Microstructural design of hard coatings

Paul H. Mayrhofer a,*, Christian Mitterer a, Lars Hultman b, Helmut Clemens a

a Department of Physical Metallurgy and Materials Testing, University of Leoben, A-8700 Leoben, Austria
b Thin Film Division, Department of Physics (IFM), Linköping University, S-58183 Linköping, Sweden

Received 2 August 2005; received in revised form 14 November 2005; accepted 8 December 2005

Abstract

Microstructural design has attracted increasing interest in modern development of hard coatings for wear-resistant applications. In plasma-assisted vapor deposited thin films, the material’s microstructure can be designed during growth or post-deposition annealing treatments. In this review, we demonstrate the correlation between microstructure and mechanical as well as tribological properties of hard ceramic coatings. This is done for single-phase coatings and composition or phase modulated layers. In the latter case, the microstructure can be designed by choice of the deposition technique chosen, understanding the growth processes taking place on a film surface, either by sequential deposition of layers or by taking advantage of newly discovered self-organization processes including segregation effects of the elements. Consequently, the effects of individual microstructural features like grain size, defect density (and hence residual stress), phase arrangements in a one-, two- or three-dimensional manner on the mechanical properties are treated. Here, especially TiN–TiB 2 is used as a model system to describe the development of two- and three-dimensional coating nanostructures. Due to their particular structures, such coatings can exhibit superhardness (H \geq 40 \text{ GPa}). The microstructural changes of hard ceramic coatings during a post-deposition annealing treatment are discussed in detail. Although the significance of heat treatments to optimize properties (by a well-designed microstructure) for specific applications is recognized in bulk material science, only a few elements have been applied for hard ceramic coatings so far. Due to limited atomic assembly kinetics during the deposition process (e.g., by using a low substrate temperature), defects (point-, line-, and area-defects), supersaturated, and metastable phases can easily be obtained. For example, growth of (Ti,Al)N and Ti(B,N) films can result in the formation of a supersaturated TiN based phase. Such films undergo age hardening processes during post-deposition annealing due to the decomposition of the supersaturated phases into their stable
The review clearly shows that nanostructure dependent hardness increase (compared to hardness of the bulk counterparts) sustains higher annealing temperatures than hardness increase due to an increased density of point- and/or line-defects. Tribological properties of hard thin films can be engineered by adding phases with lubricious properties at operation temperature (either room or elevated temperatures) and prevailing environment. Especially in high speed and dry cutting applications, low-friction and lubricating mechanisms of the thin film itself are required in addition to excellent mechanical properties.

© 2006 Elsevier Ltd. All rights reserved.

Contents

1. Introduction ............................................................. 1033
2. Microstructure-mechanical property relations .................................. 1037
   2.1. Single-phase hard coatings ........................................... 1039
   2.2. Composition or phase modulated hard coatings ....................... 1046
      2.2.1. Nanocolumnar coatings ........................................... 1047
      2.2.2. Nanocomposite coatings ........................................... 1051
3. Microstructural evolution during post-deposition annealing ................ 1053
   3.1. Recovery ........................................................... 1053
      3.1.1. Single-phase hard coatings ....................................... 1058
      3.1.2. Composition or phase modulated hard coatings ................. 1062
   3.2. Recrystallization and grain growth ................................ 1066
      3.2.1. Single-phase hard coatings ....................................... 1066
      3.2.2. Composition or phase modulated hard coatings ................. 1068
   3.3. Age hardening ..................................................... 1076
      3.3.1. Formation of supersaturated phases ............................... 1077
      3.3.2. Decomposition of supersaturated phases ....................... 1078
4. Self-lubrication of hard coatings ........................................ 1090
   4.1. Self-lubrication by solid lubricants ................................ 1090
      4.1.1. Diamond-like carbon based coatings .............................. 1090
      4.1.2. Transition metal dichalcogenide coatings ...................... 1092
      4.1.3. Future approaches ............................................... 1093
   4.2. Frictional self-adaptation by oxide-formation ........................ 1094
5. Summary and outlook .................................................. 1100
Acknowledgements .................................................................. 1103
References ............................................................................. 1103

1. Introduction

Deposition of hard thin films with a few microns in thickness is a common technology to improve the performance of tools, dies, and molds for many different applications [1]. Starting with thermally activated chemical vapor deposition (CVD), different methods have been developed including plasma-assisted physical vapor deposition (PVD) [2], plasma-assisted chemical vapor deposition (PACVD) and laser-assisted methods like pulsed laser deposition (PLD). The importance of wear-resistant hard coatings is given by the following facts. About 90% of all indexable inserts for metal cutting, based on
cemented carbide substrates, are coated for wear protection by CVD or PVD techniques, with PVD being used in about 25% of the cases. The important issue for the future is not which deposition technology is used, but instead which properties can be achieved at which production cost. This is where the need for microstructural design enters the field.

Hard coatings prepared by various deposition techniques and conditions exhibit the widest variety of microstructures among materials in terms of grain size, crystallographic orientation, lattice defects, texture, and surface morphology as well as phase composition. The recent review by Petrov et al. [3] clearly shows the microstructural evolution during film growth. Examples of the wealth of microstructures encountered in hard coatings are given in Figs. 1 and 2. Fig. 1 is a bright-field cross-sectional transmission electron microscopy (TEM) image of the center portion of a TiN film [3,4] prepared by magnetron sputtering. The film exhibits a columnar microstructure with elongated grains, which is the typical product of today’s coating processes of ceramic materials. For characterization of the ion bombardment conditions, which are commonly used to adjust the film microstructure during plasma-assisted growth, the incident ion/metal flux ratio \( J_{\text{ion}}/J_{\text{me}} \) and ion energy \( E_{\text{ion}} \) are used as main parameters. \( J_{\text{ion}} \) and \( E_{\text{ion}} \) can be determined using Langmuir-probe measurements following the procedures described in Ref. [5]. The incident metal flux \( J_{\text{me}} \) is estimated based upon measured deposition rate values \( R \), film composition, film thickness, and film density. Details about the evaluation and calculation of \( J_{\text{ion}}/J_{\text{me}} \) and \( E_{\text{ion}} \) are described in Refs. [5–7].

Fig. 1. Cross-sectional TEM micrograph obtained from the middle portion of a 3.5-μm-thick TiN layer grown by reactive magnetron sputter deposition on a steel substrate at 300 °C. \( J_{\text{ion}}/J_{\text{me}} = 1 \) while the ion energy \( E_{\text{ion}} \) was varied in steps of 40 eV [3,4].
During the growth of the film (shown in Fig. 1) the ion energy was varied in steps of 40 eV. The microstructure of layers produced at $E_{\text{ion}} \leq 80$ eV consists of dense columns with open column boundaries. As the ion energy is increased to 120 eV, the voids along column boundaries disappear and the film becomes fully dense. This is accompanied, however, by incorporation of intragranular residual damage from lattice point defects and dislocation aggregates (manifested in Fig. 1 as a darker average contrast in the $E_{\text{ion}} = 120$ eV sublayer) whose concentration increases at higher voltages (see the even darker contrast in the $E_{\text{ion}} = 160$ eV sublayer). When the energy is increased above 160–200 eV, the defect density becomes large enough to disrupt the local epitaxial growth on individual columns and hence renucleation occurs [3].

Fig. 2a and b show examples of nanoscale coatings. Such materials are generally single- or multi-phase polycrystals with a typical crystal size of a few nanometers (1–50 nm) in at least one direction. Thus, they can be layered (nanostructured in one dimension, see Fig. 2a [8]), or they can consist of a columnar (nanostructured in two dimensions) or equiaxed nanostructure (nanostructured in three dimensions, see Fig. 2b [9]). The film presented in Fig. 2b contains closely packed small (<10 nm) TiN crystallites (A–D) separated by an alleged Si$_3$N$_4$ amorphous phase. Similar hard coatings have been reported by Veprek et al. for PACVD processing [10–13].

Gradients or inhomogeneities over the film thickness are typically present by design or through process determination. Such coating synthesis is also driven by the industrial demand for low-temperature deposition, especially plasma-assisted methods. Obviously, thermodynamic equilibrium is not obtained during this kind of deposition. In fact, it is the kinetic limitation induced by low-temperature deposition that allows for controlled synthesis of metastable phases and artificial structures such as nanolaminate, nanocolumnar, and nanocomposite materials. The as-deposited coatings can in turn be subject to annealing and consequential recovery (stress relaxation), interdiffusion, recrystallization,
or phase transformation. These phenomena are technologically relevant since the resulting structure (on both nano- and microscale) has a large impact on the film properties. Therefore, the development of advanced hard thin films is preferably made using phase transformation concepts from physical metallurgy, such as primary (vapor-to-solid) and secondary (precipitation) phase transformation. It is noteworthy that many of the nitride, oxide, and carbide coatings produced for the engineering and manufacturing industry are also highly relevant for the microelectronics field. This applies for example to diffusion barriers and contact layers (e.g., TiN, TaN) and gate oxides from transition metals. Developments in all applications will benefit from a closer interaction of the different fields as many of the materials quality and reliability issues are similar, e.g., for controlling film texture, defect density, and purity.

In the deposition of compounds, a reaction, mainly at the substrate surface, occurs between their components, which may be introduced into the growth chamber as a gas or vapor. If all components of the films are generated by a vaporization of the target material in an inert atmosphere, these processes are generally termed non-reactive. Reactive processes involve the introduction of additional ‘reactive’ gases (such as N₂, O₂, CH₄, etc.) into the growth chamber (here dissociation of these gases into atomic or ionic state is necessary). This differentiation is useful, but there are of course also reactions (mainly at the substrate surface), which form compound layers by ‘non-reactive’ processes. Plasmas can be used to provide heat to the film surface to improve adhesion as well as structure and to increase reactivity [14]. A comparative study on reactively and non-reactively grown TiN films is given in [15].

Very often, in hard thin films deposited by plasma-assisted vapor deposition techniques, several of the classical strengthening mechanisms are active. Such strengthening mechanisms are caused by providing obstacles for dislocation movement, which can also be applied to some extent to hard films. Hindering of dislocation movement can be achieved by: (i) high density of point and line defects generated during growth, (ii) internal boundaries like grain, phase, and column boundaries, (iii) second phase particles, and (iv) solutes [16–18]. Most noteworthy, age hardening has recently been demonstrated in coating materials [19].

Due to limited atomic assembly kinetics during low-temperature (substrate temperature \( T_s \) below 0.2–0.3 of the melting point \( T_m \) in K [3]) deposition techniques and the extremely high cooling rates (~\( 10^{13} \) K s⁻¹ [20]), ceramic thin films prepared by plasma-assisted vapor deposition are known to have a non-equilibrium density of structural built-in defects (see Section 2). Therefore, the film hardness is increased compared to their corresponding bulk counterparts [1,15,21–58]. Here, this hardening effect will simply be called growth-defect hardening.

The development of advanced coatings follows the major trend of synthesizing multi-component and multi-phase structures that are individually developed for special applications [25–34]. As an example, in order to enhance the abrasion and erosion resistance of TiN [59], carbon is added at the expense of nitrogen which results in forming a stable solid solution Ti(C,N). Further improvement was obtained by (Ti,Al)N, which is a hard coating where a supersaturated TiN-based solid solution with additional Al is formed [25]. Such nano- and microstructurally designed hard coatings enable new machining applications, e.g., high-speed cutting or even dry cutting [60].

The structurally and/or compositionally designed hard thin films grown by plasma-assisted vapor deposition are often in a non-equilibrium state [26,27,35], e.g., characterized
by compressive stresses, small grain sizes in the nanometer range, or metastable phases.
Advanced surface engineering design approaches have led to the development of coating
materials with unique property combinations, e.g., superhardness ($H \geq 40$ GPa [36,37])
combined with high toughness [36], or chameleon-like frictional self-adaptation [61]; both
property combinations are realized by a functional nanostructure. It should be appreciated
that while the number of coating material systems explored is growing rapidly, they share
a few basic structural types for conceptual design. Their response in tribological contacts is
the least understood so far.

Most often, e.g., in machining, casting or hot-forming applications, the protective lay-
ers are subjected to severe conditions like high temperature, plastic deformation, severe
friction, and high local stresses causing mechanical and thermal fatigue. Therefore, the
required thin-film characteristics e.g., in cutting, are: high hardness, low thermal conduc-
tivity, reasonable toughness, controlled adhesion, minimum diffusion, and low friction at
operating temperature. Typically, when applied onto cemented carbide cutting tools, the
temperature at the cutting edge may exceed 1000 °C [62–64], giving rise to microstructural
changes affecting application-oriented properties. Thus, the thermal stability, in particular
of advanced microstructurally engineered coatings, is of vital importance and has a strong
impact for the choice of constituent elements to be selected for the coating. As the micro-
structural complexity grows with the power of the number of components (atomic ele-
ments), future progress will require knowledge based design concepts of the coatings
used. Present research and development are in a stage of rapid expansion for testing
and investigating a growing number of new combinations of coating systems.

This review is focused on structure–property relations of advanced wear-resistant coat-
ing. The term microstructure is used as a commonly accepted concept in materials engi-
eering, although many systems are more or less nanoscale in nature. The first section of
the review describes the most important growth defects of hard coatings generated during
plasma-assisted deposition. In the main part of the review we deal with mechanical prop-
erties, recovery effects, recrystallization, grain growth, phase separation, and self-lubrica-
tion phenomena. Finally, we present an outlook for research on hard wear-resistant
coatings including a discussion of up to now unsolved problems.

2. Microstructure–mechanical property relations

The mechanical properties of materials depend fundamentally on the nature of
bonding among their constituent atoms and upon their microstructures on a variety of
length scales. During plastic deformation of materials, atoms have to be displaced with
respect to one another. The easier this process is established, the more ductile is the
material. At ambient temperature dislocations move by a glide process. Plastic deforma-
tion can also be obtained by creep processes, which include dislocation climb, atomic
diffusion, and mechanisms like grain boundary sliding. In contrast to the glide process,
climbing of dislocations requires lattice defects such as vacancies and their diffusion, thus
this process depends on substantial thermal activation. The diffusion coefficient $D$ is
described as

$$D = D_0 \cdot \exp \left( -\frac{E_a}{k_B \cdot T} \right), \quad (1)$$
where $D_0$ is a material constant, $E_a$ is the activation energy for diffusion, $k_B$ is Boltzmann’s constant, and $T$ is the absolute temperature [65].

Creep rates in polycrystals strongly depend upon the grain size, $d$, and vary from $d^{-2}$ behavior, in cases where the mechanism is volume diffusion controlled (Nabarro–Herring creep), to $d^{-3}$ behavior, in cases where the mechanism is controlled by interface or grain boundary diffusion (Coble creep) [16,65]. Thus, the mechanical properties can change dramatically as the grain size is reduced [66,67].

In conventional metallic materials, strengthening during plastic deformation (work hardening) occurs by an increase of the dislocation density. Hence the hardness increases as the interaction between dislocations increases. This can also be obtained by increasing the grain boundary fraction (grain size reduction), which increases the barrier to dislocation motion. Since the minimum stress required to activate common dislocation sources (such as a Frank–Read source) is inversely proportional to the distance between dislocation pinning points, these stresses will increase with decreasing grain sizes into the nanophase regime owing to the limitation of the maximum distance between such pinning points [39]. Therefore, grain refinement in the nanoscale regime reduces the density of dislocations, which are necessary for plastic deformation. Hence, small-grained crystalline materials (nanocrystalline materials) show improved strengthening and the theoretical yield stress of a dislocation-free metal may be reached [16].

For ductile materials, the yield strength can be estimated using the Tabor-relation [68] which states that hardness $H$ (obtained from a Vickers indentation) is about three times the yield strength [16]. For hard materials, the hardness is better described by the Marsh-relation [69], which takes into account the elastic modulus $E$, the Poisson ratio $\nu$, and the yield strength. Considering that the theoretical strength of a solid is solely a function of the strength of the interatomic bonds between atoms, the theoretical maximum yield strength would be $E/10$ [70]. Thus, the theoretical hardness (obtained via a Vickers indentation and using the Marsh-relation) e.g., for TiN ($E = 300–400$ GPa, $\nu = 0.25$) would be 40–53 GPa and for TiB$_2$ ($E = 400–500$ GPa, $\nu = 0.25$) the theoretical hardness would be 53–66 GPa, respectively. In practice, however, the strength of inherently strong solids never meets these theoretical expectations because of the presence of crystal imperfections, microscopic defects and cracks. This is also valid for materials composed of 1–2 nm sized crystals, which are considered to be dislocation-free [71].

Two main types of atoms can be distinguished in nanocrystalline materials; crystal atoms with nearest neighbor configuration corresponding to the lattice and boundary atoms with a variety of interatomic spacings, different from boundary to boundary. A nanocrystalline material typically contains a high number of interfaces with random orientation relationships, and consequently, a substantial fraction of atoms belongs to the interfaces. The interfacial component has a lower atomic density than the crystal [72]. PVD single-phase layers usually show a nanostructure in one or two dimensions. Depending on the ion bombardment used during deposition and the thermodynamic driving force for segregation also a nanostructure in three dimensions can be formed. This can also be obtained by the simultaneous growth of a second phase or by doping with additional atoms during deposition. The amount and type of additional phases and elements can be used to control the orientation and the size of grains in the films [73]. The nanocrystalline materials may contain crystalline, quasi-crystalline, or amorphous phases of metals, ceramics or their composites. Nanocrystalline materials are of increasing interest because they exhibit mechanical, physical, and chemical properties different
from their coarse-grained counterparts. For example, the hardness and yield stress may increase 5–10 times when the grain size is reduced from the macroscopic to the nanometer range [66].

The effect of an increased strengthening with decreasing grain size $d$ is described by the well known Hall–Petch relation (HPR) [74,75]:

$$\sigma_y = \sigma_0 + k \frac{1}{\sqrt{d}},$$

where $\sigma_y$ is the yield stress, $\sigma_0$ the ‘friction stress’ and $k$ is the Hall–Petch factor which depends on the material and measures the relative hardening contribution of grain boundaries/interfaces. However, grain size reduction has a negative effect on creep strength, especially for metallic materials. For brittle materials, such as intermetallic compounds and ceramics, the ductility increases by grain refinement as a result of increased grain boundary volume fraction and, therefore, grain boundary sliding [66]. Nanocrystalline pure metals are much stronger and apparently less ductile than conventional ones. Intermetallics are also strengthened by a grain refinement, but they are also affected by the large percentage of atoms in grain boundaries and other microstructural features.

The strengthening effect by grain refinement is limited by the critical grain size, $d_c$, where the Hall–Petch effect ceases or even reverses. Reasons for this behavior can be an increased porosity at small grain sizes, suppression of dislocation pile-ups, dislocation motion through multiple grains, grain boundary sliding, enhanced grain boundary diffusion or the fact that no dislocations can form in such small grains, thus the concept of mirror forces into the next grain becomes irrelevant [39,76,77]. The reverse Hall–Petch effect is classified by negative $k$ values in the HPR. The critical grain size $d_c$ depends strongly on the nature of a material. While the premises for the Hall–Petch effect are no longer fulfilled below $d_c$, the nanoscale phenomena are not yet fully understood.

### 2.1. Single-phase hard coatings

This section addresses the influence of microstructural features, which depend on the growth conditions, on the mechanical properties of single-phase model coatings such as TiN and CrN.

![Fig. 3. (a) Effects of an ion bombardment on a growing film [78,80] in the 10–1000 eV range commonly employed in thin film processing. (b) Schematic view of possible lattice defects created by an impinging energetic atom, primary knock-on atom (P) [16].](image-url)
Ion bombardment during film growth at low temperatures is a useful tool to increase the density and to modify the morphology of the films [78]. During growth, stresses are generally induced via energetic particle bombardment. In magnetron sputtering, these particles are generated from the sputtering gas and consist of back-scattered inert gas neutrals or ions of the inert or the reactive gas (e.g., Ar$^+$ and N$_2^+$) accelerated towards the growing film by a negative substrate bias potential [79]. For the case of arc-evaporation or high-power pulsed sputtering, there are metal ions of multiple ionization states. When making predictions on the microstructure–mechanical property relations in hard coatings, it is important to know the lattice defect arrangements responsible for the stress and how they depend on the energetic species.

The effects of impinging energetic ions with a solid are illustrated schematically in Fig. 3a and possible generated lattice defects are shown in Fig. 3b. When the incoming ions or knock-on atoms possess enough kinetic energy, they will knock atoms out of their lattice positions and create secondary collisions (see Fig. 3a). In this way, collision cascades develop, which are – in contrast to high energetic bombardment – not isotropic for low energetic bombardment. The primary knocked atom absorbs most of the energy. The resulting strong atomic motion along the trajectory of the ions leads to a rearrangement of the lattice atoms [80]. The impinging particles may also be physically implanted and surface species may be recoil-implanted into the subsurface lattice. If the colliding particles have high energies, the effective diameter of the lattice atoms appears to be reduced and the lattice seems to be empty in certain directions (‘channels’). Lattice channeling processes can carry the mentioned effects several additional atomic layers below the surface.

As energy and flux of the impinging ions are increased, atomic displacements produced in the collision cascades result in an increasing number of residual interstitials and vacancies (Fig. 3b). An additional atom in a close-packed lattice direction due to focused

Fig. 4. Cross-sectional TEM overview image of a TiN film grown at 300 °C, $E_{\text{ion}} = 120$ eV, and $J_{\text{ion}}/J_{\text{me}} = 0.2$ [83].
collision cascades (focussons) is known as crowdion, which is a one-dimensional defect of split-interstitial nature. It moves by exchange collisions, which can destroy the order in an alloy along its path. The extensive focused transport of matter occurs for atoms only and not for vacancies. An example of crowdion formation in bcc metals under low-energy ion bombardment was given by Carlberg et al. [81] using molecular-dynamics studies. At the end of a displacement cascade a zone of reduced density (diluted zone) can be formed which is surrounded by an increased density of interstitial atoms [16]. The point defects

Fig. 5. Cross-sectional HRTEM lattice image of the TiN film presented in Fig. 4. (a) Dislocations and the generated strain fields, indicated by dashed lines. (b) Detail of (a) showing a Burgers circuit and the bended lattice planes around a dislocation [83].

Fig. 6. Cross-sectional HRTEM lattice image of the TiN film presented in Fig. 4. (a) Dislocations and strained areas and a low-angle grain boundary. (b) Detail of the low-angle grain boundary [83].
can, in turn, lead to an increased density of extended defects and form one-, two- or three-dimensional defects. Two-dimensional defects are, for example, grain, subgrain, and twin boundaries. Clusters of point defects such as voids, cracks, bubbles, and particles of a different orientation or structure than the surrounding matrix, count to three-dimensional defects [82]. The recombination of point defects, however, reduces the total strain energy of the lattice [47, 65, 80].

Fig. 7. Cross-sectional HRTEM lattice image of the TiN film presented in Fig. 4, showing a high-angle grain boundary [83].

Fig. 8. XRD of as-deposited (a) TiN reactively grown at 300 °C, $E_{\text{ion}} = 30$ eV, and $J_{\text{ion}}/J_{\text{me}} = 6.6$ and [15]. (b) CrN reactively grown at 300 °C, $E_{\text{ion}} = 15$ eV, and $J_{\text{ion}}/J_{\text{me}} = 2.5$ [21].
Some examples of microstructural defects in TiN films [15] non-reactively grown at 300 °C, \( E_{\text{ion}} = 120 \text{ eV} \), and \( J_{\text{ion}}/J_{\text{me}} = 0.2 \), investigated by transmission electron microscopy (TEM), are given in Figs. 4–7. Fig. 4 is a bright-field cross-sectional TEM image of a TiN film, where defects and hence the strained areas are the origin of different image contrasts. Along the column boundaries an accumulation of voids due to kinetic limitations of the deposition process can be seen [3,83].

The structural defects in the films can be investigated using high-resolution TEM (HRTEM). Examples are given for dislocations (Fig. 5) and low- and high-angle grain boundaries (Figs. 6 and 7, respectively). Fig. 5a and b are bright-field HRTEM lattice images of the TiN film presented in Fig. 4. The dislocations generate long-range stress fields causing diffuse regions within the lattice micrographs. A detailed view at the dislocation indicated at the bottom of Fig. 5a is given in Fig. 5b. Here, bent lattice planes around the dislocation due to the additionally introduced lattice plane can clearly be seen, giving rise to strained areas.

Low-angle grain boundaries (subgrain boundaries) are built of systematically arranged dislocations. Fig. 6 is a bright-field HRTEM image of a low-angle grain boundary in a TiN film, where the two adjacent crystals are tilted by periodically inserted dislocations (similar observations are found for ZrN thin films [84]). On the left side of Fig. 6a, in a distance of some lattice planes from the low-angle grain boundary, bent lattice planes can clearly be seen. A detail of the low-angle grain boundary is given in Fig. 6b, here the inserted dislocations are indicated. Increasing orientation-difference of the adjacent crystals requires an increasing density of dislocation-structure. If this is not possible, high-angle grain boundaries are formed (see Fig. 7). In distinct cases the neighboring crystals coincide and hence a coincidence site lattice is formed [16,65].

Experimental X-ray diffraction (XRD) measurements of a material containing structural defects (point-, line-, and area-defects) result in broadened line profiles, which are often characterized by the full width at half maximum intensity \( \Gamma \) as indicated for the (2 2 0) reflections of as-deposited TiN [15] and CrN [21] in Fig. 8a and b, respectively. The asymmetrical line profiles of single-phase binary layers are a strong indication for lattice faulting/twinning and a high dislocation density. Furthermore, an orientation dependence of the diffraction line shape indicates that the lattice faulting/twinning and dislocation distributions in the crystallites are non-uniform. For asymmetric line profiles, the integral breadth \( \beta \), which is the ratio of the peak area and maximum intensity, should be preferred instead of \( \Gamma \) to describe their broadening. The strongest line broadening is caused by finite grain size and three-dimensional defects (like subgrains) [85]. In addition to the line broadening the displacement of XRD peak positions to their standard values gives information about the macrostrain and macrostress, where the unit cell distances (e.g., lattice parameter \( a \)) will become either larger or smaller resulting in a shift to smaller or larger diffraction angles, respectively. Consequently, the films exhibit compressive or tensile intrinsic macrostresses.

The microstructural information deduced from the shape and broadening of X-ray line profiles can be grouped into size- and distortion-broadening contributions. If the line profile can be described by a Cauchy function, then the crystallite size predominates, whereas a Gaussian distribution arises from distortion or local lattice strain [86]. The X-ray line profiles of the thin films investigated in [21–24] were characterized using a pseudo-Voigt profile function which is a linear combination of a Cauchy and Gaussian function [86,87]. Following the procedures described in Refs. [7,85,87,88], the average crystallite
size can be estimated from the Cauchy contribution of the pseudo-Voigt function. It has to be mentioned, that the correct term for crystallite size in columnar grown films would be column diameter, as the crystallite size is estimated from Bragg–Brentano XRD patterns. The estimated crystallite size or column diameter fits very well to TEM investigations. For

Fig. 9. Cross-sectional TEM of a TiN film, non-reactively grown at 300 °C, $E_{\text{ion}} = 120$ eV, and $J_{\text{ion}}/J_{\text{me}} = 0.2$ [83].

Fig. 10. Interrelationship between (a) hardness and biaxial coating stress and (b) hardness and grain size for sputtered single-phase (TiN, CrN), phase-modulated (CrN$_{0.6}$: CrN + Cr$_2$N) and composition and phase modulated (TiB$_{1.4}$N$_{0.65}$: TiN + TiB$_2$) hard films [21,22].
TiN, non-reactively grown at 300 °C and using an ion bombardment condition of $E_\text{ion} = 120$ eV and $J_{\text{ion}}/J_{\text{me}} = 0.2$, a crystallite size of $\sim 24$ nm is estimated by XRD line broadening [15]. The average column diameter estimated from bright-field cross-sectional TEM investigations on this film (Fig. 9) is between 20 and 30 nm.

The generated defects during growth of the films can also (indirectly) be investigated by stress measurements. In arc-deposited coatings of TiN and TiC$_x$N$_{1-x}$ a compressive residual stress builds up for increasing negative substrate bias up to $V_s = 50–100$ V. For the larger biases (up to 800 eV studied), however, the stress decreases and levels out at 2–3 GPa [89]. A maximum stress of $\sim 7$ GPa is obtained limited by interior cracking (cohesive materials failure) of the films. It is suggested that the apparent relaxation of the stress with increasing energy of the incident metal ions is determined by the defect annihilation processes occurring in the collision cascade in the growing film surface [90] and due to the fact that the effective stability of defect complexes increase with increasing carbon content in TiC$_x$N$_{1-x}$. In comparison, for magnetron sputtered coatings with inert gas discharge, the intrinsic stress continuously increases by an apparent square root dependency of the bombarding particle energy (momentum transfer), as described by the forward sputtering models [91].

Coatings with hardness values well above those of the respective bulk materials may be synthesized using high ion energies, making use of strain hardening by high defect densities where the defects responsible for the residual stresses also act as obstacles for dislocation movement. In fact, an apparent linear relationship between residual stress and hardness has been reported for several single-phase coatings, e.g., TiN [15], Ti(C,N) [89,92], or CrN [21], deposited by different PVD methods (see Fig. 10a for TiN [15] and CrN [21] thin films).

The strengthening effect of TiN and CrN coatings with decreasing grain size (or column diameter) follows the Hall–Petch relationship (HPR) as can be seen in Fig. 10b. For reactively and non-reactively sputtered TiN films, which follow the HPR, the highest hardness values could be obtained for those layers showing the smallest grain size. Contrary to these findings, the highest hardness for stoichiometric CrN was not related to the smallest grain size. As shown in Fig. 10b, the hardness decreases again if the grain size is below a critical one $d_c$, which is estimated to $\sim 12$ nm for sputtered CrN films with stoichiometric composition [21].

The mechanisms active below this critical grain size are not fully clear. Several authors report on a softening for values below the critical one (see Fig. 10b or [38,43,77]) which is related to the so-called inverse Hall–Petch effect. There are several explanations for this behavior: (i) Including the dislocation stress field, where the upper cut-off distance is supposed to equate the grain size in nanocrystalline materials making the radius of the dislocation loop sensitive to the grain size [39,93]. (ii) Shiøtz et al. [76] found that softening is mainly a result of grain boundary sliding where a few atoms undergo a large number of uncorrelated sliding events with respect to each other. In the grain interiors, only little dislocation activity is seen using atomic-scale simulations. The limitation of the deformation to the grain boundaries leads to a hardening effect as the grain size is increased. Summing up, a transition from dislocation-dominated plastic deformation to grain boundary sliding occurs at a critical grain size when the stress required for dislocation motion or formation becomes larger than for grain boundary sliding. As shown in Fig. 10b, this transition corresponds to a maximum in hardness of stoichiometric CrN, where thin film stresses close to zero are observed [21]. It should be noted, however, that those single-phase layers (CrN,
TiN) which do not follow the Hall–Petch relationship exhibit a correlation between their hardness and macrostress (see Fig. 10a). In such a case, the microstructural defects, which result in compressive stresses, hinder dislocation movement. Consequently, for PVD-grown hard coatings both the grain size and residual stress state needs to be determined, and their respective roles need to be analyzed with respect to actual mechanical deformation mechanisms, before conclusions can be drawn for the structure–property relationships of the material.

2.2. Composition or phase modulated hard coatings

If the neighboring grains in a material are of different lattice structures (hence different phases) they probably have different slip systems, so it becomes more difficult for dislocations to run through the respective grains. Such coatings belong to the group of phase-modulated hard layers. For example, a CrN0.6 coating, which consists of CrN phase (fcc, number of possible slip systems: 12) and Cr2N phase (hcp, number of possible slip systems: 3) shows an extremely high hardness of 39 GPa (see Fig. 10b). This substantial hardening compared to bulk (CrN ≈ 11 GPa [94]) occurs even for a tensile residual stress of ∼0.7 GPa and a grain size of 8 nm, which is smaller than dc obtained for stoichiometric CrN [7,21].

If grain boundary sliding or the propagation of dislocations can be avoided or suppressed by an appropriate design of nanocrystalline materials with strong grain boundaries, the strengthening of materials can be adjusted by extremely small grains as dc decreases [95]. Such materials can reach hardness values higher than 40 GPa, thus they count to superhard materials [36,37]. This can especially be obtained by materials, which are composed of at least two different phases in the nanoscale regime, either obtained by composition or phase modulation. An example of a compositional and phase modulated hard film is TiB0.80N0.83 (see Fig. 10b), having a grain size of ∼3 nm. Such films will be discussed in detail in Section 2.2.2. According to Holleck [96] the diffusion length of a condensed ad-atom via surface diffusion (which is assumed to dominate in low-temperature deposition, typically between 300 and 500 °C) is a few nanometers, making formation of large columns without coalescence of several growing crystals impossible [97].

An alternating layer-by-layer arrangement of two phases is obtained in a multilayer or, when the periodicity is in the nanoscale range (i.e., one-dimensional nanostructure), in the superlattice form. The superlattice approach, originally presented in the late 1980s [30], provides a challenging method for the synthesis of materials with superhardness and has been transferred to production in the last years [98]. Such lamellar materials, either designed by an alternating growth of individual layers in an artificial superlattice (e.g., TiN/VN [30]), or inherently nanolaminated phases (e.g., the so-called Mn+1AXn (n = 1–3) phases [99], where M is a transition metal, A represents a group A element such as Al or Si, and X is C and/or N) are promising materials as they combine high hardness and toughness.

The formation of a nanostructure in two dimensions, i.e., a nanocolumnar arrangement of alternating phases, is more difficult to control by deposition as the structure strongly depends on a segregation-driven growth-process. TaN [100] and TiCeN [101] are known to form such nanocolumns at specific growth conditions. B-rich TiB2, where the segregation of excess B is responsible for the formation of ∼5 nm-diameter TiB2 nanocolumns, will be discussed in Section 2.2.1 as a model-film for a two-dimensional nanostructure.
Materials which are composed of alternating different phases in the nanoscale regime, where at least one phase is nanosized in all three dimensions, are termed nanocomposites. In the idealized case, they can be composed of a nanocrystalline-(nc) and amorphous-(a) phase (i.e., nc/a like TiN–Si$_3$N$_4$ [10,36]), a combination of different nc-phases (i.e., nc/nc like TiN–TiB$_2$ [34]) or a combination of a nc-phase and a metallic phase (i.e., nc/metal like ZrN–Cu [102]) [22–24,36,37,102–110]. All these nanocomposites contain at least one hard nc-phase. In order to produce superhard films from such multi-phase systems, both phases should be immiscible (i.e., thermodynamically segregated) and the cohesive energy at the interface between the phases must be high [107].

Especially PVD techniques, where different materials can simultaneously be evaporated from targets by sputtering either in a reactive or non-reactive atmosphere, are very useful tools for the preparation of nanocomposites with different amounts of nanocrystalline, amorphous or metal phases. By the simultaneous growth of immiscible phases, nanocrystalline compound materials can be obtained by vapor deposition. For this, the films must grow in the miscibility gap of quasi-binary (or multinary) compound systems. There, continuous nucleation and growth by segregation of insoluble elements and formation of thin segregated layers of the second phase on top of the growing nuclei take place preventing coalescence and repeated nucleation, thus limiting grain size (see Section 2.2.2). The amount and type of additional phases and elements can be used to control the orientation and the size of grains in the films [73]. The nanocrystalline materials may contain crystalline or amorphous phases and can be metals, ceramics or their composites. During deposition of such composite films the atoms, condensing in an intermixed state, try to find a stable configuration characterized by a low free energy of formation [108]. Non-reactive sputtering of a segmented TiN/TiB$_2$ or TiC/TiB$_2$ target results in the formation of nc/nc nanocomposite films [23] with the nanocrystalline phases of TiN and TiB$_2$ or TiC and TiB$_2$, respectively. The individual phases are separated by boundaries having a mixed chemical composition [23,24,34,108,110]. These nanocomposites exhibit strong interfaces which hinder grain boundary sliding, thus hardness values higher than 40 GPa can be obtained by grain refinement to the nanoscale regime [36]. In more detail, an example for a three-dimensional nanostructure obtained in a nc/nc nanocomposite TiN–TiB$_2$ film is given in Section 2.2.2. Here, segregation-driven renucleation of 2–3 nm sized TiN and TiB$_2$ nanocrystals during growth results in the formation of such a three-dimensional nanostructure.

### 2.2.1. Nanocolumnar coatings

An ideal candidate to describe nanocolumnar coatings is TiB$_x$ if $x > 2$. TiB$_2$ crystallizes in the hexagonal C32 structure in which B is located in the interstices between (0 0 0 1) close-packed Ti planes [111,112]. Thus, B and Ti planes alternate along the c-axis [111,112]. Reported lattice parameters are $a = 0.3038$ nm and $c = 0.3220$ nm [113,114]. Primarily due to the strong covalent bonding in the B network, TiB$_2$ has a relatively narrow single-phase field with composition ranging from 65.6 to 66.7 at.% B, a high melting point (3225 °C), and a high hardness ($H = 25$ GPa) [94,112,115].

Recently, TiB$_x$ thin films deposited by non-reactive dc (direct current) sputtering from compound targets have been shown to have much higher hardnesses, 48–77 GPa, which are not simply due to correspondingly high residual stresses and growth-defect hardening [54,116]. The mechanism giving rise to this superhardness effect [51] is based upon the nanocolumnar structure of B supersaturated TiB$_2$ layers showing a pronounced (0 0 0 1)
texture [54,111,117–130], when produced by non-reactive sputtering of stoichiometric compound TiB$_2$ targets. The nanocolumnar structure is formed by a self-organization process during simultaneous growth of TiB$_2$ and B segregation [117].

Investigations on how the B/Ti at.-%-ratio ($x$) is influenced by the deposition conditions showed that increasing the Ar pressure results in a reduction of the B/Ti at.-%-ratio, even below 2 [124]. Increasing the target-to-substrate distance or using Xe instead of Ar as sputtering gas also results in a B/Ti at.-%-ratio below two [124]. An increasing B/Ti at.-%-ratio, even above three, is found for increasing $J_{Ar^+}/J_{Ti}$ [117], see Fig. 11. For radio-frequency

![Fig. 11. The B/Ti at.-%-ratio ($x$) for sputtered TiB$_x$ films as a function of $J_{Ar^+}/J_{Ti}$ [117].](image1)

![Fig. 12. XRD pattern of as-deposited TiB$_x$ for the different B/Ti at.-%-ratios ($x$) [131].](image2)
sputtered films an increase of B/Ti at.%-ratio from 2.1 to 2.8 with increasing the negative bias voltage from 0 to \(-50\) V, respectively, occurred [130].

---

**Fig. 13.** Bright-field cross-sectional TEM image with a corresponding SAED pattern of an as-deposited TiB$_{2.4}$ layer. The insert is a higher resolution image showing the \((0\ 0\ 0\ 1)\)-oriented, \(~5\) nm wide, subcolumnar nanostructure [117].

**Fig. 14.** Bright-field plan-view \((0\ 0\ 0\ 1)\) HRTEM image of an as-deposited TiB$_{2.4}$ layer. The insert is a lower resolution Z-contrast image [117].
The only XRD peaks obtained from as-deposited TiB\textsubscript{x} layers grown on steel substrates over the 2\theta range 20–90° are the (0 0 0 1) and the (0 0 0 2) reflex (see Fig. 12) [131]. With increasing B/Ti at.%-ratio a broadening of the XRD line profiles can be seen, indicating an increasing microstructural defect density (where also grain and column boundaries are counted as microstructural defects).

The biaxial compressive stress, \(\sigma\), measured using the cantilever beam method [23], in these TiB\textsubscript{x} films grown on Si(0 0 1) ranges from −0.03 GPa for \(x = 2.4\) through −1.82 GPa for \(x = 3.0\) to −2.40 GPa for \(x = 3.2\). Their thermal expansion coefficient is \(\sim 6.5 \times 10^{-6} \text{ K}^{-1}\) [23], giving rise to differential thermal contraction stresses of +0.90 GPa on Si(0 0 1) and −3.10 GPa on the steel substrates upon cooling from the growth temperature. Combining the results of the cantilever measurements with the thermal contraction stresses, the intrinsic stresses during growth are estimated to be −0.93 GPa for \(x = 2.4\), −2.72 GPa for \(x = 3.0\), and −3.30 GPa for \(x = 3.2\). Hardnesses \(H\) of the films on steel substrates are obtained by computer controlled microindentation at 50 mN load for which the Vickers indentation depth is \(\leq 10\%\) of the film thickness. \(H\) values for all of these B-rich TiB\textsubscript{x} films, irrespective of \(x\) or \(\sigma\), remain constant at 60 ± 3 GPa.

Fig. 13 is a bright-field TEM image with a corresponding selected area electron diffraction (SAED) pattern from an as-deposited TiB\textsubscript{2.4} layer. The film has a dense columnar structure with an average column diameter of \(\sim 20\) nm and a smooth surface with an average root-mean-square roughness essentially equal to that of the polished substrate surface of \(\sim 15\) nm. SAED patterns obtained as a function of film thickness show an [0 0 0 1] texture near the film/substrate interface with increased preferred orientation near the film surface. The columns with a diameter of 20 nm are composed of bundles of subcolumns, which extend along the film growth direction (see insert in Fig. 13) throughout the entire coating thickness. This is confirmed by HRTEM and high-angle annular dark-field (Z-contrast) investigations (see Fig. 14). The \(\sim 20\)-nm-wide columns are composed of smaller stoichiometric TiB\textsubscript{2} subcolumns with an average diameter of \(\sim 5\) nm, separated by a thin tissue phase of thickness 1–2 monolayer (ML). The modulation results from atomic number and/or strain contrast caused by a lateral compositional variation during film growth, similar to the surface-initiated spinodal decomposition [132] observed in Ti\textsubscript{0.5}Al\textsubscript{0.5}N [56].

Fig. 15. Schematic plan-view of the (0 0 0 1) textured \(\sim 20\) nm wide columns, indicating the formation process of the B-rich tissue phase by preferred B-diffusion on the (0 0 0 1) plane [164].
The insert in Fig. 14 is a high-angle annular dark-field (Z-contrast) plan-view image where ~5 nm sized bright areas are embedded in darker areas (elucidated by white lines). Regions with lower average atomic number Z appear darker and are therefore B-rich compared to the brighter TiB₂ regions. Quantitative analyses of electron energy-loss spectroscopy (EELS) measurements obtained from plan-view samples using a 1 nm spot size reveal that the subcolumns are stoichiometric TiB₂, whereas the disordered regions are highly B-rich [117].

As TiB₂ has a relatively narrow single-phase field (65.6–66.7 at.% B), the excess B in the films segregates to interfaces [117]. Due to the limited diffusivity at the low deposition temperature of 300 °C B-atoms that cannot reach the column boundaries accumulate to form a tissue phase within the columns (see Fig. 15). This process is preferred along the (0 0 0 1) plane as this is the preferred B-diffusion plane in TiB₂. Consequently, columns are encapsulated in excess B and are themselves composed of smaller stoichiometric TiB₂ subcolumns with an average diameter of ~5 nm, separated by a thin B-rich tissue phase exhibiting a thickness of 1–2 ML.

Due to the small dimension across the TiB₂ nanocolumns, i.e., the (0 0 0 1) plane, nucleation and glide of dislocations is inhibited during hardness indentation measurements (the primary dislocation glide planes in TiB₂ are {0 0 0 1} planes), while the high cohesive strength of the thin B-rich tissue phase prevents grain boundary sliding. Together, these two effects explain the observed superhardness of B-rich TiB₂ layers.

2.2.2. Nanocomposite coatings

An example to obtain a three-dimensional nanostructure is given for a TiN–TiB₂ thin film, which is non-reactively co-sputtered from a segmented TiN–TiB₂ target [24,110]. The
The bright-field HRTEM micrograph in Fig. 16a shows randomly oriented 2–3 nm sized crystals of the film in the as-deposited state. The Z (atomic number) contrast image (Fig. 16b) clearly indicates a random distribution over the whole film of at least two different phases (brighter and darker dots). The brighter dots indicate the appearance of a phase with a higher density (higher Z) than compared to the phase represented by the darker dots. It is shown in Refs. [24,110] that although the chemical composition of the film with Ti = 38 at.%, B = 30.5 at.%, and N = 31.5 at.% is located in the TiN–TiB₂–BN ternary phase field of the equilibrium Ti–B–N phase diagram, the film only consists of stoichiometric TiN and TiB₂ crystals encapsulated by B-rich disordered regions.

The diffraction rings in Fig. 16c confirm the mixture of TiN and TiB₂ nanocrystals with random orientation and a considerable amount of disordered regions. Due to the limited local resolution and a possible lattice deformation, no clear relation to the individual phases is possible. Considering the extremely short diffusion path for point and line defects in 2–3 nm grains to the nearest sink (i.e., grain and phase boundaries, interfaces), almost perfect crystals should form during deposition. In such nanocomposites with an average grain size of 2–3 nm, the volume fraction of grain and phase boundaries can be ∼50%. This estimate is based on a nanostructure where 2.5-nm-sized cubes are separated by ∼0.7 nm thin interfaces. Excess boron, which is known to segregate easily to boundaries, favors the formation of B-rich disordered boundary regions as discussed above [133]. Also the formation of amorphous BN and/or TiBₓ surrounding the nanocrystals is likely [134–136]. However, the disordered boundary regions and the small grain size of 2–3 nm are responsible for an increased interfacial volume fraction explaining the contribution of amorphous-like diffraction rings to the SAED pattern (see Fig. 16c).

Sputtering of a ceramic TiN–TiB₂ target causes a flux of Ti, N, and B atoms arriving on the substrate, resulting in the formation of TiN and TiB₂ nuclei (see Fig. 17). Boron has a low solubility in TiN and segregates during film growth to surfaces and interfaces forming B-rich disordered regions, which cover TiN surfaces and reduce boundary mobility. The presence of these regions affects film growth and hinders grain coarsening during coalescence and further grain growth [3,137].

The B-enriched areas, covering TiN crystals, promote nucleation of TiB₂, which itself has a low solubility for N. Thus, similar segregation-driven processes occur for N with TiB₂ as mentioned above for B with TiN. These processes repeatedly interrupt the homo-epitaxial growth of individual crystallites and cause renucleation during deposition as indicated in Fig. 17. The periodically interrupted growth explains the extremely small
crystallite size of 2–3 nm obtained for TiN–TiB₂ films. The conclusively high amount of interfacial fraction indicates that their constitution plays an important role in determining the mechanical properties. Yip [138] describes the grain size \( d_c \) where the Hall–Petch effect ceases or even reverses [76,138] as the smallest size that can sustain a dislocation pile-up. Below this size, dislocation blocking breaks down and intergranular sliding becomes important, as the proportion of interfacial atoms increases. Thus, the constitution of interfaces themselves plays an important role in determining the properties of such materials. Nevertheless, the mechanical properties of many materials can be improved by additions of B, which was proven to enhance the cohesion of boundaries by affecting the local bonding at the interface (assuming that there is no chlorine, oxygen or hydrogen uptake during deposition or exposure in air, which would weaken the interfacial strength) [139,140]. This agrees with the results where TiN–TiB₂ films with B-rich interfaces, although having extremely small grains of 2–3 nm in the as-deposited state, exhibit a high hardness of \(~42\) GPa, which is stable over time. Hardness values obtained after exposure in air for more than two years showed only a deviation within the measurement error (less than 10%). This suggests that the films contain no, or almost no amorphous BN, which is known to easily take up H and O to form hydroxides [141] with a larger molar volume as a result. This, we infer, causes an effective softening of the film from a weakening of grain boundaries and eventually might lead to delamination from the substrate.

3. Microstructural evolution during post-deposition annealing

In bulk materials, annealing treatments are well known and powerful practices to adjust their microstructure. In the following sections the main microstructural changes of hard coatings during post-deposition annealing will be discussed in some detail. These can be divided into recovery, recrystallization, and grain growth, and secondary phase formation/transformation (age hardening). We project that these microstructural changes will be of vital importance for the hard coating industry as the coating properties can either be adjusted by a proper heat-treatment or they change during application at elevated temperatures.

3.1. Recovery

Annealing of non-equilibrium materials with a high concentration of microstructural defects causes recovery effects which change the properties of the material. This is highly valid for thin films grown under ion irradiation. Recovery involves all annealing phenomena that occur before the appearance of new recrystallized strain-free grains, i.e., migration, recombination and annihilation of point defects, rearrangement and annihilation of dislocations, and growth and coalescence of subgrains [17]. It is a relatively homogeneous process, where the whole volume of the material is involved, if the local defect density does not vary significantly. Thus, recovery effects can be investigated by the measurement of intrinsic coating stresses.

Recovery starts with the recombination of vacancies and interstitial atoms or their diffusion to the boundaries [142,143]. For these processes, the different point defects need a corresponding thermal activation. At higher temperatures the stored energy within the films can also be lowered by dislocation movement. This results in annihilation of
dislocations and rearrangement into lower energy configurations, e.g., in the form of regular arrays (polygonization) or low-angle grain boundaries (see Fig. 6b) causing the formation of subgrains \[16,65,143\]. These processes are achieved by glide, climb, and cross-slip of dislocations. Therefore, recovery strongly depends on the stacking fault energy of a material, which determines the rate of dislocation climb and cross-slip (in metals of low stacking fault energy, climbing is difficult). Coarsening and coalescence of the substructure during further annealing can be obtained by subgrain boundary migration, where the local driving forces arise from the energy and orientation of adjacent boundaries \[17\]. These recovery effects result in a reduction of the overall stress state of the material. Consequently, the XRD film peak positions shift towards their unstrained values, the X-ray line broadening is reduced and the distribution of structural defects is more uniform (and hence symmetric line profiles are generated). This is shown in Fig. 18a and b for TiN and CrN layers, respectively. Both films were deposited at 300 °C and subsequently annealed at temperatures \(T_a\) for 30 min at pressures \(\leq 10^{-3}\) Pa.

Crystallites with grain sizes of a few nanometers often give broad XRD peaks of low intensity preventing stress measurements by X-ray techniques. Thus, the cantilever beam method may beneficially be applied to evaluate the thermal resistance of a coating against softening by stress relaxation. The biaxial stress \(\sigma\) in the coating can be calculated from the substrate-curvature radius \(r\) (e.g., measured by the deflection of two parallel laser beams) using the modified Stoney equation \[144\]:

\[
\sigma = \frac{E_s}{1 - \nu_s} \cdot \frac{r_s^2}{6 \cdot t_c} \frac{1}{r}.
\]

Here, \(E_s\) and \(\nu_s\) are Young’s modulus and Poisson’s ratio of the substrate, \(t_s\) and \(t_c\) are the thickness of substrate and coating, respectively. It should be mentioned that both intrinsic \(\sigma_{\text{int}}\) (i.e., growth induced) and thermal stresses \(\sigma_{\text{th}}\) (i.e., due to the mismatch of thermal expansion coefficients of substrate and coating) contribute to \(\sigma\) calculated via Eq. (3).
Fig. 19 shows an example of biaxial stress temperature measurement (BSTM) cycles for a sputtered nanocomposite TiB$_{0.6}$N$_{0.7}$ coating deposited onto a Si substrate. During heating of the film–substrate composite, the compressive stresses increase because of the higher thermal expansion coefficient of the coating ($\alpha_{\text{Ti-B-N}} = 6-7 \times 10^{-6} \text{K}^{-1}$) [23] with respect to the Si substrate ($\alpha_{\text{Si}} = 3.55 \times 10^{-6} \text{K}^{-1}$). In coatings having tensile stresses at room temperature, heating first relaxes these tensile stresses and then causes the film to go into a state of compression (Fig. 19). This thermoelastic behavior as a result of differential thermal expansion is only valid until recovery occurs. During the cooling segment, the stress–temperature curve again shows a linear-thermoelastic behavior. A second annealing treatment immediately after the first one does not show any significant deviation of the heating segment from the cooling segment of the previous run, provided that no tensile cracks have been formed after cooling down from the first cycle [145]. A deviation from the straight-line behavior during the heating portion of the second run appears if the annealing temperature exceeds the maximum of the first run. For coatings without tensile cracks formed in the cooling phase, plastic deformation in the coating or substrate during this heat treatment can be excluded, enabling to extract information on recovery from these BSTM curves [146].

From the BSTM cycles, essentially two types of information on the thermal stability of coatings may be obtained. The first one is the onset temperature for recovery $T_{\text{rec}}$, representing a measure for the thermal stability of the coating itself. The second one is the amount of stress relaxation $\Delta\sigma$ for a given maximum temperature and heating and cooling rate, respectively, which is related to the hardness loss after the annealing treatment. The extent of recovery depends on the annealing time and temperature, and the energy stored [16,17] within the films, which acts as an additional driving force [21,22]. In Refs. [21,22] and Section 3.1.1, it is shown that the onset temperature for recovery effects, $T_{\text{rec}}$, during thermal annealing decreases with increasing compressive stresses of the films in the as-deposited state. The dependence of recovery effects on time and temperature can be seen.
in Fig. 20 where two different heating rates (HR) are used to investigate identically grown TiN films (non-reactively deposited at 300 °C, $E_{\text{ion}} = 60$ eV, and $J_{\text{ion}}/J_{\text{me}} = 1$) onto silicon [15].

When recovery effects start, the additional recovery-time by using a lower heating rate results in a more pronounced stress relaxation than compared to the higher heating rate (see Fig. 20 for HR = 2.5 and 5 K min$^{-1}$, respectively). However, at the maximum annealing temperature of 700 °C used in this experiment, both BSTM curves meet at a similar stress value. During cooling down from 700 °C the BSTM curve of the film–substrate composite follows the linear-thermoelastic behavior, therefore, plastic deformation in the thin film or substrate during this heat treatment can be excluded. Within the linear-thermoelastic behavior (where no recovery effects occur) there are almost no measurable differences of the BSTM curves (and hence no microstructural changes with temperature) by using the different heating rates of 2.5 and 5 K min$^{-1}$. The stresses obtained from BSTM by measuring the bending of the film–substrate composite correspond to macrostresses, which are responsible for the displacement of the XRD peaks.

The apparent activation energy, $E_a$, for the recovery processes can be determined depending on the type of measurement performed for the relaxation of stress. Using the method of Damask and Dienes [147], $E_a$ can be calculated from XRD peak broadening data. Assuming that defect annealing occurs by a single, thermally driven process with an activation energy $E_a$ and rate constant $K_0$, the defect density $n$ is described by

$$\frac{dn}{dt} = F(n) \cdot K_0 \cdot \exp \left( -\frac{E_a}{k_B \cdot T} \right),$$

where $F(n)$ is a continuous function of $n$ and $k_B$ is Boltzmann’s constant. Using isochronal annealing curves, the two times $t_1$ and $t_2$ necessary to reach a given value of $n$ at temperatures $T_1$ and $T_2$, respectively, are related by
\[
\ln \left( \frac{t_1}{t_2} \right) = \frac{E_a}{k_B} \cdot \left( \frac{1}{T_1} - \frac{1}{T_2} \right).
\] (5)

In case where the defect annealing occurs via a single process, with a constant activation energy for all defect concentrations, performing this calculation at different values of \( n \) should yield a constant activation energy. If either or both of these assumptions are not valid, however, variable apparent activation energy will be obtained by this method. In this example [147], the X-ray structural broadening is due to inhomogeneous strains, which in turn are associated with defects in the crystal lattice. Assuming this broadening to be proportional to the defect density, Eq. (5) can be used along with data for the variation of X-ray structural broadening, \( \beta \), for a given peak with tempering temperature to determine the apparent activation energy for defect relaxation in a coating. Values for \( E_a \) are determined at various defect concentrations (values of \( \beta \)) by defining the fractional amount of defects remaining in the coating, \( \rho \), as

\[
\rho = \frac{\beta_f - \beta_0}{\beta_f},
\] (6)

where \( \beta_0 \) and \( \beta_f \) represent the structural broadening in the as-deposited and tempered coatings, respectively. This approach was used by Almer et al. [148] for determining \( E_a \) in Cr–N coatings made by arc-deposition.

In the method of Mittemeijer et al. [149], \( E_a \) is determined from the rate of stress recovery. The model does not depend on any specific kinetic mechanism, but assumes that the fraction transformed, \( f \), is fully determined by a state variable, \( \beta \), i.e., \( f = F(\beta) \). For isothermal annealing, \( \beta = k \cdot t \) where \( k = k_B \cdot \exp(-E_a \cdot (R \cdot T)^{-1}) \). The final expression is

---

Fig. 21. Onset temperature for recovery for non-reactively and reactively sputteres TiN coatings as a function of the biaxial stress in the as-deposited state. The coating, deposited at an ion energy of 30 eV and an ion/atom flux ratio of 0.5, is characterized by open-voided column boundaries allowing easy stress relaxation. The deposition temperature was 300 °C [15].
\[
\ln(t_{f2} - t_{f1}) = \frac{E_a}{k \cdot T} - \ln k_B + \ln(\beta_{f2} - \beta_{f1}),
\]

where \( t_{f2} - t_{f1} \) is the time for the recovery to proceed from fraction \( f_1 \) to \( f_2 \) at temperature \( T \). The apparent activation energy, \( E_a \), for the process can be obtained by measuring the slope of the \( \ln(t_{f2} - t_{f1}) \) versus \( (k \cdot T)^{-1} \) plot. This method was applied by Karlsson et al. [92] and will further be discussed in Section 3.1.1.

### 3.1.1. Single-phase hard coatings

Generally, recovery of coatings deposited by plasma-assisted deposition techniques occurs during annealing above the deposition temperature. For example, magnetron sputtered TiN [15] and CrN [21] exhibit annihilation of point defects created by ion irradiation during deposition that starts more or less immediately above the deposition temperature. The intention here is to discuss the stability of homogeneous single-phase coatings such as TiN, CrN, and TiC\(_x\)N\(_{1-x}\) with respect to composition and compressive intrinsic (growth-induced) residual stress. Oettel et al. [150] observed a decrease in compressive residual stresses in arc-evaporated TiN and Ti\(_{1-x}\)Al\(_x\)N films after annealing. Herr and Broszeit [151] reported a decrease of the apparent hardness of TiN films from 36 to 27 GPa when the compressive stress decreased from \(-6.7\) to \(-2.1\) GPa upon annealing for 1 h at 650 °C. Perry [152] studied the change in XRD peak widths of electron-beam-evaporated TiN films deposited between 400 and 500 °C. By annealing the films up to 900 °C an activation energy of 2.1 eV was obtained, which was attributed to diffusion of self-interstitials into excess vacancies.

Fig. 21 shows an apparently linear relation between the onset temperature for recovery and the biaxial stress in the as-deposited state of non-reactively and reactively magnetron sputtered TiN coatings [15]. Using high ion/atom flux ratios at moderate ion energies, the onset temperature may be shifted slightly towards higher values. This may be explained by the promoted relaxation of stress-active defects due to a more intense low-energy ion...
bombardment during growth [15]. The stored internal energy associated with the defects due to the growth process provides a driving force for changes in microstructure when heated to a temperature sufficient to promote diffusion. However, it should be mentioned that coatings with open-voided column boundaries deviate from this linear dependence (see Fig. 21) because of the possibility of stress relaxation at their column boundaries, making predictions of coating behavior difficult.

Since structural defects act as obstacles for dislocation motion, a direct relationship between defect density and hardness is expected [143]. This is shown in Fig. 22 for TiN films reactively deposited at 300 °C under an ion bombardment of $E_{\text{ion}} = 60$ eV and $J_{\text{ion}}/J_{\text{me}} = 1$. These films exhibit intrinsic compressive macrostrain in the as-deposited state ($a > a_0$ of bulk TiN which is 0.4240 nm) changing to an almost unstrained condition after annealing at $T_a > 800$ °C (see Fig. 22). In addition to $a$, also the diffraction peak widths $\Gamma$ (only $\Gamma_{200}$ is shown in Fig. 22) decrease with $T_a$ for annealing at temperatures above the deposition temperature. The results suggest that the increased diffusivities lead to defect annihilation and atomic rearrangement, resulting in less defects and lower stresses due to recovery effects. Consequently, the hardness of the films approach the hardness for bulk TiN ($H \sim 20$ GPa) as growth-defect induced hardening diminishes. The slightly lower onset-temperature for these recovery effects obtained by structural investigations (Fig. 22) compared to BSTM (Fig. 20) is due to the different annealing conditions. The structural investigations were obtained after annealing at $T_a$ for 30 min at pressures $< 10^{-3}$ Pa, whereas for BSTM at pressures $< 10^{-3}$ Pa a constant heating and cooling rate of 5 K min$^{-1}$ was used.

The effects of annealing up to 900 °C on the intrinsic stress, $\sigma_{\text{int}}$, and hardness of arc-evaporated TiC$_x$N$_{1-x}$ films have been addressed recently by Karlsson et al. [92]. Fig. 23 shows the residual stress measured in TiC$_x$N$_{1-x}$ films as a function of annealing time and temperature. Films with $x = 0, 0.15, \text{and } 0.45$, each having an initial compressive intrinsic stress $\sigma_{\text{int}} = -5.4$ GPa, were deposited by varying the substrate bias $V_s$ and the
gas composition. Annealing above the deposition temperature leads to a steep decrease in the magnitude of $\sigma_{\text{int}}$ to a saturation stress value, which is a function of the annealing temperature. For temperatures above the substrate temperature during deposition $T_{\text{s}}$, $\sigma_{\text{int}}$ decreases steeply in the first 10–20 min, after which the stress relaxation process essentially saturates. The magnitude of the saturation stress level progressively decreases with increasing $T_{\text{a}}$. For instance, annealing TiC$_{0.45}$N$_{0.55}$ at 500 °C for nearly 32 h decreased $\sigma_{\text{int}}$ by only 0.5 GPa (see Fig. 23a). The intrinsic stress decreases in a roughly linear fashion with increasing annealing temperature, $T_{\text{a}}$, and deviates from linearity with increased stress relaxation (see Fig. 23b). For comparison, the $\sigma_{\text{int}}$ values of as-deposited films grown at different $T_{\text{s}}$ obtained from Ref. [89] are also shown. In order to compare the annealing

Fig. 24. $\ln t_{50\%}$ versus $(k_B T_a)^{-1}$ plot for TiC$_x$N$_{1-x}$ films. The activation energy $E_a$ is determined by measuring the slope [92].

![Graph](image)

Fig. 25. Influence of the biaxial stress in the as-deposited state on stress relaxation during BSTM up to 700 °C for sputtered CrN and TiN coatings deposited at different ion bombardment conditions [21].

![Graph](image)
behavior of films with different as-deposited residual stress states TiC$_{0.45}$N$_{0.55}$ films with a lower residual stress (‘low $\sigma$’) are added, which were grown at high bias voltage [89]. For TiC$_{0.45}$N$_{0.55}$ films with different initial stress values, the final stress values converge with increasing $T_a$, and are indistinguishable at $T_a \geq 800$ °C. Note that a film deposited at a low $T_s = T_1$ and annealed at $T_a = T_2$ ($T_2 > T_1$) has a higher stress level than an as-deposited film of the same composition deposited at $T_s = T_2$. This observation indicates that stress relaxation during deposition is more efficient (with more degrees of freedom and energetic conditions prevailing in the collision cascade), i.e., it causes a larger stress recovery than a post-thermal annealing treatment at the same temperature.

The corresponding apparent activation energies for stress relaxation are $E_a = 2.4$, 2.9, and 3.1 eV, for $x = 0$, 0.15, and 0.45, respectively, as shown in Fig. 24 which is an Arrhenius plot for the stress relaxation versus inverse temperature [92]. TiC$_{0.45}$N$_{0.55}$ films with a lower initial stress $\sigma_{int} = -3$ GPa, obtained using a high substrate bias, show a higher activation energy $E_a = 4.2$ eV. In all the films, stress relaxation is accompanied by a decrease in defect density indicated by the decreased width of XRD peaks and decreased strain contrast in TEM. Correlation of these results with film hardness and microstructure measurements indicates that stress relaxation is a result of point-defect annihilation taking place both during short-living metal–ion surface collision cascades during deposition and during post-deposition annealing by thermally activated processes. The difference in $E_a$ for the films of the same composition deposited at different $V_s$ proves the existence of different types of point-defect configurations and recombination mechanisms.

Recovery was found to be activated for arc-evaporated Cr–N coatings with energies in the range of 2.1–3.1 eV depending on the initial stress state $\sigma$, which in turn was a function of the deposition conditions [148]. It was also concluded that for the Cr–N coating system, possessing a relatively low melting temperature, defect diffusion during growth at 400 °C was active. The values are similar to those observed for bulk diffusion of nitrogen in CrN. However, also the variation of the chemical composition of Cr–N and other systems during annealing has to be considered. For example, CrN transforms into Cr$_2$N and finally Cr by desorption of nitrogen [153].

The amount of recovery occurring during thermal annealing strongly depends on the amount of defects in the coating, i.e., on the biaxial stress in the as-deposited state, providing the driving force for recovery. This is illustrated in Fig. 25 for reactively sputtered stoichiometric CrN coatings as well as reactively and non-reactively sputtered TiN coatings, where a linear relation between the amount of recovery, $\Delta \sigma$, and the biaxial stress, $\sigma$, in the as-deposited condition was found [15,21,154]. The graph shows that recovery effects are material-dependent. If a coating shows compressive stresses in the as-deposited state (substitution and displacements of atoms on interstitial lattice sites, Frenkel pairs and anti-Schottky defects), recovery effects cause stress reduction. However, intrinsic tensile stresses as a result of voids or vacancies (Schottky defects), which are promoted by an insufficient ion bombardment, especially if the total working gas pressure is high, may even increase during annealing. The thermal activation causes these defects to anneal out or move to grain boundaries, which are favored sites for lattice imperfections, determining the total amount of stress relaxation due to annealing.

Almer et al. [148] studied the thermal stability of Cr–N coatings prepared by arc-evaporation. The as-deposited coatings consisted of slightly nitrogen-deficient cubic-structure CrN and were under compressive residual stress. Significant reductions in the defect density accompanied by the formation of equiaxial grains of equilibrium phase fractions of
β-Cr₂N and stoichiometric CrN were found to take place after furnace tempering between 400 and 550 °C for 270 min in an Ar atmosphere. Correspondingly, Herr and Broszeit [151] have shown that significant stress relaxation occurs in sputter-deposited CrN coatings after annealing between 400 and 650 °C.

For the studies of stress recovery in a range of other nitrides, Perry [152] has reported on the effects of tempering for TiN, ZrN, and HfN films under compressive stress. The expanded lattice parameter in the as-deposited state contracts above 400 °C, where the contraction is completed after 1 h annealing at 800 °C. However, the lattice parameters found were still above the equilibrium values indicating defects, which are still resistant to thermal treatment at this temperature. For sputtered low-stress (i.e., −0.4 GPa) VN coatings deposited at a substrate temperature of 300 °C, an extremely low onset temperature of 330–350 °C for recovery has been reported [155].

Residual stress engineering can be used for improving film performance by applying strain hardening, but for industrial application the thermal stability is very important. The gradual decrease of the contribution of strain hardening results in a hardness loss at temperatures above the deposition temperature, thus limiting the use of highly stressed coatings in high-temperature applications. Although a hardness loss at elevated temperatures due to vanishing strain hardening is unavoidable, a constant high hardness well above the bulk hardness up to 700 °C has been reported for sputtered B-rich TiBₓ coatings [117] (see also Section 3.1.2.2). There, the excess boron results in a compositional modulation exhibiting a boron-rich tissue phase surrounding TiB₂ nanocolumns during growth. This modulation contributes to the hardness enhancement and is not affected up to 700 °C (although stress recovery during the same annealing treatment occurs), thus explaining both the extremely high hardness of B-rich TiBₓ coatings and their high thermal stability.

3.1.2. Composition or phase modulated hard coatings
3.1.2.1. Artificial superlattice films. The deposition conditions for superlattices with sequential fluxes offers additional conditions for the generation of residual stresses via energetic particle/surface interactions during nucleation and growth. There are examples in literature of both stress increase and decrease in a given layered system depending on the periodicity. For example, in TiN/NbN nitride superlattices a decreased compressive stress state was found compared to the homogeneous nitride films [8]. The exact mechanisms of the decreased stress generation in superlattice films are unknown. It is proposed that the strain level built up by point defects created in the collision cascades can partly be relaxed along with coherency-strain relaxation of the nitride overlayers during misfit dislocation formation. Surface tension effects at the interfaces, however, may also be present in the superlattice [156]. Furthermore, for different mass number of the elements in the constituent layers, the ion–surface interactions – in terms of recoil, forward sputtering, and Frenkel pair production – will be different, depending on which layer is growing. The Linköping group recently introduced ‘atomic scale interface engineering’ for growth of sub-nanometer-scale multilayers [157,158]. This technique utilizes an intense ion assistance during growth \( (J_{\text{ion}}/J_{\text{me}} > 5) \) of a superlattice film achieved by dual target magnetron sputtering. The ion energy is modulated within each layer to reduce roughness and intermixing. The initial \(~0.3 \text{ nm}\) of each layer is deposited with low energy \((<10 \text{ eV})\) to eliminate forward recoil mixing of the interfaces, while slightly higher ion energies \((20–30 \text{ eV})\) are used for the remaining parts of each layer to densify and smoothen the layer surfaces.
Although superlattice coatings with different stress levels (up to –9 GPa have been reported in Ref. [32]) have been studied extensively, there is only limited information available on their stress recovery behavior. It is significant that initial annealing of as-deposited superlattice structures or multilayers can exhibit apparent interface sharpening. This phenomenon is little studied while it may be quickly transient during annealing experiments. It can be due to phase separation of constituent layers that are in a chemically intermixed state from the synthesis or arise from interface sharpening due to a surface energy minimization of structurally rough interfaces. It is termed differently in literature including ‘chemical cleaning’ or ‘reverse diffusion’ as studied in the Ni–Nb/C system [159]; ‘negative interdiffusion’ for the CoMoN/CN system [160]; and ‘preannealing stabilization’ for the W/C system [161]. A typical gauge for the effect is XRD peak sharpening and intensity increase for reflectivity or diffraction measurements of the film structures during initial annealing. Eventually interdiffusion or layer coarsening takes place for extended annealing depending on themiscibility of the system. For example, Setoyama et al. [162] reported on a slight increase of the intensity of the superlattice peak in high-temperature XRD (HT-XRD) at 300 °C for TiN/AlN superlattices with a periodicity of 2.9 nm, which indicates sharpening of the interfaces due to faster interfacial diffusion.

3.1.2.2. Nanocolumnar coatings. This section describes the annealing behavior of TiB$_2$ coatings presented in Section 2.2.1. as an example for nanocolumnar coatings. XRD investigations of the films show that the (0 0 0 1) peak position continuously shifts from 27.54° (lattice constant $c = 0.3239$ nm, which is in the out-of-film direction) in the as-deposited state to 27.71° ($c = 0.3219$ nm) after annealing at 800 °C in vacuum (pressure $\leq 10^{-3}$ Pa) for 1 h as indicated in Fig. 26a and b. The lattice constant in the [0 0 0 1] direction for bulk TiB$_2$ is indicated by an arrow [117,164].
TiB$_2$ is $c_0 = 0.3220$ nm [114]. Thus, the TiB$_{2.4}$ films exhibit compressive in-plane macrostrain in the as-deposited state ($c > c_0$), changing to an essentially unstrained condition after annealing at $T_a \geq 700$ °C (see Fig. 26b). The full width at half maximum intensity $I_{0001}$ of the (0 0 0 1) reflection remains constant at $0.28 \pm 0.01^\circ$ as a function of annealing temperature for $T_a \leq 800$ °C, suggesting that there are no significant changes in either nanostructure or local lattice microstrain [85].

As already discussed in Section 3.1.1 and shown for TiN and CrN (see Figs. 20, 21 and 25) increased diffusivities during film annealing commonly lead to defect annihilation and atomic rearrangement resulting in lower stresses [49,143]. Thus, lattice parameter, diffraction peak widths $\Gamma$, and the hardness of stoichiometric binary nitride and carbide layers typically decrease during annealing [15,19,21,26,58,163,164]. However, for TiB$_{2.4}$, it is shown [117] that while the film lattice constant (and hence macrostress) decreases with annealing temperature $T_a \leq 800$ °C, $\Gamma$ and $H$ remain constant (see Fig. 26b).

TEM investigations reveal no observable differences in the nanostructure of as-deposited TiB$_{2.4}$ layers and samples which were annealed at 700 °C [117]. This is consistent with the XRD results showing that $I_{0001}$ remains constant with $T_a$. Fig. 27 is a bright-field cross-sectional TEM image with a corresponding selected area electron diffraction (SAED) pattern from a TiB$_{2.4}$ layer annealed at 700 °C for 1 h. Again, the compositional modulation, as already discussed in Section 2.2.1, can be found (see Fig. 13). Essentially, the 20-nm-diameter columns, which are composed of bundles of subcolumns with an average coherence length of $\sim 5$ nm (see insert in Fig. 27) did not change with annealing. Resulting from the thermal stability of this nanostructure also the hardness is unaffected for $T_a \leq 800$ °C (see Fig. 26b).

Fig. 27. Bright-field cross-sectional TEM image with a corresponding SAED pattern obtained near the upper portion of a TiB$_{2.4}$ film after annealing for 1 h in vacuum (pressure $\leq 10^{-3}$ Pa) at 700 °C. The insert is a higher resolution image showing the 0 0 0 1-oriented, $\sim 5$ nm wide, subcolumnar nanostructure [117].
3.1.2.3. Nanocomposite coatings. Nanocomposite coatings are materials consisting of at least two nanoscale phases where phase separation typically occurs already during synthesis. A motivating factor for the research on nanocomposite coatings has been to achieve superhard materials. The hardness determination, however, becomes dubious at levels close to that of diamond and, therefore, measurements must be interpreted with care. Design concepts for nanocomposites were postulated by Veprˇek et al. [10,11]. The constituting phase systems should be selected for strong segregation (immiscibility), strong interface bonding and high shear moduli. It is well known that nm-grains exhibit no dislocation activity, thus, plasticity is obtained by means of grain boundary sliding/rotation [165].

Most coatings made by PVD processing are prone to residual (compressive) stress formation. Strain hardening is, however, often assumed to have a minor contribution to the extremely high hardness values of nanocomposites because of easy migration of defects to phase boundaries during growth [36] and, consequently, often relatively low stress values (e.g., below 1 GPa of compressive stress) have been reported [12]. Thus, only relatively limited stress relaxation effects at elevated temperatures have been found for low-stress nanocomposite coatings, for example in the TiN–Si3N4 system [166,167]. There, stress relaxation is not only an effect of recovery mechanisms (i.e., annihilation of point defects and dislocations), but also grain boundary sliding/rotation events [165] are assumed to contribute. In agreement with the less pronounced stress relaxation, only limited hardness losses after annealing below the recrystallization temperature have been reported for several nanocomposite coatings. For example, Männling et al. [46] found no hardness decrease for TiN–Si3N4 nanocomposite coatings up to an annealing temperature of 800 °C. For the systems TiN–TiB2 and TiC–TiB2, the formation of nanocomposite films with stresses below 1 GPa by magnetron sputtering has been reported [23]. The onset of stress recovery was found to vary between 400 and 480 °C, which again is only slightly above the deposition temperature of 300 °C. The maximum onset temperature was found for comparable amounts of TiN (or TiC) and TiB2 and well-defined phase boundaries obtained by low-energy ion irradiation during growth. Again, no hardness loss was reported for an annealing cycle up to 700 °C [168].

Up to now, only very limited information is available on the thermal stability of nanocomposite coatings consisting of a hard and a soft phase. Karvánková et al. [169] reported softening of highly stressed (compressive stress up to several GPa) sputtered ZrN–Ni and CrN–Ni films upon annealing up to 400–600 °C. They attribute this hardness loss to stress recovery, evidenced by the decreasing lattice parameter of the nitride phase and found no significant effect of the annealing temperature on the grain size up to 700 °C. Although the lower thermal stability against stress relaxation compared to low-stress nanocomposite coatings can be explained by the high compressive stress and consequently high defect density representing a higher driving force, the origin for the high stress levels in these coatings is still unclear. Zeman et al. [170] reported on the stress recovery of complex Al–Si–Cu–N films, which were classified as nanocomposite (consisting of Al, Al2Cu, and AlN phases in the as-deposited condition) and amorphous coatings. BSTM cycles up to 700 °C showed that low-nitrogen containing crystalline films (i.e., coatings with Al and Al2Cu phases) showed stress relaxation due to recovery and recrystallization at temperatures above 300 °C. Here, stress relaxation is related to changes in phase structure and texture of the films. In contrast, amorphous low-stress high-nitrogen containing films showed no apparent stress relaxation and no hardness loss after thermal cycling. However, these coatings show small deviations from thermoelasticity during the BSTM heating and cooling.
cycles, indicating that small sliding events accommodate the thermal stresses introduced due to the mismatch of the thermal expansion coefficients.

3.2. Recrystallization and grain growth

Recrystallization involves the formation of new strain-free grains and the subsequent growth of these grains leads to a further reduction of energy [17]. In nanostructured materials, an extremely high interfacial energy is stored due to the high amount of grain and phase boundaries [171], which can reach ~70 vol.% for grain sizes of 2–3 nm [172]. This stored energy represents a high driving force for recrystallization and subsequent grain growth and, consequently, rapid recrystallization and coarsening kinetics have been reported for nanocrystalline metal films for relatively low recrystallization temperatures [173].

For detailed investigations of recrystallization and other microstructural changes with temperature and time, differential scanning calorimetry (DSC) is an ideal complement to techniques such as XRD and TEM. For example, phase transformation temperatures, enthalpies of transformation, thermal hysteresis data, and melting temperatures of a wide range of samples can be determined from one DSC measurement [174,175]. This is achieved by measuring the difference in heat flow to (endothermic) or from (exothermic) a sample and a thermally inert reference material as a function of temperature or time. If applied to coatings, the substrate material has to be removed to avoid superposition of microstructural changes of substrate and coating. This may be done by chemical dissolution of the substrate, e.g., thin low-alloyed steel substrates may be dissolved in nitric acid. To achieve a suitable heat flow, a minimum mass of coating material is required (e.g., 30 mg). DSC measurements may be performed in inert or aggressive environments.

The DSC method shows high potential to study microstructural changes during thermal annealing, which will further be discussed in Sections 3.2.2.2 and 3.3.2.1.

3.2.1. Single-phase hard coatings

While ultra-pure In, Pb, and Sn film materials exhibit recrystallization and subsequent grain growth at or even below room temperature and Al films are sensitive at above 80°C, refractory carbides and nitrides are usually not considered in this context due to the large difference between melting temperature of the material and the deposition temperature or the temperature of application for the film [26]. However, this section will show, that a small grain size and a large level of compressive intrinsic stress, which are typical for PVD films, offer significant driving forces leading to recrystallization of refractory carbides and nitrides even at relatively low temperatures. It is of significant technological relevance for making predictions of coating performance and stability that exposing a PVD coating to a temperature above its deposition temperature will trigger recovery processes within the material. The relaxation of films with deposition-process-induced intrinsic residual compressive stress (from growth defects) may thus occur not only through effective point defect annihilation, but also by creep or recrystallization. In the following the different technologically relevant nitride systems Cr–N, TiC_xN_1–x, and Ti_1–xAI_xN, which exhibit a large variation in thermal stability with respect to recrystallization, are compared. The lowest apparent stability is observed for Cr–N films as indicated in the previous section. Significant reductions in the defect density accompanied by the formation of equiaxial grains of equilibrium phase fractions of β-Cr_2N and stoichiometric CrN were found to
take place during tempering at 400–550 °C for 270 min [148] for films deposited at 240 or 340 °C. In fact, precipitation of β-Cr$_2$N took place inside CrN columnar grains. Heau et al. [176] reported on the decomposition of Cr$_2$N and CrN films prepared by magnetron sputtering. For both coating types an increase of the grain size obtained by HT-XRD was found above 400–450 °C. The onset temperature for formation of the β-Cr$_2$N phase was in the same temperature range. It should be mentioned that O incorporation in the CrN lattice up to 25 at.% has almost no effect on the onset temperature for grain coarsening, whereas the coarsening rate is dramatically reduced by O incorporation [177].

As an example for the TiC$_x$N$_{1-x}$ system plan-view TEM micrographs from an arc-evaporated TiC$_{0.45}$N$_{0.55}$ film are shown for the as-deposited state ($T_s = 550$ °C, see Fig. 28a) and after annealing at 900 °C for 2 h (Fig. 28b) [92]. The as-deposited film exhibited a dense microstructure with no porosity and a residual stress of −5.5 GPa. Due to the presence of strong strain contrast (a high lattice defect density) grain boundaries could not be imaged easily. However, from following the shift of bending contours (seen as bright and dark areas in Fig. 28a) over the sample during tilting it was possible to obtain a measure for the cell size of the film. During film growth, dislocations have aligned themselves within the emerging grains resulting in the formation of subgrain boundaries (polygonization) which partly relax intrinsic stresses [17]. This is similar to dynamical recovery occurring during hot rolling of metal sheets. Characteristic cell sizes were 100–300 nm similar to those obtained from cross-section images of samples grown under the same condition.

The annealed sample, however, exhibited a granular structure with grain sizes in the range of 25–100 nm (see Fig. 28b). This is lower than the cell size of the as-deposited film, which is a result of the high nucleation density for recrystallization, where the cell boundaries with high dislocation density may act as nuclei within the original grains. There was an apparent reduction in defect density, however, isolated dislocations do not appear. Annealing also resulted in the formation of voids between grains and at grain boundary

---

**Fig. 28.** Representative plan-view TEM micrographs from TiC$_{0.45}$N$_{0.55}$ films (a) in the as-deposited state and (b) after annealing at 900 °C for 120 min. The insert is a high magnification micrograph showing grain boundaries (120° angles) as well as void formation at the triple junctions [92].
triple points. A larger magnification of such an area is shown in the insert of Fig. 28b. The investigation by TEM showed that annealing changed the film microstructure from a material, exhibiting a high dislocation density (similar to a cold-worked material), to a recrystallized material with well-defined grain boundaries. The presence of grain boundary voids shows that vacancies (or any excess N) have segregated during the annealing treatment.

3.2.2. Composition or phase modulated hard coatings

3.2.2.1. Artificial superlattice films. The high density of interfaces makes artificial superlattices or nanolaminates susceptible to thermodynamically driven microstructural changes upon annealing [178]. Four different types and mechanisms can be distinguished: (i) interdiffusion, (ii) coarsening of the layering, (iii) reactions between the layers to produce a new phase, and (iv) transformation within one or both layer types. Examples from the first two effects are given below.

Interdiffusion of thin film layers results in a modification of the composition profile. As an effect, this can remove the precondition for mechanical hardness enhancement in nitride superlattices or loss of reflectivity in optical mirrors based on nanolayered coatings, both of which show a reciprocal dependence upon the interface width [178–183]. Superlattice or nanolaminate thin films show a high potential as hard protective coatings on cutting tools. However, during the cutting process the tools may be subjected to temperatures as high as 1000–1200 °C [64]. TiN/NbN and Ti$_1$–$_x$Al$_x$N/CrN superlattices show constant mechanical behavior up to 700 and 750 °C, respectively [184,185].

Recently, the metal interdiffusion in TiN/NbN superlattices [186,187] was investigated by annealing samples in a purified He atmosphere at 1 atm. The TiN/NbN system exhibits solid solubility [188,189]. Using the decay of X-ray superlattice satellite intensities in both low-angle (reflection measurement) and high-angle (diffraction) regime as shown in Fig. 29a and b, a range of apparent activation energies were found; $E_a = 1.2$ eV for annealing temperatures $T_a < 830$ °C [186], $E_a = 2.6$ eV for $T_a$ up to 875 °C [186] and $E_a = 4.5$ eV for $T_a$ up to 930 °C [187] (see Fig. 30). While the lower energy should correspond to defect-
mediated diffusion (as will be discussed below), the 2.6 and 4.5 eV values may correspond to grain boundary or even bulk diffusion. During interdiffusion in this system, Ti diffuses at a faster rate into NbN than vice versa while retaining a compositionally sharp interface.

Engström et al. [190] studied the high-temperature stability (up to 1100 °C) of epitaxial Mo/NbN nanolaminates which represents a non isostructural system (bcc metal and fcc nitride). After 3 h at 1000 °C, an interfacial reaction resulted in the formation of a tetragonal MoNbN phase. The superlattice satellite peaks were correspondingly reduced in intensity and interpreted as a gradual coarsening of the Mo, NbN, and MoNbN phases.

Other superlattice systems produced by PVD methods have also been investigated with respect to thermal stability, including Ti$_{1-x}$Al$_x$N/CrN [185], TiN/CN$_x$ [191,192], and ZrN/CN$_x$ [193]. For Ti$_{1-x}$Al$_x$N/CrN superlattices with a compositional modulation period of 3.6 nm, annealing at 750 °C resulted in effectively interdiffused layers after 16 h [185]. The Ti(Zr)N/CN$_x$ couples are thermodynamically unstable with respect to the formation of TiN–TiC and ZrN–ZrC solid solutions.

The lifetime of TiN/NbN superlattices showing an 8 nm period at 850 °C was calculated to be 1 h [190]. During annealing the superlattice hardness was reduced from >30 to 27 GPa. Further annealing at 950 °C resulted in a drop to 23 GPa. Such values represent a useful base for selecting cutting parameters in application.

For crystalline/amorphous superlattice structures, the group of Barnett studied TiN/TiB$_2$ and ZrN/ZrB$_2$ systems [194]. High-angle XRD on as-deposited polycrystalline superlattices showed amorphous boride peaks and crystalline nitride peaks. On the other hand, monolithic boride films generally showed both (0 0 0 1) and (0 0 0 2) X-ray reflections, indicating that a crystalline structure was present, with the hexagonal basal planes oriented out of plane. Thus, boride crystallization was inhibited for nanometer-thick layers. After annealing at 1000 °C for 1 h, XRD results showed clear boride (0 0 0 1) peaks, indicating that these layers had crystallized. Superlattice reflections were observed in low-angle reflectivity scans and were retained after annealing, indicating that the nanolayers are stable. TiN/TiB$_2$ and ZrN/ZrB$_2$ are thus excellent examples of nanolayers that can exhibit...
good high-temperature stability, which is in agreement with the results of TiN–TiB$_2$ nano-composites [7,23,164,195,196]. In these superlattice structures, the rocksalt nitride (1 1 1) planes match well with the hexagonal boride basal planes.

3.2.2.2. Nanocomposite coatings. Grain sizes of a few nanometers and a high amount of interfaces provide a high driving force for recrystallization and grain growth. Although this is an important aspect determining applicability of nanocomposite coatings with superhardness, only limited information on their resistance against softening by recrystallization and grain growth is available. Literature data focus on the systems TiN–Si$_3$N$_4$ and TiN–TiB$_2$, and the most important results obtained will be summarized in the following.

In case of nanocrystalline TiN–Si$_3$N$_4$ composite films (consisting of crystalline TiN separated by an amorphous Si$_3$N$_4$ tissue phase), grain growth of the TiN phase takes place at temperatures between 800 and 1200 °C with an increasing onset temperature for grain sizes decreasing from ~10 to 2 nm [12,36,46]. This has been attributed to the stabilizing role of the Si$_3$N$_4$ tissue phase. The crystallization of the amorphous Si$_3$N$_4$ phase in thin films made by reactive rf magnetron sputtering takes place during annealing at temperatures between 1300 and 1700 °C leading to the formation of α-Si$_3$N$_4$ and β-Si$_3$N$_4$ [197]. The crystallization process of silicon nitride was also described as three-dimensional, interface controlled growth from pre-existing nuclei. The determined rate constants followed a thermal activated behavior with a single activation energy of 5.5 eV.

The evolution of the dual-phase structure of nanocomposite TiN–TiB$_2$ coatings has recently been studied using HRTEM investigations after annealing at temperatures between 700 and 1400 °C for 1 h [110,195]. Fig. 16 shows that in the as-deposited state the coating consists of TiN and TiB$_2$ nanocrystals with a grain size of 2–3 nm. The micro-structural changes during annealing were investigated by DSC measurements in Ar up to 1450 °C. The solid line in Fig. 31 is a typical RT to 1400 °C DSC spectra from an as-deposited TiN–TiB$_2$ layer. Each DSC measurement (run) has to be immediately followed by a second run (re-run), which serves as a baseline (i.e., where no microstructural changes
occur anymore) for the first run (Fig. 31a). The deviation between the first measurement and the re-run shows all exothermic and endothermic irreversible reactions of the coating (Fig. 31b). The broad peaks suggest that several exothermic reactions have occurred over the investigated temperature range. The individual reactions can be quantified, e.g., by the onset temperature $T_o$, the peak temperature $T_p$ and the peak area $A_p$ (see Fig. 31b). The enthalpy of the individual reaction can be calculated by dividing $A_p$ with the used heating rate $B$.

EPMA and XRD analyses of as-deposited layers annealed to different temperatures $T_a$ are carried out to identify the structural and chemical changes giving rise to the DSC peaks. EPMA results show that the ratio of B/Ti at.% remains constant at $\sim$0.80 for annealing temperatures up to 800 °C and than decreases rapidly at higher $T_a$ values. However, the N/Ti ratio remains constant at $\sim$0.83 up to 1100 °C to reaction with residual oxygen to form a surface B$_2$O$_3$ layer which transforms, in the presence of residual moisture, to volatile boric acid (H$_3$BO$_3$) [198]. Thus, the broad DSC signal between 700 and 1000 °C (see Fig. 31b) is assigned to B loss via the formation of H$_3$BO$_3$.

Fig. 32 shows XRD results for as-deposited and annealed films as a function of $T_a$. XRD peaks from as-deposited layers are very broad and consist of overlapping TiN and TiB$_2$ peaks which are already quite broad due to the nanometer-scale grain size. Previously we have shown [23] that Ti–B–N layers deposited under these conditions exhibit only a relatively modest biaxial compressive stress, $\sim$0.5 GPa. Thus, strain is not the major contribution to XRD peak broadening but grain size [85].

XRD scans from annealed films exhibit an increase in intensity together with peak narrowing with increasing $T_a$ indicating crystallite growth. Separate TiN and TiB$_2$ peaks are resolved at $T_a \geq$ 800 °C. However, there is no evidence for the formation of additional crystalline phases. The films continue to lose B at $T_a \geq$ 800 °C until their composition

Fig. 32. XRD evolution of a TiN–TiB$_2$ film (Ti = 38 at.%, B = 30.5 at.%, and N = 31.5 at.%) with annealing temperature $T_a$. Isothermal annealing for 30 min was conducted in vacuum (pressure $\leq$ 10$^{-3}$ Pa) [110]. The steel substrate reflections are at 43.62° and 74.70° [199].
reaches the quasi-binary TiN–TiB₂ tie line at \( T_a = 1100 \) °C. At this point, the XRD peaks reach the positions expected for unstrained bulk TiN and TiB₂ [199]. At \( T_a \geq 1000 \) °C also more TiN and TiB₂ reflections can be detected than compared to lower annealing temperatures. Thus, the large and broad exothermic reaction detected by DSC with a peak temperature near 1150 °C (Fig. 31b) is assigned to grain growth processes [23,24]. The reduction of the grain boundary fraction during grain growth and consequently the reduction of the energy associated with the grain boundaries causes a decrease of the total enthalpy of the material which can be detected by DSC. The XRD pattern recorded after the DSC measurement up to 1400 °C, shows well pronounced reflexes for TiN as well as TiB₂ at the positions of their unstrained bulk counterparts [199].

Energy dispersive X-ray analysis during scanning cross-sectional TEM analyses of films annealed at \( T_a \geq 900 \) °C reveal the presence of Cr and Fe interdiffusion. The diffusion zone extends to \( \sim 150 \) nm at \( T_a = 1100 \) °C.

SAED analyses (Fig. 33) of as-deposited and annealed layers are in good agreement with XRD results. The as-deposited film shows diffuse SAED rings due to a high volume fraction of disordered regions and 2–3 nm sized TiN and TiB₂ crystals with random orientation as discussed above. With increasing annealing temperature sharper diffraction rings are obtained indicating enhanced crystallization and ordering. For \( T_a = 900 \) °C, the position of the diffraction rings, while still overlapping, match those of bulk TiN and TiB₂. The SAED patterns of samples after annealing at 1000 and 1100 °C indicate TiN and TiB₂ grain growth as the diffraction rings continue to sharpen with \( T_a \). The (1 1 1) TiN and (1 0 0) TiB₂ diffraction rings appear as separate features after annealing at \( T_a = 1100 \) °C. Furthermore, the presence of small discrete spots along the diffraction rings is consistent with continued grain growth. However, there is still no evidence, even after 1100 °C anneals, for crystalline BN in agreement with the XRD results.

The cross-sectional HRTEM images in Fig. 34a–d show the nanostructural evolution following annealing at \( T_a = 700, 900, 1000, \) and 1100 °C. Comparing Fig. 34a

![Fig. 33. Development of the SAED diffraction rings of a TiN–TiB₂ film with annealing temperature \( T_a \). Isothermal annealing for 30 min was conducted in vacuum (pressure \( \leq 10^{-3} \) Pa). The standard diffraction rings for fcc TiN and hcp TiB₂ are added [110].](image-url)
(T_a = 700 °C) with Fig. 16 (as-deposited) reveals no significant difference in average grain size d consistent with the observations that XRD and SAED patterns remain approximately the same. However, at higher annealing temperatures, d increases from 2–3 nm for as-deposited layers to ~4, 7, and 9 nm at T_a = 900, 1000, and 1100 °C, respectively.

Average grain sizes d are also estimated from XRD results by fitting appropriate diffraction peaks with Cauchy functions [85] whose maxima are at bulk TiN and TiB_2 diffraction positions. The results yield average grain sizes of 2–3 nm for T_a ≤ 800 °C with d increasing at higher T_a in good agreement with HRTEM results as shown in Fig. 35a.

The nanostructural changes in the films during annealing lead to changes in mechanical properties as shown in Fig. 35b. The hardness H of the TiN–TiB_2 films increases from 36.6 ± 0.8 GPa in the as-deposited state to a maximum of 42.7 ± 2.1 GPa at T_a = 800 °C and than decreases at higher annealing temperatures. For T_a = 1000 °C,
$H = 38.3 \pm 0.7$ GPa. This behavior is in contrast to stoichiometric binary nitride and carbide layers, in which the hardness typically decreases during annealing due to relief of compressive stresses as discussed in Section 3.1.1.

The nanoindentation moduli $E$ of our films increases from $\sim 332$ GPa in the as-deposited state to $\sim 375$ GPa after annealing at $T_a \geq 900$ °C (see Fig. 35a). This increasing elastic modulus is due to a reduction of grain boundary fraction and extrinsic defects like pores and cracks [200]. The XRD and SAED results of the Ti–B–N layers indicate a decreasing fraction of disordered regions during annealing at $T_a \geq 700$ °C (see Figs. 32 and 33). Film-thickness $t_f$ measurements (see Fig. 35a) which are obtained from cross-sectional HRTEM investigations corroborate these conclusions. The constant grain size and chemical composition during annealing up to $\sim 800$ °C (see Fig. 35a) suggest that the reduction of $t_f$ from 2.83 to 2.51 µm (i.e., $\sim 10\%$ reduction in film thickness and volume fraction) is due to a reduction of the grain boundary fraction (i.e., formation of compact interfaces). Hence the thickness of the interfaces would be reduced from $\sim 0.7$ to $\sim 0.6$ nm. Calculations of nanocrystalline Fe yield higher elastic constants for materials with a grain boundary thickness of 0.5 nm than compared to 1.0 nm [200].

A schematic representation of the nanostructural modification during annealing of the TiN–TiB$_2$ film is given in Fig. 36 [110]. Upon annealing, the grain size increases and the volume fraction of the amorphous phase encapsulating TiN and TiB$_2$ crystals decreases. The hardness maximum is obtained after annealing the film at 800–900 °C, where $\sim 4$ nm sized TiN and TiB$_2$ crystals are separated by $\sim 0.5$ nm disordered region. Similar mechanisms could explain the hardness increase by 40–50% for Ti–B–N coatings after annealing at 400 °C as reported by Mollart et al. [201].

Recently, recrystallization and grain growth in nanocrystalline dual-phase TiN–TiB$_2$ films as a function of their chemical composition has been studied by a combination of XRD and DSC measurements using the approach described above [23,24]. The onset temperature for recrystallization increases from 1030 °C for TiN-rich coatings with a grain size of 6 nm to 1070 °C for TiB$_2$-rich coatings with a grain size of 2 nm (see Fig. 37a).
A slightly higher thermal stability has been found for dual-phase TiC–TiB$_2$ coatings [23]. For nanocomposite coatings consisting of crystalline TiN and amorphous BN, coarsening of the TiN phase has also been found to occur above 1000 °C [135]. Contrary, amorphous TiN films prepared by magnetron sputtering start to crystallize at 400 °C with defect-assisted diffusion as the rate-limiting step [202].

The huge difference in the onset temperatures for recrystallization and grain growth of TiN-based coatings in pure single-phase form and dual-phase coatings with a Si$_3$N$_4$, TiB$_2$, or BN phases added knowledge towards a stabilizing mechanism of grain and phase boundaries. Pure single-phase nanocrystalline coatings are thermodynamically unstable,
whereby a lower grain size corresponds to a higher driving force and, consequently, to a reduced thermal stability against grain growth due to the well-known Gibbs–Thomson effect. Gleiter [171] applied classical concepts of physical metallurgy for nanocrystalline materials. Thereby, grain growth may be prevented on the one hand by fine-dispersed inclusion of second phase particles acting as pinning sites for grain boundaries and on the other hand by slowing down the growth kinetics by reducing the driving force (i.e., the grain boundary energy) or the grain boundary mobility. This can be achieved by segregation of insoluble elements to the interfaces between nanocrystals, either already during growth (as it is the case for nanocomposite coatings) or during post-deposition annealing. As an example for the latter, in the Fe–P system the grain boundary energy decreases linearly with the logarithm of P content and should approach zero, where the kinetics of P segregation determines the kinetics of grain coarsening [203].

To determine the activation energy $E_a$ of a microstructural reaction by means of the Kissinger equation [204] different heating rates $B$ during the DSC experiment are necessary:

$$\ln \left( \frac{B}{T^2} \right) = - \frac{E_a}{k_B \cdot T} + \text{constant.}$$

In the Kissinger equation $T$ represents a specific temperature such as the onset temperature $T_o$ or the peak temperature $T_p$ of the reaction peak. By using $T_o$ or $T_p$ values for the different heating rates, plots of $\ln(B \cdot T^{-2})$ versus $T^{-1}$ allow the determination of $E_a$ [205].

Fig. 37b shows Kissinger plots for four different chemical compositions of nanocrystalline dual-phase TiN–TiB$_2$ coatings obtained by using different heating rates during DSC experiments [24]. The apparent activation energies $E_a$ for grain growth were determined from the slope of these lines, yielding values of 7.9, 6.9, 6.4, and 4.4 eV for the compositions TiB$_{0.55}$N$_{0.95}$, TiB$_{0.8}$N$_{0.85}$, TiB$_{1.0}$N$_{0.75}$, and TiB$_{1.25}$N$_{0.7}$, respectively. For comparison, $E_a$ for vacancy diffusion in TiN films is 2.09 eV [206], whereas $E_a$ for self-diffusion of N in TiN is 2.1 eV [152]. $E_a$ for diffusion of metal atoms in the nitride phase is in general higher than that of N atoms [26]. However, only for pure metals there exists consistency between self-diffusion and the activation energy for the movement of grain boundaries [17]. For TiN–TiB$_2$ coatings, it is reasonable that grain size, grain, and phase boundary structure and the different bulk and interface diffusivities of Ti, B, and N atoms, play major roles in determining the activation energy for grain growth. $E_a$ decreases with increasing TiB$_2$ content, which is related to the decreasing grain size (see Fig. 37a) due to the higher amount of stored interfacial energy. Likewise, the heat released during grain growth increases [24,154]. However, $E_a$ shows significantly different values only for coatings having a predominant TiN or TiB$_2$ phase which is assumed to be related to the competing influence of interfacial energy and phase structure as well as chemistry on the driving force for grain coarsening.

### 3.3. Age hardening

The concept of age hardening [207,208], i.e., the formation of supersaturated phases, their decomposition by post-deposition annealing treatments and the associated increase in hardness has successfully been adopted to thin film materials. In Sections 3.3.2.1 and 3.3.2.2 the decomposition process via the formation of intermediate phases is demon-
strated for Ti$_{0.34}$Al$_{0.66}$N [19] and Ti(B)N [163], respectively. As-deposited films show a dense columnar microstructure of a supersaturated NaCl structured phase based on fcc TiN. For Ti$_{0.34}$Al$_{0.66}$N, the metallic part (Ti) of the TiN lattice is substituted by Al [45,58,209], whereas for Ti(B)N mainly the non-metallic part (N) of the TiN lattice is substituted by B [163]. For comparison, TiB$_{2.4}$, TiN–TiB$_2$, TiN–Si$_3$N$_4$, and similar nano-composite thin film materials that also attract a large interest [10–13,36,51,110,117,134,135,210] do not undergo such decomposition during annealing, assuming that they are ‘equilibrium phase materials’ consisting of ideal nanograins already in the as-deposited state. These systems, however, may have potential for age hardening if deposited in a ‘quenched’ metastable solid solution state.

3.3.1. Formation of supersaturated phases

Due to limited atomic assembly kinetics during the deposition process (e.g., by using a low substrate temperature), defects, supersaturated, and metastable phases can easily be obtained in thin films. This results in a broad range of unusual properties, which cannot be achieved via other processing methods. During vapor deposition insufficient activation energy for kinetics leads to material conditions far from equilibrium [25]. Assisting ion fluxes or energetic condensing species can enable more equilibrium-like structures achieved by additional activated atomic assembly. Therefore, strong relationships between deposition conditions and the growth mechanisms/kinetics exist. As a result, structures and properties can be effectively influenced by the deposition techniques.

By using the Gibbs free energy as a function of composition of any mixture $\Delta G_{\text{mix}}$ it can be calculated whether the mixture is stable, metastable or unstable. The Gibbs free energy of mixing $\Delta G_{\text{mix}}$ is defined by

$$\Delta G_{\text{mix}} = \Delta H_{\text{mix}} - T \Delta S_{\text{mix}},$$

where $\Delta H_{\text{mix}}$ is the enthalpy of mixing, $T$ the absolute temperature and $\Delta S_{\text{mix}}$ is the entropy of mixing. $\Delta S_{\text{mix}}$ is always positive and $\Delta H_{\text{mix}}$ can be negative or positive or even zero if an ideal system is prevailing [16–18]. For exothermic solutions (i.e., $\Delta H_{\text{mix}} < 0$) mixing results in a free energy decrease at all temperatures. The mixture is stable for all compositions since any phase separation raises the total free energy [16–18].

For $\Delta H_{\text{mix}} > 0$, the situation is more complicated. It depends on the temperature whether the mixture is stable (over all compositions) or not. This case is illustrated in Fig. 38a and b showing the Gibbs free energy as a function of composition, $x$, for fcc Ti$_{1-x}$Al$_x$N [211] for 4100 and 1000 °C, respectively. At higher temperatures, where for all compositions $T \Delta S_{\text{mix}} > \Delta H_{\text{mix}}$, the free energy of mixing has a positive curvature, i.e., a positive second derivative of the Gibbs free energy $\partial^2 \Delta G_{\text{mix}}/\partial x^2$ (see Fig. 38a).

At lower temperatures, where $T \Delta S_{\text{mix}}$ is smaller than $\Delta H_{\text{mix}}$, the free energy of mixture $\Delta G_{\text{mix}}$ develops a negative curvature in the middle as indicated in Fig. 38b. Here the mixture is not stable for all compositions. At any composition between $x_1$ and $x_2$ the total free energy is reduced if the mixture splits into two phases with the compositions $x_1$ and $x_2$, hence the original mixture is supersaturated. The local points of the coexisting compositions with $x_1$ and $x_2$ defined by the common tangent construction are known as binodal. Increasing temperature results in a decreasing difference between $T \Delta S_{\text{mix}}$ and $\Delta H_{\text{mix}}$. Thus, the coexisting compositions $x_1$ and $x_2$ come closer. Consequently, a miscibility gap is formed in a phase diagram for $T \Delta S_{\text{mix}} < \Delta H_{\text{mix}}$ [16–18,207,208,211].
A material under such supersaturated conditions seeks and achieves equilibrium by separating the excess atoms into another phase of different composition and/or structure if diffusivity is assisted by time and temperature.

3.3.2. Decomposition of supersaturated phases

The total free energy of a system is lowered if the supersaturated phase splits into its stable constituents. The reason for decomposition by means of diffusional processes is always to reduce the Gibbs free energy of the system.

To understand the decomposition path of supersaturated phases, which is essential for the development of age hardenable materials, it is necessary to know the equilibrium conditions, the occurring phase transitions and their kinetics. A system, which is not in equilibrium, need not directly adopt its minimum free energy state when it undergoes a phase transformation. The necessary rearrangement of atoms requires time and therefore the systems can be arrested in intermediate non-equilibrium states [16–18,207,208].

Fig. 39 shows a detail of the free energy curve for $T \Delta S_{\text{mix}} < \Delta H_{\text{mix}}$ (Fig. 38b) indicating the important distinction that the curvature of free energy may be positive or negative. If $\frac{d^2 \Delta G_{\text{mix}}}{dx^2}$ is negative, the mixture is intrinsically unstable and will rapidly separate into two phases by the so-called spinodal decomposition [132]. This occurs because fluctuations in a homogeneous mixture actually lower the free energy and will thus be amplified (for a composition $x_{\mu}$, phase separation into compositions close to $x$ results in a lowering of free energy from $\Delta G_{\text{mix},\mu}$ to $\Delta G_{\text{mix},\nu}$, see Fig. 39) [16–18,207,208]. However, if the second derivative of $\Delta G_{\text{mix}}$ is positive, the mixture is metastable and separation will take place by nucleation and growth. At a composition $c_a$ a small compositional fluctuation increases the free energy from $\Delta G_{\text{mix},a}$ to $\Delta G_{\text{mix},a'}$. Thus, the system is locally stable with respect to such small fluctuations, even though it is globally unstable with respect to separation into the two coexisting phases $x_1$ and $x_2$ (Fig. 38b). There is an energy barrier, which needs to be surmounted in order to achieve the global energy minimum associated with phase separation. Therefore, this composition is metastable.

The limit of local stability with respect to small fluctuations is clearly defined by the condition $c^2 \Delta G_{\text{mix}}/c^2 = 0$. The locus of these points (the inflection point) is known as the spinodal. Since the free energy of mixture changes with temperature and concentra-
tion, the binodal and spinodal meet at the critical or consolute temperature and composition [16–18,207,208,211].

Depending on the composition of the supersaturated phase its separation can be implemented through the spinodal mechanism (if $\frac{\partial^2 \Delta G_{\text{mix}}}{\partial x^2} < 0$) or the nucleation and growth mechanism (if $\frac{\partial^2 \Delta G_{\text{mix}}}{\partial x^2} > 0$). The major difference between these two mechanisms is the direction in which atoms move in a concentration gradient. Uphill diffusion in a binary system is the dynamic hallmark of spinodal decomposition which has been verified in a number of alloy systems [16,132,207,208,212]. The increased defect density in vapor deposited thin films may enhance their kinetics of decomposition. For detailed investigations of the decomposition processes and to investigate the microstructural modifications during annealing, DSC measurements may be conducted [19,23,24,110].

Spinodal decomposition [132] is known from studies of the bulk. A bulk material example relevant to hard coatings is the decomposition observed in TiMoCN [213]. However, little is known for thin films. Here, both surface and in-depth decomposition can take place during synthetic growth by deposition processes that give effectively quenched alloys from the supercooling induced by vapor condensation on a substrate kept at low temperature as in PVD. The shape and properties of the so-formed components might also be controlled by anisotropic conditions. For example, surface-initiated spinodal decomposition of Ti$_{1-x}$Al$_x$N takes place during growth, to form a rod-like nanostructure of fcc TiN and fcc AlN domains with a period of 2–3 nm (see Fig. 40) [56,214].

Other thin film systems exhibiting spinodal decomposition include the group III–V systems and quaternary InGaAsSb epitaxial layers deposited deep into the miscibility gap [215]. The occurrence of spinodal decomposition during annealing of as-deposited ternary ceramic coating materials, in particular the Ti–Zr–N system, has been proposed by Knötek and Barimani [42] and Andrievski et al. [43,216,217] and later by other researchers [40,44–46]. However, a critical review gives that the corresponding microstructural changes have not been observed under the conditions applied. Spinodal decomposition with a connected increase in film hardness between 700 and 1000 °C has, however, recently been confirmed in metastable Ti$_{0.34}$Al$_{0.66}$N thin films [19,58], which will be discussed in detail in the following section.
3.3.2.1. Hard coatings with phases supersaturated with metallic atoms. The best-known example for hard coatings consisting of supersaturated phases where the metallic part is substituted can be found in the system TiN–AlN. Here, Al has been added to TiN, originally to enhance the oxidation resistance due to the formation of a protective Al-rich oxide layer at the film surface. Low-temperature ion-assisted (20–50 eV) PVD growth processes are useful to stabilize supersaturated phases with compositions within the miscibility gap in a metastable state. PVD phase field arguments for a range of pseudo-binary nitride systems including TiN–AlN were presented early by Holleck. A stability regime in terms of temperature and AlN content for fcc phase solid solution in the Ti$_{1-x}$Al$_x$N system at temperatures normal for PVD processes (300–700 °C) was predicted. Recently, modified metastable and still single-phase fcc Ti$_{1-x}$Al$_x$N coatings with small Cr and Y additions have been suggested with further improvement in oxidation and corrosion resistance.

Compared to TiN, the incorporation of Al (and likewise Cr and Y) into the TiN lattice causes an increase of compressive stress due to hindered annihilation of point defects and thus increasing lattice distortion. Donohue et al. reported on a reduction of compressive stress from −6.8 GPa for Ti$_{0.43}$Al$_{0.52}$Cr$_{0.03}$Y$_{0.02}$N in the as-deposited condition to −2.3 GPa upon annealing at 900 °C and attributed this to either thermally activated plastic deformation or to recovery resulting from annihilation of stress-active defects produced during the coating process. Suh et al. found a gradual reduction of compressive stress in Ti$_{1-x}$Al$_x$N after annealing up to 600 °C. BSTM measurements of Ti$_{1-x}$Al$_x$N coatings with low Al contents (up to 8 at.%) deposited by PACVD at 500 °C yielded an onset temperature for recovery of ~620 °C (which is slightly lower compared to PACVD TiN coatings grown at the same substrate temperature). Likewise, the amount of stress relaxation after a BSTM cycle up to 700 °C was significantly higher compared to TiN. Both effects are related to the higher driving force for recovery in the case of the higher defect density of Ti$_{1-x}$Al$_x$N coatings in the as-deposited condition with respect to TiN. The activation energy $E_a$ for recovery of an arc-evaporated Ti$_{0.34}$Al$_{0.66}$N coating is ~3.4 eV as reported in [19].
As already mentioned, as-deposited Ti$_{1-x}$Al$_x$N coatings show a supersaturated solid solution. The maximum solubility limit within the NaCl structure strongly depends on the deposition conditions [209]. Ab initio calculations [222–224] showed that the transition from NaCl-structure to wurtzite-structure occurs for an AlN mole fraction ($x$) of $\sim 0.7$ [225,226]. XRD and TEM studies showed that arc-evaporated Ti$_{1-x}$Al$_x$N films are single-phased with NaCl-structure for $x \leq 0.66$ and mixed-phased with NaCl/wurtzite structure for $x = 0.74$ [227]. For magnetron sputtering, the composition limit for single-phase fcc Ti$_{1-x}$Al$_x$N films seems to be similar, however, there is a tendency for higher aluminum concentrations in the films compared to the respective target [228].

The lattice parameter, $a$, of fcc Ti$_{1-x}$Al$_x$N was found to decrease with increasing Al content (Fig. 41) [226,229–231]. To investigate the decomposition path of supersaturated fcc Ti$_{1-x}$Al$_x$N films, TiN, Ti$_{0.75}$Al$_{0.25}$N, Ti$_{0.5}$Al$_{0.5}$N, and Ti$_{0.34}$Al$_{0.66}$N were deposited onto cemented carbide substrates in a high-vacuum arc-evaporation system and subjected to furnace annealing [58]. The depositions were made using Ti$_{1-x}$Al$_x$ targets in a 99.995% pure N$_2$ atmosphere. The substrates were negatively biased and kept at a temperature of 500 °C throughout the deposition. A constant cathode evaporation-current power supply was employed to give a deposition rate of 3 μm h$^{-1}$. The films were grown to a thickness of $\sim 3$ μm. Isothermal annealing of all samples was carried out in a hot-wall quartz-tube furnace with a 0.40 m long constant temperature zone ($\pm 5$ °C). The annealing experiments were performed in flowing Ar at atmospheric pressure for a duration of 120 min.

During annealing, exothermic reactions are seen in DSC curves (Fig. 42) [19]. Fig. 42a shows a comparison of the DSC measurements obtained for the investigated Ti$_{1-x}$Al$_x$N coatings. For TiN, just one small exothermic reaction at $\sim 750$ °C can be detected by DSC. XRD and TEM investigations prior to and after this exothermic reaction indicated recovery processes of deposition-induced lattice point defects [58]. Such structural relaxations explain the hardness decrease at temperatures above the growth temperature (see Fig. 22) [15,58]. The DSC measurements for Ti$_{0.75}$Al$_{0.25}$N, Ti$_{0.5}$Al$_{0.5}$N, and Ti$_{0.34}$Al$_{0.66}$N show the presence of several pronounced exothermic reactions. A closer look to the individual measurements clearly shows the appearance of at least four reactions for

![Figure 41. Lattice parameter $a$ of fcc Ti$_{1-x}$Al$_x$N films as a function of the AlN mole fraction ($x$) [19,226].](image-url)
Ti$_{0.75}$Al$_{0.25}$N, Ti$_{0.5}$Al$_{0.5}$N, and Ti$_{0.34}$Al$_{0.66}$N labeled as A, B, and C for recovery including lattice defect annihilation with residual stress relaxation and decomposition of the supersaturated Ti$_1$C$_0$Al$_x$N phase to form cubic Ti-rich and Al-rich domains, and D for hcp AlN (wurtzite structure) formation, recrystallization and grain growth. Initial results using small-angle neutron scattering (SANS) on a Ti$_{0.50}$Al$_{0.50}$N film after annealing at 860°C (i.e., slightly above the temperature of the DSC peak B shown in Fig. 42) indicate an average domain size of 1.2 nm [233].

Clarification of the origin of the DSC reactions is obtained by XRD investigations of films annealed slightly beyond the corresponding peak temperature (Fig. 43). XRD analysis showed that initially the Ti$_{0.34}$Al$_{0.66}$N films were single-phase NaCl-structured. Comparison of XRD patterns of the as-deposited condition and after annealing at $T_a = 760$ °C shows no significant difference in peak position and peak width. Consequently, in addition to recovery processes, which would result in a decrease of the peak widths as already
discussed in Section 3.1.1 and shown for TiN and CrN (see Fig. 18), also decomposition induced effects occurred which give rise to a still broad XRD reflex [19]. Thus, the exothermic reaction at $\sim 720 \, ^\circ\text{C}$ detected by DSC (Fig. 42d) is due to recovery and the onset of decomposition processes of the supersaturated $\text{Ti}_{0.34}\text{Al}_{0.66}$N phase [19,58].

After $T_a = 890 \, ^\circ\text{C}$ an increase of the right-hand shoulder (close to fcc AlN) of the (200) peak was found. The XRD peak can no longer be described by one symmetric function, indicating the formation of AlN domains (Fig. 43). The coherency strains cause a shift of the XRD matrix-peak to the position of fcc AlN and vice versa [19]. After annealing at $T_a = 1030 \, ^\circ\text{C}$ the films are composed of three different phases: fcc $\text{Ti}_{1-x}\text{Al}_x$N (matrix), fcc TiN, and fcc AlN. The AlN domains which were most probably formed by spinodal decomposition [18,132] (at $T_a \sim 890 \, ^\circ\text{C}$) coarsen, resulting in a hardness decrease, which is retarded by the formation of TiN domains at $T_a = 1030 \, ^\circ\text{C}$ (see Figs. 42d and 43). After annealing at $T_a = 1400 \, ^\circ\text{C}$ the decomposition process is completed and the fcc AlN phase transformed into its stable hcp structure. XRD indicates that the $\text{Ti}_{0.34}\text{Al}_{0.66}$N films now are composed of fcc TiN and hcp AlN phases (Fig. 43). In addition, the phase transformations in this system were supplied by further experimental results [232] and ab initio studies [19,225,226,234].

The cross-sectional TEM investigation of as-deposited $\text{Ti}_{0.34}\text{Al}_{0.66}$N films shown in Fig. 44a revealed a dense and columnar microstructure with a high defect density and overlapping strain fields from ion-bombardment-induced lattice defects corresponding to a compressive residual stress of 2–3 GPa. With respect to the spinodal decomposition initialized at 900 $^\circ\text{C}$ observed by XRD, the micrographs of the annealed sample revealed a structure similar to the as-deposited condition except for column boundaries appearing more clearly defined and showing a reduced contrast from lattice-defect-induced strain fields (Fig. 44b). Annealing at 1100 $^\circ\text{C}$ resulted in phase separation of the metastable fcc $\text{Ti}_{1-x}\text{Al}_x$N phase into hcp AlN precipitates in an fcc $\text{Ti}_{1-x}\text{Al}_x$N matrix. Original
column boundaries were also dissolved at this temperature and a fine-textured structure consisting of subgrains of diameter in the range of 50–100 nm evolved in which spinodal decomposition progressed (Fig. 44c). After annealing at 1250 °C grains of both hexagonal and cubic phase were found to coarsen (Fig. 44d).

The corresponding age hardening effect can be seen in Fig. 45. Hardening does not occur for TiN [19,58,92] which instead softens at $T_a \geq 400$ °C due to stress recovery by annihilation of deposition process induced defects and recrystallization [15,26,92]. For comparison, the compressive stress in the Ti$_{1-x}$Al$_x$N films decreased only from $\sim$2.5 to $\sim$1 GPa during annealing [227]. In Fig. 45 the hardness of Ti$_{0.34}$Al$_{0.66}$N as a function of
$T_a$ is directly compared to the DSC results. A hardness increase between 600 and 950 °C occurs, indicating age hardening effects due to the mentioned decomposition of the supersaturated Ti$_{1-x}$Al$_x$N phase forming fcc AlN domains. The hardness decrease at $T_a > 950$ °C, due to the growth of these domains, is retarded by the formation of fcc TiN domains as indicated in Fig. 45. The reaction at $\sim 1210$ °C, which is due to recrystallization effects including the transformation of fcc AlN into its stable hcp form [19, 58], causes a further decrease in hardness (Fig. 45).

The observations made by XRD, TEM, and DSC provide evidence for spinodal decomposition operating in the bulk of Ti$_{1-x}$Al$_x$N thin films at temperatures starting at 800–900 °C [19]. For comparison, the surface-initiated spinodal decomposition in the Ti–Al–N system discussed above (cf. Fig. 40) took place during growth at $T_s \geq 540$ °C [56]. The much lower $T_a$ for that process is explained by the difference in activation energy for surface and bulk diffusion.
For the determination of the activation energy of the individual exothermic reactions, Ti$_{0.34}$Al$_{0.66}$N was investigated by DSC using different heating rates of 20, 25, 32, 40, and 50 K min$^{-1}$ (Fig. 46). The results clearly show that the reactions become more pronounced with increasing heating rate and as expected they shift to higher temperatures. Fig. 47 shows the Kissinger plots for the individual reactions A, C, and D (Fig. 42d). As the reaction B is overlapped by the neighboring reactions A and C a clear determination of the peak temperature and, thus, to establish a Kissinger plot was not possible.

The activation energies for the reactions A and C are determined to be 3.4 eV. Recovery and spinodal decomposition are diffusion-controlled within one phase, thus explaining the similar values. For the transformation of NaCl to wurtzite structure of AlN and subsequent recrystallization (reaction D), additional nucleation is needed resulting in $E_a = 3.6$ eV. Considering the similarity of these values with that reported for surface diffusion of Ti on TiN of 3.5 eV [235], it is tempting to speculate that spinodal decomposition in this case is facilitated by defect-assisted diffusion processes.

The formation of domains as a consequence of the decomposition process is schematically illustrated in Fig. 48a. If a domain is formed with a lattice parameter, $a_B$, different from the matrix, $a_a$, both domain and matrix must be strained by equal and opposite forces for full coherency. Therefore, the hardness increases as additional stress is required to propagate dislocations through these domains [236]. Adequate mechanisms occur for Ti$_{0.34}$Al$_{0.66}$N. The findings for Ti$_{0.34}$Al$_{0.66}$N point to a coherency of the formed AlN and TiN domains. Verification is obtained by HRTEM showing strained cubic domains. After annealing at 1100 °C, the coherency strain between fcc TiN and AlN domains became relaxed by the introduction of misfit dislocations as shown in Fig. 48b [19].

Effectively, the Ti$_{1-x}$Al$_x$N coatings harden as a nanocomposite in which dislocation activity ceases for a given small grain size [10,21,36,237–241]. The material adapts in strength to the annealing temperature conditions and can thus be considered to be functional for its application. The reasons for this recently revealed behavior lie in the coherence strain from the 4.2% lattice mismatch between the phases ($a_{fcc\text{TiN}} = 0.424$ nm,

![Graph](image-url)

Fig. 47. Kissinger plots of peak temperatures A, C, and D for Ti$_{0.34}$Al$_{0.66}$N. A denotes recovery processes, C refers decomposition of the supersaturated Ti$_{1-x}$Al$_x$N phase, and D stands for recrystallization including the fcc hcp phase transformation of AlN [195].
The lattice parameter of the matrix and the newly formed domain are indicated by $a_a$ and $a_b$, respectively. High-angle annular dark-field (HAADF) image of the AlN tile showing the {001} projection of the AlN lattice with dissociated {110} misfit dislocations due to relaxation of coherency strains (isothermal annealing at 1100 °C was conducted in a flow of nitrogen for 120 min) [19].

Figure 48. (a) Schematic drawing showing the origin of coherency strains. The lattice parameter of the matrix and the newly formed domain are indicated by $a_a$ and $a_b$, respectively. (b) HRTEM micrograph showing the [001] projection of an fcc Ti$_{0.34}$Al$_{0.67}$N lattice with dissociated {110} misfit dislocations due to relaxation of coherency strains (isothermal annealing at 1100 °C was conducted in a flow of nitrogen for 120 min) [19].

A note for consideration is that the phase transformation of fcc AlN into hcp AlN, which formed from metastable fcc Ti$_{1-x}$Al$_x$N during tools service at elevated temperatures, can be detrimental to the coating cohesion and adhesion [227]. This is due to the 20% expansion in molar volume associated with the fcc to hcp structure transformation of AlN. It is proposed that this mechanism provides an explanation for the observed drop in hardness for TiN/AlN multilayer films with longer periodicities, e.g., 16 nm, containing a mixture of cubic and hexagonal structure AlN phases during annealing between 800 and 1100 °C [251]. The cubic AlN-based nanolayers, however, generally exhibit higher hardness values than coatings with hcp AlN structures [178,252] and are stable during high-temperature annealing up to 1000 °C [178,251].
3.3.2.2. Hard coatings with phases supersaturated with non-metallic atoms. A model system to demonstrate precipitation hardening at elevated temperatures in thin films having phases supersaturated in non-metallic atoms is Ti–B–N if the B-content is <18 at.% [253]. Here we give an example of a Ti(B)N film deposited using PACVD. In the as-deposited state, the Ti(B)N film contains 43.4 at.% Ti, 17.4 at.% B, 36.0 at.% N, and 3.2 at.% Cl (which is due to the utilization of TiCl₄ and BCl₃ pre-cursors in the PACVD process [254]). The film is close to the quasi-binary TiN–TiB₂ tie line of the ternary Ti–B–N phase diagram [254]. Experimental (i.e., EPMA and XRD) and ab initio results [163] show that the Ti(B)N film is composed of a metastable NaCl-structured Ti(B)N phase, where B substitutes mainly for N and excess B is located at grain boundaries.

Fig. 49 indicates that the XRD film-peak positions shift towards the standard values of stoichiometric TiN during annealing at \( T_a < 800 \, ^\circ C \). The steel substrate reflections are at 43.62° and 74.70° [199]. After annealing at \( T_a = 800 \, ^\circ C \), where the XRD film peak positions are close to that of stoichiometric TiN an additional small XRD reflection at \( \sim 34.34^\circ \) can be detected. This peak which matches \((10\overline{1}0)\) TiB₂ gradually increases (and hence the TiB₂ fraction) as the film is annealed at higher temperatures. The XRD results show that an additional hcp TiB₂ reflex can be detected for \( T_a \geq 800 \, ^\circ C \). After annealing the film at \( T_a \geq 1100 \, ^\circ C \), the preferred \((10\overline{1}0)\) orientation of the precipitated TiB₂ phase and the preferred \((200)\) orientation of the film-matrix change into a more random orientation (Fig. 49). This loss in preferred orientation of precipitate and film-matrix indicates the onset of recrystallization in the Ti(B)N film. After annealing at \( T_a = 1400 \, ^\circ C \) the Ti(B)N film is fully recrystallized with fcc TiN and hcp TiB₂ phases in random orientation (Fig. 49).

Fig. 50 shows that the lattice parameter \( a \) is \( \sim 0.4315 \) nm for \( T_a \leq 500 \, ^\circ C \) and gradually decreases from \( \sim 0.4315 \) to \( \sim 0.4245 \) nm \( (a_0 = 0.4240 \) nm for bulk TiN) for annealing between 500 and 800 °C. The B/Ti film ratio is unaffected by the annealing treatments at \( T_a \leq 1100 \, ^\circ C \). Consequently, the decreasing lattice parameter suggests that the B-solute...
concentration of the film-matrix decreases from the as-deposited condition to zero. In as-deposited condition the B-solute and excess B at grain boundaries is $17.4 \text{ at.}\%$ [104].

The full width at half maximum intensity $\Gamma_{200}$ of the (2 0 0) reflection remains constant at $1.12 \pm 0.02^\circ$ as a function of annealing temperature for $T_a \leq 500 \, ^\circ\text{C}$ (Fig. 50), suggesting that there are no significant changes in either nanostructure or local lattice microstrain [85]. As soon as $T_a$ is higher than the growth temperature, the B-solute concentration of the film-matrix decreases and $\Gamma_{200}$ increases to $1.20$ (and is above $1.05$ unless $T_a$ exceeds $900 \, ^\circ\text{C}$, see Fig. 50). Annealing the film between $1000$ and $1200 \, ^\circ\text{C}$ causes a gradual decrease of $\Gamma_{200}$ from $1.05$ to $0.39^\circ$, respectively, which is connected with a distinct formation of hcp TiB$_2$ precipitates (compare Figs. 49 and 50). The increase in $\Gamma_{200}$ for $T_a$ where the lattice parameter starts to decrease (and hence the decomposition starts) as well as the behavior that the decrease of $\Gamma_{200}$ starts at $900 \, ^\circ\text{C}$, indicate the formation of intermediate metastable B-rich phases which contribute to $\Gamma_{200}$ in the temperature range of $500$–$800 \, ^\circ\text{C}$. For example, orthorhombic TiB has the strongest XRD line at $42.19^\circ$, which is very close to the (2 0 0) reflex of the Ti(B)N film. Furthermore, to minimize the activation energy the precipitation process to form hcp TiB$_2$ might start with the formation of coherent and/or semi-coherent domains, according to observations made for Ti$_{1-x}$Al$_x$N films discussed above.

In the as-deposited state the Ti(B)N film hardness is at a very high value of $42 \pm 2 \, \text{GPa}$ due to the small grain size, growth-defects, solid solution hardening, and excess B, which is also located at grain boundaries and thereby increases their cohesive strength [117,255]. The hardness for bulk stoichiometric TiN is $\sim 20 \, \text{GPa}$. The Ti(B)N film hardness remains at $\sim 43 \, \text{GPa}$ for $T_a \leq 900 \, ^\circ\text{C}$, although the growth-defect density and the concentration of solute B-atoms in the film-matrix decrease (and hence the lattice parameter $a$) if the annealing temperature exceeds the deposition temperature (see Fig. 50). Consequently, the decreasing growth-defect and solid solution hardening effects are compensated by the onset of age hardening, as simultaneously B-rich precipitates are formed. For higher annealing temperatures, the film hardness decreases as the fine-dispersed particles grow. If the precipitation process is completed and the TiB$_2$ particles coarsen, the particle

![Fig. 50. Hardness $H$, lattice parameter $a$, and full width at half maximum intensity $\Gamma_{200}$ of the (2 0 0) XRD reflection of the Ti(B)N film-matrix as a function of $T_a$. Isothermal annealing was conducted in vacuum (pressure $\leq 10^{-3} \, \text{Pa}$) for 30 min. For comparison, the lattice parameter $a_0$ of TiN is indicated by an arrow [163].](image-url)
strengthening and hence film-hardness decrease (see Fig. 50 for $T_a \geq 1000$ °C). The hardness further decreases when the whole film starts to recrystallize and grain growth occurs (see Figs. 49 and 50 for $T_a \geq 1000$ °C).

4. Self-lubrication of hard coatings

Tribological properties of hard thin films for tooling applications are the topic of many research works [61,98,256–262]. Especially, Ti$_{1-x}$Al$_x$N is well known for its excellent wear and oxidation resistance enabling improved machining processes like high speed and dry cutting [1,98,209,263–265]. Due to their supersaturated metastable phase, films within the Ti–Al–N system show the previously described age hardening effects, which effectively increase their hardness and wear resistance at higher temperatures [19]. However, the presently used transition metal nitride coatings lack lubricating properties and show, consequently, relatively high friction coefficients in the range between 0.4 and 1.0 against steel [266].

Low friction coefficients can effectively reduce the contact temperatures during dry sliding and thus have a huge potential to reduce the thermal load for coating and substrate material. A lot of effort is being made to develop so-called low-friction protective layers to act as solid lubricants with the goal to replace the commonly used expensive and hazardous coolant lubricants. Especially in high speed and dry cutting applications, low-friction and lubricating mechanisms of the coating itself are required in addition to excellent mechanical properties. Here, the intrinsic solid lubricant films based on diamond-like carbon (DLC) or MoS$_2$ that have been successfully transferred to industry, will be reviewed in Section 4.1. However, since the solid lubricant used has to be chosen according to the contact conditions of the given application, several new approaches are taken including self-adaptive lubrication mechanisms of microstructurally engineered hard coatings as will be summarized in Section 4.2.

4.1. Self-lubrication by solid lubricants

Generally, no single lubricant phase can provide reasonably low and continuous friction coefficients over broad test or application loads, i.e., over a wide temperature range and changing environmental conditions. Each lubricant operates properly under specific test conditions, but there is no universal lubricant [267]. In the following, the state-of-art in solid lubricant coatings is summarized. Presently, mainly solid lubricants based on diamond-like carbon and MoS$_2$ are industrially applied. Although hexagonal boron nitride (h-BN) in bulk form provides lubricious properties, up to now there is no indication for low-friction properties of h-BN containing films deposited by plasma-assisted vapor deposition [268].

4.1.1. Diamond-like carbon based coatings

DLC coatings constitute a class of artificial metastable materials with a wide range of composition, properties, and performance. DLC is the term commonly accepted for carbon-based coatings which have mechanical, optical, electrical, and chemical properties similar to natural diamond, but which do not have a dominant crystalline lattice structure. The H content determines film structure, which is amorphous and consists of a mixture of sp$^3$ (diamond) and sp$^2$ (graphite) carbon [269]. The most common type of DLC is the
amorphous hydrogenated a-C:H film with only moderate sp³ content and a relatively high H content up to ~50 at.%. Thus, the DLC notation can be correspondingly inappropriate, however, now used more as a trade name for various C-based coating products. The non-hydrogenated films can be divided into amorphous C showing a high amount of sp² hybridization and tetragonal amorphous C (ta-C) with a predominance of sp³ hybridization of up to 85% [270].

DLC films can be deposited by a huge variety of PVD and CVD techniques. Today, the main processes used for deposition of non-hydrogenated DLC films are magnetron sputtering, cathodic arc-evaporation and pulsed laser deposition. The most popular method to grow a-C:H films is plasma-enhanced CVD. There, the sp³ and H content are strongly controlled by the average impact energy, which is used to control film structure between polymer-like C films at low energies and DLC type at intermediate energies [267,270].

The transformation of DLC to its more stable structure, graphite, may start at temperatures of ~250 °C. The graphite formation takes place rapidly at 400 °C and proceeds by loss of hydrogen, which is responsible for stabilizing the sp³ configuration. This graphitization process might be a key factor to describe and explain the tribological properties of DLC films [271]. For application as tribological coating, DLC films have to adhere well to the substrate material, which implies that the adhesive forces have to overcome the often high internal stresses within DLC films [272].

Friction coefficients of DLC films, which are typically ranging from 0.01 to more than 0.5, are depending on the nature of the film and the conditions used for friction testing [269,272]. The friction behavior of DLC is controlled by an interfacial transfer layer formed during sliding. This transfer layer of low shear strength (sp²-type) is formed from the top layer of the DLC coating and is responsible for the low friction coefficients [269]. In ambient humid air at a relative humidity of 20–60%, the friction generally ranges between 0.05 and 0.30 with wear rates strongly depending on the nature of the films. Hydrogen-free (ta-C) films generally exhibit lower friction coefficients (<0.15) when compared with hydrogenated a-C:H coatings. In ta-C films, friction easily causes a local shear-induced graphitization (sp³ → sp²) in the contact zone localized at so-called hot-spots with high contact flash temperatures, resulting in reduced friction due to formation of a thin graphitized tribolayer [273]. The increase of the friction coefficient with increasing humidity can be explained by a condensed water layer at asperity contacts, which has a ‘cooling effect’ at the small hot-spots, so that necessary temperatures may not be attained and, therefore, the graphitization process is expected to be suppressed. At very high humidity, no graphitization was observed [273], resulting in relatively high friction coefficients. The trend of increasing friction with humidity is attributed by to the saturation of micro-contacts with water molecules, causing increased bonding at the interfaces [274]. Previous investigations have also shown that surface oxidation of DLC films in humid air forming C=O bonds causes an increase in friction [275].

In inert environments such as dry nitrogen and vacuum, friction coefficients can reach either ultra-low values (less than 0.01) [276,277] or high values (above 0.50). This binary friction behavior under ultra-high vacuum conditions was explained by the different hydrogen content of the coating [269], where the hydrogen atoms determine the contact bonding between the DLC film and the counterface [273]. Higher hydrogen contents thus result in low adhesion to other counterpart materials yielding low friction coefficients.

The friction characteristics of DLC films can be summarized as following [269]: (i) DLC films display a behavior similar to diamond in UHV, but not in humid atmosphere.
(ii) The friction coefficient of hydrogenated DLC (a-C:H) is low in humid nitrogen and oxygen, extremely low in dry nitrogen and UHV, and very high in dry oxygen atmosphere. (iii) The loss of hydrogen through annealing at high temperatures causes a marked increase in the friction coefficient in UHV, and (iv) both friction and wear can be affected by the transfer layer which may form during friction.

To overcome the disadvantageous humidity sensitivity of DLC, metal atoms in the range of 10–40% can be added to the DLC film to obtain less humidity degradation compared to pure DLC films [273,278]. Addition of metals like tungsten [279], titanium [280,281], chromium [282], or silicon [273] to DLC films forming nanocrystalline metal carbides embedded in the DLC matrix thus improve their performance in humid and dry conditions.

High potential for use of DLC coatings is found in tribological applications, in particular in non-lubricated, in dry, and in vacuum environments, e.g., for components used in mechanical engineering or automotive and engine applications. Generally, humid or oxygen containing environment are a limiting factor of usage of these coatings. The lack of high-temperature resistance and thickness limitation due to high intrinsic stresses of these coatings restrict their use in cutting industry.

4.1.2. Transition metal dichalcogenide coatings

In the class of transition metal dichalcogenides, molybdenum disulphide (MoS 2) is the most popular lamellar solid lubricant. The hexagonal structure of MoS 2 (and, analogously, WS 2) exhibits individual sheets of transition metal and sulphur atoms. There, atoms within a layer are bonded by strong covalent bonds, while the interaction between layers is a weak van der Waals interaction. Due to these weak bonds between the layers, sliding occurs easily between sulphur layers [283,284]. In the following, we will mainly focus on MoS 2, however, similar results are valid for WS 2 [285,286].

The microstructure and morphology of MoS 2 films depend on the deposition conditions, where sputtering is used most often. The H 2O residual pressure in the deposition chamber seems to be the most important parameter affecting microstructure and morphology. High levels of H 2O lead to amorphous structures, whereas at lower partial pressures of H 2O crystalline MoS 2 coatings can be deposited [284].

MoS 2 shows good friction performance in vacuum and under dry running conditions, but degrades quickly in moist and oxidizing environments. Friction coefficients of 0.002–0.05 can be observed in vacuum, dry or inert atmospheres, which increase rapidly to 0.2 in humid air. The mechanisms yielding these ultra-low friction coefficients of MoS 2 are [283]: (i) the formation of an MoS 2 transfer film on the frictional counterface, (ii) friction-induced orientation of the (0 0 0 1) basal planes of the MoS 2 grains in the contact region, parallel to the sliding direction, and (iii) the absence of contaminants.

It is well known that humidity or oxygen has an inverse effect on the friction properties of MoS 2. This is attributed to the formation of oxides in the contact zone [252]. Other investigations have shown that the presence of oxygen and water vapor leads to physical processes, such as oxidation and water condensation, which determine friction properties [283].

Increased friction and wear properties of MoS 2 coatings can be obtained by the addition of titanium atoms (up to ~20 at.%) to the MoS 2 structure. In this so-called MoST™ structure, titanium is thought to be in solid solution within the MoS 2. HRTEM investigations [287] show that the Ti atoms actually also form nanocrystallites in a matrix of MoS 2 in the form of bundles of curved MoS 2 basal planes. The distortion due to the titanium
atoms is responsible for an increase in hardness from \( \sim 4 \) GPa (pure MoS\(_2\)) up to 10–20 GPa, yielding improved wear resistance [288]. Another advantage of titanium-doped MoS\(_2\) compared to conventional MoS\(_2\) is a reduced sensitivity of the coating to water vapor. Therefore, the applicability of these coatings is extended from dry conditions up to 50% humidity [267,289,290].

The environmental temperature has a strong influence on the friction behavior of MoS\(_2\) coatings. The films with open-ended basal planes as in turbostratic arrangement tend to oxidize at elevated temperatures, and thus lose their lubricity. MoS\(_2\) can provide lubrication up to 400–500 °C, where serious oxidation occurs [267,284]. MoS\(_2\) based coatings have found application in non-lubricated conditions for metal-forming and cutting operations, usually deposited as a top layer on hard underlayers (such as TiN, Ti(C,N), CrN) [290] or as a nanocomposite, where MoS\(_2\) forms a nanodispersion in a TiN matrix [260]. For fullerene-like MoS\(_2\) films where the basal plane ends close on each other [286,291] the reactivity to the atmosphere, however, is passivated, but the coating strength is much reduced by the relative absence of cross-linkage of the fullerene-like domains.

4.1.3. Future approaches

To combine the superior friction properties of DLC in humid air and WS\(_2\) in dry conditions with the high wear resistance of transition metal carbides, a nanocomposite concept for tribological applications has been proposed by Voevodin et al. [61]. The nanocomposite coating within the W–C–S system consisted of 1–2 nm WC and 5–10 nm WS\(_2\) grains embedded in a DLC matrix. The WC/DLC/WS\(_2\) nanocomposite exhibited chameleon-like self-adaptation to operations that occur in aerospace systems, providing friction reduction in both dry and humid environments. Recently, this concept was expanded to high temperatures, where DLC was combined with Au, providing high-temperature lubrication, and embedded in an yttria-stabilized zirconia (YSZ) matrix [292].

Other lines of developments are to add F to DLC for reduced friction [293,294] or to use nitrogen to obtain carbon nitride CN\(_x\) \((0 < x \leq 0.3)\) compounds, that are different from the DLC coating family in that they contain essentially no sp\(^3\)-coordinated carbon (see Ref. [295] for a recent review). In particular, the fullerene-like CN\(_x\) \((0.1 < x < 0.3)\) coatings generally exhibit a low work of indentation usually connected to high-hardness materials. Yet, the CN\(_x\) shows a low-to-moderate resistance to penetration depending on deposition conditions. Since the deformation energy is predominantly stored elastically, the material possesses an extremely resilient character. This new class of materials consist of sp\(^2\)-coordinated basal planes that are buckled from the incorporation of pentagons and cross-linked at sp\(^3\)-hybridized C sites, both of which arise due to structural incorporation of nitrogen. Carbon nitride thus deforms elastically due to a bending of the structural units. The orientation, radius of curvature of the basal planes and the degree of cross-linking in between them is shown to define the structure and properties of the material. Due to the unique deformation behavior the hardness requires special care to assess, but can be very high for films with a large degree of cross-linking. Similar to DLC, CN\(_x\) coatings exhibit a large variation in tribological properties. Generally, however, the CN\(_x\) exhibits a coefficient of friction that is similar to non-hydrogenated DLC. Friction tends to increase as the film nitrogen content is increased. This can be explained by a corresponding increase in surface energy from the extra electron supplied by the N substitution in the graphene sheets. This has impact for the choice of any lubricants in applications.
Low friction combined with a damage-tolerant structure has been proposed for the so-called $M_{n+1}AX_n$ phases [99,296–303]. These MAX phases form thermodynamically stable nanolaminates, which result in damage tolerance and low friction, as measured by lateral force microscopy for the basal planes of single-crystal Ti$_3$SiC$_2$ [304]. However, these low friction coefficients do not translate to polycrystalline samples, where values of 0.15–0.45 have been obtained in the initial stage of pin-on-disc testing against stainless steel, reaching finally a steady-state value of 0.8 [305]. We project that recently introduced thin film processing of MAX materials in phase-pure crystalline or nanocomposite [306] form will stimulate intensive research on their tribological functioning.

4.2. Frictional self-adaptation by oxide-formation

Intrinsic solid lubricants like DLC, MoS$_2$, and h-BN often begin to fail in their tribological effectiveness with increasing temperature, in humid atmosphere or due to oxidation [252,290,307,308]. Recently, several approaches have been suggested to improve the friction properties of hard coatings by self-adaptive mechanisms occurring during application, e.g., during dry cutting, by tailoring their oxidation behavior. Examples for these developments, which could be used to add functional self-adaptive features to the structural properties of hard coatings, are summarized below. All approaches are based on the assumption that the contact zone between, e.g., cutting tool and workpiece is exposed to oxygen, which is true for interrupted cutting, but not verified in other cases yet.

Boric acid formation on boron carbide as a potential mechanism for reaching ultra-low friction has been suggested by Erdemir et al. [198,309]. Boron carbide (B$_4$C) films are characterized by high hardness and stress resulting in excellent wear resistance, however, they show relatively high friction coefficients of 0.3–0.4. After annealing at 800 °C, reduced friction coefficients of 0.03–0.05 due to oxidation of B$_4$C can be obtained. The (now reached) low-friction mechanism is based on the reaction of the boric oxide (B$_2$O$_3$) with ambient humidity to form a thin boric acid (H$_3$BO$_4$) film. The low friction coefficients of boric acid is associated with its layered triclinic crystal structure [267,310]. The layers consist of closely packed and strongly bonded boron, oxygen, and hydrogen atoms, but the layers are widely separated and attracted by van der Waals forces only. During sliding, these atomic layers can align themselves parallel to the direction of relative motion and slide easily over one another [198]. The temperature sensitivity of boric acid, however, restricts the applications of low-friction B$_4$C coatings at elevated temperatures. Above ~170 °C, boric acid tends to decompose, thus losing its layered crystal structure and hence its lubricity [267]. The use of boric acid as a low-friction coating is also limited to reasonably humid conditions, precluding its use in dry or vacuum applications [198,309].

The unavoidable Cl impurities in TiN coatings deposited by PACVD using TiCl$_4$ as pre-cursor cause a significant decrease of the oxidation resistance compared to PVD coatings (i.e., without Cl) [145]. However, it has recently been shown that these Cl impurities have some concurrent beneficial effects on the coating constitution. On the one hand they exhibit grain refinement by continuous renucleation during growth [311], whereas on the other hand the formation of a thin rutile layer, in humid air, is stimulated at the topmost surface [312]. These rutile layers provide easy-shearing crystallographic lattice planes and thus a lubrication effect in ball-on-disc testing, reducing the friction coefficient against steel from 0.7 (without Cl impurities) to 0.15 for Cl contents above 3 at.% [313]. The low-friction effect is obtained by self-adaptation after a certain run-
ning-in distance in ball-on-disc testing, where Cl which segregated to the grain boundaries is released from the coating by abrasion and then became accommodated at the contact zone. The length of the running-in distance is either determined by the Cl-content itself [313], by the hardness of the counterpart determining coating abrasion and thus Cl release [314] as well as by the sliding contact conditions [312,313]. This Cl-induced low-friction effect has been demonstrated for several transition metal nitride coating systems, e.g., TiN [311–316,315,317] and nanocomposite Ti(B)N [254], where the Cl addition was provided by the PACVD process or by Cl ion implantation, respectively. Since one of the pre-conditions for this effect is the adsorbed water film on top of the coating [312], low friction is provided for temperatures below 60 °C [318]. However, the effect might be beneficial for cold working applications like deep drawing or metal forming [319,320].

Analogous phenomena happen to other lubricants that are a product of chemical reactions between coating and moisture of the ambient atmosphere, e.g., the above mentioned boric acid formation [198], limiting applicability at elevated temperature. Thus, a new concept of high-temperature lubrication was found in the use of lubricious oxide materials with easy moveable shear planes, also referred to as Magnéli phases [321]. These phases exhibit good thermal stability, high resistance against tribo-oxidation and low adhesion [322,323]. Especially, the oxides of W, Mo, V, and Ti form homologous series with planar faults according to the common Magnéli phase principles $\text{Me}_n\text{O}_{2n-1}$, $\text{Me}_n\text{O}_{3n-1}$ and $\text{Me}_n\text{O}_{3n-2}$ [323]. Such crystalline structures based on the rutile structure contain rutile-like chains of edge-sharing octahedrals, interrupted by shear planes every $n$th octahedron. Generally, these shear planes exhibit reduced binding strength. In comparison to common solid lubricants such as MoS$_2$, where every second layer offers crystallographic slip ability, Magnéli phases might exhibit less promising lubrication performance, because here only every $n$th layer has a crystallographic shear structure [323]. Solid lubricants based on these Magnéli phase oxides have not yet found wide use, because of the difficulty in achieving and maintaining the very narrow range of oxide stoichiometry [262,324], which is necessary for good lubricity [267].

Thermo- and/or tribo-oxidatively formed Magnéli oxide phases are ideal high temperature solid lubricants in combination with additional liquid self-lubrication due to melting of the formed oxide phases, with their usually low melting points [323]. The system V–N is known to form oxygen deficient Magnéli phases with planar lattice defects at elevated temperatures [155].

In [155] it is shown, using DSC, thermo-gravimetric analyses (TGA), BSTM, and dry sliding tests, that due to the formation of numerous low-melting, easy-shearable Magnéli phases the friction coefficient of VN films against austenitic stainless steel as well as alumina balls decreases remarkably with increasing the temperature from room temperature to 700 °C (see Fig. 51).

The results obtained by DSC and BSTM in air correlate well with respect to oxidation threshold temperatures (and hence the Magnéli oxide phases formation temperature) from 500 to 525 °C and melting temperatures of the formed V-oxides from 625 to 660 °C. The formed Magnéli oxide phases, which were identified by subsequent XRD measurements, belong to the $\text{V}_n\text{O}_{2n-1}$ and $\text{V}_n\text{O}_{3n-1}$ homologous series. No limitation of the films for tribological applications due to possible oxide phase evaporation within the tested temperature range was found. The friction coefficient of the films shows a remarkably low value during testing at 500 and 700 °C.
Tribometer testing against austenitic stainless steel balls (Fig. 51a) shows at room temperature, where no Magnéli oxide phases are present, significant differences of the frictional contact compared to testing against alumina. The relatively low hardness of the stainless steel balls compared to alumina allows considerable ball material transfer to the wear track [314]. Consequently, room temperature tests result in a highly scattered friction curve with a friction coefficient of \( \sim 1.35 \) after a short running-in period. The high scatter of the friction curve indicates several sliding friction mechanisms due to material transfer from the austenitic stainless steel ball [314]. However, from the investigations presented in [155] it can be concluded, that VN thin films offer self-lubricious abilities as the concept of solid/liquid self-lubrication at high temperatures yields very low friction coefficients during high-temperature dry-sliding against different counterpart materials (see Fig. 51a and b).

As VN is a potential candidate to add self-lubricious properties to existing hard thin films for advanced cutting applications, some effort has been made to develop TiAlN/VN superlattice films [32] where the beneficial properties of both types of layers are combined. TiAlN/VN has proven to be an excellent candidate in protecting machine wear parts and cutting tools due to its high wear resistance and low friction [98,257–265].

In [325] the formation of oxides of TiAlN/VN films having a superlattice period of \( \sim 3.5 \) nm and the oxide-effects on the tribological properties at elevated temperatures are reported. Main emphasis was laid on the formation and stability of \( V_2O_5 \) as it is known to transform to \( VO_2 \) via different V-oxides [326] which partly count as Magnéli phases. Furthermore, the effect of the low melting point \( V_2O_5 \) phase on tribological properties and constitution of the TiAlN/VN superlattice films during testing above the melting point of \( V_2O_5 \) was investigated.

Friction coefficients of the TiAlN/VN superlattice films in ambient atmosphere against an alumina ball using a normal load of 1 N are presented in Fig. 52 for three different
testing temperatures. At room temperature, the TiAlN/VN superlattice film [264] shows a friction coefficient which is relatively low compared to other hard films [98,257,265]. The value of ~0.55 exactly fits the measurements of Constable et al. [257] using the same tribological conditions except for the normal load which was 5 N. The 3-D observation of the wear track after a sliding distance of ~1 km showed just a few tiny radial grooves of maximum 0.16 μm in depth. Almost no transfer material from the alumina ball could be found. Also, the ball itself showed no remarkable abrasion after the ball-on-disk test.

Increasing the test temperature to 500 °C causes the friction coefficient to rise to 0.96. After the test, the depth of radial grooves of the wear track is below 0.17 μm and also adherent material (indicating the formation of oxide products due to tribo-oxidation) can be detected. As the testing temperature is further raised to 700 °C, the friction coefficient drops sharply from 0.96 to 0.18 (Fig. 52). This drop in friction is an indication of a sudden change in the controlling friction mechanism. After half of the sliding distance the friction coefficient increased nearly abruptly to ~0.55 and remained at this value. Since the coating was still intact after the sliding experiment at 700 °C, reasonable changes of the top layer seem to be responsible for this transition.

The low friction coefficient of TiAlN/VN superlattice films at room temperature has already been reported in the literature [98,257,265]. During sliding, a V₂O₅ containing oxide with lubricious properties is formed at asperity contacts. In the wear debris formed on the wear track, Constable et al. [257] found firm evidence for the presence of V₂O₅ using Raman microscopy. The excellent tribological properties of TiAlN/VN superlattice films can thus be attributed to the formation of these particular oxides.

Usually, as the temperature is increased the viscosity of the oxide formed on top of the TiAlN/VN superlattice film should decrease. Consequently, it should become easily shearable signified by a lower friction coefficient. Although this was observed for a pure or doped V₂O₅ film [327], TiAlN/VN shows a remarkable increase of its friction coefficient at 500 °C when compared to room temperature. Also, the investigation of the wear track indicated a higher amount of adherent material, although comparable grooves were
formed on the disk. This indicates that the dominant wear mode changes from mild abrasion to adhesion [311]. Obviously, the V$_2$O$_5$ formed at the TiAlN/VN surface [257] plays no significant role in the used testing configuration up to 500 °C. More exciting is the remarkable friction drop between 500 and 700 °C (see Fig. 52) indicating to be a melting effect of a V$_2$O$_5$ containing oxide on top of the film. Bulk V$_2$O$_5$ melts at ~685 °C and its boiling point is 1750 °C [323]. DSC measurements conducted on TiAlN/VN superlattice films also indicate a melting effect of a previously formed oxide as the exothermic oxidation–reaction is overlapped by an endothermic (melting) reaction (see Fig. 53).

The first heating run (i.e., using the virgin film sample) in an oxidizing environment shows a pronounced exothermic reaction indicating oxidation of TiAlN/VN. The onset of oxidation is ~450 °C, however, rapid oxidation occurs for temperatures above 550 °C. This is in agreement with results of the ball-on-disk test at 500 °C, where after testing no considerable oxide formation was found by scanning electron microscopy (SEM), XRD and 3-D profilometry.

The oxidation peak in Fig. 53 shows a distinct cut-off in the temperature range between 630 and 750 °C. This is attributed to a competition between an exothermic (i.e., oxidation) and an endothermic (i.e., melting) reaction. Verification is obtained by a second and third measurement of the same sample, which was now already oxidized. Consequently, only the endothermic melting reaction of the previously formed oxide can be detected by DSC. The melting point was determined to be ~635 °C. Evidence for this is also given by an exothermic crystallization peak in the cooling segment (see the cooling segment for the third run in Fig. 53). The temperature delay between melting and crystallization results from the high heating and cooling rate of 50 K min$^{-1}$. Another endothermic reaction occurs at ~1400 °C. At this high temperature the V$_2$O$_5$ containing oxide evaporates, confirmed by an observed sample-weight loss. The lower melting and boiling point compared to bulk V$_2$O$_5$ [323] may be due to a lower-melting eutectic formed with the other film elements [257]. It should also be considered that the oxidation product of nm sized layers may have a lower melting point than bulk V$_2$O$_5$.

![Fig. 53. DSC heating and cooling curves for a TiAlN/VN superlattice film in an argon/oxygen atmosphere with a maximum temperature of 1450 °C [325].](image)
The endothermic melting reaction during heating the oxidized TiAlN/VN film diminishes with each following run (and hence with increasing the exposure-time) even if the maximum temperature is below the boiling point. During a fourth run up to 1200 °C almost no exothermic or endothermic reaction could be observed. Consequently, V₂O₅ transforms into lower-oxidized V, showing higher melting points [325,327]. Formation of a liquid phase and its loss with annealing time explains the low friction coefficients at 700 °C and the observed increase with exposure-time (see Fig. 52) [325]. The observed steady-state friction coefficient of ~0.55 after this transformation is, however, extremely low for a hard thin film at 700 °C [328]. Evidence for a liquid phase on top of the TiAlN/VN superlattice film during a tribo-test at 700 °C is given by subsequent cross-sectional SEM investigations (Fig. 54).

Fig. 54a shows a smooth surface of the wear track as expected if it had been liquefied. A more detailed view presented in Fig. 54b indicates that melting occurs just for the surface-near region of the TiAlN/VN superlattice film. The results indicate that the formation of a liquid phase during exposure at elevated temperatures (i.e., above ~650 °C) can easily destroy the superlattice structure. Nevertheless, the concept may be advantageous for lubrication at intermediate to high temperatures where other solid lubricants begin to fail in their effectiveness due to environmental degradation, providing functional behavior of hard coatings. Furthermore, to benefit (in terms of friction) from the liquid phase formed due to melting of V₂O₅, the vanadium containing hard phase should be incorporated in form of nanoparticles into a hard film-matrix. There, the amount of Magnéli phase oxides and the possible formation of liquid phases on top of the film can be controlled by the
amount of V incorporated in the film-matrix \[329\]. Thereby, disintegration of the film structure by uncontrolled formation of \(V_2O_5\) containing film-oxides and liquid phases can be avoided.

5. Summary and outlook

This review has presented the state-of-the-art of microstructure–property relationships known for hard coatings in mechanical and tribological applications. It is significant that microstructural design can be provided both during deposition and post-deposition annealing stages, e.g., during operation at elevated temperature. The term microstructure is used as a commonly accepted concept for the materials structural engineering, although this review deals also with thin films fabricated by design concepts being nanoscale in nature in order to provide enhanced functional or even multifunctional properties.

Generally, the hardness of a material is determined by its resistance to bond distortion as well as dislocation formation and motion, which themselves depend on the amount of obstacles provided by the material. Hindering of dislocation movement can be achieved by a high density of point and line defects generated during plasma-assisted growth at relatively low temperatures, internal boundaries like grain and column boundaries, second phase particles, and solutes. Consequently, there is a direct relation between hardness and structural defects.

Microstructural defects like point defects and dislocations easily anneal out during thermal treatments at temperatures above the growth temperature. Thus, compressive micro- and macrostresses generated by non-equilibrium structural defects of high-density decrease by atomic and dislocation rearrangements into lower energy configurations. These effects are responsible for the decrease in hardness of binary carbide and nitride films observed during annealing at temperatures above the growth temperature, as their growth-defect induced hardening decreases and no additional hardening effect can be built up.

For materials with a microstructural arrangement of immiscible phases in the nanoscale range, their hardness is mainly determined by the created nanostructure, as grown-in point defects easily anneal during deposition due to the short diffusion-paths to the nearest sink (grain-, phase-, column-boundaries, in general the interfaces). Consequently, the annealing influence on hardness is mainly determined by a concomitant nanostructural modification. In materials with a grain size of 2–3 nm the volume fraction of the interfaces can be \(\sim50\%\), thus their constitution with respect to their cohesive strength is of vital importance in determining the mechanical properties. Nanostructures in thin films can be generated during growth by plasma-assisted vapor deposition or during post-deposition annealing treatments if the films contain supersaturated metastable phases. Such phases can readily be formed by deposition at low temperatures due to limited atomic assembly kinetics.

Softening and loss of performance of wear-resistant nanostructured coatings when exposed to elevated temperatures occur via diffusion-controlled mechanisms like stress relaxation by recovery, recrystallization and subsequent grain coarsening, interdiffusion of the individual layers or phases in nanolaminate or nanocomposite coatings, or oxidation. Diffusion measurements in ceramic materials with very high melting points are, however, difficult to perform. Furthermore, many of these materials are largely non-stoichiometric and contain impurities or growth-induced defects, depending on the synthesis process. From this perspective, even TiN represents a quite complex system,
not mentioning carbonitrides or metal alloy nitrides, e.g., Ti(C,N) and Ti$_{1-x}$Al$_x$N, as the most simple examples, which are widely used in the coating industry.

The stress relaxation rate in coatings during thermal annealing is a strong function of the as-deposited material stress level which is influenced by the deposition temperature and bias voltage in PVD or PACVD processing. The final stress state of the film is due to short-range momentum transfer processes in the ion collision cascade during deposition and thermally activated point defect transport as well as agglomeration. For contemporary PVD processing, films are typically in a compressive stress state that effectively hardens the material. However, essentially nothing is known about the nature of point defects or defect clusters that form in coatings synthesized by ion-assisted deposition techniques. This is potentially limiting the full exploitation of nanostructured coatings as residual stress engineering, phase selection, or grain size design is always of considerable concern. Stress recovery during deposition is generally more pronounced than during post-deposition annealing at the same temperature due to the larger degrees of freedom for ad-atom diffusion and defect annihilation near a free surface subject to growth. The kinetics of stress recovery has only been described for few but convincing cases and materials. Typically, there is a range of defects with different activation energies for effective recovery. For the compressive stress relaxation in nitrides, a range of apparent activation energy values are found below that for self-diffusion in bulk. In particular, residual stress engineering with a retained thermal stability is an important field for future research in coating technology.

Hard thin films possess a high potential to achieve superior properties which even improve during high-temperature applications due to diffusion-driven microstructural rearrangements. In nanocrystalline materials especially the constitution of interfaces is of vital importance. Films with supersaturated phases have large implications for the exploration of wear-resistant films as several pseudo-binary nitride or carbide systems exhibit large miscibility gaps. They can be synthesized by state-of-the-art vapor deposition processes to yield effectively quenched supersaturated solid solution coatings ready for age hardening. For optimized films, the nanostructure and hence the interfaces, either generated during growth or post-deposition annealing treatments are responsible for high or extremely high film hardness even after annealing at $\sim$1000 °C.

The revelation that thin films can be hardened through spinodal decomposition or by nucleation and growth of precipitates has vast technological importance for the design of next generation’s of wear-resistant coatings by advanced surface engineering. Decomposition processes are expected to occur in several of the ternary, quaternary, and multi-nary transition metal nitride-based systems as they exhibit miscibility gaps and can be processed by state-of-the-art vapor deposition processes.

Especially in high-speed cutting, where the commonly used expensive and hazardous coolant lubricants begin to fail, low-friction and lubricating mechanisms of the coating itself are required in addition to excellent mechanical properties. Whereas intrinsic solid lubricant films based on diamond-like carbon (DLC) or MoS$_2$ have been successfully transferred to industry, several new approaches utilizing self-adaptive lubrication mechanisms of microstructurally engineered hard coatings are discussed in this review. The functionality of such coatings is based on the formation of so-called low-friction protective layers during tribological contact, due to which the contact temperatures and thus the thermal load for coating and substrate can effectively be reduced. Future developments in the hard coating community may include hybrids of these compounds and further alloying (doping) as for fluorinated DLC and fullerene-like carbon nitride that are designed to
match the lubricants used in tribological contacts by surface energy (wettability) tuning or to match the stiffness of the machining tool or engineering component used for improved adhesion and fracture toughness of the system.

The results presented in this review clearly show that mechanical, thermal, and tribological properties of hard thin films can significantly be optimized by well designed microstructures. From industrial interest in increased cutting speed and feeding rate as well as for dry cutting there is consequently a demand for coatings exhibiting optimized microstructures, which can withstand high temperatures. To enable further developments in this area, diffusion-related phenomena controlling recovery, recrystallization, segregation, grain growth, interdiffusion, and oxidation have to be studied in much more detail. Such investigations would benefit from the generation of data for the diffusion in ceramics, which are presently lacking in most cases. In particular, for the exploitation of non-equilibrium state wear-resistant coatings, future investigations will give important input to predict the lifetime of a coating and to select cutting data for a given application. Furthermore, a detailed understanding of the microstructure–property relations is indispensable to support the massive and collective move in coating industry towards the use of quaternary and multinary nitrides to be followed possibly by carbides, borides, oxides, and oxy-nitrides.

In conclusion, from the scientist’s point of view, rewarding activities for research on advanced hard coatings may be focused on:

(i) Increasing the thermal stability and oxidation resistance by fine-tuning the composition and deposition condition.
(ii) Application of age hardening schemes by secondary phase transformations in metastable multinary systems. Two possibilities exist, self-organization during deposition or self-adaptation during service.
(iii) Increased understanding of the friction and wear mechanisms dominating in different applications in combination with fundamental mapping of existing and new coating materials, providing a suitable basis for knowledge-based material’s selection enabling multifunctional tribological properties.
(iv) Computations in materials science today can be used for both making precise predictions and offering explanations to phenomena in materials processing and properties, e.g., ab initio calculations applied to the phase stability of several binary, ternary, quaternary, and multinary systems.

Apart from these scientific tasks, future industry-related opportunities for improved performance of coated products will be based on:

(i) For micro- and nanostructurally optimized coatings, the importance of sample fixture rotation and variation of coating flux by angular position on microstructural development has to attract attention.
(ii) New combinations of individual layers and phases will satisfy the increasing demand for multifunctional coatings tailor-made for specific applications.

In addition, since arc-evaporation is the most widely applied tool coating technology, the single most important factor, which slows down the development of new arc-evaporated coatings, is actually related to difficulties finding high quality targets alloyed to your demand at a reasonable price. Also, simple things such that the PVD system suppliers all
develop their own arc sources, using different target dimensions, limits a faster development due to lack of standardization.

Acknowledgements

We acknowledge the use of the Center for Microanalysis of Materials, University of Illinois at Urbana—Champaign, which is partially supported by the US Department of Energy. Part of the work done at the University of Leoben was financially supported by the Technologie Impulse GmbH in the frame of the K-plus Competence Center program. The Erwin Schrödinger Program (project J2469-N02) of the Austrian Science Fund (FWF) is acknowledged by Paul H. Mayrhofer for financial support. Lars Hultman acknowledges support from the Swedish Research Council (VR) and the Swedish Foundation for Strategic Research (SSF) Strategic Research Center on Materials Science for Nanoscale Surface Engineering (MS²E).

References


Willmann H, Ernst W, Mayrhofer PH, Hultman L, Mitterer C. preliminary results.


Mayrhofer PH, Music D, Schneider JM. Influence of the Al distribution on the structural properties and phase stabilities of supersaturated Ti$_{1-x}$Al$_x$N. J Appl Phys, submitted for publication.


Kutschey K, Mayrhofer PH, Kathrein M, Poleik P, Tessadri R, Mitterer C. Structure, mechanical and tribological properties for sputtered Ti$_{1-x}$Al$_x$N coatings with 0.5 ≤ x ≤ 0.75. Surf Coat Technol 2005;200 (7):2338–65.


Face milling tests of low-C steel (42CrMo4) workpieces with a width of 80 mm were performed using 12 × 12 × 3 mm$^3$ pressed and sintered WC-13 wt% Co milling inserts (SEKN 1203 AFTN-M14), 125 mm diameter of milling cutter, 300 m min$^{-1}$ cutting speed, 0.2 mm rev$^{-1}$ feed, 2.5 mm cutting depth, and 4 min time of engagement for the cutting edges.


