are best which equally distribute the internal stresses resulting from different CTE, from lattice mismatch and from different solubilities among a sufficiently large volume without generating positions of reduced strength or increased brittleness. These conditions are best met by diffusion joints which can be achieved often by heating of the substrate during and/or after deposition.

Microscopic Bond Types

As mentioned above the adhesion between substrate and film depends on the binding energy between substrate material and coating material and on the structure of the interface zone. Chemical, electrostatic and van der Waals bonds as well as a combination of these bond types may be formed. With a binding energy of 0,5 - 10 eV the chemical bond is the strongest among these. Electrons are transferred between film and substrate atoms and are newly distributed. This requires an intimate contact between the partner atoms. In the

case of covalent, ionic or metallic bonding the bond strength depends on the degree of the electron transfer. In the first two cases compounds form which generally are brittle, in the last case mostly ductile alloys are generated.

Van der Waas bonding is based on a polarization interaction which does not require an intimate contact of the partner atoms but is significantly weaker (0,1 - 0,3eV) than the chemical bond and also decays rapidly with distance.

It can be shown that a typical chemical bond (4 eV) should withstand a mechanical stress of about 10^4 Nmm^{-2} and a typical Van der Waals bond (0,2 eV) should bear a stress of 500 Nmm^{-2}. Adhesion measurements often yield lower values than expected by these estimates. This has three reasons: first, intrinsic stresses have to be taken into account which are added (according to tensorial calculus) to the external stress and which may approach the theoretical adhesion value. Second, the strength of the interface layer is determined by the defects and the microstructure and therefore lower than expected for an ideal atomic bond. Finally, a low adhesion value may also be due to the structure of the interface, e. g. because of a low contact area which results from a low nucleation density. In this case the contact area is significantly lower than the geometric interface surface.

Since ion bombardement increases the nucleation density and therefore the contact area to the substrate and zone boundaries within the Thornton model are shifted towards lower T/T_m values the adhesion of coatings which are produced by bias sputtering is higher than the one for evaporated coatings. Even higher adhesion values are found in the case of an interface zone which is generated by diffusion and/or chemical reaction as e. g. in coatings produced by ion plating or CVD.

4.3.5.3. Substrate Pre Treatment

General

The purity of the substrate surface has a decisive influence on the microstructure and on film adhesion. The criteria which define a clean, or better, a suitable substrate surface are dependent on the given application of the coating. If gas adsorption or desorption, surface diffusion or similar interface processes shall be investigated a surface is demanded which only is covered by a tiny fraction of a monolayer. Therefore an atomically clean surface is required as it can only be produced by UHV techniques.

If, on the other hand, a steel construction shall be protected against corrosion by a Zn layer, sand blasting is a perfectly adequate pre treatment. Rust and scale are removed and a mechanical transition zone is prepared by roughening.

In the following special cleaning methods shall be discussed which are applied in the case of PVD processes. In this case, cleaning consists of two steps: one happening outside the vacuum chamber and one within the chamber as part of the deposition process.

The vacuum chamber is often positioned in a way that the process chamber is part of a clean room while all other components are located in a service compartement.

Pre Cleaning

In principle a coarse cleaning step of the substrates has to be performed out of the deposition chamber. Often they are heavily contaminated or covered by protective substances (e. g. corrosion protection oil) and have to be cleaned by manual scrubbing, sand or glass blasting, polishing, etching in diluted acids or bases or other methods. Many

different procedures for a subsequent cleaning are given in literature for different substrate types. Often a combination of multiple solvents and ultrasonic cleaning is chosen. The excellent cleaning efficiency of ultrasonic treatments is based on a "microscrubbing" effect caused by the implosion of cavitation bubbles within a liquid.

Glow Discharge Cleaning

After sufficient HV is achieved in an evaporation plant an inert gas (e. g. Ar) with a pressure of approx. 10 Pa is istroduced into the chamber at reduced pumping speed. A glow discharge is ignited between an insulated ring or rod shaped Al cathode and the substrate as anod. The substrate holder is usually kept at 2 kV voltage and a current of approx. 1 $mAcm^{-2}$ is flowing.

If the substrate is an insulator it is charged negatively due to the higher mobility of the ectrons compared to the heavy ions. An ion current from the plasma to the substrate is generated and the substrate is cleaned by sputtering. Electrically conductive substrates are hit by electrons drawn from the plasma which triggers the desorption of impurities and heats the substrate. In addition sputterered electrode material is deposited on the substrate and forms nucleation sites.

Sputter Cleaning

To clean an insulating substrate prior to deposition by sputtering, an RF voltage has to be supplied to prevent positive charging of the substrate. In the case of coating by ion plating continuous sputter cleaning is an integral part of the deposition process.

DC sputter cleaning ("Sputter Etching") is the adequate method to clean conductive substrates by sputtering.

Improvement of Adhesion

As mentioned before, bias sputtering and ion plating results in better adhering films than simple evaporation. Apart from in situ cleaning by ion or electron bombardement there are also other methods for adhesion improvement:

... Adhesion promoting lavers: a thin film of reactive material as e. g. Ti, Cr etc. is deposited initially on a glass or ceramic substrate. Then the desired metal layer is produced (e. g. Au, Ag). The adhesion promoting layer reacts chemically or by diffusion into the substrate and alloys with the film.

... Graded oxide transition: a metal layer is sputtered under addition of oxygen with an initially high partial pressure. Then the oxygen pressure is reduced gradually to zero.

... Implantation or diffusion of adhesion promoting atoms: For adhesion improvement of e. g. W on SiO_2 surfaces Al, which has a lower valence number (+3) than Si (+4) can be introduced into the surface by diffusion or implantation. An unsaturated oxygen bond is formed which is free for bonding W.

4.3.5.4. Adhesion Measurement

General

Under ideal conditions a device for the determination of adhesion should yield quantitative and reproducible results, should enable a non destructive and fast measurement on samples of arbitrary shape, should have a wide range of measurement and should be easily standardized with results independent of the operator. In reality, unfortunately, one is far from this ideal state.

For adhesion measurements on thin films the situation is complicated by the fact that the interface may contain inhomogeneities and flaws or that parameters change during deposition which may lead to significant spatial variations within the film. Additionally one has to take into account that in most test methods the connection to the sample is not ideal which limits the measurement range of a given method and influences the results uncontrollably.

Resulting from the above problems the fact arises that mainly empirical results can be derived. Nonetheless some of the many available test methods shall be discussed here.

Tape Test

An adhesive tape is applied to the film surface and pulled off again. The tape test is a subjective test which is not only dependent on the type of tape but also on the pull off velocity and the pull off angle. If the coating delaminates completely or partially bad adhesion is given in each case (see Fig. 4.34.). If the coating adheres to the substrate it is not possible to decide if adhesion is moderate, good or very good.

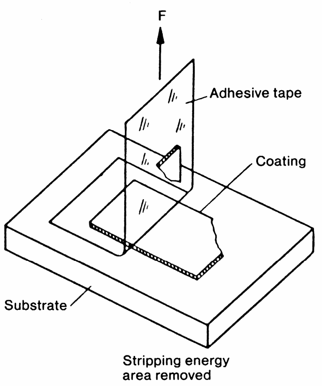


Fig. 4.34.: Schematic of the tape test [4, p. 80]

Cross Cut Test

A somewhat better significance than the tape test can be assigned to the cross cut test. With a suitable cutting instrument (DIN norm) a cutting band with 6 cuts, which reaches to the substrate, is drawn. A second one is drawn perpendicular to this cutting band, so that a grid consisting of 25 squares is formed. Afterwards, a tape test is performed on the grid and adhesion is judged in dependence on the percentual fraction of the removed film area. 6 steps can be distinguished (see Fig. 4.35.).

Gitterschnitt-kennwert	Beschreibung	Bild
Gt 0	Die Schnittränder sind vollkommen glatt, kein Teilstück des Anstriches ist abgeplatzt.	–
Gt 1	An den Schnittpunkten der Gitterlinien sind kleine Splitter des Anstriches abgeplatzt; abgeplatzte Fläche etwa 5% der Teilstücke.	
Gt 2	Der Anstrich ist längs der Schnittränder und/oder an den Schnittpunkten der Gitterlinien abgeplatzt; abgeplatzte Fläche etwa 15% der Teilstücke.	
Gt 3	Der Anstrich ist längs der Schnittränder teilweise oder ganz in breiten Streifen abgeplatzt und/oder der Anstrich ist von einzelnen Teilstücken ganz oder teilweise abgeplatzt; abgeplatzte Fläche etwa 35% der Teilstücke.	
Gt 4	Der Anstrich ist längs der Schnittränder in breiten Streifen und/oder von einzelnen Teilstücken ganz oder teilweise abgeplatzt; abgeplatzte Fläche etwa 65% der Teilstücke.	
Gt 5	Abgeplatzte Fläche mehr als 65% der Teilstücke.	–

Fig. 4.35.: Criteria for judging adhesion by the cross cut test [3, p. 304]

Quantitative Pull Off Test

For this method a cylindrical stud is glued to the coated side of the sample (see Fig. 4.36.). The force necessary to pull off the stud from the sample is measured by a force measurement cell.

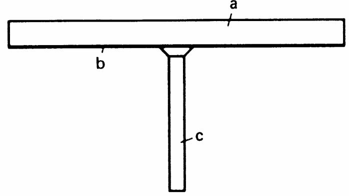


Fig. 4.36.: Coated substrate with glued stud [3, p. 305]

The significance of this method is limited by the fact that the glue, which is often applied under elevated temperatures, can diffuse into the film or can change internal stresses and therefore modifies the initial adhesion value between substrate and coating.

Of course reasonable results can only be achieved if the adhesion of the glue to the coating and to the stud is significantly higher than the one of the coating to the substrate. Given careful handling measurements up to adhesion forces of $5 kNcm^{-2}$ can be performed.

Ultra Centrifuge

The main problem for all pull off tests is the influence of the compound formed by measurement device and sample on the adhesion value. This problem is circumvented by the ultra centrifuge. The coating to be tested is applied to a rotor which turns so fast that film delamination happenes solely due to centrifugal forces.

For the determination of the adhesion force one only has to know the thickness and the density of the coating applied to the rotor. Given a small thickness compared to the diameter of the rotor

$$A = 4\pi^2 N^2 R\rho d \tag{4.35}$$

is valid, with

A = Adhesion force [Nm^{-2}]
R = Rotorradius [m]
d = Schichtdicke [m]
N = Umdrehungen [s^{-1}]
ρ = Dichte der Schicht [kgm^{-3}]

Adhesion values up to $50kNcm^{-2}$ can be measured. Apart from the problem of high rotation speeds for thin coatings (low total coating mass) the big disadvantage is that adhesion cannot be measured on technical parts but only on a specific rotor made from the same material which has to be coated separately.

Scratch Test

A rounded cone made of steel, tungsten carbide or diamond is moved across the sample perpendicular to the surface and loaded with increasing force L. At a critical load L_C the film cannot follow the deformation of the substrate anymore and cracks or delaminates from the substrate. L_C is dependent on the material of the coating and the substrate and is a measure for the adhesive strength.

The analysis of the scratch formed is mostly done in a SEM or, for automatic devices, by an acoustic emission detector which records the noises generated upon cracking of the film. (see Fig. 4.37.).

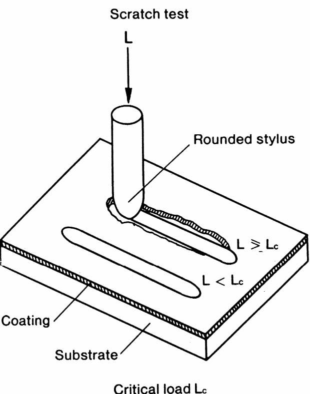


Fig. 4.37.: Scratch test [4, p. 83]

Theory yields the adhesion strength σ_A as

$$\sigma_A = K\sqrt{\frac{L_c H}{\pi R^2}} \qquad (4.36)$$

with

R = Radius of tip curvature [mm]
H = Sibstrate hardness [Nmm^{-2}]
L_C = Critical load [N]
K = experimentally determined coefficient [dimensionless]

In principle the scratch test is well suited for a quantitative assessment of adhesion strength ad is therefore widely used. Unfortunately the scratching process is complex so that there is no valid theory which comprises all quantities that influence an absolute measurement up till now. In the case of relative measurements the present method is well reproducible and is well suited even for coatings with excellent adhesion.

4.3.6. Friction and Wear

The coefficient of friction and the wear rate ($\Delta d/\Delta s$ = reduction in film thicknes Δd per friction distance Δs) are mostly measured by a pin on disk device under defined conditions (load on pin and material, constant turning velocity of the disc) in a well controlled ambient (dry nitrogen, water vapor, UHV).

For selected applications also test rigs are used which check the wear resistance of tribological coatings by special test procedures.

4.3.7. Ductility

Ductility is an important criterion for the amount of plastic deformation a thin film can sustain and serves as a guideline how far a part can be deformed after coating without film failure. With the increasing amount of decorative coatings and hard material films produced by PVD and CVD methods the measurement of ductility gains increasing importance.

The ductility of massive metals is evaluated by the well known tensile test. A measure of ductility is the relative breaking elongation given in percent. Ductile metals have a high, brittle materials a low breaking elongation. It is

$$\varepsilon_B = \frac{l_Z - l_0}{l_0} \qquad (4.37)$$

with

l_{ZZ} = length of the sample at the point of fracture
l_0 = length of unloaded sample
ε_B = breaking elongation in %

The breaking elongation of thin films ε_B can be determined by a bending test. The film which is deposited onto a ductile metal strip is bent around cylindrical bodies with decreasing diameter and the bending radius R is determined for which the film shows the first cracks. Under the constraint of $d << R$

$$\varepsilon_B = \frac{100d}{2R + d} \tag{4.38}$$

where *d* is the film thickness. A variation of this technique is the three point bending test where the ductile metal strip is supported in the middle and the ends of the strip are loaded by two edges with increasing force until the first cracks show up. This test can be performed in situ e. g. in a SEM.

4.3.8. Hardness

4.3.8.1. General

Hardness is defined as the mechanical resistance which a body opposes to a penetrating second, harder test body. The measured values depend on the elastic and plastic properties of the material, on the shape and properties of the test body and on the specific set-up. One distinguishes e. g. hardness according to Vickers, Rockwell and Brinell where these three types of hardness measurement show different shapes of the test body.

4.3.8.2. Micro Hardness Measurement

For the measurement of the micro hardness of thin films one mostly uses the microhardness measurement according to Vickers or Knoop. The terminus "microhardness" does not, as one might think, imply small hardness values but hardness measurement procedures producing wery small test impressions. A generally accepted (but not standardized) definition of microhardness means test loads of approx. 0,02 - 2 N, where practically the range from 0,05 N to 5 N is preferred. Utra microhardness measurements use test loads from 10^{-2} down to approx. 10^{-5} N.

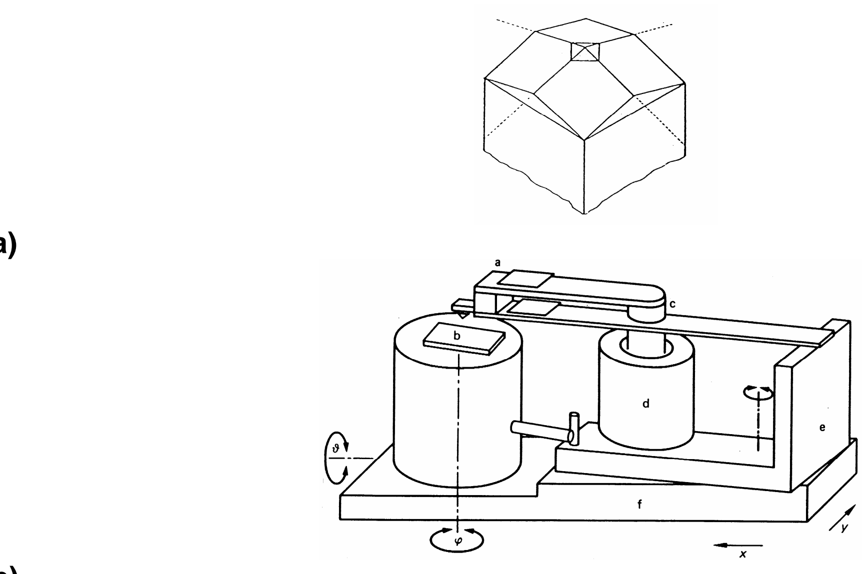


Fig. 4.38.: **a)** Vickers pyramid **b)** Schematic of an ultra microhardness tester [3, p.301]
a strain gauge; **b** sample; **c** double spring; **d** coil;
e clutch; **f** base plate

In the case of micro and ultramicrohardness measurement the test impression in the coating to be measured is produced by a pyramid shaped diamond (Vickers geometry) and its size is optically measured. The base area of the Vickers pyramid is quadratic and the opening angle between two opposite pyramid surfaces amounts to 136° (see Fig. 4.38a).

At given test load *p* [N] and length of the diagonals of the impression *d* [µm] the microhardness *HV* can be calculated by Eqn. (4.39).

$$HV = \frac{18192 p}{d^2} \tag{4.39}$$

It is easily understandable that a hardness measurement for very thin films which are used in optical and electronical applications is not possible with a commercially available microhardness tester which uses an optical microscope. Because of the low film thickness (<1 µm) one has to work with very low test loads which leads to impressions with diameters far below the diffraction limit. Bangert and Wagendristl developed a Ultramicrohardness measurement system which can be implemented into a SEM and allows measurements with test loads down to 10^{-4} N (see Fig. 4.38b). The state of the art, finally, is the so-called "nano indenter" where also the dependence of the tip displacement on the load can be measured and where the geometry of the impressions is determined by an AFM.

4.4. Electrical Properties

4.4.1. Introduction

The global economic and technological relevance of micro electronics is out of question. The development, production and investigation of micro electronic devices is also a significant part of thin film technology. Understanding conductivity, finally is the basis for developments which happened during the last few decades, starting from a simple diode to a VLSI circuit or a Gibabit RAM chip.

The electrical conductivity of thin films deviates from the bulk conductivity due to the size effect (film thicknesses < 1 µm) and because of other effects as e. g. grain size or incomplete continuity of the film. The conductivity can be determined by the sheet resistance R_N.

4.4.2. Definition and Measurement

The sheet resistance is the resistance which one measures at two electrodes which are located at opposite sides of a film with a square area. R_N (which is also often termed $R_{\square}$)is independent on the size of the quadratic area because the electric resistance of a cuboid with thickness d, width b, length l and specific resistivity ρ (see also Fig. 4.39.) is given by:

$$R = \rho \frac{l}{db} \qquad (4.40)$$

For a quadratic surface ($l = b$)

$$R = R_{\square} = \frac{\rho}{d} \qquad (4.41)$$

Independent on the size of the square.

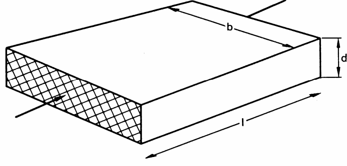


Fig. 4.39.: Definition of the sheet resistance [3, p. 262]

The sheet resistance is measured by the four point method. Four point shaped electrodes are put in contact with the film surface. The electrodes may be arranged linear with equal distance from each other or on the edges of a square (see Fig. 4.40.). For the linear arrengement a current I is generated by the outer electrodes (taken from a constant current source in the mA to µA range) and the voltage drop between the inner electrodes is measured.

For the sheet resistance

$$R_{\square} = 4{,}532\frac{U}{I}. \qquad (4.42)$$

In the case of the quadratic electrode arrangement the current is supplied through two neighboring electrodes and the voltage drop is measured at the electrodes opposite these two electrodes:

$$R_{\square} = 9{,}06\frac{U}{I} \qquad (4.43)$$

Eqns. (4.42) and (4.43) are only valid for a film surface which is large when compared to the electrode distance. If this is not the case geometric correcting factors have to be used which are supplied in literature.

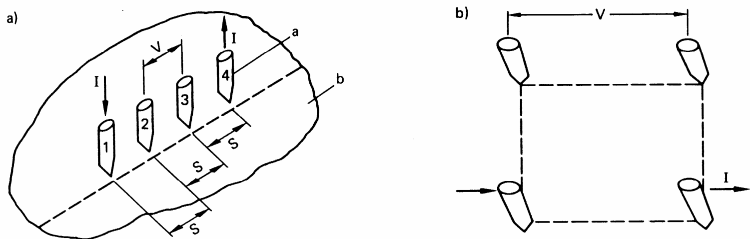


Fig. 4.40.: Determination of the sheet resistance [3, p.263]
a: linear four point method
b: quadratic four point method
I constant current; **U** measured voltage drop;
s electrode distance; **a** electrodes; **b** sample

4.4.3. Theoretical Fundamentals

4.4.3.1. Leitfähigkeit durchgehender dünner Schichten

The description of the electrical conductivity in thin films can be attributed to Fuchs and Sondheimer (approx. 1940). The starting point is Drude's theory of the electrical conductivity of metals.

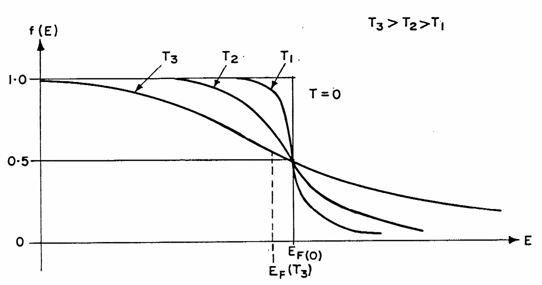


Fig. 4.41.: *Fermi Dirac statistics: distribution function f(E) [10, p. 30]*

The conduction electrons are treated as free electron gas which is described by Fermi Dirac statistics. This yields the distribution function in equilibrium:

$$dN = 2 f_0(E)(V/h^3)m^3 dv_x dv_y dv_z \tag{4.44}$$

with

$$f_0(E) = \frac{1}{1 + e^{(E-E_f)/kT}} \tag{4.45}$$

where E is the energy and E_f is the fermi energy. All other quantities have their usual meaning. Without an external field no net current **j** *is observed due to symmetry (see appendix):*

$$\mathbf{j} = -e\int_N \mathbf{v}dN/V = 2e\int_{\mathbf{v}} f_0(E)\mathbf{v}(m^3/h^3)dv_x dv_y dv_z = 0 \tag{4.46}$$

An external **E** *field changes the energy distribution. Its spherical symmetry is lost. This change can be described by the Boltzmann equation.*

Assuming small changes for the stationary stete

$$\frac{df}{dt} = \frac{e\mathbf{E}}{m} grad_v f - \mathbf{v} grad_r f + \left(\frac{df}{dt}\right)_{coll} \tag{4.47}$$

is valid. With the ansatz

$$f(t) - f_0 = Ce^{-t/\tau} \tag{4.48}$$

the collision term an be expressed by

$$\left(\frac{df}{dt}\right)_{coll} = -\frac{f(t) - f_0}{\tau} \tag{4.49}$$

under the assumption that the distribution function exponentially approaches equilibrium after the the dusturbance. The time constant τ is the mean time between two collisions. In an isotropic metal in the stationary state df/dt = 0 is valid. Therefore, Eqn (4.47) yields:

$$\frac{e\mathbf{E}}{m} grad_v f - \mathbf{v} grad_r f - \frac{f(t) - f_0}{\tau} = 0 \tag{4.50}$$

Assuming the E-field only to have one component in x-direction and that the distribution of electrons is only influenced in the z-direction due to the film interfaces one obtains:

$$\frac{eE}{m}\frac{df}{dv_x} - v_z\frac{df}{dz} = \frac{f(t) - f_0}{\tau}. \tag{4.51}$$

To consider scattering of electrons at the film interfaces one writes the distribution function f as

$$f = f_0 + A(E, z) \tag{4.52}$$

with A a small distortion. Eqn. (4.51) then yields

$$v_z \frac{dA}{dz} + \frac{A}{\tau} = \frac{eE}{m} \frac{df_0}{dv_x} \tag{4.53}$$

which has the general solution

$$A = \frac{eE\tau}{m} \frac{df_0}{dv_x} (1 + Ke^{-z/\tau v_z}) \tag{4.54}$$

with the integration constant K which can be determined by the boundary conditions.

The current density j is given by

$$j = -e \int v_x f dv_x dv_y dv_z \tag{4.55}$$

and the cinductivity σ is given by

$$\sigma = j / E . \tag{4.56}$$

All further results will be referred to the conductivity of the bulk, σ_0. The reflection of electrons at the film interfaces can either be treated as completely diffuse scattering or as partially specular.

<u>*Diffuse Scattering*</u>

Given the assumption of completely diffuse scattering one can assume that electrons scattered at an interface start with the equilibrium distribution function, i. e. that the field drift generated by **E** *is completely lost. From this, after doing some calculus, one obtains for the conductivity σ*

$$\frac{\sigma}{\sigma_0} = 1 - \frac{3}{8k} + \frac{3k}{4}\left(1 - \frac{k^2}{12}\right)\int_k^{\infty} \frac{e^{-x}}{x} dx + \left(\frac{3}{8k} - \frac{5}{8} - \frac{k}{16} + \frac{k^2}{16}\right)e^{-k} \tag{4.57}$$

where: $k = d/\lambda_0$, with d = film thickness and λ_0 = mean free path of the conduction electrons in the bulk (10 to 40 nm at room temperature).

For specialcases Eqn. (4.57) can be simplified:

$$\frac{\sigma}{\sigma_0} = 1 - \frac{3}{8k} \qquad \text{for } k>>1 \tag{4.58}$$

and

$$\frac{\sigma}{\sigma_0} = \frac{3k}{4} \ln \frac{1}{k} \qquad \text{for } 0<k<<1 \tag{4.59}$$

Partial Specular Reflection

For the case of partial specular reflection of the electrons on an interface it can be said first that completely elastic collisions (corresponding to completely specular reflection) at the interface do not change the conductivity since the drift remains unchanged. Therefore only the case is of interest, where partially inelastic scattering can be assumed. With p as the part of elastically scattered electrons (p ≤ 1) one obtains

$$\frac{\sigma}{\sigma_0} = 1 - \frac{3(1-p)}{8k} + \frac{3}{4k}(1-p)^2 \sum_{n=0}^{\infty} p^{n-1} \left[E_1(kn)\left(k^2n^2 - \frac{k^4n^4}{12}\right) + e^{kn}\left(\frac{1}{2} - \frac{5kn}{6} - \frac{k^2n^2}{12} + \frac{k^3n^3}{12}\right)\right] \quad (4.60)$$

with

$$E_1(k) = \int_k^{\infty} \frac{e^{-x}}{x} dx \,. \quad (4.61)$$

For the simplification k > 1

$$\frac{\sigma}{\sigma_0} = 1 - \frac{3}{8k}(1-p) \quad (4.62)$$

is valid.

Fig 4.42. shows the dependence of the specific resistivity ratio ρ/ρ_0 on k.

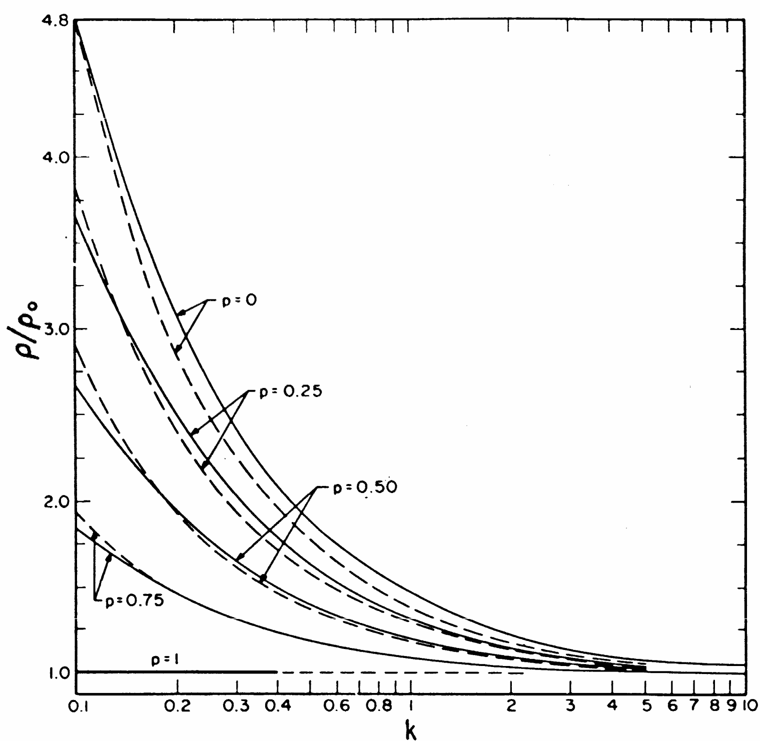


Fig. 4.42.: *Dependence of the specific resistivity ratio ρ/ρ_0 on k [10, p. 155]:*
___ for Eqn. (4.60)
- - - for Eqn. (4.62)

Although the above considerations include partially directed reflection and are an improvement when compared to the case of totally diffuse reflection one has to admit that these models still do not asllow for an exact description of conductivity. The reasons for that are:

... Scattering of electrons happenes on both interfaces of the film, where p might differ.

... Electrons can be scattered at the crystal lattice (τ_G) at grain boundaries (τ_K) and at impurities (τ_V).

Then the relaxation time τ defined in Eqn. (4.25) can be written as

$$\frac{1}{\tau} = \frac{1}{\tau_G} + \frac{1}{\tau_K} + \frac{1}{\tau_V} \qquad (4.63)$$

which is called "Matthiessen's rule

4.4.3.2. Conductivity of Discontinuous Thin Films

If no continuous metallic coating is present then the conductivity of a discontinuous metallic film on an insulating substrate is determinded by the mean thickness D (see Fig. 4.43.).

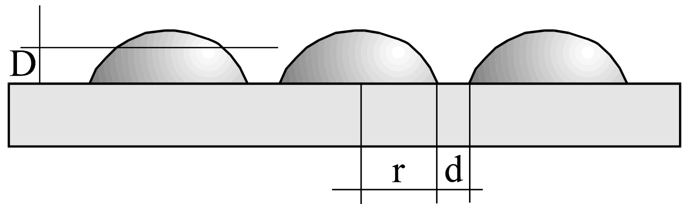


Fig. 4.43.: *Schematic of a discontinuous film*

The experimental findings are:

1. $\sigma_{dis} << \sigma_{kont}$
2. σ proportional. $e^{-A/kT}$
3. $\sigma = \sigma(\mathbf{E})$
4. $\sigma = \sigma(r,d)$

They can be explained by the following mechanisms.

1. *Thermal emission of electrons into the vacuum or into the conduction band of the insulating substrate. This mechanism is dominating for d > 10 nm und T > 300 K.*

2. *Quantum mechanical tunneling: possible via the substrate and via vacuum.*

4.4.4. Practical Aspects

4.4.4.1. General

The demands on thin film resistors are various. One of the most important features is the lerge area of resistance values which has to be covered. It extends from approx. 100Ω/□ to 100MΩ/□. A variation of the sheet resistance which comprises this many orders of magnitude cannot be realized by changing the geometrical parameters of the coatings alone. Different materials with selected specific resistivities have to be used. Fig. 4.44. shows an overview of the dependence of the Temperature Coefficient of Resistivity (TCR) on the sheet resistance for the most common materials used for thin film resistors.

For films with relatively low resistance values metals or metal alloys are selected. For higher resistance values so-called cermets (**ce**ramic **met**al combinations) are used. For extremely high resistance values the classical insulators as e. g. quartz are the natural choice.

Other important demands are a low TCR, a good stability of the electrical properties and cost effective production. Therefore not only the choice of material, but also the selected deposition processes and process parameters are important, because they also influence the coating properties.

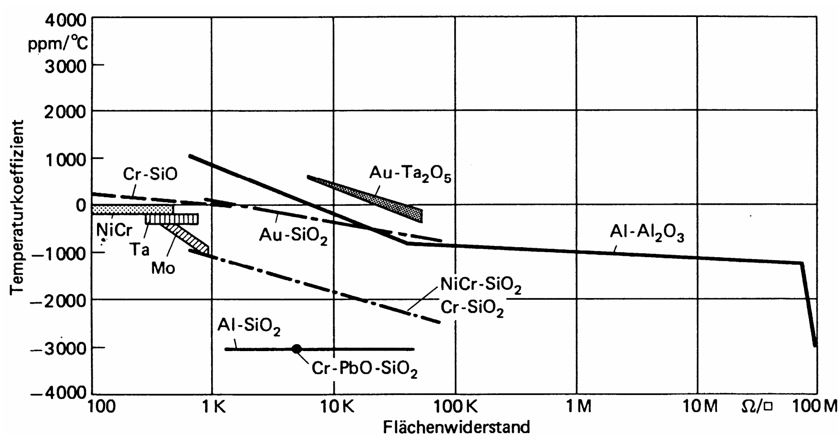
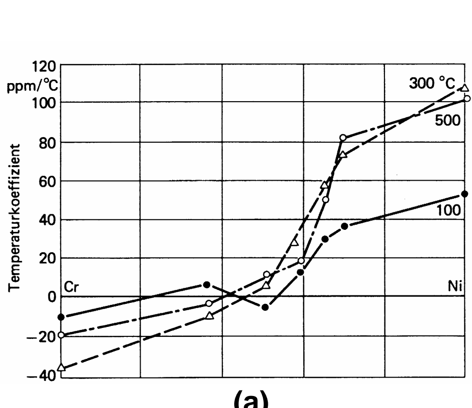
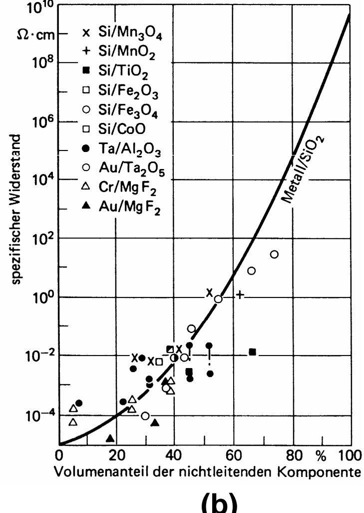


Fig. 4.44.: Temperature Coefficient of Resistance (TCR) of different materials in dependence on the sheet resistance [3, p. 403]

4.4.4.2. NiCr-Thin Film Resistors

For low resistivity values NiCr films are preferred because of their excellent properties in regard to low noise, stability and high durability. The sheet resistance and the TCR depend on the composition of the alloy. The electrical properties are critically influenced by the residual gas pressure and the composition of the background gas in the case of PVD processes. In Fig. 4.45a the TCR of NiCr films is given in dependence on the substrate temperature and on the alloy composition.





(a) (b)

Fig. 4.45.: **(a)**TCR or PVD deposited NiCr films (PVD) in dependence on their composition and at different substrate temperatures
(b) Resistance values of typical cermets
[3, pp. 406, 409]

4.4.4.3. Cermet Thin Film Resistors

The field of application of thin film resistors made of NiCr extends to about 1000Ω/□. This limit is given by the fact that for very high resistance values the fims are so thin that extreme demands on substrate roughness are imposed. Also the reproducibility of film properties decreases and the TCR increases.

For higher resistance values cermets are chosen. These are mixtures of metals and metal oxides or other chemical compounds. These materials allow for a sheet resistance from 1Ω/□ up to 1MΩ/□ accompanied by a low TCR. Variaous material combinations have been investigated and applied (see Fig. 4.45b).

4.4.4.4. Semiconductor Films

R&D worldwide is permanently investigating semiconducting materials and semiconductor technologies to comply with emerging demands posed on electronic and optoelectrinic devices. This is especially valid for thin film technology without which modern semiconductor industry would be unthinkable.

Si is the most common and best investigated semoconducting material (90% of all devices). The importance of compound semiconductors (GaAs, AlAs, InP) is steadily increasing in respect to electronic and optoelectrinic applications.

Deposition processes for semiconductors range from high purity CVD via Molecular Beam Epitaxy (MBE) to Metallo Organic CVD (MOCVD).

4.4.4.5. Magnetic Thin Films

Of course also magnetism is a direct consequence of the electronic structure of a material. In connection with thin film technology it only shall be mentioned here that by using PVD methods as e. g. magnetron sputtering the storage density of magnetic storage media could be tremendously increased from 1 MByte/dm^2 (3 1/4" diskette) to some 10 GByte/dm^2 (modern hard disc) due to the possibility to deposit very fine grained magnetic coatings.

4.5. Optical Properties

4.5.1. Introduction

One of the first technologically relevant applications of thin films was optical coatings. Because of its huge importance the invention of antireflex coatings in the first half of the 20th century was indicated to be one of the few funfamental advances in applied optics. The field of interference optics has also considerably widened by the development of new coating technologies, by improvements in instrumentation and by the development of new scientific and technological fields.

4.5.2. Theory

4.5.2.1. Fresnel's Equations

For optically active films energy conservation is valid:

$$T + R + A + S = 1 \qquad (4.64)$$

with

T = Transmission
R = Reflection
A = Absorption
S = Scattering

Generally, scattering may be neglected so that it is possible to calculate transmission from reflection and vice versa if a film is not absorbing. Therefore, the following considerations can be limited to calculating one of these quantities.

An statement about the reflected and the transmitted part and about the polarization state of a plane wave with known direction of oscillation which hits an interface can be made by the equations which were first derived by A. J. Fresnel.

$$r_k^p = \frac{n_{k-1}\cos\varphi_k - n_k\cos\varphi_{k-1}}{n_{k-1}\cos\varphi_k + n_k\cos\varphi_{k-1}} \qquad (4.65)$$

$$r_k^n = \frac{n_{k-1}\cos\varphi_{k-1} - n_k\cos\varphi_k}{n_{k-1}\cos\varphi_{k-1} + n_k\cos\varphi_k} \qquad (4.66)$$

$$t_k^p = \frac{2n_{k-1}\cos\varphi_{k-1}}{n_{k-1}\cos\varphi_k + n_k\cos\varphi_{k-1}} \qquad (4.67)$$

$$t_k^n = \frac{2n_{k-1}\cos\varphi_{k-1}}{n_k\cos\varphi_k + n_{k-1}\cos\varphi_{k-1}} \qquad (4.68)$$

r and t are the ampuitude vectors (Fresnel coefficients) for reflection and transmission. p - vectors are perpendicular to the paper plane and the positive direction is up. φ_k are the angles which a light ray within the k-th medium with refractive index n_K is including with the perpendicular to the k-th interface.

For perpendicular impingement $r_k^p = r_k^n$ und $\varphi_k = 0$. Given a bulk surface

$$r = \frac{n_0 - n_1}{n_0 + n_1} \tag{4.69}$$

and

$$t = \frac{2n_1}{n_0 + n_1} \tag{4.70}$$

wher n_0 is the refractive index of the surrounding medium and n_1 is the refractive index of the solid body which fills the half space. The energy flux *R,* which quantifies the ratio of reflected and impinging energy, is given at perpendicular impingement by

$$R_k^p = R_k^n = (r_k^p)^2 = (r_k^n)^2 \tag{4.71}$$

and

$$R_k = (r_k)^2 = \left(\frac{n_{k-1} - n_k}{n_{k-1} + n_k} \right)^2 \tag{4.72}$$

For transmission at perpendicular incidence, given the fact that the light ray travels through two media with different refractive indices, one obtains

$$T_k = \frac{n_{k-1}}{n_k} (t_k)^2 = \frac{4n_{k-1}n_k}{(n_{k-1} + n_k)^2} . \tag{4.73}$$

The above relations are only valid for absorption free media and if scattering can be neglected.

4.5.3. Anti Reflection Coatings

4.5.3.1. Single Layer

If an isotropic, absorption free thin coating with a thickness d_1 in the region of the wavelength of light and a refractive index n_1 is deposited onto a substrate with refractive index n_2 and the whole system is is located within a medium with refractive index n_0 then multiple reflections will be present at the interfaces n_2/n_1 and n_1/n_0 (see Fig. 4.46.).

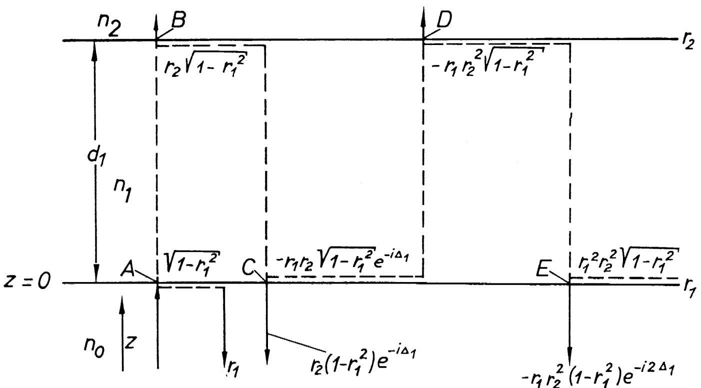


Fig. 4.46.: *Sketch illustrating the derivation of the reflection on an absorption free, homogenous single layer at perpendicular light incidence [11, p.13].*

For perpendicular incidence summing up all partial beams yields

$$r = \frac{r_1 + r_2 e^{-i2\delta_1}}{1 + r_1 r_2 e^{-i2\delta_1}} \qquad (4.74)$$

and

$$t = \frac{t_1 t_2 e^{-i2\delta_1}}{1 + r_1 r_2 e^{-i2\delta_1}} \qquad (4.75)$$

where r_1, r_2, (see Eqns. (4.81) and (4.82)!) t_1 and t_2 are the Fresnel coefficients of the two interfaces and $\delta_1 = (2\pi/\lambda)n_1 d_1 \cos\varphi_1$ is the phase thickness of the film. n_1 and d_1 are the refractive index and the thickness of the film, respectively. λ is the wavelength of the incoming light and φ_1 is the angle between the perpendicular and the light ray.

Taking into account Eqns. (4.72) and (4.73) as well as $T_k+R_k = 1$, one obtains

$$R = \frac{r_1^2 + 2r_1 r_2 \cos 2\delta_1 + r_2^2}{1 + 2r_1 r_2 \cos 2\delta_1 + r_1^2 r_2^2} \qquad (4.76)$$

and

$$T = \frac{n_0}{n_2}\frac{t_1^2 t_2^2}{1 + 2r_1 r_2 \cos 2\delta_1 + r_1^2 r_2^2} \qquad (4.77)$$

Given Eqn. (4.74) reflection either vanishes at infinite denominator, which is unphysical, or if the enumerator is zero. In this case, real part and imaginary part have to be zero.

Because

$$e^{+/-i2\delta_1} = \cos 2\delta_1 +/- i \sin 2\delta_1 \qquad (4.78)$$

$$r_1 + r_2 \cos 2\delta_1 = 0 \qquad (4.79)$$

and

$$r_2 \sin 2\delta_1 = 0 \tag{4.80}$$

have to be valid to get R = 0. with

$$r_1 = \frac{n_0 - n_1}{n_0 + n_1} \tag{4.81}$$

and

$$r_2 = \frac{n_1 - n_2}{n_1 + n_2} \tag{4.82}$$

Eqns. (4.79) and (4.80) are satisfied, if $2\delta_1$ is an even multiple π. This solution with $n_0 = n_2$ is practically relevant for "invisible" protective layers ("λ/2-coatings"). For the second solution ($n_0 \neq n_2$) $2\delta_1$ is an odd multiple of π. Following Eqns. (4.79) to (4.82) one obtains

$$n_1 = \sqrt{n_0 n_2} \tag{4.83}$$

with

$$\delta_1 = (2m-1)\frac{\pi}{2} \qquad , m=1,2,3,... \tag{4.84}$$

Therefore, to obtain zero reflection for a single layer with refractive index n_1 on a substrate with refractive index n_2 the so called amplitude condition (4.83) and phase condition (4.84) have to be fulfilled.

For perpendicular incidence one obtains upon insertion of Eqns. (4.81) and (4.82) into (4.76) and (4.77)

$$R_{max} = \left(\frac{n_0 - n_2}{n_0 + n_2}\right)^2 \qquad \text{for} \qquad \lambda_{max} = \frac{2n_1 d_1}{m} \tag{4.85}$$

and

$$R_{min} = \left(\frac{n_1^2 - n_0 n_2}{n_1^2 + n_0 n_2}\right)^2 \qquad \text{for} \qquad \lambda_{min} = \frac{4n_1 d_1}{2m-1} \tag{4.86}$$

In Fig. 4.47. the reflectivity R of single layers with refractive index n_1 deposited on glass with n_2 = 1,52 is drawn for perpendicular incidence. The coated substrates are located in air (n_0 = 1).

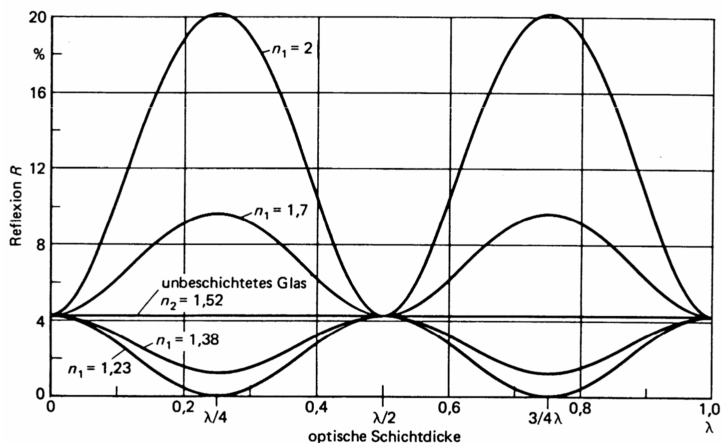


Fig. 4.47.: Reflection at absorption free single layers with different refractive indices n_1 on glass substrates (n_2 = 1,52) [3, p. 459].

For $n_1 = n_2$ there is no change in R, $n_1 < n_2$ yields a decrease and $n_1 > n_2$ leads to an increase in reflectivity. A zero in reflection can only be achieved if the amplitude and the pahase condition (4.83) and (4.84) are fulfilled simultaneously. If only the phase condition is fulfilled then one obtains a minimum in reflectivity which is closer to zero if $(n_1^2 - n_0 n_2)$ is smaller.

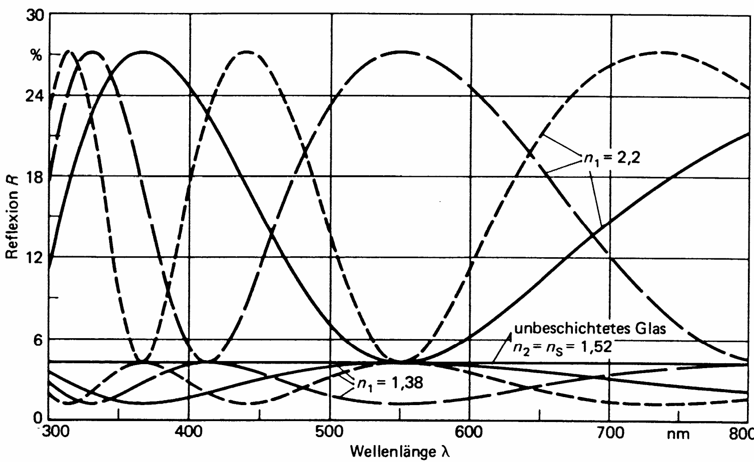


Fig. 4.48.: Reflection at absorption free single layers with different refractive indices and thicknesses on glass in dependence on the wavelength of impinging light; wavelength of reference: 550 nm [3, p. 459]

At non-perpendicular incidence r^p and r^n are different. For the calculation of the reflectivity the angle φ_1 has to be considered. Zeroes in reflection are then obtained for $n_1 d_1 \cos\varphi_1 = (2m - 1)\lambda/4$. The film thickness has to be larger by a factor $1/\cos\varphi_1$ in the case of oblique incidence when compared to the perpendicular case. Therefore, for oblique incidence the maxima and minima of reflectivity are shifted towards smaller wavelengths with increasing angle of incidence (see Fig. 4.49.).

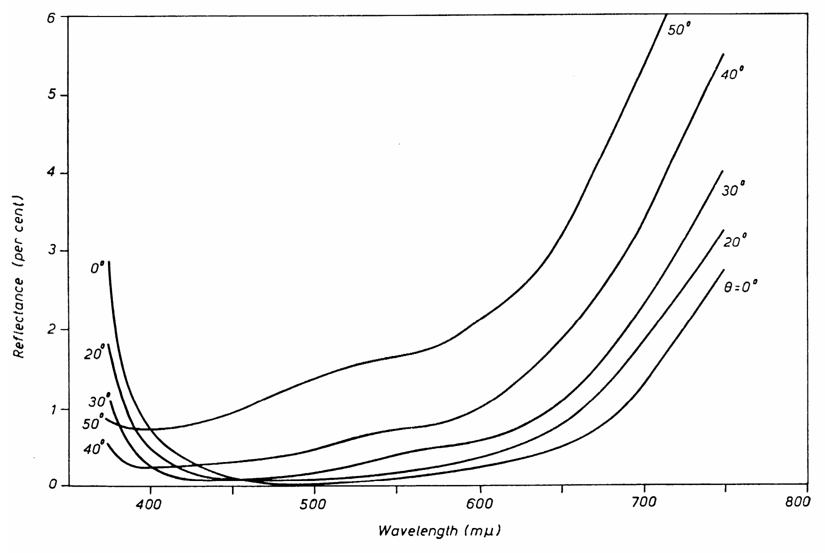


Fig. 4.49.: Reflectivity of a reflection suppressing film system at different angles of incidence [12, p. 57]

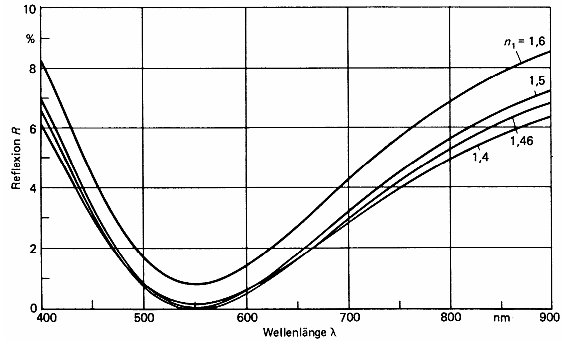
From the above considerations one can see that it is impossible to obtain full transmission for an extended range of wavelengths for different angles of incidence for a single layer system. Nonetheless it is possible - with the aid of modern calculation-, coating- and measurement tools - to manufacture optical systems with high transmission values. This is possible only by employing multilayered systems.

4.5.3.2. Double Layer

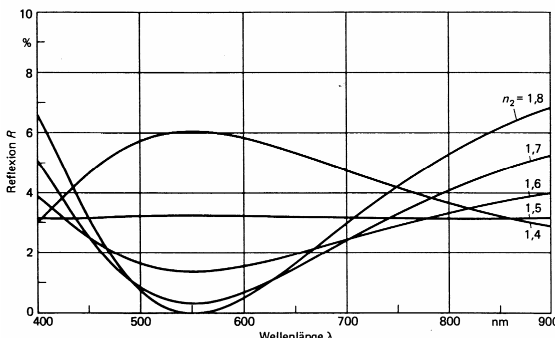
Zeroes in reflection can be achieved for single layers only on substrates with a refractive index higher than the one of the coating. From the available coating materials which satisfy todays demands in respect to durability towards influences from the ambient and towards mechanical loads MgF_2 has the smallest refractive index of n = 1,38. After (4.38) zero reflectivity can be achieved only for a substrate with n_s = 1,9 if the phase condition (4.84) is met.

The glasses usually employed in optics have refractive indices ranging from 1,5 - 1,7 which is far below 1,9. Zero reflection can therefore not be achieved by a MgF_2 single layer. With double or, even better, with multilayers one can achieve zero reflection also on substrates with low refractive indices. For their calculation the same methods as for the derivation or reflectivity and transmittivity of single layers can be employed.

In Fig. 4.50. the dependence of the reflection of $\lambda/4$-double layers on glass (n=1,52) on the wavelength at different refractive indices is displayed.



a



b

Fig. 4.50.: Dependence of the reflectivity of a $\lambda/4$-double layer on the wavelength at variable refractive index of **a** n_1 (n_2=1,8, n_3=1,52) or **b** n_2 (n_1=1,46, n_3=1,52) [3, p. 462]

In real life mostly coating materials with the calculated refractive indices to match the amplitude condition and therefore allow for a zero in reflectivity are unavailable.To obtain an optimum approximation of the real reflectivity to the desired one often coatings with film thicknesses other than $\lambda/4$ are used. This allows to tune the width of the region with low reflectivity as well as the intensity of suppression of reflectivity within this region.

4.5.3.3. Multilayer Coatings

In comparison to double layers the suppression of reflectivity within a given region is by far stronger for multilayer coatings (see Fig. 4.51.). Because of this very small deviations from the amplitue or phase condition are easily visible by a change in the reflected color.

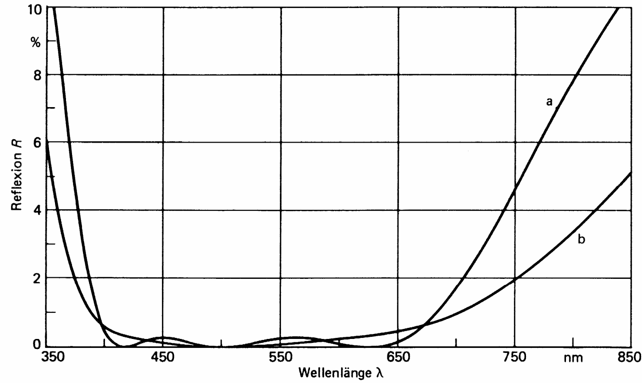


Fig. 4.51.: Reflection suppression of glass by triple layers with different thickness combination: **a** $\lambda/4$ - $\lambda/2$ - $3\lambda/4$; **b** $\lambda/4$ - $\lambda/2$ - $\lambda/4$ n_0 = 1; n_1 = 1,38; n_2 = 2,1; n_3 = 1,7; n_4 = 1,52 [3, p. 464]

Even higher degrees of transmissivity can be achieved with multilayer systems. Without coating modern optical systems with up to 100 interfaces between glass and air would be unthinkable. Termini like "fully coated" or "multi coated" can be found on nearly every optical component.

4.5.4. Reflection Enhancement

4.5.4.1. Metal Mirrors

For highly reflecting coatings metal mirrors are the easiest choice. In addition they can be covered by absorption free interference layers which enhance reflectivity and act as protective layer at the same time (see Fig. 4.53.). The reflectivity of metallic layers is not only dependent on the material but also on the deposition method and on the roughness parameters of the substrate since a smooth surface is prerequisite for high reflectivity.

For mirror layers preferrably Al is used since it shows good reflectivity from UV to IR (see Fig. 4.52.). Upon storage on air a thin oxide layer is formed which reduces reflectivity but enhances chemical stability.

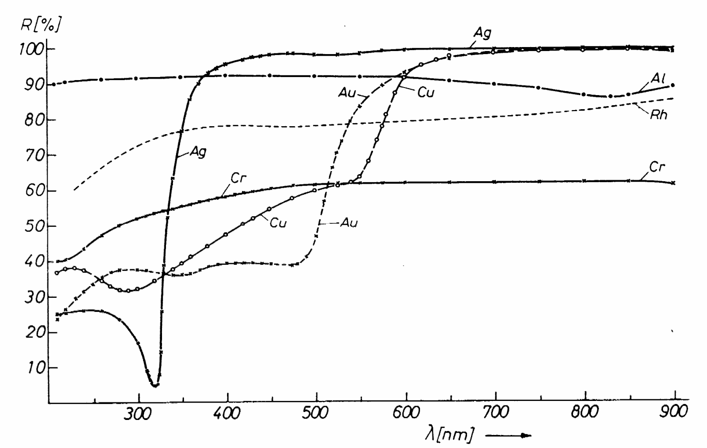


Abb. 4.52.: Reflectivity of freshly deposited mirror coatings [11, p. 93]

Silver mirrors exhibit a higher reflectivity than Al mirrors, but have the disadvantage that they discolorate after short times due to sulfide formation and lose their reflectivity. Gold and copper have a high IR reflectivity while their optical properties are very wavelength dependent in the visible region.

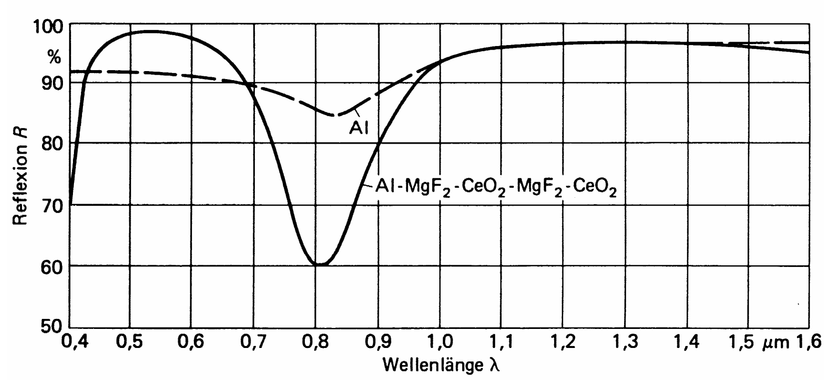


Fig. 4.53.: Reflectivity of an evaporated Al coating with and without reflection enhancing layer [3, p. 476]

4.5.4.2. Dielectric Mirrors - Multilayer Systems

An additional possibility for reflectivity enhancement is given by multilayered systems. Eqn. (4.84) yields that reflection enhancement is given when a material is deposited on a substrate with a refractive index higher than that of the substrate. The reflectivity increase is higher the higher the refractive index of the coating is when compared to that of the substrate. An additional result of (4.84) is that for a given refractive index reflectivity reaches a maximum if the optical thickness is a multiple of $\lambda/2$.

An amplification of this effect can be reached if $\lambda/4$ layers with alternating low and high indices of refraction are deposited. Maximum reflection is achieved if the number of films is odd and in the beginning and the end of the stack a film with the higher refractive index is located.

In Fig. 4.54. the reflectivity for two different material combinations is calculated for different layer numbers. Scattering and absorption are neglected.

Anzahl der Schichten	Reflexion in %	
	$n_L = 1{,}38$, $n_H = 2{,}3$, $n_s = 1{,}51$	$n_L = 1{,}47$, $n_H = 2{,}3$, $n_s = 1{,}51$
3	53,89	53,23
5	85,20	80,84
7	94,67	92,15
9	98,08	96,79
11	99,31	98,68
13	99,75	99,46
15	99,91	99,78
17	99,97	99,91
19	99,99	99,63

Fig. 4.54.: Reflection of multiplayer systems; H: high refractive index, L: low refractive index [3, p. 467]

The application of highly reflecting coating systems are laser mirrors on the one hand and so-called cold light mirrors on the other hand. Laser mirrors with 19 or more films of high quality show losses below 0,01%. Cold light mirrors show high reflectivity in the visible region and high transmittivity in the IR region. So a high light yield at low warming is achieved (e. g. for projectors). Also Infrared optics and X-ray mirrors use multilayers as reflecting elements since the reflectivity of single layers is too low in this spectral region.

4.5.5. Filters

4.5.5.1. Fabry-Perot-Filter

In the simplest case a Fabry Perot filter consists of two partially reflecting metal layers (mostly Ag) between which a spacer layer made from an absorption free material is located. The optical thickness of the spacer layer determines the location of the transmission band.

$$\lambda = \frac{2nd}{k} \qquad \text{with } k=1,2,3,... \qquad (4.87)$$

is transmitted, where *nd* is the optical film thickness. *k* is the so-called order of the Fabry perot filter.

The finite absorption of the metal layers leads to disadvantages as e. g. side maxima or difficulties regarding production in respect to the position of the transmission region.

These disadvantages can be avoided by using dielectric multilayer systems. The mode of operation is the same, only that the metallic reflectors are replaced by highly reflecting multilayers. A reflection enhancing film stack consisting of an odd number of λ/4 layers and a layer with a high refractive index at the bottom and the top of the stack is put on each side of the spacer layer (see Fig. 4.55.).

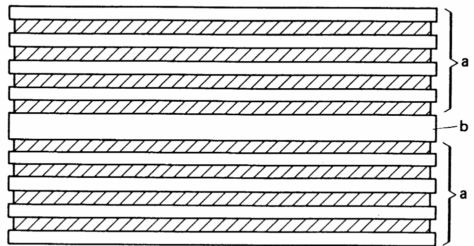


Fig. 4.55.: Fabry Perot filter made of dielectric films [3, p. 480]:
a reflection enhancing λ/4 coating system
b λ/2 spacer layer

By properly chosing the parameters it is possible to tune the width, shape and position of the transmission area as well as the transmittivity in a wide range. Transmissivities of more than 90% and half widths of 0,05 nm can be achieved (see Fig. 4.56). If one uses this kind of filters one has to take into account that, for unprotected systems, the peak is slightly shifted due to humidity and/or temperature.

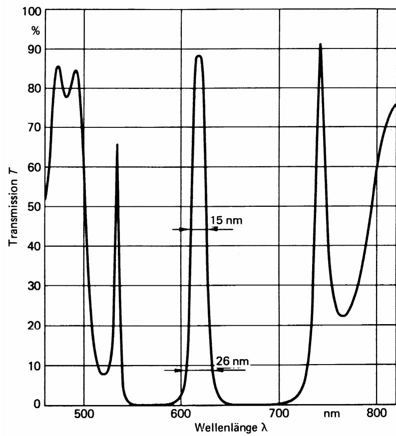


Fig. 4.56.: Transmission curve of a narrow band interference filter [3, p. 482]:
Assembly: $(HL)_3$ H $(LH)_3$ L $(HL)_3$ H $(LH)_3$
H: high refractive index λ/4 layer, L: low refractive index λ/4 layer

4.5.5.2. Edge Filter

Edge filters are optical coating systems which have an excellent transmissivity in one spectral region and an excellent reflectivity in an adjacent spectral region. The transition region between these two areas is very narrow (see Fig. 4.57.). Examples are temperature protection filters and blocking or transmission filters for certain spectral ranges (IR, VIS, UV).

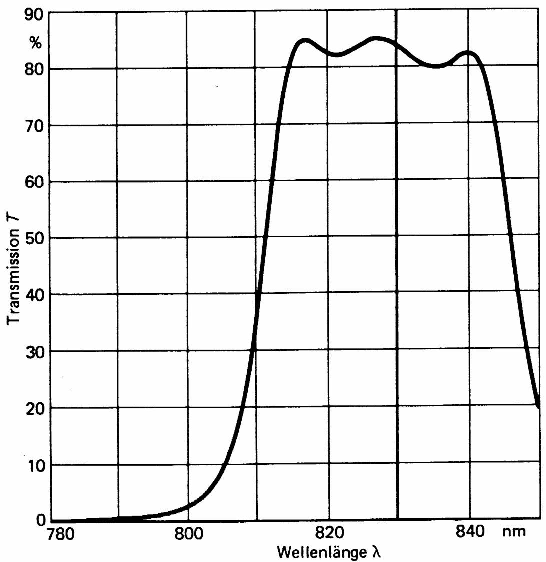


Fig. 4.57.: Transmission curve of an edge filter [3, p. 482]:
Assembly: $(HL)_2$ 4H $(LH)_2$ L $(HL)_2$ 4H $(LH)_2$ 4H $(LH)_2$
H: high refractive index λ/4 layer, L: low refractive index λ/4 layer

By a proper choice of the coating materials, of their respective thicknesses and of their sequence it is possible to realize nearly every desired filter design in respect to bandwidth, edge position and transmission.

4.5.6. Characterization of Optical Coatings

4.5.6.1. General

The reproducible production of thin films for optics is much easier to achieve today due to the available measurement methods, the advanced deposition plants and the higher degree of aufomatization.

Nonetheless the control of the final products is necessary even if it is only by random inspection. For optical coatings sticking to the pre defined optical values is often not enough because also selected mechanical properties as e. g. hardness, adhesion or chemical stability have to be given for a coating subjected to application.

4.5.6.2. Characterization Methods

The measurement methods for film thickness, mechanical properties and chemical composition were treated in detail in the proper chapters. The measurement of optical constatnts is mostly done with an ellipsometer.

Ellipsometer

There are many possible set-ups for an ellipsometer, but only the RAE (Rotating Analyser Ellipsometer) set up is of practical relevance (see Fig. 4.58.).

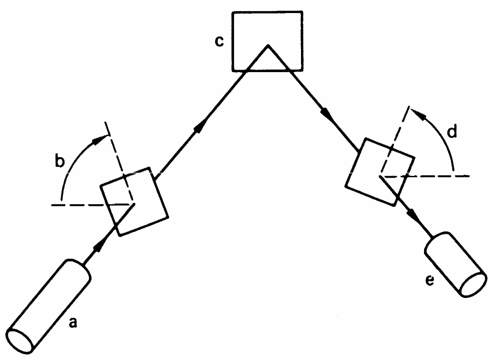


Fig. 4.58.: *RAE-ellipsometer:* ***a*** *light source;* ***b*** *polarisator;* ***c*** *substrate;* ***d*** *analysator;* ***e*** *photo detector [3, p. 394]*

Typical values for the angle between a and e are 130 to 140°. If one designates the angle between polarisator and the plane of impingement with P and the corresponding angle of the analysator with A then the signal intensity of the photo detector for the ellipsometric values Ψ and Δ is given by

$$I(P,A) = C\left[1 - \cos 2\Psi(\cos 2A + \cos 2P) + \cos 2A \cos 2P + \sin 2\Psi \cos\Delta \sin 2A \sin 2P\right] \quad (4.88)$$

where C is an apparative constant, Ψ is the amplitude ratio and Δ is the phase difference.

Rhe dependence of the ellipsometrisc quantity Ψ in dependence on Δ for different coating materials is given in Fig. 4.59.

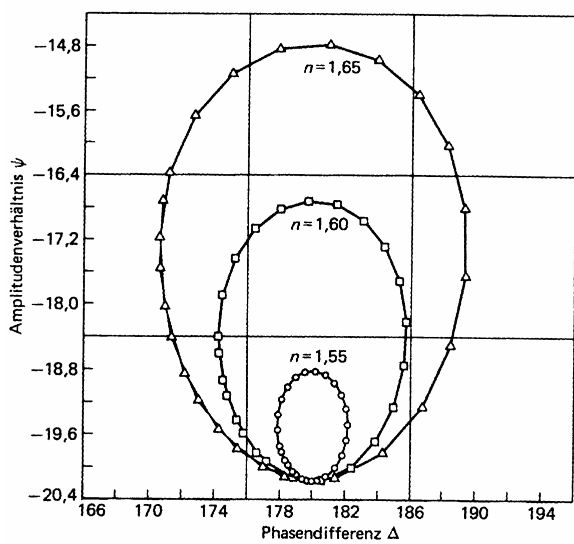


Fig. 4.59.: *Ellipsometric quantities Ψ and Δ [3, p.397]*

The big advantages of optical methods can fully been exploited if one chooses a direct measurement on the substrate with the real film system rather than an indirect measurement on a test sample.

Other Test Methods

Optical coatings are subjected to a wide variety of mechanical and chemical tests to check their performance in the field. Some selected examples shall be mentioned here:

Humidity change climate: 12 hours at 40°C and 100% humidity, and then 12 hours cooling corresponds to one cycle. The number of cycles is varied.

Spray fog test: At 50°C a aqueous solution with 5% NaCl is sprayed onto the coating via a nozzle.

Sand trickle test: 3 kg sand of a grain class of 0,4 to 0,8mm trickle through a pipe from 1650 mm heigth onto the sample which is positioned on a turntable tilted by 45°.

Boiling in salt water and dionized water: the sample remains for 5 to 60 min in a saline solution.

4.6. Chemical Composition

4.6.1. Introduction

Conventional methods of chemical analysis as atomic emission, atomic absorption spectral analysis, X-ray fluorescence and mass spectrometry play an important role for the production of the coating materials used in several processes. In most cases samples of some milligram are necessary. Given this amount the detection limit may well be in the ppm-region.

On the other hand the increased demands to micro analysis of thin films in respect to spatial and areal resolution as well as information depth have led to the development of many new methods. They are mostly based on the interaction of photons, electrons, ions or other particles with the coating or surface to be analyzed. Depending on the exciatation and on the type of analyzed material particles are either scattered elastically or inelastically or secondary particles are emitted. These particles are analyzed by a suitable detector system. A summary of the methods, ordered in respect to exciting and detected particles is given in Fig. 4.60.

Anregung durch		Nachweis durch				
		Photonen		Elektronen		Ionen
		optisch	Röntgen			
Photonen	optisch	AA UV IR		ESCA	UPS	LIMA
	Röntgen		XRF XRD		XPS	
Elektronen			EPM	SEM TEM STM	AES SAM LEED RHEED	
Ionen		SCANIIR	IIX			SIMS IPM SNMS RBS ISS

Erklärung der Abkürzungen:

AA	Atomic Absorption
AES	Auger Electron Spectroscopy
EPM	Electron Probe Microanalysis
ESCA	Electron Spectroscopy for Chemical Analysis
IIX	Ion Induced X-Rays
IPM	Ion Probe Microanalysis
IR	Infrared Spectroscopy
ISS	Ion Scattering Spectroscopy
LEED	Low Energy Electron Diffraction
LIMA	Laser induced Ion Mass Analyzer
RBS	Rutherford Backscattering Spectroscopy
RHEED	Reflexion High Energy Electron Diffraction
SAM	Scanning Auger Microanalysis
SCANIIR	Surface Composition Analysis by Neutral and Ion Impact Radiation
SEM	Scanning Electron Microscopy
SIMS	Secondary Ion Mass Spectrometry
SNMS	Secondary Neutrals Mass Spectrometry
STM	Scanning Tunnel Microscopy
TEM	Transmission Electron Microscopy
UPS	UV-Photoelectron Spectroscopy
UV	UV-Spectroscopy
XPS	X-Ray Photoelectron Spectroscopy
XRD	X-Ray Diffraction
XRF	X-Ray Fluorescence Spectroscopy

Fig. 4.60.: *Methods for the determination of the chemical composition and the structure of surfaces and thin films [1, p. 46]*

4.6.2. Electron Probe Microanalysis (EPM)

Electron Probe Microanalysis (EPM), also called X-ray micro analysis, is the eldest among these methods. The excitation is performed by electrons and X-ray quanta are detected. The method is suitable for analyzing films of at least 1 µm thickness. Two basic set-ups are existing: the "micro probe" with wavelength dispersive crystal spectrometer (WDX) which often is combined with a TEM. In the SEM mostly energy dispersiveSi(Li) semiconductor spectrometers are used. The method os then called Energy Dispersive X-ray analysis, EDX.

These methods fail for light elements (z < 8) and for films which are thinner than the information depth given by the penetration depth of the electron beam (approx. 1 µm).

4.6.3. Auger Electron Spectroscopy (AES)

The sample is probed by electrons with energies between 1 and 10 kV. By this bombardement electrons are removed from inner shells which are refilled by electrons transiting from higher energy levels (see Fig. 4.61.). The energy released during this process can either be transferred to a X-ray quantum which is emitted with the energy $h\nu_F$ *(X-ray fluorescence) or can be transferred to a third electron within the atom (Auger effect) which also is released with a characteristic energy.*

The information depth is given by the escape depth of the Auger electrons and amounts to approx. 1 - 10 nm. For analysis the number of Auger electrons emitted per unit time is measured as a function of their kinetic energy (determined e. g. by a cylinder mirror energy analyzer) by a secondary electron multiplier. The maxima of the so gained AES spectrum are tabulated and are characteristic for the type and concentration of the atoms within the sample. To suppress the background resulting from scattered electrons the spectrum can be differentiaied by electronic means.

The atomic number of detectable atoms is Z > 3. The detection limit is about approx. 0,1%. Depth profiles can be measured by removing material by sputtering, but this method is obviously destructive.

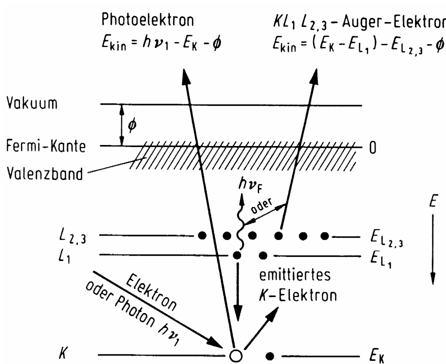


Fig. 4.61.: *Energy level diagram for Auger-electron (AES)- and photoelectron (ESCA)-spectroscopy. Q: work function of electrons [1, p. 48]*

4.6.4. Photo Electron Spectroscopy (UPS, XPS)

For excitation either X-rays (then the method is calles XPS) or UV light (then the method is called UPS) can be used. For XPS e. g. Aluminium-K-radiation with an energy of $h\nu_1 = 1486$ *eV (see Fig. 4.61.) is used. The emitted photoelectron therefore carries a kinetic energy of* $E_{kin} = h\nu_1 - E_k - F$ *which is* $h\nu_1$ *reduced by the binding energy characteristic for the*

element, E_k, and the work finction F. Similar to AES analysis the number of electrons emitted per unit time is measured in dependence on E_{kin}.

The atomic number of detectable elements is also Z > 3, the information depth is 0,5-10 nm and the detection limit is 0,1%. An advantage of ESCA is that the analytical information contains evidence about the binding state of the atoms which can be obtained from the line shift (chemical shift) which is dependent on the type of bond (e. g. metallic or oxidic).

4.6.5. Secondary Ion Mass Spectroscopy (SIMS)

The sample is bombarded with positively charged ions, mostly Cs (for the detection of electronegative elements) or O_2 (for the detection of electropositive elements) of an energy of approx. 10 keV (see Fig. 4.62.). Also Ga, Au or Bi ions may be used. They are emitted from low intensity sources and therefore allow for vertically and laterally highly resolved SIMS. Between the impinging ions and the target atoms collision processes occur which lead to the formation of collision cascades within the target. Some of them reach the surface and lead to the emission of single atoms or of molecular fragments (sputtering effect).

From the emitted particles some are positively or negatively charged and can be detected e. g. within a quadrupole mass spectrometer. Newer devices are also using Time of Flight mass spectrometry (TOF-SIMS) which can detect even very low ion currents. The secondary ion mass spectrum is characteristic for the distribution of elements on the sample surface and also (due to the molecular fragments) for their binding state. The mean escape depth is some nm. The SIMS spectrum has no background. Therefore the detection limit is very low (fractions of particle number down to 10^{-4}%). All elements (even H) and their isotopes can be detected.

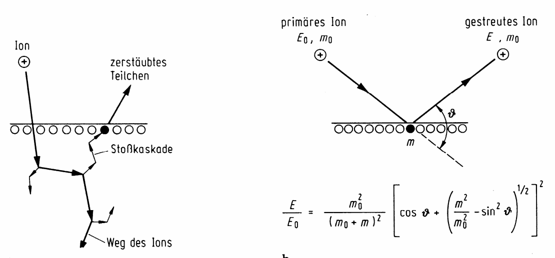


Fig. 4.62.: *Principle of **(a)** Secondary Ion Mass Spectroscopy (SIMS), and **(b)** Ion Scattering Spectroscopy (ISS) [1, p. 49]*

While AES and ESCA are non destructive (if the sample is not subjected to additional sputtering) the SIMS method is destructive, but this has the intrinsic advantage that depth profiles of the atomic concentration can be measured.

4.6.6. Secondary Neutral Mass Spectroscopy (SNMS)

In the case of SIMS the processes of ionization and emission of the particles to be analyzed are coupled ("matrix effect"). Therefore the intensities of the single lines of the mass spectrum are not representative for the coposition of the investigated surface. To obtain a quantitative analysis a calibration using samples of known composition is necessary. To increase the accuracy of measurement, especially in the case of sophisticated multilayer structures and diffusion or implantation profiles, it is preferrable to separate ionization and emission events (see Fig. 4.62.).

This is the case for the SNMS method where the neutral particles generated due to ion bombardement, which are the main species emitted for sputtering, are post ionized by suitable means and analyzed in a quadrupole mass spectrometer. The atomic concentration of a given mass is then a function of the sputter yield and the ionization probability for a

given material in a given set-up. The sputtered particles can be ionized by a RF low pressure plasma (Ar) or by laser radiation (Excimer laser, e. g. group Husinsky at TU Wien).

4.6.7 Ion Scattering Spectroscopy (ISS)

This method allows the analysis of the uppermost atomic layer of a sample whose surface is probed by ions with known energy E_0 (some 100 to 10^3 eV) and mass m_0 (see Fig. 4.62.). The primary ions elastically scattered under a given angle (e. g. $\Theta = 90°$) are analyzed in respect to their energy E in a detector. The obtained ISS spectrum exhibits maxima at values at E/E_0 which correspond to mass ratios m/m_0 according to the theory of elastic scattering.

4.6.8. Rutherford Backscatter Spectroscopy (RBS)

RBS analysis is a high energy version of ISS (see Fig. 4.63). It is executed by Hydrogen, Helium or ions of other light elements within the energy region between 0,1 to 5 MeV. Energy spectra of the backscattered ions are obtained by variation of the energy E_0 of the primary ions. From these spectra the depth profiles of the atomic concentrations of all light elements within depths of 0,1 - 10 µm can be determined non-destructively. A big advantage of the RBS method which exhibits a detection limit of approx. 10^{-3}% is that the method is absolute while the other methods mentioned above require calibration by standard samples.

As displayed in Fig. 4.63., high energy ions may be used to realize two other important analysis methods of solid state physics: the method of Ion Induced X-ray radiation (IIX) and the method of nuclear reactions triggerd by ions (Nuclear Reaction Analysis, NRA)

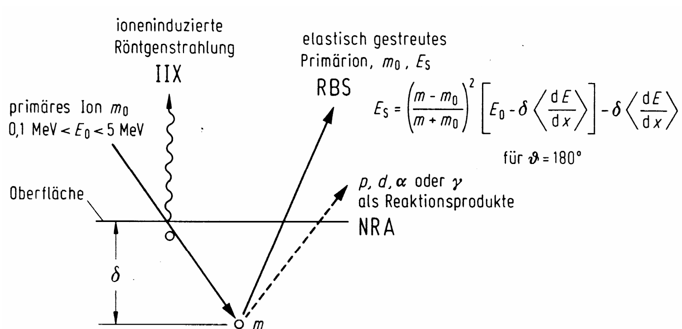


Abb. 4.63.: *Schematic of the interaction of high energy light ions with solids and of their application in analysis [1, p. 51]:*

a: *Ion induced X-rays: IIX*
b: *Elastic backscattering of primary ions (Rutherford Backscattering: RBS)*
c: *Νυχλεαρ ρεαχτιονσ υνδερ εμισσιον οφ χηαραχτεριστιχ προδυχτσ λικε ε. γ. προτονσ, δευτερονσ, γ-quanta or α particles (nuclear reaction analysis: NRA)*