In-situ monitoring towards monolayer precision for III-V semiconductor growth

Dr. Markus Pristovsek

Habilitationsschrift an der Fakultät II
Mathematik und Naturwissenschaften
der Technischen Universität Berlin

Lehrgebiet:
Experimentalphysik

Eröffnung des Verfahrens: 16.2.2011
Verleihung der Lehrbefähigung: 23.11.2011
Ausstellung der Urkunde:
Aushändigung der Urkunde:

Gutachter:
Prof. Dr. M. Kneissl
Prof. Dr. D. Aspnes (NC State University)
Prof. Dr. J. McGilp (Trinity College Dublin)
Prof. Dr. R. Goldhahn (Otto-von-Guericke-Universität Magdeburg)
Prof. Dr. M. Sugiyama (University of Tokyo)

Berlin, 2012

D 83
Abstract

The purpose of this review is to introduce the state of the art of in-situ monitoring of III-V compound semiconductor epitaxy especially to the operators of epitaxial equipment in industry and research. Therefore, the sections are focussed on typical tasks i.e. in-situ measurements of the growth rate, the composition, the doping level, and the roughness or topography of a layer. Information on the underlying surfaces physics is only given in context for the understanding and interpretation of the data. The measurement principles are introduced shortly, with two exceptions: multi-channel reflectance anisotropy spectroscopy and in-situ scanning tunneling microscopy, since for those no reviews are available.
1 Introduction

Electronic devices based on III-V semiconductors are nowadays mass produced mainly with two techniques, Molecular Beam Epitaxy (MBE) and Metal-Organic Vapour Phase Epitaxy (MOVPE). Historically, MBE was the technique of choice, since the high purity and low background doping was first achieved using ultra-high vacuum and ultra-pure elemental sources. MBE was developed with in-situ monitoring by Reflection High Energy Electron Diffraction (RHEED) from the beginning, and several books on RHEED have been published (e.g. [1, 2]).

Over the years, MOVPE has surpassed MBE in industrial use, since it can deliver similar or even better device performance, allows for simple regrowth and masking, and can be scaled up easier. Nowadays nearly all telecommunication lasers or hetero-bipolar-transistors (HBT) on InP are produced using MOVPE as well as all nitride based light emitting diodes. In the latter the quantum wells are thinner than 3 nm. Control of such thin layers and the surrounding layer structure is the most important task of in-situ monitoring.

MOVPE uses near atmospheric pressures above 20 hPa for its process, which makes operation and maintenance easy. Since at least 6 mm are between any analysis port and the sample, the mean free path of electrons is too short for electron based in-situ analysis with RHEED. But optical in-situ monitoring can be used. Therefore, the main focus of this review is optical monitoring of growth rate, composition, doping, and roughness. Other non-optical in-situ technique will be used for comparison. Since growth monitoring is the main target, the surface’s reconstruction and other surface processes will be covered only when relevant to in-situ monitoring.

Wavelengths of visible light are roughly 50 times larger than the dimensions of typical unit cells. Therefore optical in-situ monitoring with monolayer precisions requires special setups or analysis to be still sensitive to changes on the monolayer scale, despite large background signals. One of the methods to achieve such sensitivity is Reflectance Anisotropy Spectroscopy (RAS). While RAS is around since the mid 80ies, only recently multi-channel setups have been built. Section 2 will review RAS and multi-channel RAS detection.

Most III-V semiconductor devices are used in opto-electronic and high speed electronic devices. Since the thickness of most layers – even using binary materials – is critical to a device, most efforts of in-situ monitoring of growth has been devoted to determine the growth rate. Section 3 will review in details the current state of the art of growth rate measurements during hetero- and homo-epitaxy.

In compound semiconductor devices, band gap engineering using ternary or quaternary material is essential to realize quantum heterostructures. Hence additional control on composition of the layers is required. With data from growth rate measurements, the composition can be obtained by further analysis of the data. Composition can be also gained by RAS, which is important for ternary group V compounds like GaAsN or InAsP, and when there is no direct access via growth rate. Section 4 will give a short summary of the state of in-situ composition monitoring.

Most devices require control of the electron and hole concentrations in their layers too. Dopant (and hence carrier) concentrations are usually measured ex-situ, i.e. after growth. But often the layers in the devices are very thin, or are buried within the structure. Thus in-situ monitoring of doping is highly desired. RAS is sensitive to doping for III-V semiconductors with zincblende structure. Qualitative analysis is nowadays used in device production [3, 4] to detect deviations from the intended values. In section 5 the spectral measurements of doping by RAS will be summarized by a mathematical treatment to extract the actual electro-optic coefficients of GaAs and InP. This allows for quantitative measurements by fitting the signal shape to RAS spectra. I will also discuss the limits of the method and finally introduce in-
situ doping profiling using RAS in section 6, yielding simultaneously growth rate and carrier concentration – a convenient way to extract homo-epitaxial growth rates of thick layers by optical methods on MOVPE reactors.

As stated above, roughness can affect the function of quantum heterostructures, and the control of the topography is even more important with three dimensional structures like quantum dots. However, in-situ measurements of the topography are the least developed. For instance, a surface with a root mean square roughness of 5 nm would be still considered smooth for most lenses or mirrors, while such a roughness would inhibit the proper function of a quantum well. Even more, it is impossible to extract the topography via light scattering. Only tools like in-situ scanning tunneling microscopy (STM) can yield direct insight into the development of topography during growth. In-situ STM is reviewed in section 7 in comparison to optical roughness monitoring.
2 Multi-channel Reflectance Anisotropy Spectroscopy

Reflectance anisotropy spectroscopy (RAS, also termed RDS) probes the difference in the reflectivity \(r\) along two perpendicular eigenaxis\(^1\). For face centered cubic crystals by convention a higher reflectivity along \([1\bar{1}0]\) gives a positive contribution:

\[
\left( \frac{\Delta r}{r} \right)_{(hkl)} = 2 \frac{r_{[1\bar{1}0]} - r_{[xyz]}}{r_{[1\bar{1}0]} + r_{[xyz]}}
\]

with \((hkl), [xyz]\) and \([1\bar{1}0]\) being three perpendicular directions. For the most commonly used \((001)\) surfaces, the axis are then \([1\bar{1}0]\) and \([110]\).

In cubic crystals the bulk contributions to the RAS signal cancel themselves out due to symmetry reasons. In such crystals any anisotropy must originate from broken symmetry, either at the surface, an interface, or even in the bulk, as fig. 1 illustrates. On the surface bonds are cut and the surfaces must relax or even reconstruct to accommodate the remaining charges and dangling bonds (fig 1b). At an interface an anisotropy occurs because the upper bonds are different than the bonds to atoms in the layer below (fig 1c). Finally, a symmetry breaking in bulk can be caused by via internal strain, external stress, or an electric field (fig 1d). The latter is also the origin of the sensitivity to doping, which will be discussed in-depth in sections 5 and 6.

![Figure 1](image)

Figure 1: Sketch of isotropic bulk (a) and three different origins of the optical anisotropy (b-d) for a zincblende III-V compound in (001) orientation. In the bulk (a) the bonds along the \([1\bar{1}0]\) and \([110]\) directions are equal. Thus the reflectivities of those axes are equal and no anisotropy is caused. At a surface (b) or and interface (c) the symmetry is broken, because the perpendicular bonds upward and downwards differ due to the different environment above and below. The broken symmetry results in different reflectivity along \([1\bar{1}0]\) and \([110]\) and causes an anisotropy. Finally a vertical electrical field (or vertical strain) (d) also breaks the symmetry and causes an anisotropy.

The technologically interesting semiconductors like Si, Ge, GaAs, and InP as well as many metal are of the cubic group and thus are suited for RAS analysis especially in their most important \((001)\) orientation. If there is threefold symmetry like for the cubic \((111)\) or hexagonal \((0001)\) orientation, no RAS signal is expected. However, if the surface forms a twofold symmetry (like the Si \((111)\) single domain \((4\times1)\) reconstruction \([6]\)) the resulting anisotropy can be measured by RAS. If the bulk has twofold symmetry – as found on semipolar hexagonal

\(^1\)I will follow the convention of using capital \(R\) for the reflectance as measured on a detector and small \(r\) for the complex reflectivity. The connection between both is given by \(R = r^* \cdot r = |r|^2\).
GaN or with some superconductors – a strong bulk response is expected. In such cases RAS can be useful for bulk diagnosis.

The first measurements with RAS were made on rotating samples [7, 8]. While such a setup is very simple, the main problems are a strong sensitivity to stray light from the sample, polarisation dependent detectors, and the need to rotate the sample very homogeneously. Furthermore, only the difference in reflectance $\Delta R \approx 2\Re(\Delta r)$ can be measured. Therefore, setups using a photo-elastic modulator (PEM) were introduced [9, 10] and are still used today, even on rotating samples. This is due to another advantage, faster measurement times with a PEM. Typical PEMs operate in the 20-60 kHz regime. Thus measurement times of 10 ms or less are possible.

Fig. 2 shows a typical RAS setup using a PEM. Polarised light is shone near vertically on the surface. The polarisation direction is 45° to the main eigenaxis of the surface, i.e. along [110] and [\bar{1}10] on (001)-surfaces of materials with zincblende or diamond structure. Because of the surface anisotropy the reflected light is elliptically polarised. The reflected light reaches the PEM, which is an oscillating quartz crystal with its ordinary axis oriented parallel to the initial polarisation direction (fig. 2). Light travelling through the PEM will split in two polarisation components. The component with the same orientation (polarized along the ordinary axis) is not changed. The light polarized along the extraordinary axis (originating from the anisotropy) will be decelerated periodically. The PEM oscillation amplitude is now controlled in such a way, that the extraordinary polarisation receives a phase shift of nearly ±$\frac{\lambda}{2}$. Behind the PEM there is the analyser prism at 45° to the PEM main axis and the monochromator and detector.

When only linear polarized light enters the PEM along its ordinary axis, no modulation occurs. If the surface rotates the polarisation then there will be also light along the extraordinary axis. This light will receive a phase shift. The resulting anisotropic component will oscillate between −90° and 90° polarisation. Since one PEM oscillation actually results in reaching 45° polarisation of the analyser prism twice, the real part of the anisotropy is an intensity modula-
Figure 3: Calculated correction to the amplitude for measurements with a PEM operating at a fixed wavelength of 300 nm.

Figure 4: Comparison of measured amplitude (thick line) and calculated amplitude (dashed). The measured signal strongly derivates for retardations above $\pi$ from the second Bessel function.

The measured signal strongly derivates for retardations above $\pi$ from the second Bessel function. The measured signal strongly derivates for retardations above $\pi$ from the second Bessel function.

A more in-depth description of the different setups and the signal analysis can be found in the reviews of Zettler [11] and Weightman et al. [12].

With the setup shown in fig. 2 a data point can be obtained in 30-100 ms with a good signal to noise ratio. For a full spectrum the monochromator must be stepped, the PEM must be tuned, and probably also the lock-in amplifier sensitivity must be adjusted. Altogether a spectrum takes at least 30 s to several minutes, which is very long compared to typical growth times of a monolayer (less than few seconds). Thus a multichannel RAS setup is desired.

The key point to realize a multichannel system is the effect of a detuned PEM. When the PEM oscillation amplitude does not match the wavelength the signal is attenuated, since the $\pm\frac{\lambda}{2}$ phase shift at longer wavelengths is not reached and at shorter wavelengths $\pm\frac{\lambda}{2}$ is exceeded. This is the situation for a multichannel RAS system, since the PEM can be only set to match $\pm\frac{\lambda}{2}$ retardation at a single wavelength and the signal at all the other wavelengths must be
corrected. The degree of attenuation (of the most interesting real part) is given by the second Bessel function $J_2$. The maximum of the second Bessel function and thus the highest signal amplitude is gained for a retardation of 3.0521 ($J_2(3.0521) \approx 0.4862$). If the PEM deviates from this retardation, the signal is reduced by the factor $\frac{J_2(3.0521)}{J_2(3.0521 \lambda_{PEM})}$.

Fig. 3 shows the spectral dependence calculated for a PEM with retardation set for 300 nm. This results in the spectral range between 2 and 6 eV in an acceptable maximum loss of 50% of the signal.

Since at shorter wavelengths than the PEM set point the signal is attenuated even more (fig. 4), a compromise must be found between stronger attenuation at longer wavelength and slight deviations from the ideal signal at shorter wavelengths.

For a multichannel system the PEM is set to a fixed frequency at the UV end of the spectrum. Instead of a monochromator a spectrometer with a fast multichannel detector is used. Such a setup was first realized in 1996 by Kipp et al. [13]. While the non-modulated component is gained easily, the modulated component requires further analysis. Hence, true simultaneous multi-channel RAS detection was first achieved in 2001 using a fast detector array and a Fourier transformation based data analysis [14]. The main problem with Fourier analysis for the RAS signal is the insensitivity to the phase of the signal. If there is noise, the Fourier transformation can detect the wrong phase which results in a higher background noise especially for small signal amplitudes. Typical noise levels achieved by Fourier analysis are in the order of $10^{-3}$. With lockin amplifiers easily a noise level below $10^{-4}$ is reached, an improvement of a magnitude. Thus our multichannel setup uses 16 lockin amplifiers to demodulate the signal at 16 wavelengths.

Fig. 5 shows the slightly advanced version of our multichannel setup (first published in [E1]). Using a special Glan-air polarizer prism allows for simultaneous operation of scanning RAS and
multichannel RAS. The ordinary beam enters a scanning monochromator with high spectral and temporal resolution and the extraordinary side beam is fed into a multichannel system at 16 fixed wavelengths. (The multichannel extension is indicated by the dotted beam in fig. 5). During epitaxy the multichannel system takes a spectrum typically every 100-300 ms while for high-resolution static spectra the scanning RAS is used. A nice example of the capabilities has been the discovery of the doping induced Fabry-Pérot oscillation (see sec. 6).
3 Growth Rate

The growth rate is one of the most fundamental properties which need to be controlled during layer growth. Growth rate (and thus thickness) determines the operation of devices as well as growth mode transitions, like quantum dots. Consequently growth rate measurements have been done since the beginning of semiconductor epitaxy, using various methods.

3.1 Fabry-Pérot oscillations

The oldest optical in-situ tool is optical reflectance \( R \), which is quite simple to perform in a non- or semi-quantitative manner. Any RAS setup measures also the reflectance signal additionally to the anisotropy, to normalize the anisotropy signal, as described in the previous section.

Growth rate measurement by reflectance relies on the interference of light reflected at the top and the bottom interface of the growing layer. The resulting intensity is modulated with an oscillation corresponding to the optical thickness (layer thickness \( d \) multiplied by the refractive index \( n \)) which reads for perpendicular incidence:

\[
R \approx R_{01} + R_{12} \exp \left( - \frac{8 \pi d k_1}{\lambda} \right) + \left[ 2 \sqrt{R_{01} R_{12}} \sin \left( \frac{4 \pi d n_1}{\lambda} + \arctan \frac{A_+}{B_+} \right) \right] \exp \left( - \frac{4 \pi d k_1}{\lambda} \right) \tag{2}
\]

with

\[
R_{\mu \nu} = \frac{n_\nu^2 - 2 n_\mu n_\nu + n_\mu^2 + k_\mu^2 - 2 k_\mu k_\nu + k_\nu^2}{n_\mu^2 + 2 n_\mu n_\nu + n_\nu^2 + k_\mu^2 + 2 k_\mu k_\nu + k_\nu^2} = r_{\mu \nu} \cdot r^{\ast}_{\mu \nu}
\]

and

\[
r_{\mu \nu} = \frac{n_\mu + i k_\mu - n_\nu + i k_\nu}{n_\mu + i k_\mu - n_\nu + i k_\nu}
\]

and

\[
A_\pm = 2(\text{Re}(r_{12})\text{Re}(r_{01}) \pm \text{Im}(r_{12})\text{Im}(r_{01}))
\]

and

\[
B_\pm = 2(\pm \text{Re}(r_{12})\text{Im}(r_{01}) - \text{Im}(r_{12})\text{Re}(r_{01})).
\]

The resulting reflectance signal will approximate a damped oscillation with a thickness period given by \( d = \frac{\lambda}{2 n_1} \) during growth (fig. 6). By fitting a reflectance transient one obtains a growth rate in units of \( n_1 \). If at least a full oscillation (i.e. half a period) of the signal is observed, \( R_{01} \) and thus \( n_1 \) and \( k_1 \) can be fitted in a self consistent way. For typical values – like in fig. 6 – \( n_1=3, k_1=0.3 \), and \( \lambda=400 \text{ nm} \) this yields \( \approx 40 \text{ nm} \) as a lower limit. If the refractive index of substrate and growing layers are known, a 5% signal change might be already enough for a quantitative analysis of the growth rate. In this case the lower limit would be around 2.5 nm.

Unfortunately for many ternary materials no data for the refractive index at growth temperatures are available. Hence, reliable data from reflectance is usually obtained for layers with a thickness of ten nanometers and above. If there is a material combination which allows for the growth of thick layers – like Al\(_{1-x}\)Ga\(_x\)As on GaAs (as in fig. 7) or GaN on sapphire – then this method is very powerful and a more detailed analysis can be used to extract composition or surface temperature [E2].

When only qualitative data is needed, a change of the initial slope can be detected for thin layers even without knowledge of the refractive index. An example is the delayed incorporation

\(^2\) Using denominator\( \approx 1 \) and the identity \( A \cos(x) + B \sin(x) = \sqrt{A^2 + B^2} \sin(x + \arctan \frac{A}{B}) \) for \( A, B > 0 \).
Figure 6: Left side shows the ideal curve of a reflectance transient of a layer stack indicated on the right side. The envelope (dash) is nearly proportional to \( \exp(-\frac{4\pi k_1 d}{\lambda}) \) since the term \( R_{12} \exp(-\frac{8\pi k_1 d}{\lambda}) \) (dash-dot) is small.

Figure 7: Left side diagram shows that the growth rate of GaAs obtained from fitting the reflectance signal agrees nicely to those from RAS monolayer oscillation (see sec. 3.2) and from ellipsometric data. Right side shows an example of a reflectance transient during growth of AlAs on GaAs (0-400 s) and subsequent GaAs growth on this AlAs layer (600-1200 s) (from [15]).

of Sb during growth of GaAsSb. In this case X-ray diffraction indicated a delay in incorporation. A careful analysis of the reflectance transient during GaAsSb growth in fig. 8 showed that the slope changed after 157 s (4.7 nm of GaAs according to the XRD data). The change of slope is due to the different complex index of diffraction of GaAsSb compared to GaAs. Fig. 8 shows also the signal of the GaAsP strain compensating layers in the structures, which were 20 nm thick. For the initial GaAsP layer even a clear maximum is visible. Since the refractive index of the GaAs surface before growth is known at growth temperatures, a quantitative fit was possible for this GaAsP layer.

Using a technique that yields the absolute values of real and imaginary part of the reflected intensity (like ellipsometry) allows for a quantitative analysis of very thin layers, especially with
spectral resolved measurements. Even changes of the thickness within 0.2 nm can be detected, i.e. monolayer resolution is possible (e.g. [109, E4, E5, E6]).

To observe Fabry-Pérot oscillations, a hetero-interface is needed. If none is present, a thin layer of a nearly lattice matched material like AlGaAs on GaAs, InGaAs on InP, or AlInN on GaN could be inserted at the beginning of the growth procedure. If this is not possible, then the techniques described in following sections must be used.

### 3.2 Monolayer growth oscillations

A very accurate way to determine the growth rate even during homo-epitaxy without prior knowledge of the dielectric functions is the direct observation of the monolayer by monolayer growth in the two dimensional island growth mode. This growth mode was first observed by RHEED during MBE of GaAs by Joyce and coworkers [17].

A real substrate surface always possesses regularly spaced steps, either due to some intentional miscut, or by chance. At high enough temperatures all species on the surface (called subsequently monomers) can diffuse long enough such that they can reach a preferred site for incorporation, a kink or a step-edge. This growth mode is called step-flow growth, since the steps move over the surface but the surface roughness and step spacing remains constant.
Reducing the temperature and/or increasing the growth rate will reduce the diffusion length of monomers. When the diffusion length becomes shorter than the spacing between steps (i.e. the terrace width), the growth mode will change to the so-called 2D island growth mode. Growth in 2D island mode performs via four stages, diffusion of monomers, nucleation of islands, growth of islands, and finally coalescence of islands. Monte-Carlo simulations have show that the number of monomers at all times is very small. Therefore monomers are hardly observed experimentally [16] and only the growth and coalescence of the islands (where unfortunately no good analytical theory exists) is observed in the experiment.

In case of in-situ RHEED monitoring the intensity of diffraction spots during the growth of a monolayer anti-correlates with the number of island step edges scattering the electrons out of the Bragg directions [18]. At half monolayer coverage (i.e. maximum island density and thus maximum roughness) the scattering is high and consequently Bragg reflected intensity is low. After completion of a monolayer the surface is again relatively smooth and the reflected intensity increased nearly to its initial value. Since the diffusion length increases with higher surface temperatures, more and more atoms will reach a step-edge during diffusion, and fewer islands will nucleate on the terraces. Above a certain temperature all atoms are incorporated at the steps-edges: the growth mode has transitioned to step-flow mode. Since the surface is no longer changing periodically, intensity oscillations cannot be observed any more.

During MOVPE 2D island growth mode was observed using in-situ gracing incidence X-ray scattering (GIXS) on (001) GaAs [19, 20, 21] and InP [22]. In GIXS the intensity of the crystal truncation rods changes due to scattering by islands. From the scattering data the dimensions of islands can be estimated as a function of temperature [19], resulting in an energy for diffusion of around 2.7 eV for GaAs [20].

Oscillations of the RAS signal with monolayer periodicity were observed first in MOVPE of GaAs using tri-methyl gallium and AsH$_3$ at 500-550°C [23] (left side fig. 9). The growth rate determined from the RAS oscillations agrees nicely to the one determined from the growth of thick layer using reflectance and ellipsometry of thick layers (fig. 7).

At first, periodic roughening might be expected to cause also the RAS oscillations. Roughness induced RAS signals were first calculated by Aspnes [24]. For steps with height $d$ and dis-
tance $L$ (and $d \ll L$) the resulting anisotropy at perpendicular polarisation of an air/semiconductor interface is:

$$\text{Re} \left( \frac{\Delta r}{\langle r \rangle} \right) = \frac{2\pi d}{\lambda} \frac{2\varepsilon_2}{\varepsilon_1^2 + \varepsilon_2^2}$$

with

$$G = \frac{d}{2\pi L} \left[ \frac{1}{2} + \ln \pi + \ln \left( \frac{L}{d} \right) \right].$$

Slightly different geometries result in slightly different prefactors $G$. For instance a sinusoidal undulation with amplitude $d$ and period $L$ would give $G = \frac{\pi d}{2\pi L}$. For $d=0.285$ nm and $L=10$ nm the prefactor is 0.169 nm for steps and 0.080 nm for a sinusoidal shape. The calculated RAS signal at 2.65 eV for GaAs at 550°C ($\varepsilon = 19.56 + i14.36$) is then in the order of $10^{-4}$...$10^{-5}$. However, as fig. 10 shows, the measured oscillations are completely different. Not only the calculated anisotropies are nearly two magnitudes smaller, also the spectral shape is wrong, no matter what size and shape of islands are assumed [25].

Since the expected anisotropy from step induced roughness is too small, the appearance of oscillation in the RAS signal must originate from a change of the reconstruction near the step edges. Spectrally resolved measurements of the RAS signal during signal oscillation on GaAs in MOVPE (like in fig. 9) found oscillations between spectra with a minimum at 2 eV and at 2.6 eV. RAS spectra with a minimum around 2 eV are typical for gallium rich surfaces [26, 21, E8, E9]. Thus after half monolayer growth the surface is more gallium rich, and after completion of a monolayer the surface is again more arsenic rich. Quantitative fits at 510°C indicated an oscillation between 40% and 60% gallium rich ($n \times 6$) surface area fraction [25, 27].

The direct correlation between steps and gallium rich surface area was demonstrated on vicinal surface, where the miscut angle determines the step density. At higher miscut angle the surfaces became less arsenic rich due to enhanced arsenic desorption near step-edges [28, 29]. The increase of gallium rich areas near steps during growth is also linked to kinetics. Steps are preferred incorporation sites. After incorporation of a gallium atom, the steps became more gallium rich. If the gallium attaches faster than arsenic can diffuse to step-edges, gallium-rich areas will increase at the step-edges until those gallium-rich areas are large enough to catch enough arsenic. Since an isolated island provides additional steps, the oscillating number of
islands will cause oscillations in step density, which causes oscillations in gallium rich areas and thus oscillations of the RAS signal. Fig. 11 demonstrates this mechanism schematically.

Oscillation can be observed too (with a very small amplitude) in ellipsometry [30, 31, 32]. Their origin is similar, since the dielectric function of different surface domains changes not only the anisotropy but also the mean reflectance [31].

The indirect signal relation to steps has an important consequence. Increasing the temperature will increase the reaction rates, accelerating AsH$_3$ decomposition and surface diffusion while the growth rate is constant. Thus the gallium rich areas at the steps will shrink. Finally, when kinetics are fast enough, the surface will not change even near steps and the resulting RAS signal will be constant. In other words: when during growth and before growth RAS signal are similar, no oscillations can be monitored, even in case of 2D island growth mode.

In MOVPE this is the case during growth in the so-called phase I in the phase diagrams (for GaAs see [36, 28, 29, 37, E11, E10]). As fig. 12 shows phase I is observed during growth on GaAs above 630°C for 1 Pa tri-methyl gallium (TMGa) and above 550°C for 1 Pa tri-ethyl gallium (TEGa). Thus for temperatures below $\approx$630°C (with TMGa) respective $\approx$500°C (with TEGa) the surface reconstructions changes and would allows to observe oscillations, if the growth is in the 2D island mode. The maximum oscillation amplitude for the growth of GaAs (001) using TMGa and AsH$_3$ was observed around 505°C (fig. 13) but oscillations are visible between 450°C and 600°C, in the phase II region of the phases diagram (fig. 12).

In MBE there are two large regimes where the surface of arsenides differ at the step edges from the during growth surface. In the $\beta 2(2x4)$ regime the step edges are more $\alpha(2x4)$-like,
Figure 12: Phase diagram during growth using TMGa (left) and TEGa (right) on a GaAs (001) 2°AB substrate. The crossed diamonds in the diagrams mark transition points from transport limited to flow to kinetic limited growth (from [33, 34, 23, 35, E2] and own data). The lines mark the phase boundaries (dotted lines for TMGa boundaries in the right figure) (from [E10]).

Figure 13: RAS oscillation amplitude around 2.65 eV for growth of GaAs (001) with various miscuts using TMGa at different temperatures measured between 1994 and 1996 with two different RAS setups at our reactor. The thick line is a median of all points. (from [E7])

while during growth in the c(4x4) regime the steps are more (nx6) like. In both regimes RAS oscillations were observed [25, 38].

RAS oscillations can only occur in the 2D island growth mode. Then the structure of the step-edges determines if a material will show RAS oscillations or not. For instance, RAS oscillations are not observed for InP in MOVPE, even though GIXS experiments in MOVPE
confirmed 2D-island growth mode at low temperatures [22]. Furthermore, at those temperatures in MBE RAS oscillations were observed [5]. Apparently the higher V-III ratio in MOVPE prevents a different reconstruction near the step-edges.

All arsenide based semiconductors show RAS oscillations in MOVPE and MBE at or near typical growth temperatures. RAS oscillations were observed during growth of AlAs [40], strained InGaAs on GaAs up to 50% indium content [41, 32, 42], and even at 400°C one or two oscillations were observed during growth of InAs [43]. Also during growth of GaAsN with up to 7% nitrogen, RAS oscillations are present.

In case of rotating substrates, monolayer oscillation measurements are nearly impossible. A single RAS measurement can be only obtained after at least half a wafer rotation (more than half second). At typical rotation speeds of ≈1 Hz are too slow for typical growth rates of monolayer/second. Together with the higher noise level of such measurements, monolayer oscillations were found not suitable for growth rate determination in industrial multiwafer reactors.

### 3.3 Absolute optical thickness by interferometry

If one knows the absolute height of surface and backside, one would automatically have access at the growth rate even during homo-epitaxy. Furthermore via thermal expansion of the substrate, the temperature could be derived too. Finally, if the wafer strain results in a concave shape (i.e. the middle is pushed upwards) then the surface will have a different position in relation the susceptor and the strain can be accessed too.

![Figure 14: Schematic setup of an in-situ low-coherence tandem interferometer (from [44]).](image)

This is possible in the transparent region of a wafer using the so-called low-coherent tandem interferometry [44]. In this technique the signal of one interferometer is connected to the signal of the second interferometer which contains the sample. This allows for very precise measurements of the position of reflecting interfaces in the sample [45]. Moreover, at typical telecom wavelength many substrates like GaAs, InP, GaN, or Si are transparent, allowing to use mature components from telecom fibre optics. In the setup in fig. 14 the beam of a low-coherence source (here a superluminescent diode at 1550 nm) is first split and half is passed through delay lines (first interferometer). After merging again it passes a second beam splitter to detector and sample (second interferometer).
The optical thickness can be gained from the initial tuning of the delay lines which gives two signals if their length is close to the distances of an interface. A first signal will be for the surface and a second from the wafer backside (and probably additional signals from inner interfaces, if index contrast and thickness allows). The widths of those signals are given by the coherence length, the shorter the better. The absolute position is sensitive on a micron scale which is enough to measure substrate temperature by its expansion. By using the shift of the interference fringes (phase signal) a change of the surface position (and hence the growing layer’s thickness) with a resolution of about 2 nm could be measured [46]. Since the beam travels through air (or nitrogen/hydrogen/vacuum) until it reaches the first interface, the phase signal does not depend on the unknown index of refraction of the sample stack. Since the fibre optics itself is a small diaphragm, any partial reflective material will give a signal. Only for very rough surfaces (in the order of the wavelength) the phase signal cannot be observed.

Since also a signal from the substrate wafer backside is obtained, the thickness of the substrate can be converted directly into a temperature using an empirical calibration (or even absolute values, if the index of refraction and expansion coefficients are known at growth temperatures.) Since one can even detect a signal from the susceptor, the height of the substrate backside over the susceptor can be measured and used to calculate the wafer bowing (if the layer is compressively strained).

So far there has no application of this technique to a commercial multi-wafer machine. One challenge is gas foil rotation, which can result in a change of the absolute position of the interface during a rotation and requires fast delay line tuning. Alternatively, for motor driven susceptor only tuning at a certain position is possible. If fast enough tuning can be realized, even the surface position over the whole wafer could be monitored, which would yield directly the wafer curvature. To conclude, this interesting technique has a high potential, but further investigations needs to be undertaken.
4 Composition

Since III-V semiconductors require group V stabilisation at growth temperatures, the growth rate is determined by the supply of group III atoms. Thus, for a ternary material like InGaAs or AlGaAs, which only mixes group III elements, the increase of growth rate by adding additional In or Al to Ga can be converted to solid composition. In case of InGaAs it would be like:

$$x_{In} = \frac{g^T_{InGaAs}}{g^T_{GaAs}} - 1.$$  \hspace{1cm} (4)

This method is common for MBE growth using RHEED monolayer oscillations (described in the previous section). RAS monolayer oscillations can be used in the same way. An InGaAs/GaAs superlattice (SL) grown only via this calibration resulted in 30 periods each consisting of 5 monolayers In$_{0.11}$Ga$_{0.89}$As and 10 monolayers of GaAs, which was verified by X-ray diffraction \[41\]. Using related ellipsometry monolayer oscillations measuring the In content up to 37\% was achieved \[32\]. When using monolayer oscillation, the composition is known after growing about 5 ML, thus with near monolayer resolution.

An increase of growth rate is not expected to be observed on group V ternary like GaAsP, if the growth rate is limited by the group III supply. In such cases the surface sensitivity of the RAS signal can be used to extract the solid composition, since arsenic terminated surfaces have very different reconstruction compared to phosphorous terminated surfaces. This was first demonstrated in 1994 by Zorn et al. \[41\]. The surface sensitivity also works with InGaAsP \[47\] or with diluted nitrides like GaAsN (fig. 15). Again the surface changes very fast and the composition can be extracted after a few monolayers.

In case of nitride semiconductors like AlGaN or InGaN the increase of growth rate is not clearly correlated with solid composition. The high growth temperatures (for AlGaN above 1100°C, for InGaN above 650°C) result in strong desorption of Ga or In from the surface, and the binding energy of those atoms also depends on the composition. For instance, the growth rate of AlGaN at 1100°C can be higher than expected (when the reduced Ga desorption due to the higher binding energy of Ga to AlGaN is dominating) or lower (when gas-phase pre-reactions are dominating) \[E12\]. For InGaN increasing the growth rate at fixed In/Ga ratio increases the In content due to less desorption losses. Thus for AlGaN and InGaN a different analysis is needed.

As mentioned in sec. 3.1, the reflectance signal of a thick layer yields the growth rate and the complex index of refraction. If there is a database of the dielectric function for various compositions at the growth temperature, the refractive index can be converted straight forward into a composition. The only remaining error is the accuracy of the database and the model used to fit the data (probably including roughness). For AlGaN, which can be growth thick enough layers on GaN, a quite accurate database is available. Hence, fitting transient of various samples give good agreement to ex-situ measurement of X-ray diffraction (XRD), transmission spectroscopy and photoluminescence (fig. 16). The accuracy that can be reached is close to 5\% for photon energies above 3 eV where GaN is not transparent any more at growth temperature. In case of AlGaAs the database is very precise, and not only composition but additionally the surface temperature can be extracted \[E2\].

To determine the refractive index at least half of a Fabry-Pérot oscillation must be monitored, thus at least 20-40 nm must be grown. This may be impossible for highly strained materials like InAsP or InGaN because the material undergoes growth mode transitions either to 3D growth or the layer relaxes, changing solid composition and strain state (and thus dielectric function). If the growth rate is known (for instance from previous experiments using ex-situ X-ray diffraction) a fit of the refractive index is possible for thinner layers (as mentioned
Figure 15: RAS spectra for GaAsN (001) grown with three different tertiary-butyl hydrazine partial pressures in comparison to GaAs (001) (left top) and the difference spectra (right top). The nitrogen content correlates with the difference in the RAS signal $\delta \text{RAS}$. At 2.7 eV (left bottom) additional doping effects are overlaid (see sec. 5), thus best correlation of RAS signal change and solid composition is above 3.6 eV (right bottom). (data from [48])

in sec. 3.1). This has been demonstrated using reflectance transients for GaAsN multiple quantum wells. At 635 nm the nitrogen content only correlates with with the imaginary part of the refractive index [52]. Assuming a fixed growth rate, the N content could be determined for 5 nm quantum wells.

In case of the quaternary compound InGaAsN the indium content is affecting both real and imaginary part of the refractive index, while the nitrogen changes only the imaginary part at certain wavelengths (like 635 nm [52]). In this case In as well as N content could be determined from a single reflectance measurements [52]. But in general, for quaternary materials a single measurement is not enough to determine the composition and a combination of more than one method must be used.

For classical III-V semiconductors the obvious combination is RAS and reflectance, since reflectance is obtained simultaneously with RAS. First control of quaternary composition was demonstrated for InGaAsP, where the As/P ratio was given by the change of RAS signal at 1.75 eV and 2.65 eV while the In/Ga ratio was gained from the increase of growth rate relative to the InP substrate [47]. In case of lattice matched AlGaInP the increase of growth rate cannot be used, since any change of Al, Ga, or In will alter the growth rate. It was found that shape of the RAS spectra depends strongly on the Al content in the structure. The In content could be extracted from the refractive index, i.e. from the amplitude of the maximum during growth.
Figure 16: AlGaN solid composition from fitting Fabry-Pérot oscillations obtained from in-situ reflectance at different photon energies (left) and comparison of solid AlGaN composition using different TMGa/TMAI ratios by in-situ reflectance and several other ex-situ techniques (right). The GaN band gap was extrapolated from [49, 50]. The database used for evaluation was from the EpiR tool from LayTec company [51].

Figure 17: Principle of optical curvature measurements (from [58]).

on GaAs. Together this allows for control of lattice match as well as Al content [53]. Again the reflectance signal requires at 20-30 nm of thickness, which is no problem for these lattice matched materials.

Recently another technique came into focus, measurements of composition via strain of the layer. This is possible by using in-situ X-ray diffraction together with MOVPE either at a synchrotron (first pioneered by Kisker et al. [54]) or in a lab (for InGaAs [55] and GaN [56] more details on the instrument see [57]). While a very interesting research tool the problems associated with combining a radiation source with toxic gases and the special reactor setup makes this technique currently not suitable for industrial applications.
Strain can be measured optically by detection of the wafer curvature. This technique requires only optical access to the sample, which is anyway needed for in-situ reflectance. There are several realisation of in-situ curvature measurements (also called deflectometer), using either multiple beams [59] or just a pair of reflection [60, 61]. The principle is the same (fig. 17). When a layer with a different lattice constant is grown on a substrate, it will cause strain which will bend the substrate. For mechanically isotropic round wafer (with four or three-fold symmetry) like zincblende (001), (111), or wurzite (0001) orientation, the wafer will assume a concave or convex shape, which is detected by the change of position of the reflected beam(s). The bending radius is given by

$$ r(t) = \frac{\cos \alpha \cdot d_0 - d}{2L} \cdot \frac{d_0}{d(t)} $$

with $\alpha$ angle of incidence, $L$ distance, $d_0$ the initial distance on the screen, and $d(t)$ the current distance. Using two etalons also the curvature in two directions can be obtained to detect asymmetric curvature in case of two-fold symmetry. A different approach to detect concave curvature is the use of interferometry to detect the absolute position of the wafer backside to calculate curvature [44] (see sec. 3.3).

In principle, the slope of the curvature signal can be used to extract the lattice mismatch (if the elastic constants, growth rate, and the sample structure is known), similar to the shift of the layer peak in an $\omega - 2\Theta$ X-ray diffraction scan [62]. In practice defects, complex sample structure, and uncertainties of the elastic constants of ternary materials at growth temperatures make quantification difficult. Currently curvature measurements are mostly used with empirical calibration for a certain layer stack. Especially for lattice matched materials calibration is straightforward, since one targets for a flat wafer. This works well for (Al$_x$Ga$_{1-x}$)$_{1-y}$In$_y$P [63]. In all cases, one must take into account that the strain at room temperature is different when the expansion coefficients are different of substrate and layer. Thus the target level is often set empirically for a certain device stack, like no bending during quantum well growth or for flat wafers after growth on large area GaN on Si [61].
5 Doping

Additionally to composition and thickness most devices require control of the electron and hole concentrations in their layers too. Furthermore, the carrier concentration also indicates about the state of the sample, like the mobility at low temperatures or the background carrier concentration. An example is InN where the electron concentration depends e.g. on nitridation of the sapphire before growth [E4].

Control of dopant incorporation (and hence the carrier concentration) is usually done via ex-situ measurements, i.e. after growth. For calibration thick (typically more than 1000 nm) layers are grown, which are characterized afterwards by Hall effect or other resistive measurements. Complete structures can be analysed by destructive C/V profiling, i.e. an alternating etching and C/V measuring process. Device performance can also be related to doping levels.

In-situ electron diffraction techniques with RHEED cannot monitor the carrier concentration, since they sense the surface symmetry. Certain characteristic changes of the diffraction pattern and their intensities occur during delta doping, i.e. supplying up to a single monolayer of Si on GaAs [64]. But the carrier concentration itself cannot be monitored by RHEED.

The most common optical in-situ technique – reflectance – cannot measure doping directly, since it relies on changes of the refractive index of at least a percent in thicker layers (as discussed in sec. 3.1). Only by delta doping or for very high carbon concentrations in GaAsFabry-Pérot interferences have been observed [65, 66]. However, in actual devices the carrier concentrations are below $10^{19} \text{cm}^{-3}$ and thus are not accessible by reflectance.

For classical III-V semiconductors with zincblende structure, RAS can measure carrier concentrations below $10^{19} \text{cm}^{-3}$. The sensitivity of RAS to doping was discovered early using rotating samples [8] in 1987. Two years later also the basic origin of the signal, the change of symmetry due to an electric field due to Fermi level pinning was understood [67]. The first application to in-situ doping determination of GaAs:Si was done in 1991 by Aspnes et a. [68]. Further in-situ application has been done in MBE on delta doping [69] or for low temperature GaAs growth [70]. Due to the difficulties in doping of II-VI compounds RAS was applied here too, first ex-situ on ZnSe [71, 72]. The relatively low growth temperature of II-VI results in relatively large RAS signal amplitudes. Thus also in-situ RAS measurements on ZnSe [73, 74] and ZnTe [75] have been reported.

The first application of RAS to control the doping level in an electronic device was done in 1999 for HBTs [3], and in 2002 for vertical cavity surface emitting lasers [4]. In the same year Schmidt et al. showed that the calculated anisotropic response of the surface is modified when the surface is under an electric field gradient [76].

Nowadays, most works still use RAS signal as a qualitative measure, i.e. join empirical values with certain carrier concentrations. But the effect of doping on in-situ RAS spectra has been published for GaAs or InP, and also for ternaries like AlGaAs or InGaP [77].

5.1 Quantitative description of doping in RAS

An electric field gradient in z-direction breaks the bulk symmetry in x-y-direction (rightmost sketch in fig. 1). For zincblende structure of a compound AB semiconductor in (001) orientation, the electrons in the upper bonds will be pushed towards the A atoms and the electrons in the lower bonds towards the the B atoms in each layer. The accumulated effect from several layers results a noticeable anisotropy. Doping generates such a field due to band bending. The most simple description of band bending is the Schottky model, assuming a linear decreasing field (Fig. 18).
Since RAS is a linear method in the electric field, the RAS response from the doping ($\delta$RAS in the following) is added to the surface anisotropy, assuming that the surface is not altered by the doping. For instance fig. 19 shows the anisotropy caused by doping of GaAs (001), first on the whole spectrum and then only the doping related contribution $\delta$RAS, the latter obtained by subtraction of the RAS spectrum of the undoped sample. That the change of the anisotropy $\delta$RAS is indeed linear to the effective electric field by doping has been experimentally verified using comparison of differential electro-reflectance and RAS by Acosta-Ortiz and Lastras-Martinez [67] and using photo-reflectance and RAS by Chen and Yang [78].

The doping induced field by band bending is always normal to the interface or surface and constitutes a symmetry reduction for cubic materials. The field modifies the dielectric properties in lowest order via the linear electro-optic effect. By definition the electro-optic coefficients describe the changes in the optical indicatrix $\Delta B_{ij}$ under an applied electric field $E_k$. In the linear approximation the coefficient is defined by:

$$\Delta B_{ij} = d_{ijk}E_k.$$  

This is called the linear electro-optic effect, also abbreviated in the literature as LEO. (Sometimes $d_{ijk}$ is written as $z_{ijk}$. ) More common is the contracted notation using $B_{ii} = B_i$, $B_{23} = B_4$, $B_{31} = B_5$, and $B_{12} = B_6$:

$$\Delta B_i = \Delta \left( \frac{1}{\hat{n}^2} \right)_j = \sum_{i=1}^{3} d_{ji}E_i.$$  

where $\hat{n}_j = n_j + ik_j$ is the complex index of refraction along the direction $j$.

For zincblende type semiconductors with the pointgroup $T_d$, $d_{ij}$ has three identical non-zero components ($d_{41} = d_{52} = d_{63}$)[79]. Thus there is only one non-zero $\Delta \hat{n}_i(E_j)$ and I can omit the index in the following. With the definition of $\Delta B$ finally the change of the complex refractive index $\Delta \hat{n}(E)$ can be obtained for a field $E$ in z-direction:

$$\Delta B_6 = d_{41}E = \Delta \left( \frac{1}{\hat{n}^2} \right)_i = \frac{1}{\hat{n}^2(E)} - \frac{1}{\hat{n}^2(0)} = \frac{\hat{n}^2(0) - \hat{n}^2(E)}{\hat{n}^2(0)\hat{n}^2(E)}.$$
\[ \frac{1}{2} \delta \text{RAS} = \frac{r(E) - r(0)}{< r >} = \frac{\hat{n}(E) - 1}{\hat{n}(E) + 1} - \frac{\hat{n}(0) - 1}{\hat{n}(0) + 1} = \frac{2(\hat{n} + 1)(-\Delta \hat{n})}{\hat{n}(E) + 1(\hat{n}(0) + 1)} \approx -\frac{2\Delta \hat{n}}{\hat{n}^2 - 1} \]

Figure 19: RAS spectra (upper part) and \(\delta\)RAS spectra (undoped spectrum subtracted) for n- (Si) and p-type (C) doped GaAs at room temperature on (4x3) reconstructed arsenic rich surfaces. The grey lines mark the position of the \(E_1\), \(E_1 + \Delta_1\), \(E'_0\) and \(E'_0 + \Delta'_0\) transitions. Note the different y-scales for upper and lower figures. (from [E13])

For vertical incidence the reflectivity is \(r = \frac{\hat{n} - 1}{\hat{n} + 1}\) and thus the change of the RAS signal can be directly calculated:

\[ \frac{1}{2} \delta \text{RAS} = \frac{\hat{n}(E) - 1}{\hat{n}(E) + 1} - \frac{\hat{n}(0) - 1}{\hat{n}(0) + 1} = \frac{2(\hat{n} + 1)(-\Delta \hat{n})}{\hat{n}(E) + 1(\hat{n}(0) + 1)} \approx -\frac{2\Delta \hat{n}}{\hat{n}^2 - 1} \]

since \(\hat{n}(0) + 1 \approx \hat{n}(E) + 1 \approx < \hat{n} > + 1\) i.e. the change of the complex refractive index \(\Delta \hat{n}\) is small compared to the value of \(\hat{n}\). The equation yields only \(\frac{1}{2} \delta\)RAS because the changes along [110] and [1\overline{1}0] direction both contribute to \(\delta\)RAS.
Putting eq. (9) and (8) together, a direct linear relation between $d_{41}$, the electric field $E$ and the RAS-signal is obtained:

$$\frac{1}{2} \delta \text{RAS} \approx \frac{\tilde{n}^3}{\tilde{n}^2 - 1} d_{41} E_3. \quad (10)$$

In order to calculate the optical anisotropy, the electric field due to doping must be known. This electric field originates from the band bending in the surface space charge layer. In the most simple model the space charge layer is described within the Schottky model [80] which yields a linear dependence of the electric field amplitude on the distance from the surface ($z$):

$$E_{\text{Schottky}}(z) = \frac{eN}{\varepsilon_0 \varepsilon} (z - z_d)$$

with $z_d = \sqrt{\frac{2 \varepsilon_0 \varepsilon}{eN} \Phi_0}$, \( (11) \)

where $N$ is the carrier concentration, $\Phi_0$ the pinning position of the Fermi level at the surface, $z_d$ the depletion width, $e$ the elementary charge and $\varepsilon_0 \varepsilon$ the dielectric constant (here at the $E_1$ critical point). Hence the electrical field decreases linearly over the the penetration depth of the light.

In order to calculate the change of reflectivity, a three layer system of vacuum/linear decreasing field region/bulk region needs to be solved. The lowest interface does not contribute since the field is exactly zero there and thus the refractive index does not change. Solving the Maxwell equations for the light field yields Airy-functions for a linearly decreasing field in the depletion region or Bessel functions for an exponentially decreasing field [81].

Most groups used a different approach. The decreasing field region is replaced by a layer with an effective field $<E>$ [68, 69, 70]. This is a good approximation if the space charge region is much thinner than the wavelength, which is the case for moderate and high doping concentrations. Moreover, the assumption of the Schottky model – a delta like state distribution and a linear decreasing field – are probably a much larger generalisation than using an effective field.

To obtain the effective field $<E>$, the linear decreasing field is multiplied by the losses of light due to absorption given by $I(z) = I_0 \exp(-z/z_p)$ with $z_p = \frac{\lambda}{4\pi k}$ the penetration depth of light. This term is then integrated from $z = 0$ to $z = z_d$, since the field is only non-zero between the surface at $z = 0$ and the end of the space charge layer $z = z_d$ region (similar to ref. [69, 70]).

$$<E> = \int_0^\infty E_{\text{Schottky}}(z) \frac{I(z)}{I_0} dz = \frac{e N 1}{\varepsilon_0 \varepsilon z_d} \int_0^{z_d} (z - z_d) \exp \left( -\frac{z}{z_p} \right) dz = -\frac{e N}{\varepsilon_0 \varepsilon z_p} \left\{ 1 + \frac{z_p}{z_d} \left[ \exp \left( -\frac{z_d}{z_p} \right) - 1 \right] \right\}. \quad (12)$$

Since $\exp(x) = \sum_{n=0}^\infty \frac{x^n}{n!}$, the term $\frac{z_p}{z_d} \left[ \exp \left( -\frac{z_d}{z_p} \right) - 1 \right]$ is identical to $\frac{z_d}{z_p} \sum_{n=1}^\infty \frac{1}{(n+2)!} \left( \frac{z_d}{z_p} \right)^n$, i.e. for large penetration depths the sum can be ignored. Thus, for $z_p \gg z_d$ the resulting field is equal to the Schottky electric field in eq. (11). This means that in the near transparent regime slightly above the fundamental gap (large $z_p$) or for high carrier concentrations (small $z_d$) the Schottky electric field is a good approximation (see fig. 20). Beyond these extremes the mean field decreases faster than the Schottky field with decreasing carrier concentration (see fig. 20).
Figure 20: Electric field according to Schottky model (eq. 11) and mean electric field $<E>$ including the effect of light absorption (eq. 12). Parameters were $\varepsilon = 12.9$, $\Phi_0 = 0.67$ V, and $z_p = 18$ nm (i.e. GaAs at 2.9 eV) for both.

Figure 21: Effective electric field $<E>$ from eq. (12) on GaAs for three carrier concentrations and $\Phi_0=0.67$ eV. The effective field is always larger at larger penetration depth in the near transparent region.

Since the penetration depth is a function of the optical wavelength, the effective field changes over the spectral range too. The left side of fig. 21 shows the effective electric field $<E>$ for three different carrier concentrations in GaAs at room temperature. For instance the effective field in the 3 to 4.5 eV is much smaller for low carrier concentrations than in the photon energy region below 2.5 eV. Therefore different spectral regions will have a different effect on the field used for the calculation of the RAS response.

Combining eq. (10) and (12) the change in the RAS signal is of the following form:

$$\delta \text{RAS} = 2 \frac{\tilde{n}^3}{n^2 - 1} d_{41} <E>$$
Figure 22: Real part of the linear electro-optic coefficient $d_{41}$ below the fundamental band gap for GaAs [84] and InP [85].

\[
\begin{align*}
\text{GaAs} & \quad \text{Berseth et al.} \quad \text{JAP 72 (1992) 2821} \\
\text{InP} & \quad \text{Nobuo Suzuki and Kunio Tada} \quad \text{JJAP 23 (1984) 291}
\end{align*}
\]

For quantitative determination two unknown values must be known: $z_d$ and $d_{41}$. $z_d$ depends on the carrier concentration $N$ and the square root of the band bending $\Phi_0$ from Fermi level pinning (eq. 11). $\Phi_0$ is some value near midgap on most compound semiconductors. For various metals on GaAs Mönch reported 0.55 V for p-type and 0.7 V for n-type material [82]. On clean GaAs surfaces the experimental Fermi level was always found near midgap (0.7 eV) with less than 0.1 V difference between n-type and p-type material [83]. In our RAS data no systematic differences in the amplitudes for n- and p-doping were observed for InP and GaAs within the errors. Thus, a pinning at midgap is a reasonable approximation, at least for GaAs and InP. Furthermore, in eq. (11) only the square root $\sqrt{\Phi_0}$ enters. Even using a 10\% wrong band bending will only change the depletion width $z_d$ by 3\%.

The second value is $d_{41}$, the complex linear electro-optic coefficient. This value depends strongly on the wavelength. Most determination of the linear electro-optic coefficients were done in the transparent region, where $d_{41}$ is supposed to be a real number. Typical values are around 1 pm/V (fig. 22).

The actual relation between the RAS signal and $d_{41}$ is given by inverting eq. (13):

\[
\begin{align*}
\text{Re}(d_{41}) & = \frac{1}{2 < E > } (\alpha \text{Re}(\delta\text{RAS}) + \beta \text{Im}(\delta\text{RAS})) \\
\text{Im}(d_{41}) & = \frac{1}{2 < E > } (-\beta \text{Re}(\delta\text{RAS}) + \alpha \text{Im}(\delta\text{RAS}))
\end{align*}
\]

with $\alpha$ and $\beta$

\[
\alpha = \frac{n^5 + (2k^2 - 1)n^3 + (k^4 + 3k^2)n}{(n^2 + k^2)^3}
\]

\[
\beta = \frac{n^5 + (2k^2 - 1)n^3 + (k^4 + 3k^2)n}{(n^2 + k^2)^3}
\]
Figure 23: The optical components $\alpha$ and $\beta$ used in eq. (16) for GaAs at room temperature in linear and logarithmic scaled (dielectric function from [86]). Below 2.5 eV $\beta$ is a magnitude smaller than $\alpha$.

$$\beta = \frac{(n^4 + (2k^2 - 3)n^2 + k^4 + k^2)k}{(n^2 + k^2)^3}$$

as abbreviations. For nearly transparent media or below the fundamental gap (i.e. no absorption $k \approx 0$) the prefactors reduce to $\alpha = \frac{n^2 - 1}{n^3}$ and $\beta = 0$. For moderate absorption $\beta$ is still small (right side in fig. 23). Therefore, the influence of the imaginary part of the RAS spectrum on the real part of $d_{41}$ and vice versa is small in this region. Since $\text{Re}(\delta \text{RAS})$ and $\text{Im}(\delta \text{RAS})$ are Kramers-Kronig consistent (and also $n$ and $k$) the complex linear electro-optic coefficient must be Kramers-Kronig consistent over the full spectral region, and the imaginary part of $d_{41}$ should show characteristic structures similar to the real part. However, no measurements of the imaginary part have been reported so far.

To calculate $d_{41}$ from several doped oxidized samples an undoped reference spectrum was subtracted and the results divided by the effective field (eq. 12). The results were then averaged and from real and imaginary part of $\frac{\delta \text{RAS}}{<E>}$ the respective real and imaginary part of $d_{41}$ is obtained using the dielectric function from Aspnes et al. [86]. The critical points were assigned after T. Trepk [88].

Acosta-Ortiz calculated $d_{41}$ around the $E_1$ transition of GaAs from a single sample, neglecting the imaginary part, the factor $\alpha/2$, and only using Schottky fields [87]. Moreover, the reported anisotropies $\delta \text{RAS}$ of Acosta-Ortiz were nearly 20 times smaller than ours, less than $5 \times 10^{-5}$. Consequently he obtained values around 1 pm/V, nearly 25 times smaller. Otherwise the shape match our result very well (fig. 24).

Therefore, we carefully calibrated our RAS amplitudes by measuring the well-known extraordinary refraction of a quartz wedge and also by comparing the anisotropy of the Si (110) surface to data from Aspnes [7]. Furthermore, the doping signals $\delta \text{RAS}$ reported by other groups were in the order of $10^{-3}$ (for instance [89] or [92]). Hence, the origin of the very small anisotropies of Acosta-Ortiz is unclear.

Fig. 24 shows the complex linear electro-optic response for oxidized GaAs (001). The main contributions to $d_{41}$ in both real and imaginary part are close to the bulk critical points, starting at $E_0 + \Delta_0$. In the visible/UV region above 2 eV $d_{41}$ is nearly ten times larger than in
Figure 24: Complex linear electro-optic coefficient $d_{41}$ for GaAs obtained from averaging of several spectra of differently doped oxidized surfaces using eq. (15) with $\Phi_0 = 0.67$ V (from [E14]). The vertical lines indicate the bulk critical points. The values below the band gap are taken from [84]. The diamonds ♦ at the $E_1$ transition are the data from ref. [87] corrected for $\alpha$ and scaled up by a factor of 25 and follow the shape given by the full analysis.

Figure 25: Complex linear electro-optic coefficient from InP at room temperature together with the values below band gap from ref. [85]. (from [E14])
the transparent region below the band gap. Close to the fundamental band gap \( \delta_{11} \) approaches the value obtained by Berseth et al. \[84\]. Fig. 24 shows also the first determination of the imaginary part \( \text{Im}(\delta_{11}) \). Near the band gap it is comparable to the real part and shows strong features above 2.5 eV, i.e. also in the region around the \( E_1/E_1 + \Delta_1 \) transition.

For InP the anisotropies of oxidized surfaces depended on their treatment. Wafers “out of the box“ showed very broad and weak features, while freshly etched wafers or freshly oxidized epilayers showed much better defined features. Therefore we used the latter for the determination of \( \delta_{11} \) in fig. 25. Similar to GaAs or InP, \( \delta_{11} \) in the visible range was up to ten times larger than in the IR region obtained by Suzuki and Tada \[85\]. The IR data of \( \delta_{11} \) decrease towards the fundamental gap (see also fig. 22), which is continued by our data from above the fundamental gap (fig. 25).

It is interesting to compare the linear electro-optic coefficients between GaAs (fig. 24) and InP (fig. 25). For InP \( \delta_{11} \) is very small below the \( E_1 \) transition and increases strongly towards the UV region. On GaAs \( \delta_{11} \) shows the most pronounced oscillation at \( E_1/E_1 + \Delta_1 \). InP shows the reverse situation: The oscillation at \( E_1/E_1 + \Delta_1 \) is smaller than the one at \( E_0/E_0 + \Delta_0 \). Finally the real part of \( \delta_{11} \) is positive for GaAs and slightly negative for InP below 2.8 eV. These differences may originate from the different band structure, namely from the weaker \( E_1/E_1 + \Delta_1 \) transition of InP, and from the different ionisity of In and P in relation to Ga and As. A detailed understanding might be achieved from DFT calculations.

### 5.2 Reproducibility

The main error in the determination of the \( \delta \text{RAS} \) signal are offsets of the spectra since two spectra are subtracted. For non-rotating samples (as in this study) the offset of the RAS system depends on the position of the light entering the photo-elastic modulator, which depends again on the sample position. The reproducibility is usually better than \( 2.5 \times 10^{-4} \). For rotating samples, the offset is automatically corrected. However the fitting of the RAS signal introduces some noise in the order of \( 2 \times 10^{-4} \). After averaging resulting error is about \( \pm 1 \) pm/V for \( \delta_{11} \) in both cases.

The second point are systematic errors in the determination of \( \delta_{11} \). The model of an effective medium with a mean effective electric field is very simple. Using Airy functions a linear decreasing field could be solved without an effective medium. However, the main assumption, a strictly linear decreasing field is very likely an oversimplification. Additionally, a constant band bending is unlikely – as states are filled the band bending should decrease.

Finally, the effect of a finite amount of surface state was not taken into account. A finite density of surface states leads to a maximum band bending. Upon exceeding a certain carrier concentration, the integrated field and thus the \( \delta \text{RAS} \) signal do not increase any more. The latter is observed in the experiments. Fig. 26 shows the calculated signal of \( \text{Re}(\delta \text{RAS}) \) at 2.0 eV with saturation for \( 4 \times 10^{12} \text{cm}^{-2} \) and \( 7 \times 10^{12} \text{cm}^{-2} \) surface states. The assumption of finite surface states approximates the experimental data for oxides and \( c(4x4) \) reconstructed GaAs surfaces well. Hence, doping can be determined by RAS on GaAs at room temperature up to an upper limit of about \( 3 \times 10^{18} \text{cm}^{-3} \) for oxides and \( 6 \times 10^{18} \text{cm}^{-3} \) on \( c(4x4) \) reconstructed surfaces. This agrees nicely with finding on oxidized GaAs (001) surfaces where the saturation of the change in spectral ellipsometry was reported above \( 3 \times 10^{18} \text{cm}^{-3} \) \[90\].

Apart from these fundamental limitations, there are other effects. Fig. 27 shows the field normalized \( \text{Re}(\delta \text{RAS}) \) for a typical series of oxidized GaAs at room temperature and \( c(4x4) \) reconstructed samples at 650°C. Because the \( \delta \text{RAS} \) signal is smaller for lower carrier concentrations, those spectra are more noisy. But the general shape is very similar over the whole spectral range with the exception of the amplitude of the \( E_1/E_1 + \Delta_1 \) transition, which seems
Figure 26: Calculated and measured $\delta$RAS signal on several oxidized and reconstructed c(4x4) GaAs (001) surfaces. The reconstructed surface signal increases until $1.75 \times 10^{-3}$ while the doping induced anisotropy does not exceeds $10^{-3}$ on oxidized surfaces. These end-levels can be directly translated into a density of surface states.

Figure 27: Field normalized real part of the doping induced change $\delta$RAS for a series of five oxidized GaAs samples (left, [E14]) and for reconstructed samples at 650°C. Clearly the amplitude at $E_1$ scales differently from the rest.

to be systematically smaller at higher doping concentrations (indicated by arrows in fig. 27). This effect is less prominent at 650°C, since the amplitude at the $E_1$ transitions is smaller, but is also there.

Since $\delta$RAS signal at the $E_1/E_1 + \Delta_1$ transition scales differently, other effects related to doping must reduce the $\delta$RAS amplitude in this region. The most likely explanation is the broadening of bands at higher doping concentrations. It is well known from ellipsometry that at higher carrier concentrations the $E_1/E_1 + \Delta_1$ transition broadens until at very high
concentrations the $E_1$ and the $E_1 + \Delta_1$ peak cannot be distinguished any more [91, 90]. The broadening is caused by the redshift of bands at higher carrier concentrations. Due to the strong linear decreasing electric field at the surface, a continuum of $E_1$ transitions are overlaid, causing a broadening and reduce the amplitude of the $E_1/E_1 + \Delta_1$ oscillation in the $\delta$RAS signal.

In principle, additional effects could influence the $\delta$RAS signal at the $E_1/E_1 + \Delta_1$ transition like alloying, strain, or disorder. Alloying means a change of the spectral shape due to a change of the composition material and will be discussed in sec. 5.5. Moreover, doping levels below $10^{18}\text{cm}^{-3}$ are too low for alloying.

Stress and strain influence the symmetry of the crystal and its band structure and may result in anisotropy. For uniaxial stress along [110] the RAS signal change is most pronounced at the $E_1/E_1 + \Delta_1$ transition [98]. Typical dopants like silicon or carbon are much smaller than gallium or arsenic atoms and can cause a change of the lattice constant large enough to be detected by X-ray diffraction for concentrations in the higher $10^{18}\text{cm}^{-3}$ range. Since the reduction of $\delta$RAS occurs already in the mid $10^{17}\text{cm}^{-3}$ range, strain seems unlikely. Even more, a change of the lattice constant causes isotropic strain, which does not cause an optical anisotropy in the bulk.

Finally, doping causes a disorder on the surfaces (e.g. for reconstructed GaAs surfaces see ref. [99, 100]). The effect is also seen on the oxidized samples. It might be that this roughness is also present at the interface from bulk to oxide to neutralize charged dopant sites at the interface. Since the RAS signal mainly originates from the interface between oxide and crystal [E17, 101], any interfacial disorder influences the signal. Calculations of Lastras-Matrínez et al. showed that a roughness would indeed reduce the amplitude of the $E_1/E_1 + \Delta_1$ transition [94]. Disorder would also further reduce the amplitudes at high temperatures like 650°C, since a hot surface has also thermal disorder. However, as discussed in sec. 3.2 using eq. (3), a typical roughness on the monolayer scale is too small to contribute to $\delta$RAS in the visible range.

Therefore, I conclude that the stronger than expected reduction of the amplitude at the $E_1$ transition at higher doping levels is due the broadening of the dielectric function by band bending. Since the $\delta$RAS signal at the $E_1/E_1 + \Delta_1$ transition is the most pronounced doping related feature on GaAs, it has been used frequently in the past to determine the carrier concentration in-situ by the peak amplitudes or calculating the area enclosed by the oscillation. Because this signal has a different behaviour from the rest of the spectrum, I suggest instead to use the broad plateau region between 3.5 eV and 4 eV or even a full spectral fit to determine the carrier concentration from $\delta$RAS.

5.3 Surface contributions

Since RAS is surface sensitive, one might expect that the surface signal is modified by the electric field. Indeed, there is one report on DFT-LDA (density functional theory with local density approximation and energy shift by scissors operator) calculation of different GaAs surfaces in an electric field of 50 MV/m [76]. The calculations did not include spin-orbit splitting and thus no $E_1 + \Delta_1$ signal is expected, making direct comparison to experimental data difficult. The calculated $\delta$RAS spectra in fig. 28 showed a strong dependence of the spectra on the surface reconstructions. For instance at the $E_1$ transition the signal for $c(4\times4)$ and $\beta 2(2\times4)$ were negative, while for the $\zeta 2(4\times2)$ it was positive.

There are few experimental reports on surface dependent doping signatures. Most of the experimental investigations focussed on the signal of GaAs surfaces at the $E_1$ transition. Lastras-Martínez et al. found different amplitudes in photo-reflectance difference spectroscopy around the $E_1$ of GaAs (001) for (2x4) reconstructed and $c(4\times4)$ surfaces and attributed this to a reconstruction induced strain causing an electric field sub-surface layer [93]. In a newer publication
Electro-static potential of a β2(2x4) surface

Change of electron density on β2(2x4)

Figure 28: Electrostatic potential, change of electron density by an external field and induced changes of the RAS signal on a GaAs (001) surface with β2(2×4) reconstruction calculated, by DFT-LDA from ref. [76].

Figure 29: Change of RAS spectra of different GaAs (001) reconstructions observed at room temperature in MBE together with the bulk data calculated with $d_{41}$ from fig. 24. The observed anisotropies are all similar to the bulk data above 2.7 eV (thick blue line).

they reported a similar behaviour of the amplitude at $E_1$ for both c(4x4) and (2x4), but the amplitudes of the (2x4) were smaller. Since the (2x4) reconstruction was measured at higher temperatures, the differences where attributed to thermal effects [92].

For MBE prepared n-type doped GaAs c(4x4), (2x4) and (nx4)/(4x2) surfaces the author reported difference in amplitudes below the $E_1$ transition [E15] with a peak ration of 1.8:0.9:1 for c(4x4):2x4):nx6). Above the $E_1+\Delta_1$ transition the $\delta$RAS spectra were identical. To further investigated this, the procedure in [E15] was repeated with a new set of reference samples. Again a c(4×4), β2(2×4), and ζ2(4×2) reconstruction was prepared at several temperatures. Then, on the same sample an n-doped layer was grown and the three reconstructions were prepared again in the same way to extract the doping induced $\delta$RAS. This was repeated for two slightly different temperatures of the silicon Knudsen cell, which resulted in carrier concentrations of $8 \times 10^{17}$ and $1.2 \times 10^{18}$ cm$^{-3}$. The real parts of $\delta$RAS of the different surfaces is shown in fig. 29.
There is some scattering between the different preparation runs. But in general the $\delta$RAS signal is very close to the bulk response calculated in the previous section. Especially above 3.4 eV there are certainly no surface contributions to the RAS signal (as already reported in [E15]), which is quite different to the predictions of DFT calculations in fig. 28 from ref. [76]. The remaining differences to the bulk signal in fig. 29 are around the $E_1$ critical point oscillation, a shoulder just below $E_1$, and a peak at 2.2 eV for the gallium rich $\zeta(2 \times 4 \times 2)$.

In case of the gallium rich $(4 \times 2)$ the peak at 2.2 eV is neither present at higher temperatures (fig. 30) nor in the first run (see [E15]). Moreover, its amplitude is not changed by different doping levels. Therefore, the peak at 2.2 eV for the gallium rich reconstruction is most likely an artifact from the preparation of the gallium rich reference surface (the relatively high arsenic background in MBE causes $(n \times 6)$ contributions to the spectra, see [E8]). The shoulder between 2.5 eV and $E_1$ is not seen for the $\beta_2(2 \times 4)$ surface with $8 \times 10^{17}$ cm$^{-3}$ doping. Moreover, in this region the slope of the RAS signal is very steep, thus a slight misalignment of the monochromator due to a variation of the sample position can cause a considerable difference signal. To conclude the MBE data in fig. 29 cannot confirm a surface contribution above 3 eV. The highest scattering of the data is in the region between 2-3 eV, thus any surface contribution should be expected there.

A different approach to identify a surface contribution is the different dependence on carrier concentration. Any surface contribution would be additional to the bulk contribution given by eq. (13). Since the surface field directly couples with the linear electro-optic coefficient for the surface $d_{\text{surface}}$, absorption can be neglected. At the surface, the field is the Schottky field at the surface $E_{\text{Schottky}}(z = 0)$:

$E_{\text{surface}} = E_{\text{Schottky}}(z = 0) = \frac{eN}{\varepsilon_0 \varepsilon} (0 - z_d) = 2 < E_{\text{Schottky}}>$. 

Then the surface field induced signal adds to the bulk signal of eq. (13):

$\delta \text{RAS} = 2 \frac{\tilde{n}^3}{\tilde{n}^2 - 1} (d_{\text{surface}} 2 < E_{\text{Schottky}} > + d_{41} < E >) = \ldots$ 

![Figure 30: Doping induced $\delta$RAS signal for three GaAs (001) reconstructions at 400°C after doped layer growth at 540°C in MBE. Apart from a linear offset the spectra are identical for all reconstructions.](image-url)
Figure 31: Experimental $\delta$RAS signal of GaAs(001) at 2.9 eV of oxides and disordered c(4x4) surfaces at room temperature. The signal can be fitted with only $d_{\text{surface}} (0.7 \, E_{\text{Schottky}})$ and as well as $d_{41}$ (effective field from eq. (12)). The scattering of the experimental values is larger than the difference between both fitting curves.

$$
= -2 \frac{\tilde{n}^3}{\tilde{n}^2 - 1} \frac{e N}{\varepsilon_0 \varepsilon} \left( d_{\text{surface}} z_d + d_{41} z_p \left\{ 1 + \frac{z_p}{z_d} \left( \exp \left( -\frac{z_d}{z_p} \right) - 1 \right) \right\} \right),
$$

As shown in fig. 31 for GaAs at 2.9 eV the doping induced change of the RAS signal can be approximated either by pure surface $d_{41}$ as well as by pure bulk $d_{41}$ contributions. The experimental data show much stronger scattering than the difference between these two curves. Therefore, subtracting a known bulk signal is the only way to determine the surface signal.

In MOVPE the only surface that can be prepared reproducibly at room temperature is a disordered c(4x4) or (4x3) surface [E16]. Therefore, the signal has to be compared to the oxide signal, which is assumed to be free of surface contributions and resulting in the bulk $d_{41}$.

The right hand side of fig. 32 shows the surface contribution $d_{\text{surface}}$ calculated by subtracting the oxide signal (left side, similar to fig. 24) from 12 differently doped c(4x4) surfaces at room temperature. The surface contribution $d_{\text{surface}}$ is smaller than $\pm 1 \text{pm/V}$, which is the typical scattering of the data. Furthermore $d_{\text{surface}}$ only reproduces the shape of the bulk signal $d_{41}$. This hints at small offset errors in the reference surface. Those can occur very easily in MOVPE, since to determine the RAS signal baseline, the reactor pressure must be raised to atmospheric pressure. This affects the bending of glass rods and bending of the optical viewports, changing the position of the light on the photo-elastic modulator and the angle the light is entering the analyser prism and monochromator. The resulting errors of the baseline could then cause this contribution.

The signal at the $E_1/E_1 + \Delta_1$ optical transition seems to have a slightly different shape when comparing real and imaginary part, thus surface contributions may be found there. However, as discussed in the sec. 5.2 the signal at the $E_1/E_1 + \Delta_1$ transition is overlaid by a systematic doping correlated shift. Since the oxide signal was generated from a larger set of samples, the resulting difference in amplitude of the $E_1$ can originate from this. Furthermore, as already
shown in fig. 26 the c(4x4) surface has a higher density of surface states and thus saturates at higher amplitudes compared to the oxide.

Therefore, the experimental data cannot support a surface contribution to $d_{41}$ exceeding 1-2 pm/V, i.e. 20% of the bulk signal at best. Such a small surface contribution can be neglected for the in-situ determination of the carrier concentration, if the fully spectral shape is used.

### 5.4 Temperature dependence

Upon heating, the dielectric function changes strongly [96, 97, 88]. Most prominent are the reduction of the amplitude at the critical transitions due to broadening and their energy shift. Some transitions like $E_0 + \Delta_0$ cannot be recognized any more, so its position can only be calculated by extrapolation [88, 96].

The electro-optic coefficients change accordingly. Fig. 33 shows the field corrected change of RAS signal for GaAs and InP at 600°C and room temperature (scaled down, dotted). At high temperature the signals are generally smaller by about 30% and shifted similar to the shifts of the bulk critical points. The $E_1$ transition is very broad, overlaying the $E_1 + \Delta_1$ transition nearly completely. The thick grey line in fig. 33 is obtained by a Gaussian broadening of the room temperature data scaled down by 0.6× and red shifted by 0.37 eV (GaAs) respectively 0.40 eV (InP). It approximates the experimental high temperature data very well, showing that these are indeed mostly influenced by energy shifts, amplitude reduction and broadening,
similar to the dielectric function. The small deviations in the UV region in fig. 33 are most likely due to some systematic errors of the monochromator/PEM above 4.7 eV, as discussed in the introduction of the multi-channel RAS system in sec. 2. Even more, these surfaces have very different RAS spectra (especially for InP) at room temperature and 650°C. This further indicates negligible surface contribution.

This fact is very advantageous for practical applications. As a consequence the doping effect can be modelled by a single function with a temperature dependent broadening, red shift and amplitude reduction, allowing for quantitative fitting of RAS spectra (given a suitable reference spectrum for the underlying surface). Since industrial MOVPE machines employ sample rotation, the baseline correction of RAS spectrum is included automatically, reducing the systematic error. Moreover, the position of the $E_1$ transition is very sensitive to the sample temperature, and thus sample temperature and carrier concentration can be extracted by the same fitting procedure.

5.5 Transition from doping to alloying

Alloying means a change of the chemical and physical material properties due to incorporation of foreign atoms. Typical concentration of $10^{18}$ cm$^{-3}$ doping atoms are still thousand times less compared to the mid $10^{22}$ cm$^{-3}$ atoms in the bulk. Therefore, alloying is only expected for $10^{19}$ atoms/cm$^3$ and more. Reaching and exceeding such concentrations, the band structure of the material will be changed and modify other properties, including the dielectric function. Doping is then defined as incorporation levels that do not change the band structure notably.

The difference spectra $\delta$RAS are sensitive to the chemical composition of the material as well, since a change of the dielectric function of the bulk strongly affects the anisotropic response of the surface. (This was already discussed in sec. 4) During in-situ measurements also surfactants and segregating species might be present on the surface for certain material. Thus, careful evaluation is necessary to determine whether the change is due to a changing surface or rather from a changed bulk dielectric function.
Since at the onset of alloying also the properties of the layer change, plotting certain properties over the input ratio can give indications additional to the RAS signal. For instance, in the doping regime the carrier concentration is linked to the amount of dopants incorporated, after alloying a straight relationship is not necessarily expected.

This is seen in the right side of fig. 34 which shows the solid concentration of silicon as determined by secondary ion mass spectroscopy (SIMS) versus the free carrier concentration determined by CV-profiling [102]. For Si concentrations exceeding $10^{19}\text{cm}^{-3}$ the free electron concentration is not increasing any more, indicating the onset of alloying (in this case by Si-Si clusters and antisites). The corresponding RAS spectrum with $10^{20}\text{cm}^{-3}$ (more than 0.1 %) Si deviates strongly from the doping induced shape at lower concentrations.

A possible way to achieve high concentrations without affecting surface structure is low temperature MBE (LT-MBE) growth. Not only it allows for incorporation of very high concentrations of dopants but simultaneously strongly suppresses segregation and surface diffusion. However, for the evaluation of $\delta$RAS, obtaining an undoped reference spectrum is difficult, since there is always an n-type conductivity background due to As antisite defects [103].

Fig. 35 shows $\delta$RAS for high concentrations of Si, Mn (manganese) and Sn (tin). The effect of silicon doping (top row in fig. 35a) on the RAS spectra is similar to the one reported in the previous chapter and in literature at other temperatures and on other GaAs surface reconstructions (e.g. figs. 19,34). This indicates that the Si fluxes were not high enough to achieve alloying with the available setup. Also the measured electron concentrations in the lower $10^{18}\text{cm}^{-3}$ range support this interpretation.

For p-doping with Mn the situation is different. At low Mn cell temperatures the RAS spectrum just reflects the change from the n-type background (typical for low temperature growth due to arsenic antisite defects) to p-doping. Accordingly $\delta$RAS changes sign when the conduction changes from n-type to p-type. This is seen more clearly in the difference spectra (right side in fig. 35b, middle row), where the first spectrum is inverted with respect to the other ones. But for higher Mn fluxes (i.e. cell temperatures above 650°C) a broad peak centered around 4 eV emerges. This indicates that at these Mn concentrations (above 2-3 % Mn) alloying starts to take place.

The RAS spectra for Sn in Fig. 35c (lower row) show a similar behaviour. For low fluxes, $\delta$RAS on the right graph show only doping effects, i.e. the amplitudes just increase at all positions nearly linearly with cell temperatures until 850°C. At higher Sn fluxes, $\delta$RAS at the $E_1$ transition decreases again, indicating the onset of alloying (fig. 36 left). At 1000°C Sn cell...
Figure 35: The effect of cell temperature of Si (upper row), Mn (middle row), and Sn (lower row) on the RAS spectrum during LT-MBE growth of GaAs (001) at 275°C. The right columns are the differences to a spectrum of a compensated surface. Please note the different y-scales in the right hand figures. Si shows only doping related signals, while for Mn and Sn alloying occurs too. (from [E17])
temperature, the resulting RAS spectrum is very different from the initial one. The new alloy surface possesses an additional RAS minimum around 3.5 eV.

The onset of alloying is also seen in the ex-situ measured carrier concentration. This carrier concentration increases exponentially up to a cell temperature of 800°C (fig. 36 right). At 850°C (the onset of alloying) the carrier concentration is reduced compared to 800°C. But for 950°C Sn cell temperature the carrier concentration increases again. It nicely extrapolates of the data below 850°C, indicating that the fraction of doping and compensating Sn atoms is nearly constant. Apparently self-compensation is enhanced near the transition to alloying, probably caused by disorder from the simultaneous presence of GaSnAs alloy and GaAs:Sn doped domains.
6 Doping Profiling

RAS can determine the doping concentration during layer growth using transients. This yields also the growth rate during homo-epitaxy and thus provides the possibility for full in-situ doping profiling [E18]. During growth of a layer with a doping concentration different than the underlying layer, the RAS signal will oscillate (fig. 37). These oscillations are Fabry-Pérot like interferences due to the change of the refractive index through the electro-optic effect of the electric space charge field at the interface. These anisotropic changes are small (about $10^{-3}$), and thus manifest themselves only in the RAS signal.

![Figure 37: RAS (left) and reflectance (right) transients recorded simultaneously during growth of GaAs on GaAs:Si $5 \times 10^{18}$ cm$^{-3}$. The oscillation periods in RAS are longer for longer wavelength. No oscillations are visible in the reflectance signal. The black line on the RAS transient at 1.99 eV shows a fit using eq. (33).](image)

Fig. 37 shows a typical example. The variations in the reflectance signal are present in all channels. Only the RAS signal in the near transparent region below 3 eV shows clear wavelength depending oscillations. Similar to Fabry-Pérot oscillations their period is shorter for shorter wavelengths.

To calculate the anisotropy change with varying thickness, the surface signal can be ignored. It cannot cause oscillations when the layer below them changes. Even more, the growing layer (1) in fig. 38 can be up to several microns thick which is much thicker than the surface region.

Similar to the treatment in sec. 5.1, the space charge region can be approximated by an effective medium, resulting in an idealized four layers stack: vacuum, isotropic volume, space charged interface with an effective field, isotropic volume (fig. 38 left). In the case of a transition between a doped and an undoped layer the depletion width of the undoped layer would be infinite, reducing this effectively to a three layer model with a single interface inside, if the width of the depletion layer is much smaller than the wavelength (fig. 38 right). Since a priori it is not clear whether the effective medium interlayer (four layer stack) or the single step (three layer stack) is a better approximation, calculations will be performed for both models.
Figure 38: Sketch of the four layers stack for a transition between two different carrier concentrations where the space charge layer has a finite thickness (left) and for the transition from doped to undoped where the space charge layer is assumed to be very small compared to the wavelength while the field in the undoped region would extend infinitely (right). In the latter there is only one interface expected which corresponds to a three layer stack.

6.1 Four layer stack model

First the four layer stack will be treated analytically. The complex reflectivity of a four layer stack is given analytically by [104] (For vertical incidence $\Phi_0 = 0^\circ$):

$$r_{tot} = \frac{r_{01} + r_{12} \exp(2\beta_1) + (r_{12} + r_{01} \exp(2\beta_1)) r_{23} \exp(2\beta_2)}{1 + r_{01} r_{12} \exp(2\beta_1) + (r_{12} + r_{01} \exp(2\beta_1)) r_{23} \exp(2\beta_2)}$$

with

$$r_{ij} = \sqrt{\varepsilon_i - \varepsilon_0 \sin^2 \Phi_0} - \sqrt{\varepsilon_j - \varepsilon_0 \sin^2 \Phi_0} = \frac{\sqrt{\varepsilon_i - \varepsilon_j}}{\sqrt{\varepsilon_i + \varepsilon_j}}$$

and

$$\beta_i = \frac{2\pi d_i}{\lambda} \sqrt{\varepsilon_i - \varepsilon_0 \sin^2 \Phi_0} = \frac{2\pi d_i}{\lambda} \sqrt{\varepsilon_i}$$

$r_{ij}$ is the reflectivity at the interface between layer $i$ and layer $j$. The zeroth layer is vacuum, the first layer is isotropic bulk, the second layer is the space charge region in the interface between the different doping regions, and the third layer is the bulk underneath (see fig. 38). $d_i$ is the thickness of layer $i$. Note that $r_{ij}$, $\beta_i$, and $\varepsilon_i$ are complex numbers.

For a non-magnetic material ($\mu_r = 1$) the dielectric function is directly correlated with the complex index of refraction by $\sqrt{\varepsilon_i} = \tilde{n}_i$. Furthermore, region 1 and region 3 are field free, thus $\tilde{n}_1 = \tilde{n}_1$ and $r_{12} = -r_{23}$. The final formula for $r_{tot}$ is:

$$r_{tot} = \frac{r_{01} + r_{12}(\exp(i2\beta_1) - \exp(i2(\beta_1 + \beta_2))) - r_{01} r_{12}^2 \exp(i2\beta_2)}{1 + r_{01} r_{12}(\exp(i2\beta_1) - \exp(i2(\beta_1 + \beta_2))) - r_{12}^2 \exp(i2\beta_2)}$$

with $r_{ij} = \frac{\tilde{n}_i - \tilde{n}_j}{\tilde{n}_i + \tilde{n}_j}$ and $\beta_i = \frac{2\pi \tilde{n}_i d_i}{\lambda}$

For the RAS response the actual change is needed. Hence we calculate the total differential times the change of refractive index: (Note that $r_{01}$ and $\beta_1$ do not depend on $\tilde{n}_2$ and thus their derivatives are zero.)

$$\Delta r_{tot} = \frac{\partial r_{tot}}{\partial \tilde{n}_2} \Delta \tilde{n}_2$$

$$= \left[ \frac{\partial r_{tot}}{\partial \tilde{n}_2} \cdot \frac{\partial r_{12}}{\partial \tilde{n}_2} + \frac{\partial r_{tot}}{\partial \beta_2} \cdot \frac{\partial \beta_2}{\partial \tilde{n}_2} \right] \Delta \tilde{n}_2$$
\[
\begin{align*}
&= \frac{(1 - r^2_{01}) \exp(i2\beta_1)}{
(1 + r_{01}r_{12}(\exp(i2\beta_1) - \exp(i2(\beta_1 + \beta_2))) - r^2_{12} \exp(i2\beta_2))^2
\times \left[
(1 - \exp(i2\beta_2)) \left(1 + r_{12}^2 \exp(i2\beta_2)\right) \frac{\partial r_{12}}{\partial n_2}
- i2r_{12} \exp(i2\beta_2)(1 - r^2_{12}) \frac{\partial \beta_2}{\partial n_2}\right] \Delta \tilde{n}_2
\end{align*}
\]

The reflectivity at the internal interface \(r_{12}\) are very small compared to \(r_{01}\) on the surface (in the order on \(10^{-3}\)). Thus they can be approximated as zero:

\[
\Delta r_{tot} \approx 0 \quad (1 - r^2_{01}) \exp(i2\beta_1) (1 - \exp(i2\beta_2)) \frac{\partial r_{12}}{\partial n_2} \Delta \tilde{n}_2
= \frac{(1 - r^2_{01}) \exp(i2\beta_1) (1 - \exp(i2\beta_2))}{(\tilde{n}_1 + \tilde{n}_2)^2} 2\Delta \tilde{n}_2
\]

\[
\tilde{n}_i \approx \tilde{n}_2 \quad (r_{01} - 1) \exp(i2\beta_1) (1 - \exp(i2\beta_2)) \frac{\Delta \tilde{n}}{2\tilde{n}}
\tag{22}
\]

The actual change of the RAS signal is now two times the change of \(\Delta r_{tot}\) (since we have the difference between two directions) divided by the total reflectivity. As stated above, the contributions from the inner interfaces are small and the reflectivity of the whole stack can be approximated by the reflectivity from the surface \(r_{tot} \approx r_{10} = \frac{\tilde{n}-1}{\tilde{n}+1} = -r_{01}\). Finally, the change of refractive index \(\Delta \tilde{n} \approx -\frac{1}{2}\tilde{n}^3 d_{41} E\) is given by eq. (8). Putting it all together (factor two for two directions!):

\[
\delta \text{RAS} = 2 \Delta r_{tot} = 2 \frac{r_{01}^2 - 1}{r_{01}} \exp(i2\beta_1) [1 - \exp(i2\beta_2)] \frac{-\tilde{n}^3}{2\tilde{n}} 1 d_{41} < E >
= 2 \frac{4\tilde{n}}{\tilde{n}^2 - 1} \exp(i2\beta_1) \frac{-\tilde{n}^2}{4} [1 - \exp(i2\beta_2)] d_{41} < E >
= 2 \frac{\tilde{n}^3}{\tilde{n}^2 - 1} \exp\left(i\frac{4\pi \tilde{n} d_{41}}{\lambda}\right) \left[1 - \exp\left(i\frac{4\pi \tilde{n} d_{41}}{\lambda}\right)\right] d_{41} < E >
\tag{23}
\]

An alternative approach is a direct calculation of \(\Delta r_{tot}\) from eq. (20). The total reflectivity is just the total reflectivity for the undoped case plus the change due to the inner layers: \(r_{tot} = \Delta r_{tot} + r_{10} \Rightarrow \Delta r_{tot} = r_{tot} - r_{10}\). The inner reflectivity can be directly calculated using \(\tilde{n}_1 = \tilde{n}\) and \(\tilde{n}_2 = \tilde{n} + \Delta \tilde{n}\). Hence \(r_{12} = \frac{\tilde{n} - \tilde{n} - \Delta \tilde{n}}{\tilde{n} + \tilde{n} + \Delta \tilde{n}} \approx -\Delta \tilde{n}\). The reflectivity at the first interface is just given by the Fresnel equation: \(r_{10} = -r_{01} = \frac{\tilde{n} - 1}{\tilde{n} + 1}\). Putting this all together gives:

\[
\Delta r_{tot} = \frac{4\tilde{n}^2 e^{i2\beta_1}(e^{i2\beta_2} - 1) \Delta \tilde{n}}{\tilde{n}^2 (\tilde{n} + 1)^2 + e^{2i\beta_1} (1 - e^{i2\beta_2}) \tilde{n} (\tilde{n}^2 - 1) \Delta \tilde{n} - e^{i2\beta_2} (\tilde{n} + 1)^2 \Delta \tilde{n}^2}.
\tag{24}
\]

Doing a Taylor expansion and neglecting higher order terms of \(\Delta \tilde{n}\) (i.e. no higher order terms of \(< E >\)) results in:

\[
\Delta r_{tot} = \frac{2 e^{i2\beta_1}(e^{i2\beta_2} - 1)}{\tilde{n}(\tilde{n} + 1)^2} \Delta \tilde{n} + \ldots
\tag{25}
\]

This is exactly the same as eq. (22). Using \(r_{tot} = r_{10} + \Delta r_{tot} \approx r_{10}\) and together with \(\Delta \tilde{n} \approx -\frac{1}{2}\tilde{n}^3 d_{41} < E >\) from eq. (8) again eq. (23) is obtained, showing that both calculations are similar.
6.2 Three layer stack modell

A three layer stack can be calculated for instance from eq. (19) without the last interface \( r_{23} = 0 \). For vertical incidence the starting formula is:

\[
r_{\text{tot}} = \frac{r_{01} + r_{12} \exp(i2\beta_1)}{1 + r_{01}r_{12} \exp(i2\beta_1)}
\]  

(26)

The change of the RAS signal is then \( \delta \text{RAS} = 2\Delta E \mathbf{r} = 4\frac{r_{01} - r_{0i}}{r_{01} + r_{\text{tot}}} \approx 2\frac{r_{01} - r_{0i}}{r_{01}} \):

\[
\delta \text{RAS} = 2\left(1 - r_{0i}^2\right)r_{12} \exp(i2\beta_1) \frac{r_{01} + r_{0i}^2 + r_{12} \exp(i2\beta_1)}{r_{01} + r_{0i}^2 + r_{12} \exp(i2\beta_1)}
\]

(27)

Putting in \( r_{12} \approx -\frac{\Delta E}{2n} = \frac{1}{4} \tilde{n}^2 d_{41} < E > \) (from the previous section) yields:

\[
\delta \text{RAS} = \frac{8 \exp(i2\beta_1) \tilde{n}^3 d_{41} < E >}{(\exp(i2\beta_1) \tilde{n}^4 - 2 \exp(i2\beta_1) \tilde{n}^3 + \exp(i2\beta_1) \tilde{n}^2) d_{41} < E > + 4\tilde{n}^2 - 4}
\]

Finally a Taylor expansion of the field \( < E > \) results in:

\[
\delta \text{RAS} = 2 \tilde{n}^3 \frac{\tilde{n}^2}{n^2 - 1} \exp\left(\frac{4\pi \tilde{n}d(t)}{\lambda}\right) d_{41} < E > + \cdots
\]

(29)

\[
2 \tilde{n}^3 \exp\left(\frac{4\pi \tilde{n}d(t)}{\lambda}\right)
\]

is also prefactor of the four layer stack in eq. (23). The difference between the three layer stack in eq. (29) and the four stack in eq. (23) is the term \( [1 - \exp\left(\frac{4\pi \tilde{n}d}{\lambda}\right)] \), i.e. the finite thickness of the space charge region. This will be further discussed in sec. 6.5.

6.3 Calculation of the effective field

In case of a pn-junction the effective electric field can be directly taken from solving the Poisson equation with a Schottky model. Care must be taken, because the two space charge regions will have different electric fields, since the depletion widths \( z_{dn} \), will be different for different carrier concentrations \( N_i \ (i \neq j \in (1,2)) \):

\[
z_{dn_i} = \sqrt{\frac{2\varepsilon_0 \varepsilon}{\varepsilon_0 \varepsilon}} \frac{\Phi_0}{\varrho} \frac{N_i}{N_1 + N_2}
\]

(30)

\[
E_i(z) = \frac{eN_i}{\varepsilon_0 \varepsilon} (z - z_{dn_i})
\]

assuming the interface at \( z = 0 \). Since the field is continuous, at \( z = 0 \) the maximum field must be equal i.e. \( E_1(z = 0) = E_2(z = 0) \). Neglecting light absorption, the mean field will be just \( < E > = \frac{1}{2} \frac{eN_i}{\varepsilon_0 \varepsilon} z_{dn_i} \). Since the light is absorbed, the effective field is reduced similar to eq. (12). For light traveling from \( z = -z_{dn_1} \) this results in:

\[
< E > = \frac{e}{\varepsilon_0 \varepsilon} \left[ \int_{-z_{dn_1}}^{0} N_1(z_{dn_1} + z) \exp\left(\frac{-z + z_{dn_1}}{z_p}\right) dz \right.
\]

\[
+ \int_{0}^{z_{dn_2}} N_2(z_{dn_2} - z) \exp\left(\frac{-z + z_{dn_1}}{z_p}\right) dz \right]
\]

\[
= \frac{e}{\varepsilon_0 \varepsilon} \left\{ \frac{N_1}{z_{dn_1}} \left[ z_p - (z_{dn_1} + z_p) \exp\left(\frac{-z_{dn_1}}{z_p}\right) \right] \right.
\]

\[
+ \frac{N_2}{z_{dn_2}} \left[ z_p \exp\left(\frac{-z_{dn_1} + z_{dn_2}}{z_p}\right) + (z_{dn_2} - z_p) \exp\left(\frac{-z_{dn_1}}{z_p}\right) \right] \right\}
\]

(31)
The thickness $d_2$ in eq. (23) is then given by $z_{dn1} + z_{dn2}$.

For the actual calculation of $z_{dn}$, the band bending $\Phi_0$ must be known. In case of a pn-junction, the band bending $\Phi_0$ would be the band gap. For the change of undoped to doped the Fermi level is at midgap in the undoped part. For $N_2 = 0$ the second term is zero and the effective electric field with light absorption is:

$$< E > = \frac{eN_1}{\varepsilon_0 \varepsilon} \frac{z_p}{z_{dn1}} \left[ z_p - (z_p + z_{dn1}) \exp \left( -\frac{z_{dn2}}{z_p} \right) \right]$$

for a transition from undoped to doped. The depletion length is then given by the Schottky depletion length from eq. (11).

Fig. 39 compares the effective field calculated for the oscillations from the above eq. (32) to the field calculated using eq. (12). The effective field amplitudes for transients are generally smaller than for the surface. However, in the transparent region the error is less than 10\% for carrier concentrations above $2 \times 10^{17} \text{cm}^{-3}$. Often ex-situ measurements of the carrier concentration have even higher errors.

The most difficult part to understand in this model is the origin of a field for a change between two different doping levels. Such a transition results in clear oscillations with a magnitude similar to the actual change from undoped to the difference of both doping levels, even though one would expect only a small change in Fermi levels and thus no extended space charge region at all. A possible explanation could be that the resulting electric field is mostly due to thermal diffusion of carriers from the higher doped region to the lower doped region. Unfortunately I am not aware of any data on carrier diffusion lengths near growth temperatures for GaAs or InP, thus a quantification is difficult. Hence in the following I limited my analysis of experimental data to transitions from/to undoped layers using Schottky fields.
6.4 Determination of $d_{41}$ from $\delta RAS(t)$

For the experimental analysis the RAS transients are fitted by:

$$\text{Re} \left( \frac{\Delta r}{r} \right) (t) = (\text{amp}_\text{sin} \sin(n_f t) + \text{amp}_\text{cos} \cos(n_f t)) \exp(-f_k t) + RAS(0)$$  \hspace{1cm} (33)

with $f = \frac{4 \pi \text{growthrate}}{\lambda}$ proportional to the growth rate.

To obtain the linear electro-optic coefficients $d_{41}$ we have to determine the real part of eq. (23) respective eq. (29). In the following the index 3 or 4 denotes either the three layer or the four layer stack calculations.

$$\text{Re}(\delta RAS) = 2 < E > \exp(-k_f t) \left[ \text{Re}(d_{41})(A \cos(n_f t) - B \sin(n_f t)) - \text{Im}(d_{41})(B \cos(n_f t) + A \sin(n_f t)) \right]$$  \hspace{1cm} (34)

$$A_3 = \alpha$$
$$B_3 = \beta$$
$$A_4 = \exp(-k_\phi) [\alpha \exp(k_\phi) - \alpha \cos(n_\phi) + \beta \sin(n_\phi)]$$
$$B_4 = \exp(-k_\phi) [\beta \exp(k_\phi) - \beta \cos(n_\phi) - \alpha \sin(n_\phi)]$$
$$\phi = \frac{4 \pi z_d}{\lambda}$$
$$\alpha = \frac{n^5 + (2k^2 - 1)n^3 + (k^4 + 3k^2)n}{(n^2 - k^2 - 1)^2 + 4k^2 n^2}$$
$$\beta = \frac{k(n^4 + (2k^2 - 3)n^2 + k^2 + k^4)}{(n^2 - k^2 - 1)^2 + 4k^2 n^2}$$

The factors $\alpha$ and $\beta$ are similar to the ones introduced in eq. (16) only their denominator differs. They are plotted in fig. 40 for GaAs.
Figure 41: Linear electro-optic coefficients from fitting several transients (small symbols) and some averages at certain photon energies (large symbols) obtained at 650°C for a three layer model (left) and four layer model (right). The lines are spectral data from sec. 5.4. The data calculated with the three layer model show less scattering and approximate the spectral data better.

Comparing eq. (34) to the RAS transient fit in eq. (33) for sin and cos terms, we get directly the growth rate and a linear system for the real and imaginary part of $d_{41}$:

$$2 < E > (A \text{Re}(d_{41}) - B \text{Im}(d_{41})) = \text{ampcos}$$

$$-2 < E > (B \text{Re}(d_{41}) + A \text{Im}(d_{41})) = \text{ampsin}$$

$$\text{Re}(d_{41}) = \frac{\text{ampcos} A - \text{ampsin} B}{2 < E > (A^2 + B^2)}$$

$$\text{Im}(d_{41}) = -\frac{\text{ampcos} B + \text{ampsin} A}{2 < E > (A^2 + B^2)}$$

Fig. 41 shows the linear electro-optic coefficients found via transients together with spectral data from differences of RAS spectra. Within the error bars $d_{41}$ calculated from transient data is similar to the amplitude calculated from spectral data. Comparing three layer and four layer models, the amplitudes are smaller and the deviations – especially in the imaginary part – are larger for the four layer model.

6.5 Quantitative evaluation

In the transient data large scattering occurs of the amplitude between cosine and sinus (and thus real and imaginary part) between different samples (small symbols in fig. 41). In case of the four layer model the scattering is even increased by the high sensitivity of eq. (23) to a change of the depletion layer thickness $d_2$. A small variation of $d_2$ shifts via the term $[1 - \exp(i \frac{4\pi}{\lambda} d_2)]$ the balance between real and imaginary part. Furthermore the depletion layer thickness is a function of the carrier concentration. Thus even small errors in the determination of the carrier concentration will affect the thickness of the depletion layer and redistribute amplitude between $\text{Re}(d_{41})$ and $\text{Im}(d_{41})$. This cause of systematic error is absent in the three layer model and reduces the overall scattering of the data.

For high doping in the four layer model, i.e. a depletion depth $d_2 = z_{dn2}$ much shorter than the wavelength, the phase factor $n\phi$ and thus $\sin(n\phi)$ is small. Moreover, $\delta$RAS oscillations were observed in the near transparent region below $E_1$ where $\beta$ is less than a tenth of $\alpha$. This
Figure 42: RAS signal change at 2 eV during growth of two undoped GaAs layers on doped GaAs and for one doped layer on undoped GaAs with the same growth rate. The second undoped layer (dotted) is shifted for clarity. The maxima and minima are always at the same time, indicating a negligible phase shift.

means in the case of the four layer model as well as in the three layer model that the coefficient $B_i$ in eq. (34) is nearly zero:

$$\text{Re}(\delta_{\text{RAS}}) \approx 2 < E > \exp(-k ft) [\text{Re}(d_{41})(A \cos(n ft) - \text{Im}(d_{41})(A \sin(n ft))]$$

As a result, the real part of the RAS signal transient would depend on the cosine of the oscillations for $\text{Re}(d_{41}) \gg \text{Im}(d_{41})$ for nearly transparent media and (in case of a four layer model) for high enough doping levels.

But even carrier concentrations in the middle $10^{18} \text{cm}^{-3}$ range result in depletion widths $z_{dn2}$ of a few ten nanometers, according to eq. (30). The real part of the refractive index is 2-4. Together this results in a optical thickness comparable to the wavelength: $\frac{4\pi n z_{dn2}}{\lambda} \approx \frac{\pi}{4}$. Thus in case of the four layer model the term $\left[1 - \exp(i\frac{4\pi n z_{dn2}}{\lambda})\right]$ in eq. (23) will cause an imaginary part of $\delta_{\text{RAS}}$ even if the field induced contribution is real, i.e. $\text{Im}(d_{41}) = 0$. Hence, in case of a four layer model the RAS signal could start then either with a cosine or sinus-like oscillation or anything in between. However, as fig. 42 shows, the RAS signal is nearly always a cosine signal, independent from the carrier concentration. Only the sign of the cosine oscillation depends on the direction of the field, i.e. a positive sign for the growth of n-type on undoped or undoped on p-doped layer. Thus, an inner spaced charge layer $z_{dn2}$ must be very thin to allow the phase shift induced by $z_{dn2}$ to be very small. Altogether this suggests that the three layer model is preferable for the analysis.

There is a second cause of systematic error for transient data for both models. The calculations are only valid, when there is already a fully developed space charge region. During the initial growth this space charge region is not completely built up and both surface and interface regions overlap. This can be see for instance in the fit in fig. 37, where the beginning of the
Figure 43: δRAS oscillation amplitude (\(\sqrt{ampcos^2 + ampsin^2}\)) as a function of carrier concentration at 2 eV (left) and 2.25 eV (right) for GaAs at 650°C. The solid lines are the expected amplitude, calculated from spectral data. Both data reproduce the calculated values well.

first oscillation is not reproduced well. Thus the eq. (34) can reproduce the experimental only for time after the depletion layer thickness has been grown \(t > \frac{z_{dn}}{growthrate}\).

The doping dependent maximum RAS amplitude is insensitive to the phase shift and thus also insensitive to the actual depletion layer thickness. Consequently in fig. 43 the measured RAS amplitudes agrees very well with the expected ones. The latter were calculated using the effective field from eq. (32) and \(\alpha\) and \(\beta\) from eq. (34):

\[
\text{Re}(\delta\text{RAS}) = 2 < E > \sqrt{\text{Re}(d_{41})^2 + \text{Im}(d_{41})^2} \sqrt{\alpha^2 + \beta^2}. \tag{36}
\]

Hence, for practical applications using the RAS oscillation amplitudes allows for carrier determination despite the scattering of the calibration measurements.

6.6 Growth rate measurements

Even though currently the carrier concentration can be estimated best using an empirical calibration, the optical thickness can be determined with great accuracy on an absolute scale, if the real part of the refractive index \(n\) is known for a certain material at growth temperatures. This method is equivalent to Fabry-Pérot oscillations of a transparent layer on a different sublayer, which was already discussed in section 3.1. However, the contrast is not by different bulk materials (hetero-epitaxy) but only from different doping levels. The oscillations in RAS are even seen during homo-epitaxy (see fig. 37) as long as there is a small change in the doping concentration (which usually occurs anyway due to scattering in the level of bulk concentration on different substrates). Thus the growth rate can be determined directly during buffer growth without the need for marker layers and without any problems due to thickness limitations by strain.

The accuracy of the growth rate evaluation is good and its absolute error depends only on the accuracy of the refractive index. Fig. 44 shows the fitted growthrates observed using the multi-channel setup and plot the difference of the growth rates for each wavelength channel relative to the mean growthrate (from all channels). The mean deviation is less than 4% with most of the errors even close to ±1% in the region from 1.5 eV to 2.5 eV. (Large deviations usually stem from small doping levels i.e. small amplitudes.)
Fig. 44: Deviation from the growth rates of GaAs from the mean growth rate of the same run, determined by fitting multi-channel RAS data. Most data deviate much less than 4%.

Fig. 45 illustrates the changes observed during homo-epitaxy of a GaAs pnp-structure on an undoped substrate during low temperature MBE. At those low growth temperatures growth rate calibration by RHEED intensity oscillations is difficult, and RAS offers a simple tool to directly access the growth rate. The first layer in this structure was GaAs:Mn with about 2-3% Mn on a GaAs buffer layer grown at high temperature. Since the high concentrations of As-antisites from low temperature growth are changing the bulk GaAs dielectric function, oscillations are seen at the beginning in the RAS signal (top) and in the reflectance signal (bottom). During the growth of the second Sn-doped layer the RAS signal changes strongly while the reflectance signal remains nearly constant. When starting GaAs:Mn overgrowth in the third layer, the oscillations of the reflectance signal are very weak, while the amplitude of the RAS signals is nearly as strong as at the beginning of the growth. From this the growth rates respective thicknesses for each layer were determined in-situ.
Figure 45: RAS and normalized (to GaAs) Reflectance transients at 2.65 eV during the growth of GaAs doped by p(Mn)n(Sn)p(Mn) at 250°C. The inset is an enlargement of the first 1000 s. Vertical lines indicate start of LT-GaAs:Mn, switch to Sn, switch to Mn, and stop growth. The oscillation in the RAS signal are due to interference of interface band bending caused anisotropies, which is clearly seen in the second transition from n-doping with Sn back to p-doping with Mn at 3150 s (frm [E17]).
7 Roughness and Topography

7.1 Optical measurements of roughness

The intensity of scattered light carries information about the surface topography, more precisely about the power spectral density $PSD(k)$, which is the amplitudes of the Fourier transformed height profile. However, with typical in-situ setups only a small angle around the sample can be accessed. Therefore only averaged information like the roughness can be obtained.

Instead the scattering loss, more often the intensity of the scattered light integrated in all direction is used. This quantity is called total integrated scattered intensity (TIS) and is connected to scattering loss via $\Delta R = R - R_{TIS}$. For a small roughness of a ideal metallic surface the TIS was first derived 1919 by Chinmayanandam [105] and connected with the roughness 1954 by Davies [106]. In such a case the roughness is given by the square root of the ratio between scattered power to specular reflected power, or in terms of reflectance $\frac{\Delta R_{scatter}}{R}$:

$$\frac{\Delta R_{scatter}}{R} = 1 - \exp \left( - \left( \frac{4\pi d_{TIS}}{\lambda} \right)^2 \right) \approx \left( \frac{4\pi d_{TIS}}{\lambda} \right)^2$$

$$d_{TIS} \approx \frac{\lambda}{4\pi} \sqrt{\frac{\Delta R_{scatter}}{R}}$$  (37)

where $R$ is the ideal reflectance of a smooth surface, and $d_{TIS}$ is the average roughness or mean height of the surface.

Unfortunately the correlation between root mean square (rms) roughness and mean height depends on the height distribution. In case of a Poisson distribution, the rms roughness is the square root of the mean height. For a sinodial shape (or linear combinations of them) the ratio

For a dielectric material and with some information about the surface structures one would rather use eq. (3) from [24] (see sec. 3.2).

Figure 46: rms roughness of several $2\times2\mu m^2$ AFM measurements ($\times$) and mean rms roughness ($\star$) compared to ex-situ TIS data at 532 nm ($\circ$) for GaAs samples grown at different V/III ratios (for further details on the samples see [E11]). Within the error bars both method reproduce the trend of higher roughness at intermediate V/III ratios.
Figure 47: Real (left) and imaginary (right) part of the effective dielectric function calculated by Bruggeman effective medium approximation (eq. 38) between GaAs at 550°C and air for filling factors of 100%, 75%, 50% and 25%. With increasing filling factor the imaginary part decreases almost everywhere but in the near IR region of the spectra.

of mean height to rms roughness is $\sqrt{8/\pi} = 0.9003$. Most other shapes like rectangular shapes (ratio 1) or triangles (ratio $\sqrt{3}/2 = 0.866$) also give mean height values close to the rms roughness. Assuming that in reality the surface topography is composed of several such shapes, the mean height gives an useful estimate of the rms roughness. In practical applications the roughness measured by TIS and AFM correlates but differ on the absolute value (fig. 46).

We can use eq. (37) to get an estimate below reflectance cannot detect a roughness. Assuming that at 400 nm the change in reflectance must be at least 1%, then a roughness smaller than 3 nm cannot be detected. This has also been found as a limit when analysing the development of roughness during etching of Al(Ga)As by in-situ reflectometry [107]. Given that quantum wells in devices are often thinner than 5 nm, a higher sensitivity to reflectivity changes is desired.

The situation is better for ellipsometry, since it is more sensitive on an absolute scale and is measured at non-vertical incidence (which increases scattering due to a longer optical path). Moreover, multi stack analysis is a standard technique for ellipsometry. In such an analysis a rough layer is modelled using a smooth layer with an effective medium dielectric function. The dielectric function of this effective medium is given by an interpolation formula with the material below and above the interface (or air in case of a surface). An often used formula is the Bruggemann effective medium approximation [108]:

$$f \frac{\varepsilon_a - \varepsilon_{eff}}{\varepsilon_a + 2\varepsilon_{eff}} + (1 - f) \frac{\varepsilon_b - \varepsilon_{eff}}{\varepsilon_b + 2\varepsilon_{eff}} = 0$$  \hspace{1cm} (38)

where $\varepsilon_i$ are the dielectric functions of the two fractions ($a$ and $b$), $f$ the volume filling factor of fraction $a$ and $\varepsilon_{eff}$ the resulting effective dielectric function of the mixture.

Fig. 47 shows the calculated effective dielectric function using eq. (38) for a mixture between GaAs at 550°C and air ($\varepsilon = 1$). The dielectric function shows a characteristic change halfway before the first maximum (for GaAs in fig. 47 at 2.2-2.5 eV). With increasing filling factor the imaginary part $\varepsilon_2$ has a much bigger contribution at 1.5 eV than the bulk values, due to the increased loss by scattering. This has been also experimentally observed. It is quite pronounced
for InN or GaN where the first higher transition are at 4.5 eV and higher. Moreover, the spectral shape allows for a quantitative analysis of the filling factor together with the thickness of the rough layer from a spectral fit of the layer stack model to the ellipsometric measurements. Multiplying both gives the material coverage, i.e. the thickness of a smooth layer with the same amount of material. This works even for very thin layers: For InN quantum dots on GaN (0001) the volume derived from ellipsometric analysis and the volume from atomic force microscopy (AFM) measurements agreed with less than 0.1 monolayer difference [E5].

Generally, a more quantitative correct treatment of roughness can be obtained using Rayleigh-Rice theory (RRT) [110]. The scattering signal in RRT is calculated from the power spectral density (PSD). The PSD is the Fourier transformed height profile and thus contains the whole surface topography. To derive only the rms roughness a Gaussian shaped PSD is assumed. RRT give a more stable approximation to the actual roughness compared to EMA, since EMA has two parameters, the mean thickness and the filling factor which can lead to ambiguity (for an application to Si see [111]).

During in-situ measurements often only the onset of roughening is important. Ellipsometry can monitor the transition between smooth and 3D growth very accurately. This has been demonstrated for InAs quantum dots [109] and more recently for the 2D-3D transition of InGaN (see fig. 48). Since a roughness usually increases very fast during further growth, in-situ reflectance can detect the roughening too with only a short delay, at least in the UV range (e.g. for InGaAs see [109, 112]).

Thus reflectance as well as ellipsometry can detect the presence of a roughness and its magnitude. Spectral ellipsometry can also quantify a very small roughness. But in general the quantitative comparison of the optical roughness to rms roughness directly measured by AFM is only possible within a factor of two, because the calculation of the amount of scattered light requires assumption of the surface topography, either its height distribution or its power spectral density. Furthermore, optical methods sample the roughness over much larger areas.
(> 10 $mm^2$) compared to scanning techniques like AFM (typical 1-25 $\mu m^2$). In the latter a slight undulation or a few particles on the 100 $\mu m$ scale are easily overlooked but would still affect the optical scattering.

7.2 In-situ scanning tunneling microscopy

Scanning Tunneling Microscopy (STM) is a relatively new technique for in-situ analysis. Invented 1982 by Binning and Rohrer, the unique possibility to measure surfaces in real space with atomic resolution made it very quickly an indispensable tool for the surface physics together with its derivate Atomic Force Microscopy (AFM).

The principle of STM is quite simple. A tip is brought in close proximity to the surface and a bias voltage is applied between tip and surface. If empty and filled state are at same energy level (adjusted by the bias voltage) a tunnel current will flow. This tunnel current depends on the distance between sample and tip and the distribution of states in tip and surface. It is very difficult to estimate the height of the sample over the surface at a given bias voltage, even though there is an analytic expression (assuming both tip and surface are metallic) [113]:

$$I = \frac{e}{2\pi\hbar\beta^2} r^2 \left[ \phi \exp \left(-A z \sqrt{\phi}\right) - (\phi + eU) \exp \left(-A z \sqrt{\phi - eU}\right) \right]$$

with $r^2$ the contact radius, $\beta$ usually close to one (see [113]), $\phi$ energetic height of tunnel barrier, $U$ bias voltage, $h$ Planck constant, and $z$ height above sample. Unfortunately, the barrier height is very often not even close to the expected values for a given material combination, and the unknown contact radius depends also on $z$. Thus, even for a given voltage the actual scanning height depends strongly on the tip, since the type of atom determines $\phi$ and the shape determines $r$. Assuming the cone into the electrons can tunnel has an opening angle of 45° ($r = z$) then the tunnel current depends exponentially on the distance. A change in distance from 0.1 nm to 0.2 nm reduces the tunnel current by a factor of 1.8. Thus all the vibrations between the tip and surface must be well below 0.1 nm. This is usually achieved by suspending the STM with springs with a strong damping and a very rigid connection between sample and tip.

STM is a serial method, i.e. an image is taken by recording the height data on each position of the image. There are two modes of operation. For very fast STM operation with up to 25 images/second constant height mode is used (see [114] for a detailed setup description). In this mode the averaged tunnel current (and thus the tip height) is constant. The actual height is calculated from the tunnel current. Because of the exponential height dependence, imaging in constant height mode works only on very smooth surfaces. Steps or quantum dots are too high corrugation for this mode. Therefore, all realized in-situ STM for semiconductor epitaxy use the slower constant current mode, where at each point the z-position is adjusted until the tunnel current is at the target value. Then the height is directly given by the z-position. Depending on the setup, scan speed, image size, and desired spatial resolution, the time to obtain an image can vary between 10 s up to 10 minutes.

Experimentally the movement of the tip (or the sample) is realized via the piezo electric effect. Applying a voltage to a piezo-electric crystal (piezo) will cause a contraction or expansion. The exact value will also depend on the temperature. Above the piezoelectric Curie-temperature the piezo material will depolarize permanently, i.e. no longer responds to an external voltage. Thus the temperature of the piezo needs to be kept low enough and as constant as possible.
The ability to determine the position of single atoms on the surface makes STM the ultimate tool for surface growth studies, if there wouldn’t be many experimental difficulties to obtain images at growth conditions. Therefore, very few in-situ STMs have been reported that are able to measure during growth. Most of them were built for Si/Ge Molecular Beam Epitaxy (MBE) or Silicon Vacuum Chemical Vapour Deposition (SiCVD) e.g. [115, 116], because the Si/Ge material system is relatively tough and the total deposited material amounts were in the few monolayer regime.

III-V compound semiconductors however require continuous group V stabilisation. Since the tunnel current must be also less than a few nanoamperes to avoid electro-migration, any evaporation on parts of the STM is very critical, because the metals as well as the semi-metals arsenic or phosphorous are good electrical conductors and easily short-circuit the tunneling current or even the piezo. This is why the only in-situ STM for III-V compound semiconductors in MBE uses a closed protection shield to avoid deposition of As on the critical parts of the STM (fig. 49).

Nearly all in-situ STMs operated in vacuum below 1 hPa or in ultra high vacuum (UHV). So far only two setups have been realized that operate at higher pressures. One is in a hot filament chemical vapour deposition for diamond (HF-CVD) [118] and the other in a MOVPE for classical III-V compound semiconductors [E20]. The main challenges are caused by the higher pressures. For the usual pressure range in MOVPE (20-1000 mbar) only rotary pumps are available, which are a strong source of vibrations. Additionally, at these pressures all gases conducts sound well. Therefore great care must be taken to have a very rigid connection between tip and sample, which must be decoupled from the rest of the setup.

Vibrations only degrade image resolution, while overheating the piezo will destroy the STM. Thus heat management is most critical for an in-situ STM. In UHV heat transport is mostly by radiation. A shadow plate and using materials with low heat conductance (like ceramic) are sufficient for MBE (see fig. 49). However, in MOVPE heat transport by heat conduction of the gas phase is much more effective. Additionally, the growth temperatures required by precursor decomposition are typically 100°-300°C higher in MOVPE. Together, the heat load on the STM system is about 100-1000 times higher. For instance, Voigtländer et al. heated their samples in UHV via direct current heating with less than 5W compared to the 2.5kW radiative heating of the MOVPE system. Even in the diamond HF-CVD system (operated a
40 hPa) the temperature regime was limited, despite the fact that the heater was a small 12 mm boron-nitride plate [118].

Therefore in-situ STM in MOVPE requires advanced cooling. Several approaches very tested (fig. 51) but only an active cooling shield (as shown in fig. 50) kept the piezo and coarse approach motor below their critical temperatures. When properly designed the active cooling allows for nearly undisturbed growth up to 650°C (see fig. 52 for operation at 600°C). The cooling shield uses nitrogen gas, which is passed through a copper tube, which is immersed in
liquid nitrogen. By adjusting the nitrogen flow, the temperature of all parts can be adjusted to avoid changes of the scan region by temperature changes.

In MBE the species are supplied by atomic or molecular beams. Therefore, a free path from the effusion cells to the surface is mandatory. By using a more oblique angle of the cells the shadowing can be pushed to a region far from the scanning region (as indicated in fig. 49). Additionally, behind the tip a shadow cone exist, where no growth will occur [119]. In MOVPE due to gas phase diffusion shadowing is not present, since the molecules move laterally with the speed of sound which is nearly 100 times faster than the gas flow velocities.

In MBE as well as in MOVPE there are expected to be different physisorbed species on the surface. In case of GaAs MBE As$_4$ or As$_2$ clusters are present and diffuse on the surface. For imaging in MBE a high scan speed is needed or only a random noise due to hopping processes of atoms/clusters diffusing under the tip is recorded. For MOVPE there are additional carrier gas molecules. A hydrogen molecule is just $0.2 \times 0.8 \text{nm}^2$ in size and can pass under the tip. Nevertheless the influence of the carrier gas on the tunneling process was not observed. There are also many STM that operate in air, so the carrier gas is not critical.

The tip of any in-situ STM for compound semiconductors must be made of tungsten, since the often used Pt forms eutectics with most III-V semiconductors above 300°C. However, tungsten forms a strong oxide at air, which must be removed before stable tunneling can be established. To alleviate tip preparation in the growth chamber, pre-annealing can help. But a reliable and quick method to prepare a tip for in-situ measurements of III-V semiconductors is still a challenge.

The need for group V stabilisation has another impact on the tunneling current. The use of a voltage independent barrier height $\phi$ in eq. (39) is an oversimplification. Instead, the tunnel current is proportional to the combined density of states of tip and surface integrated up to the current bias voltage. Thus, from current/voltage measurement at a fixed distance the local band gap (and thus composition of the top layer) or surface states can be determined. For in-situ STM of compound semiconductors the apparent height will look different if the tip is formed by a group V atom or a metallic group III atom. Group V atoms are present in

61
abundance due to group V stabilisation. Hence most of the time, the tip is made from a group V atom. Since group V atoms are only loosely bound, very often the tip loses or picks up an atom, leading to unstable tunneling. For instance, in the sequence of images in fig. 53 the first half of image (a) and the last half of image (c) shows a blurred image due to such an instable tip.

During growth the imaging is usually better, since there are many metal atoms present, which stick to the tip more strongly and provide a more stable tunneling. The same is true on gallium rich surfaces without group V stabilisation [E9]. Also using tunneling currents below 1 nA stabilizes the tip, and higher currents can lead to electro-migration [E20, E22]). Nevertheless obtaining a sharp and stable tip and fast imaging are the main experimental challenges for in-situ STM. Still, within these limits in-situ STM can give insight into processes not possible by any other method.

Tsukamoto et al. published a series of papers on in-situ STM investigations on the formation of InAs quantum dots in MBE [E23, 120, 121, 122]. The low growth rate in MBE allowed for images in 0.05 monolayer intervals, despite the fact that an image took more than a minute to register. The images and especially the correlation of the frequencies and the RHEED signal along [1\bar{1}0] direction are remarkable, as the latter proved that in-situ STM did not influence the growth at all [120].

In MOVPE, growth processes are much faster, since higher growth temperatures are needed for precursor decomposition. Because In desorbs already at 500°C within minutes, very low growth rates are not possible. Hence, in-situ STM in MOVPE is limited to processes on longer time scales. A nice example of the power of in-situ STM in MOVPE is the Ostwald ripening of InAs quantum dots at 475°C. Fig. 54 shows a sequence of 6 images recorded between 42 and 85 minutes after quantum dots growth. Fig. 55 shows the development of the density of different sized structures for the in-situ data and from ex-situ AFM measurements. The sum of structures in both set of data decreases reciprocal, which indicates that the process of leaving an island or attaching to an island is the limiting process for ripening (for details see [E24]).

Since the tip has a small diameter (the tungsten wires are less than 0.3 mm diameter), optical in-situ monitoring is not affected by in-situ STM operation. The combination of the two techniques would allow for fast in-situ monitoring (with optical techniques) and with topographic resolution (from the STM). Such a combination of in-situ STM with in-situ ellipsometry has been recently demonstrated [123].
Figure 54: Six sequential in-situ STM images in MOVPE after deposition of 5 s 0.5 Pa TMIn on GaAs (001) at 475°C under 2.5 Pa tBAs stabilisation. The image sequence runs left to right at 260 s per image, while each image covers 5000 nm × 5000 nm and is rastered from top to bottom. The different sized structures are marked by different circles. The number of quantum dots reduces strongly during annealing (from [E24]).

Figure 55: Development of the surface density of different structures measured in-situ (left) and ex-situ (right) (from [E24]). The density of all structures decreases like $t^{-1}$ indicating that attachment/detachment is the limiting process.
8 Conclusion

With the progress of compound semiconductor devices to smaller and smaller scales into the regime where even single atoms make important contributions, in-situ monitoring becomes more and more indispensable. Current optical in-situ monitoring using ellipsometry can monitor thickness changes within single monolayer resolution for hetero-epitaxy. Reflectance Anisotropy Spectroscopy (RAS) can even monitor homo-epitaxy with monolayer precision, provided the growth occurs in 2D island mode. The emerging technique of tandem-interferometry has even the potential to measure any reflecting surface/interface with nanometer precision.

Composition analysis by the increase of growth rate is well established, but fails on several material system (like group V ternaries, or InGaN). For some materials – like AlGaN – a database is established, and the composition can be extracted from the refractive index. Other works combining RAS and reflectance have been demonstrated. Bowing measurements using deflection of visible light have in principle the potential to determine solid composition via the stress imposed by a strained layer. But in practice bowing measurements are mainly used for strain management. In general in-situ monitoring of composition of strained ternary or even quaternary materials is still a challenge.

RAS can determine the carrier concentration of a given surface for zincblende III-V semiconductors. The mathematical models for both static surfaces as well as the growth induced changes have been presented in this work. For GaAs and InP reference data are provided in this publication, for other material reference data still have to be collected. Using those calibration data, carrier concentrations can be obtained by numerical analysis of RAS spectra in a similar manner to the numerical analysis of ellipsometry spectra to obtain solid compositions i.e. by spectral fitting. For concentrations higher than $10^{17} \text{cm}^{-3}$ this results in errors less than 10%.

Using transient RAS measurement during growth at a single wavelength in the near transparent region below the $E_1$ transition, doping profiles can be created with a resolution of less than 50 nm. (The resolution depends on the actual wavelength and doping concentration). Moreover, from RAS transient data the growth rate of thick layers can be determined with an error of one percent during homo-epitaxy in a very wide range of conditions. This is so far the only method of homo-epitaxial growth rate monitoring that can work even for multi wafer reactors.

All optical methods, reflectance, ellipsometry, or RAS cannot determine the surface topography, but only the roughness semi-quantitatively. In-situ measurements of the surface topography by in-situ scanning tunneling microscopy are the only way to extract those data. While in-situ STM is a very powerful technique, the experimental challenges and constraints on the growth system are large. Since optical in-situ monitoring, and especially reflectance, is very easy to adapt to almost any growth system, in-situ STM will be only applied to solve certain scientific problems.

To overcome some of the limitations of reflectance, new optical in-situ methods have been developed. In-situ curvature monitoring is already used to increase yield during growth of GaN based LEDs, but a quantitative analysis of to extract composition is still challenging. The recently demonstrated low-coherence tandem interferometry still needs to prove its operation on multi-wafer machines, but could be another important optical in-situ technique.
Acknowledgement

It is a pleasure to acknowledge the contributions of a large number of hard working diploma students, PhD students, and coworkers at the Technische Universität Berlin, the Ferdinand Braun Institute in Berlin and at the National Institute for Materials Science, Tsukuba, Japan. In particular I would like to acknowledge M. Zorn, C. Kaspari, and R. Kremzow, who performed most of the MOVPE experiments with me together in the groups of W. Richter, M. Weyers, and M. Kneissl. I thank S. Tsukamoto for his invitation to Japan, and his introduction into in-situ STM and MBE growth and for several nights spent at the STM-MBE. Without W. Richter I would have missed the exciting topic of in-situ monitoring and I would like to express my thanks to him for many helpful and stimulating discussions.

This work was supported by the Deutsche Forschungsgemeinschaft in the Sonderforschungsbereich 296 and 787 and contract Pr711-2 and the Japanese Society for the Progress of Science.
References

own references since PhD cited in this work


other references


Reprints of paper cited in this review