Chapter 2
High-Resolution Electron Microscopy of Semiconductor Heterostructures and Nanostructures

David L. Sales, Ana M. Beltrán, Juan G. Lozano, José M. Mánuel, M. Paz Guerrero-Lebrero, Teresa Ben, Miriam Herrera, Francisco M. Morales, Joaquín Pizarro, Ana M. Sánchez, Pedro L. Galindo, David González, Rafael García, and Sergio I. Molina

Abstract This chapter briefly describes the fundamentals of high-resolution electron microscopy techniques. In particular, the Peak Pairs approach for strain mapping with atomic column resolution, and a quantitative procedure to extract atomic column compositional information from Z-contrast high-resolution images are presented. It also reviews the structural, compositional, and strain results obtained by conventional and advanced transmission electron microscopy methods on a number of III–V semiconductor nanostructures and heterostructures.

2.1 Introduction

Transmission electron microscopy (TEM) and associated techniques based on the analysis of a transmitted electron beam through a specimen are basic tools in the field of Materials Science and Engineering. These techniques give information on the structure, composition, strain, and even the bonding in materials from the mesoscale to nanoscale, presently reaching even atomic column resolution as routine task. Excellent books and reviews exist in the literature on this field [1–3]. Here, we focus on some methodologies mainly based on high-resolution electron microscopy techniques developed by the authors, as well as the application to semiconductor materials of these and other methodologies based on the analysis of an electron beam transmitted through an electron-transparent sample. The studied semiconductor materials consist of III–V semiconductor nanostructures and heterostructures grown by molecular beam epitaxy (MBE).

D.L. Sales (✉) · S.I. Molina (✉)
Departamento de Ciencia de los Materiales e I.M. y Q.I., Facultad de Ciencias, Universidad de Cádiz, Campus Río San Pedro, s/n, 11510 Puerto Real, Cádiz, Spain
e-mail: david.sales@uca.es; sergio.molina@uca.es

2 High-Resolution Electron Microscopy of Semiconductor

Fig. 2.1 Signals generated from the interaction between a high energy electron beam and a electron transparent (very thin) sample. Most of them are able to be detected by TEM symmetry, assuming that the imaging and emission systems are adequate. Then, HRTEM can be exploited to image the atomic structure of thin specimens.

2.1.1.2 Scanning Transmission Electron Microscopy and Z-Contrast Imaging

This technique also collects information from the transmitted beam to characterize the micro- and nanostructure of the material, although in this case the incident beam is concentrated (focused) in a tiny probe which is scanned over the sample. Thus, each point of the image is obtained sequentially (pixel to pixel), and several detectors record multiple signals. Therefore, this technique is optimal for extracting maximum information from a single sample point and detecting multiple signals simultaneously [4], such as bright field, dark field, or energy loss spectra as shown in the example configuration in Fig. 2.2.

This basic difference in image acquisition makes scanning transmission electron microscopy (STEM) settings different with respect to conventional TEM. The most important differences are as follows:

1. There is no need of post-specimen lenses to form the image, which makes this technique less sensitive to chromatic aberrations.
2. Scanning coils are needed to move the beam across the sample.
Fig. 2.2 Schematic of an aberration-corrected STEM with electron energy loss spectrometer. (1–3) Examples of HAADF and BF images, and EELS spectrum, respectively (adapted from [5]).

STEM images can also be bright field or dark field, depending on whether the signal comes from the direct beam or one or more diffracted beams, respectively. Normally, these microscopes have installed a central detector to capture the direct beam, and annular dark field (ADF) detectors, with the hole centered on the optical axis, to capture diffracted beams.

Probably, the main advantage of the STEM configuration, in addition to the variety of available signals, is the possibility of Z-contrast imaging, where Z refers to the atomic number. A Z-contrast image is formed by collecting high-angle scattered electrons using an annular dark field detector (HAADF) [6, 7]. Images acquired in this way very well approximate the convolution of the intensity probe with the object function representing the sample [8], being considered as incoherent images—i.e., formed without interaction of the scattered waves from surrounding atoms or atomic columns. Then, high-resolution images acquired with this detector are connected in a more direct and simple way to the material structure than a phase-contrast image, being the last subjected to the complexity of coherent interference interactions [8, 9]. The underlying physical principle is that atoms with higher
atomic number scatter more prominently at higher angles than lighter ones. When the scattering is Rutherford scattering, this is proportional to $Z^2$, the image contrast being modified in this ratio. Although the precise contrast depends on experimental setup, and is affected by several factors such as the channeling effect or the detector size, it is demonstrated in the next section that this approach is adequate to perform a quantitative compositional analysis of III–V semiconductors.

### 2.1.1.3 Aberration-Corrected Electron Microscopes

Magnetic lenses of electron microscopes are far from being perfect. They are still affected by aberrations, mainly spherical and chromatic ones. The incorporation of aberration correctors in new generation microscopes has made a breakthrough in the resolution limit, which seemed to be stacked. These correctors were applied to conventional TEM microscopes (CEOS) and STEM (Nion and CEOS) for correction of geometric aberrations. One of the most obvious improvement of the incorporation of aberration correctors in STEM, additionally to the increased resolution [4, 10], is the efficiency of images acquisition with enhanced signal-to-noise ratio.

### 2.2 Compositional Quantification Column-To-Column in III–V Semiconductors

Chemical composition is a crucial attribute of functional materials. Specifically in the case of nanostructured heteroepitaxial lasers, chemical composition of the nano-objects constituting their active layers strongly influences the emission energy of the final device. Normally, the real composition differs from the nominal one (i.e., from the desired one, for which the growth process is set up) [11]. Segregation and inter-diffusion occurring during upper confining layer growth are two important processes in this scenario. Not only real or effective composition is required, but also the compositional distribution. Shumway et al. show that for InGaAs/GaAs quantum dots (QDs), a difference of 10% of In content (from 50 to 60%) leads to a wavelength redshift of around 20–40 meV [12]. This strong effect justifies the effort on developing new methodologies for compositional analysis at the nanoscale.

Energy dispersive X-ray spectroscopy (EDX) and electron energy loss spectroscopy (EELS) are common analytical techniques based on electron–matter interaction. Although these techniques have greatly improved in recent years [13, 14], they still have some limitations, namely (1) the intrinsic electron probe broadening effect, (2) a poor energy resolution, and (3) the need for high acquisition times, and hence a high stability (i.e., radiation-resistant materials, extreme stability of the microscope setup, hydrocarbons free sample surfaces, etc.). It is therefore necessary to look for another compositional sensitive signal, requiring less acquisition time with high resolution.
HAADF images provide significant information on the position of atomic columns in a crystal with sub-angstrom resolution (when the microscope is equipped with aberration correctors), requiring a low acquisition time (high-quality images in less than 16 s). Furthermore, contrast in HAADF images is approximately proportional to the square of the average atomic number of the atomic column. Therefore, this technique is also called Z-contrast imaging.

Several works in the literature show qualitative compositional mapping with high resolution from Z-contrast images of various material systems [15, 16].

In this section, a method to determine quantitatively with atomic column spatial resolution the composition of a semiconducting ternary alloy is presented. The method is based on the analysis of local integrated intensities from aberration-corrected Z-contrast images, and was first presented in [17]. It is necessary to use a set of reference samples of known composition to quantify the relationship between the Z-contrast intensity with the thickness and composition of the material analyzed. This method is applied as an example to an InAs$_x$P$_{1-x}$/InP$_x$ alloy. It uses a series of InAs$_x$P$_{1-x}$ epilayers grown on InP(001) by Molecular Beam Epitaxy (MBE) as reference samples.

2.2.1 Reference Samples Study

Three MBE samples consisting of InAs$_x$P$_{1-x}$ epilayers grown on InP(001) were considered. The As molar fraction was precisely measured by high-resolution X-ray diffraction (XRD), giving $x = 0.27, 0.59,$ and $0.87$, respectively, for each sample. Specimens for Z-contrast imaging were produced for cross-sectional view by mechanical thinning and Ar$^+$ ion milling using a precision ion polishing system (PIPS). The Ar$^+$ PIPS was used in such a way that the beam always hits the sample from the same direction, which has been found to produce cleaner surfaces in the region of interest. A beam energy less than 3.5 kV has been selected to reduce amorphization of InP. As a final step, the sample was introduced in a Fischione ion mill at 15 and 0.5 kV to reduce surface damage. The thickness of each analyzed region was determined from the analysis of the corresponding spatially resolved low-loss EELS signal, by using the mean free path determined for each $x$ composition of the three InAs$_x$P$_{1-x}$/InP reference samples, according to the method described in [18].

A characterization procedure was defined to ensure (1) the systematicity and reproducibility of the study, (2) the minimization of radiation damage, and (3) the minimization of errors by deposition of hydrocarbons during the observation. The procedure is summarized below:

- Previous checking:
  - Although, in principle, the manufacturer asserts that the intensity ratio is linear, the detector settings (gain and black level) were not changed between sessions.
The objective lens current remains constant through the analysis of the three samples, which means that distance from the virtual objective aperture does not vary.

- Acquisition of images without sample to evaluate the detector black level.
- Astigmatism correction on an amorphous close to the area of interest.
- Acquisition of spectrum image centered at the zero loss peak, for the thickness determination of the area of interest.
- Acquisition of HAADF images from various points of the sample with different thicknesses. Images are to be acquired in the interface, and both InP and the InAs\textsubscript{x}P\textsubscript{1−x} layers must always appear.
- Localization of the HAADF images within the thickness map, identifying the mean specimen thickness for every image.

### 2.2.2 Image Analysis and Comparative Index

This paragraph describes how to analyze the acquired HAADF images in order to extract chemical information. The starting point is a HAADF image from the InAs\textsubscript{x}P\textsubscript{1−x}/InP interface with atomic column resolution and known composition \((x)\) and thickness \((t)\), as the one shown in Fig. 2.3a.

HAADF detectors normally register a background signal even if the sample is not being imaged. Then, the first step is to subtract the mean value of the detector black level, which depends on the detector gain and might differ from one microscope session to another. Next, pixels with the local intensity maxima associated to In atomic columns (higher \(Z\)) are detected on the image. For this, the Peak Finding tool of the Strain Determination Software is used [19]. Once these maxima intensity

![Fig. 2.3](image) (a) Raw data high-resolution aberration-corrected \(Z\)-contrast image taken along [110] from the InAs\textsubscript{0.87}P\textsubscript{0.13}/InP interface. (b) Area selected (blue square of nine pixels) for each group V column to integrate intensities and calculate \(R\) ratios, drawn on the projected unit cell. Projected positions of In and group V atomic columns are represented with red and green color circles. (c) Integrated intensities in the area selected in (b) superimposed on the high-resolution \(Z\)-contrast image from which the intensities were determined for each analyzed group V column. Note that the intensity distribution is very homogeneous throughout the whole analyzed image.
pixels are located, it is straightforward, with the help of image processing software, to measure the intensities integrated within a selected area of the projected unit cell, as shown in Fig. 2.3b. In this case, the number of pixels that an InAs\textsubscript{x}P\textsubscript{1-x} unit cell occupies in the 500,000 \times image is 18 \times 13 pixels. Taking into account the crystal orientation, and dimensions of the unit cell, pixels that correspond to the position of the anion are localized within the image. These pixels will be, in principle, more sensitive to the change in \textit{x} (arsenic molar fraction) in an InAs\textsubscript{x}P\textsubscript{1-x} alloy with variable composition.

The resulting integrated intensities of an InAs\textsubscript{0.87}P\textsubscript{0.13} region are shown superimposed on the experimental image in Fig. 2.3c. The integrated intensity for each pair of In-AsP columns is shown in this figure by a colored circle located in the brightest pixel, near the center of the projected position of the In atomic column. The integration area used is the one marked with a blue box in Fig. 2.3b. It is noted that the dispersion of the value is very low. This process is repeated for an InP area of the same image. Finally, two integrated intensity mean values, one for the InAs\textsubscript{0.87}P\textsubscript{0.13} ($I_{S0.87}$) and one for the InP($I_0$), are determined. Defining the normalized integrated intensity parameter as given in (2.1), the ratio $R_e(0.87)$ is obtained, which defines the image in Fig. 2.3a.

$$ R_e(x) = \frac{I_{Sx}}{I_0} \tag{2.1} $$

The error is minimized in this way, as the specimen thickness variations and other local fluctuations—that may contribute to the intensity values—are small within the area covered by a single HAADF image.

The selected integration area for the analysis has to be (1) linearly dependent on the As content, (2) independent of specimen thickness, and (3) independent on the surrounding atomic columns so that the signal is essentially due to the contribution of As and P from the selected atomic column. These three aspects were evaluated in several integration areas, the marked one in Fig. 2.3b resulting the most satisfying.

The same procedure for measuring integrated intensities of hundreds of anion–cation column pairs was performed in regions of different specimen thickness in the three reference samples (different \textit{x}). Then, the values of $R_e$ for certain \textit{x} are obtained averaging the normalized integrated intensity values from different thickness areas of the same reference sample. The maximum observed standard deviation of $R_e(x)$ due to the specimen thickness was 1.5%. The $R_e - x$ plot shows a clear linear dependency (see Fig. 2 of [17]). A fitting equation with a correlation coefficient of 0.9974 is obtained.

Thus, measuring the values of $R_e$ according to the procedure in a heterogeneous alloy InAs\textsubscript{x}P\textsubscript{1-x}, it is possible to determine the composition, atomic column to atomic column, simply from the linear regression equation relating $R_e - x$ in the calibration samples.
2.2.3 Simulation of Integrated Intensities

Image simulation would help to corroborate the conclusions of experimental analysis when a good correlation between calculated and experimental images exists. However, quantitative matching between experimental and simulated aberration-corrected Z-contrast images has not been reached up to now. The inclusion of spatial incoherence has been shown to give quantitative agreement between non-aberration-corrected HAADF-STEM images and theoretical simulations [20, 21]. It is shown here that, using the same approach, a significant improvement in the correlation between calculated and experimental normalized integrated intensities is obtained in the InAs$_x$P$_{1-x}$ ternary semiconductor alloy, but residual discrepancies still remain [22].

An important difference between aberration-corrected and non-corrected Z-contrast images is the dependence of normalized intensities on the thickness of the electron-transparent specimen. This dependence is significant for non-corrected images, whereas it is very low for a range of specimen thicknesses in aberration-corrected images. As shown in Fig. 2.4, the $R$ values calculated without including the effect of spatial incoherence clearly overestimate the experimental $R$ values. On the contrary, experimental and calculated $R$ values are much closer when a Gaussian is used to take into account the effect of spatial incoherence. In order to determine

![Figure 2.4](image_url)

**Fig. 2.4** Experimental and simulated $R$ ratios vs. As atomic content ($x$), with and without introducing the effect of spatial incoherence, for three compositions of the InAs$_x$P$_{1-x}$ alloy (reproduced from [22])
the optimum value of the width (standard deviation) of the Gaussian function that gives the best fit between experimental and calculated images, an error function defined as the sum of squared differences between simulated and experimental \(Z\)-contrast images is considered. The minimum value of the error function occurs when the optimum width of the Gaussian function is 0.06 nm, which corresponds to a source size with a full width at half maximum (FWHM) of 0.14 nm.

Integrated intensities in aberration-corrected HAADF-STEM images have been simulated by including the effect of spatial incoherence. In this way, intensities in both simulated and experimental images acquired in \(\text{InAs}_x\text{P}_{1-x}\) thin layers with calibrated compositions show improved correlation. A residual error of \(\sim10\%\), however, cannot be accounted for, and the most obvious origin of this error would be the effect of non-round aberrations that were not recorded when the experimental data were obtained. These non-round aberrations are expected to be quite dependent on the exact tuning of the aberration corrector, and further work is required to determine whether they could quantitatively account for the residual discrepancy. The calculated normalized intensities are found to be almost independent of the specimen thickness in a practical range of useful thickness values, in good agreement with the experimental results, and in contrast with the observed behavior in non-aberration-corrected images. Based on dynamical simulations of the channeling process that electron beams suffer after propagating along an atomic column, our findings are explained and predicted to be valid for a wide range of materials studied by aberration-corrected \(Z\)-contrast imaging. Our results pave the way for a direct interpretation of aberration-corrected high-resolution \(Z\)-contrast images, in terms of elemental compositions with atomic column spatial resolution.

### 2.3 Strain Measurements from High-Resolution Electron Microscopy Images

Recent advances in digital imaging have offered the possibility of locally determining the elastic strain of materials from HRTEM images very precisely. Strain mapping can be used to analyze materials at the atomic column level, measuring local displacements and so revealing lattice translations, dislocations, and/or rotations.

#### 2.3.1 Techniques

Algorithms to map elastic strains from HRTEM images may be classified into two broad categories: real space methods (also called peak-finding techniques) based on the detection of peaks of intensity in real space, and Fourier methods based on Takeda’s technique for phase retrieval [23].
Peak-finding techniques were first described by Bierwolf et al. [24] and further developed by Jouneau et al. [25], Seitz et al. [26], Kilaas et al. [27], Robertson et al. [28], and Rosenauer et al. [29]. All these techniques were based on superimposing a two-dimensional reference lattice extrapolated from a non-distorted region of the material to the experimental one, built up from the set of intensity maxima in the HRTEM image, and calculating the local discrete displacement field at each node [30]. Subsequently, the strain field is calculated as the derivative of the displacement field. Peak pairs [19] can be considered an improved version of previous peak-finding techniques in the sense that the reference lattice is not extrapolated, but calculated at each pair of neighboring peaks. The neighborhood function is calculated in the affine-transformed space described by the reference basis vectors. This transformation greatly reduces potential errors in the determination of pairs and allows the proper calculation of the displacement and strain fields in the presence of defects.

Geometric phase [31] makes use of the Fourier transform fringe analysis method. The method was first introduced by Takeda et al. [23] and adapted by Bone et al. [32], who provided a full 2D version. This technique was applied a decade later to lattice-distortion measurement in crystallography [31, 33], achieving excellent results [34]. It consists of applying an asymmetric filter at the chosen peak (Bragg spot) of the Fourier transform of a HRTEM lattice image and performing an inverse Fourier transform. The resulting complex image holds the phase information which can be calculated by the logarithm of an inverse tangent calculation. The spatial resolution of the phase map in the Fourier transform fringe analysis method is determined by the size of the filter’s window. Let us note that a larger window provides a better frequency resolution but a poorer spatial resolution, and vice versa. Hence a suitable window size must be selected for a compromise between the two resolutions. When noisy strain maps are obtained, researchers should resist the temptation to produce smoother maps by decreasing window size, because spatial resolution (which is directly linked to window size) could be, in this case, too small, and accuracy of strain results would be seriously affected.

### 2.3.2 Methodology

In order to determine strain from a HRTEM image, the first step is image filtering, a critical issue in peak-finding methods. All techniques work by first determining a base vector in an unstrained region and calculating the corresponding displacement map with respect to this reference. It is important to note that the reference area should be taken on the same image, but outside deformed regions, because strain is determined from the distortions in the image compared to the reference (unstrained) vectors. By calculating the displacement map across two non-collinear directions, it is possible to determine the displacement field to derive the strain field easily.
2.3.2.1 **Basis Vectors Determination**

In real space, image filtering is usually needed not only to reduce noise in the image, but also to smooth the image around the peaks. Special attention should be paid to the filtering stage in order to avoid loss of significant information. Currently, different techniques are frequently used, such as low-pass, Wiener, and/or Bragg filtering. Once the image has been filtered, the direction of a basis vector is obtained by selecting two neighboring peaks in the desired direction within the reference area. The coordinates of the vector connecting both peaks define the basis vector. Two basis vectors are needed in order to be able to calculate the displacement map in any direction. In order to improve the accuracy, the average base vector in a desired region of the reference area is calculated.

In Fourier space, the direction is obtained by selecting a peak of intensity in the power spectra of the image. Once the peak location has been chosen, its position determines the coordinates of the base vector. In order to improve the accuracy, Takeda’s technique is used to calculate the 2D phase. A slight error in the location of the peak will produce a flat but non-constant phase in the reference area, assuming it is unstrained. Discrete character of images, where the maximum resolution is limited by pixel size, usually generates a non-constant phase because of the limited resolution of the power spectra. However, it is possible to improve the accuracy by fitting a plane to the phase and using its coefficients accordingly to modify the coordinates of the base vector so that the refined phase is constant in the chosen region (reference area). In doing this, subpixel resolution is achieved for the coordinates of the base vector.

2.3.2.2 **Displacement Field Calculation**

Once the basis vectors have been obtained with subpixel resolution, the displacement field can be calculated using different approaches.

In the peak pairs approach, all the intensity peaks are automatically located using eight-neighborhood maxima detection at the pixel level. In order to get sub-pixel resolution, a 2D quadratic function is fitted by using gray levels in the neighborhood of each peak and solving the resulting equations for the maxima. Using all pairs of peaks, the displacement field is calculated at the positions of atomic columns using the difference between local peak pairs and basis vectors.

In the geometric phase approach, the phase is calculated using Takeda’s approach to determine the displacement field directly.

2.3.2.3 **Strain Field Calculation**

The strain field components can be easily calculated by solving the following set of linear equations:
\[ u_x = a_x e_{xx} + a_y e_{xy} \]
\[ u_y = a_y e_{yy} + a_x e_{yx} \]
\[ v_x = b_x e_{xx} + b_y e_{xy} \]
\[ v_y = b_y e_{yy} + b_x e_{yx} \]

where \((a_x, a_y)\) and \((b_x, b_y)\) are the coordinates of the basis vectors; \((u_x, u_y)\) and \((v_x, v_y)\) are the coordinates of the displacement field in the direction defined by each basis vector, respectively; and \(e_{xx}, e_{yy}, e_{xy}, \) and \(e_{yx}\) are defined as follows:

\[ e_{xx} = \frac{\partial u}{\partial x}, \quad e_{xy} = \frac{\partial u}{\partial y}, \quad e_{yy} = \frac{\partial v}{\partial y}, \quad e_{yx} = \frac{\partial v}{\partial x}. \]  

(2.3)

### 2.3.2.4 Alternative Applications

Strain mapping is often used not only to determine strain and identify defects and/or distortions in the image, but also to rectify distorted images. It is well known that CTEM projector lens systems, STEM scanning systems, and CCD cameras or scanners often have distortions that affect the images they form. When these distortions are systematic, it is possible to eliminate them by measuring the displacement map of an unstrained specimen and using it to define a warp function to restore any distorted image taken in the same conditions [35].

### 2.3.2.5 Problems

Strain determination from HRTEM images depends upon the assumption that the intensity maxima in the image and the position of atomic columns in the sample are directly related. There are many situations where this assumption is violated, and image conditions should be carefully optimized. The most important rule to determine optimal conditions is choosing, when possible, thickness and defocus values where its variation does not modify the location of the intensity peaks in the image. These values should be determined by image simulation. Other practical rules are choosing thickness and defocus conditions where fringe contrast is high, and avoiding regions where the fringe contrast changes rapidly. Let us note that frequently, Scherzer defocus might not be the optimal choice. Error quantification in strain mapping methods has shown reduced accuracy near abrupt interfaces, independently of the strain mapping technique [36]. Therefore, depending on the chosen imaging conditions, strain profiles may contain severe artifacts. Nevertheless, average strain can be determined with sufficient accuracy on experimental micrographs of thicker layers [37]. The reliability of strain mapping from HRTEM images is also affected by thin foil relaxation and lattice fringes shifting due to variations in composition in the sample. In order to take into account all these effects, accurate simulation using Molecular Dynamics or Finite Element
Method is often a requirement to corroborate the measured strain from HRTEM images. Only when these factors are taken into account, the method offers a reliable way of extracting quantitative strain information at the atomic scale [38, 39].

2.3.3 Applications

Strain mapping has been used in different fields during the last few decades. Some recent applications of peak pairs include the analysis of confinement potential of III–V quantum dots [40], the determination of surface-induced disorder in magnetic FeCoPd nanoparticles [41], the determination of precise atomic-resolution compositional distribution in HAADF [42], the control of growth directions of InAs nanowires on Si substrate [43], the determination of composition of In$_x$Ga$_{1-x}$N from strain measurements [44], etc. Geometric phase has also been applied to a number of different problems, including the measurement of strain in silicon devices [45], the mapping of strain in MOSFETS at the nanometer scale [46], the measurement of strain in strained-silicon metal-oxide-semiconductor field-effect transistors [47], and the measurement of strain in aberration-corrected HAADF-STEM images of InAs/GaAs dot-in-well heterostructures [35].

Different plug-ins for DigitalMicrograph™ implementing strain mapping algorithms have been developed by HREM Research [48], including peak pairs analysis (PPA) and geometric phase analysis (GPA). Its features include high-resolution peak detection (only available in PPA), image filtering, displacement and strain map calculation, atomic column (only available in PPA) and/or continuous strain maps, histogram and lattice point’s analysis (only available in PPA), and geometric distortion determination and correction.

2.4 Results on III-Sb Hetero- and Nanostructures

Fiber-optic systems require preferably uncooled high-speed lasers and detectors which are matched to the 1.55-μm fiber transmission window. So far, data rates in telecommunication applications at 1.55 μm are limited by the use of costly InP-based high-speed devices [49]. Huge research efforts have been made to develop improved and lower cost optoelectronic devices on inexpensive large-area GaAs substrates. Thus, GaSb/InAs/GaAs heterostructures—with or without the growth of an intermediate GaAs barrier layer—are emerging technologies to lower costs and increase broadband and network speeds. Currently, there is an increasing interest in combining the electronic properties of more than two semiconductor compounds. More specifically, heterostructures which combine InAs with antimonides offer outstanding band-gap engineering possibilities which are not available in the more traditional InAs/GaAs system. Of particular interest are
GaSb/InAs/GaAs and GaSb/GaAs quantum dot heterostructures due to the type-II band alignment expected between arsenides and antimonides [50].

The semiconductor-based optoelectronic devices face significant challenges in material quality and fabrication routes which must be solved to realize commercially viable technologies. One of the main barriers is the effect of the surface segregation, which can impede the achievement of abrupt heterointerfaces. It has been demonstrated that a competitive incorporation between As and Sb atoms occurs during the growth of GaSb on GaAs [51]. Furthermore, in GaSb/InAs/GaAs heterostructures, both cation and anion segregation can occur. A clear blocking of indium incorporation by antimony has recently been reported in the literature [52]. These effects can degrade interface sharpness considerably, and therefore alter band offsets.

Significant difficulties could arise in systems where quantitative compositional measurements at the nanometer scale or below are required. The use of a very thin specimen (tens of nm or less) in TEM means that large-scale interface roughness does not affect compositional analysis, but the interpretation of images and analytical spectra is not straightforward. An exhaustive compositional analysis requires the combination of different techniques, such as EELS, HAADF images collected by STEM, and conventional dark-field imaging. This section focuses on the compositional determination of GaSb/InAs/GaAs and GaSb/GaAs systems, mainly using spatially resolved EELS and aberration-corrected HAADF-STEM.

The combination of low-loss EELS and high-resolution HAADF-STEM can clearly demonstrate Sb incorporation inside GaSb-capped InAs QDs grown on GaAs substrates by MBE. These techniques show the presence of the four elements (Ga, In, Sb, and As) in the QDs, presumably as a result of an intermixing process during the growth of the QDs and the GaSb capping layer. In this sense, the more intense contrast in the HAADF-STEM image of Fig. 2.5a is associated with high Z-number, depicting the presence of In-rich cationic columns and/or Sb anionic columns inside the observed quantum dot. For a statistical analysis of the integrated intensities associated with anionic and cationic columns, the process described in Sect. 2.2 was followed. The intensity profile along the growth direction shows a well-defined maximum which supports the data obtained by EELS (see [52]).

The intensity maxima of anionic and cationic columns are located in the same areas where EELS measurements showed maximum In and Sb concentration. The analysis also demonstrated a tendency of the Sb to extend toward the apex of the QDs. Although no clear GaSb capping layer was observed in the heterostructure, the analytical results showed a GaAs$_x$Sb$_{1-y}$ on top of the Ga$_x$In$_{1-z}$As$_y$Sb$_{1-y}$ core.

A possible route to optimize the luminescence properties of GaSb-capped InAs/GaAs quantum dot heterostructures is through the incorporation of a GaAs barrier layer between the InAs QDs and the GaSb capping layer. Composition profiles were obtained in heterostructures containing a GaAs barrier using EELS, HAADF-STEM, and dark-field 002 diffraction contrast TEM. Figure 2.6b shows EELS maps of elemental distribution for Ga, In, and Sb. The measured scenario—two In-containing layers separated by a thin Sb-containing layer—does not correspond to the desired (nominal) structure. This striking difference gives insight into the
complex interplay between segregation and desorption. Two conclusions are evident (1) surface-segregated Sb competes strongly with In, and effectively blocks In incorporation into the lattice, and (2) when the Sb concentration falls below a certain threshold, In incorporation into the lattice resumes [53]. This is the first clear evidence of a strong influence of Sb on the incorporation of In—which occurs despite the fact that In is incorporated only on group III sites, while Sb is incorporated on group V sites.

Finally, GaSb QDs grown directly on GaAs substrate were analyzed by high-resolution HAADF-STEM images, resolving the atomic columns and obtaining a Sb distribution map. Figure 2.7 corresponds to integrated HAADF-STEM intensities for each projected unit cell superimposed on the original HAADF image, where a higher intensity corresponds to unit cells with higher Sb content. This type of analysis is very helpful to determine the lateral size and height of a nano-object. In this case, it demonstrates the formation of a discontinuous layer composed of $\text{GaAs}_x\text{Sb}_{1-x}$ nanostructures with a lateral size in the region of 1–10 nm and heights of around 1 nm [54]. These measured dimensions mean that there has been a volume reduction of the nanostructure by three orders of magnitude during the GaAs capping process. This proves the strong Sb segregation occurring during the growth of the GaAs capping layer.
Fig. 2.6 (a) HAADF-STEM image of a nominal GaSb/6 ML GaAs/InAs/GaAs heterostructure viewed along [110]. The rectangle indicates the region analyzed by EELS. (b) EELS-Independent Component Analysis images of Ga, In, and Sb distribution of the marked area. (c) Composition profiles taken from (b). Note that two In-containing layers are registered in (b) and (c), showing the reduction of In incorporation by Sb (reproduced from [53]).

Fig. 2.7 Integrated intensities for each projected unit cell, superimposed on a high-resolution HAADF-STEM image of the GaSb/GaAs layer (adapted from [54]).

In summary, these studies show the formation of Ga$_x$In$_{1-x}$As$_y$Sb$_{1-y}$ in the core of QDs, likely to be a strain-driven process. A clear influence of Sb on the In incorporation in heterostructures containing a GaAs barrier layer is revealed, as well as Sb segregation during the capping process of GaSb QDs grown directly on a GaAs substrate. In all cases, the difference between the nominal structure and the one observed experimentally is very clear, thus demonstrating the need for high-resolution analysis in the development of these semiconductor systems.
2.5 Results on InAs Quantum Wires

This section summarizes our contribution to the understanding of the growth of InAs self-assembled stacked quantum wires on InP(001) substrate by MBE using different techniques. Morphological and compositional changes from the initial stages have been characterized. Experimental and simulated results of the inhomogeneous strain field in the structures have proved it as the driving force for the heterogeneous morphology and As distribution, and for their nucleation on preferential sites on each stacked layer and with them, their vertical distribution along the stacked heteroepitaxial layers.

To manufacture improved three-dimensional (3D) semiconductor nanostructures for telecommunication devices which has to respond to increasing requirements of data transmission speed and stability, the self-assembly, through the Stranski–Krastanov (SK) growth mode, is recognized as one of the most remarkable approach to be followed. For the heteroepitaxial system InAs/InP grown by MBE, the strong stress anisotropy at the InAs/InP interface along [110] directions [55,56] can lead, under particular growth conditions, to the formation of horizontal quantum wires (QWRs) instead of QDs. Due to its potential applications for telecommunication technologies, an intensive research has been carried out on this type of nanostructures during the last years. In fact, it is well demonstrated that QWR emission properties can be tailored from 1.2 to 1.9 \( \mu \text{m} \) [57] by controlling several parameters such as the substrate temperature, the distribution of the different layers during the vertical arrangement, or their density [58–60].

In this context, the control of the shape, size, position, compositional information, and strain of the semiconductor nano-objects constituting these nanostructures is crucial to improve their design and functionality [61]. The development and implementation of methods to extract this information at the nanoscale, and in some cases at the atomic scale, should help us to be able to know and relate their characteristics with their functional properties.

Results obtained by high-resolution phase-contrast and Z-contrast electron microscopy techniques, in coordination with peak-finding methods, finite elements analysis (FEA), and EELS, are reviewed next. They are applied to measure the above-mentioned structural properties at the nano- and atomic scale, to predict the preferential nucleation sites of single and stacked layers, and to contribute to understand the growth of single and stacked layers of QWRs.

2.5.1 Nucleation and Initial Growth Stages of InAs/InP(001) QWRs

With the aim of evaluating the wires’ growth at their initial stages, samples with 2.3, 2.5, and 2.7 InAs monolayers (ML) grown on InP(001) were analyzed. The first sample corresponds to the onset of the elastic relaxation process due to the
spontaneous self-assembling of QWRs, and the last sample to the well-developed QWRs. The growth conditions are reported in [58, 62].

The analysis of atomic force microscopy images [52] and high-resolution Z-contrast micrographs of capped and uncapped nanowires reveal that atomic steps located at the surfaces of substrates or buffer layers play a central role in controlling the growth mechanism. In fact, they show that most of the nanowires are positioned on the upper terraces of the growth surface steps close to the step edge. In particular, Fig. 2.8 shows a high-resolution aberration-corrected Z-contrast image of a capped nanowire located just on InAs(P)/InP(001) interfacial steps. In this case, InAs-rich regions are demonstrated to be situated on the region described below. The strain at the growth surface on the upper and lower terraces of diatomic steps during the very initial stage (2D) of InAs deposition is calculated by FEA, using anisotropic elasticity theory. Larger positive strains are found at the upper terrace surface, which explains the lower chemical potential for further growth on the upper terrace of the diatomic step, and therefore, the preferential nucleation for the nanowires.

Regarding the morphological evolution of the initial QWRs, maps of the integrated intensity around As–P columns (Fig. 2.9b, d) determined from the high-resolution Z-contrast images in Fig. 2.9a, c—as described in Sect. 2.2—display that (001) is the main facet of the nano-objects at their very initial stage of formation; however, it is gradually reduced in favor of {114} or {118}. It is clearly observable from this analysis the compositional asymmetry of the nanowires since their initial steps of growth, which has a strong effect on the final arrangement, composition, and strain of the nanowires.

Three-dimensional models built by FEA explain the facets’ evolution in terms of variation of the elastic energy. These calculations locate energy minima on (001) facets and maxima around the QWR perimeter where it joins the wetting layer (WL), favoring diffusion of arsenic atoms from these edges toward the QWR top (to increase its height) and outside (for the nucleation of new wires).
2.5.2 Simulated and Experimental Determination of Strain Map and Prediction of Nucleation Sites for the Growth of the Stacked Structures

For self-assembled nanostructures, the size, homogeneity, density, and spatial organization degree can be improved by stacking several layers. The InP-capped InAs$_x$P$_{1-x}$ nanowires of each layer underneath each new stacked layer of nanowires produce inhomogeneous strain fields that propagate toward the capping layer surface where the next nanostructures will be formed, controlling the distribution of the strain at the InP growth surface.

By combining in situ stress measurements and TEM, the effect of this strain field has been reported and the wire growth explained by considering stress-driven mass transport and P/As exchanges [60]. The thickness and the growth temperature for layers capping nanowires are critical parameters that permit the control of the heterogeneous strain field associated with the buried nanowires and the nucleation sites on each stacked layer at the growth surface.

The understanding of the wires growth can be completed following two approaches: (1) Measuring the strain field by applying peak-finding methods [19, 30] (see Sect. 2.3) to HRTEM images of stacked nano-objects. (2) Calculating the strain field by solving the anisotropic elastic theory equations by FEA using a compositional model of the studied nano-objects. This method has been proven to be one of the most efficient procedures to model the strain associated with these buried nanostructures [63–66]. Inputs for this compositional model are given by both spatially resolved EELS and aberration-corrected high-resolution Z-contrast images. Figure 2.10 depicts the process followed to find the preferential nucleation sites of stacked epitaxial nanowires.
Both approaches have been applied to InAs/InP single and stacked nanowires for two different growth temperatures (380 and 515°C) and thicknesses (between 3 and 20 nm) of the capping layer (see growth details in [59] and [60]).

Compositionally sensitive TEM images of low capping temperature (380°C) samples show nanowires randomly distributed for 20-nm-thick spacer layers. On the contrary, for 5-nm-thick layers, they are displayed well ordered, and arranged along a direction that forms 9° with the [001] vertical direction. For high capping temperatures (515°C), a similar arrangement is only found in the sample with 3-nm-thick capping layer, resulting in a stacking angle of about 16°.

The location of the minimum chemical potential at the growth surface (preferential nucleation sites) corresponds approximately to the positions of minimum strain energy calculated by FEA. Since experimental and simulated values are in good agreement, one can conclude that this estimate allows the successful prediction of the stacking angles of the studied nanowires [67, 68].

Strain from HRTEM images has also been mapped in the same stacked layers of nanowires. The InAs(P) strained area related to the wire decreases as the growth temperature of the capping layer is risen. This is due to As/P exchange process during wire capping, which is enhanced at higher temperatures. In general, the maximum strain is not in the wire center, which is associated with the existence of an inhomogeneous composition distribution in the wires. This effect confers an asymmetric profile to the strain spreading on the InP surface that explains the shifted nucleation sites found by FEA. Moreover, quantitative agreement between the strain values based on simulated maps and experimental ones is found (see Fig. 2.11) [69].

In summary, high-resolution phase-contrast and Z-contrast electron microscopy techniques, complemented by peak-finding methods, FEA, and EELS, have made possible a better understanding of the horizontal InAs/InP quantum wires’ growth
The process since its initial stages. They can be successfully applied to measure the structural characteristics of the wires at the atomic scale and to predict the preferential nucleation sites of single and stacked layers.

2.6 Analysis of the N Distribution in GaAsN

Recently, there has been great interest in extending the possibilities for band-structure engineering for GaAs-based materials by alloying with N. The addition of a small amount of N (lower than 5%) to Ga(In)As produces a substantial extension of the emission wavelength [70], allowing operation at the transmission windows of optical fibers. However, point defects such as the N–N split interstitials [71] and the AsGa antisites [72] have been identified as electron traps in Ga(In)NAs, and the local clustering of N atoms has been suggested to create a range of defect states within the energy gap [73]. Therefore, the characterization at the atomic scale of the behavior of N at the few percent level in GaAs is critical to understand and optimize the performance of this system.

In order to shed light on the behavior of N in GaAs, Ga(AsN) quantum wells with different N compositions (0.1, 0.4, 1, and 2.5%) grown between GaAs barriers by MBE were analyzed. The N composition was controlled by monitoring the optical intensity of the atomic N plasma emission. The high-resolution HAADF-STEM study was performed using an aberration-corrected VGHB501 dedicated STEM operated at 100 kV; the low magnification analysis was carried out in a JEOL 2500 (S)TEM, operated at 200 kV.

Initial HAADF-STEM analysis of the Ga(AsN) QWs at low magnification led to an interesting result. As can be observed in the inset in Fig. 2.12a, the Ga(AsN)
Fig. 2.12  (a) Plot of the evolution of the contrast 
\((I_{\text{GaAsN}} - I_{\text{GaAs}})/I_{\text{GaAs}}\) vs. N content from the low magnification images of Ga(AsN) QWs. The inset shows a micrograph of a QW with N = 2.5%;  
(b) high-resolution HAADF-STEM image of a QW with N = 2.5%; plot of the contrast measured from the high-resolution images separately from the valley between the atomic columns (b) and from the atomic columns (c) vs. N content;  
(d) plot of the experimental average dumbbell spacing ratio vs. N content (reproduced from [74])
QWs with \( N = 2.5\% \) appear brighter than the GaAs barriers when imaged in HAADF-STEM despite the reduced average atomic number. As pointed out by early studies by Perovic et al. [75] and Treacy et al. [76] and more recently by Grillo et al. [77], the local distortion of the lattice plays a major role in HAADF-STEM imaging of non-perfect crystals. Thus, the origin of the observed contrast may rely on the local distortion of the lattice due to the introduction of nitrogen, as pointed out by Wu et al. [78]. Figure 2.12a shows a plot of the contrast GaAsN–GaAs \([C = (I_{GaAsN} - I_{GaAs})/I_{GaAs}]\) vs. the N content. It is observed that the evolution of contrast with the N content does not follow a linear behavior: the slope of the curve for low N concentration is relatively large and for the higher N content, the curve levels off. The origin of this dependence is not immediately clear for what is supposed to be a random substitutional alloy.

To investigate the origin of the observed evolution of contrast with increasing N content, the QWs were analyzed by high-resolution HAADF-STEM with aberration corrector, as can be observed in Fig. 2.12b for \( N = 2.5\% \). The contrast from the valley between atomic columns and that from the atomic columns were separately measured and plotted vs. the N content in Fig. 2.12c, d, respectively. The intensity valley/peak was measured individually for each dumbbell in each image for up to six images per concentration; the error bars correspond to the standard error over all these measurements. As it can be clearly seen, the contrast from the valleys between columns shows a strong increase and then flattens off for higher N concentration, similar to the behavior obtained from the low magnification images, whereas the increase in contrast from the atomic columns is negligible. This reveals that the increased intensity found in Ga(AsN) originates specifically from the valleys between atomic columns rather than from the columns themselves. Additionally, the average dumbbell spacing ratio was also measured from individual dumbbells, finding a nonlinear reduction when increasing the N content (Fig. 2.12c). This behavior is likely caused by the introduction of an additional strain component due to the incorporation of N in the lattice.

In order to explain the origin of these results, simulations of the HAADF-STEM images were performed for different likely N configurations in diluted nitride Ga(AsN). Thus, the existence of N–N split interstitials, N–As split interstitials, or \((AsGa–NAs)_{\text{nn}} \) pairs in Ga(AsN) has been suggested theoretically [79] and verified experimentally [80,81]. Also, the low solubility of N in GaAs [82] has been reported to induce the clustering of N atoms [83] in the alloy. Both interstitial N and N clustering could produce a change in the thermal diffuse scattering (TDS) from the Ga(AsN) region. According to this, the simulation of Ga(AsN) structures was performed, N atoms being incorporated in the group-V lattice sites (substitutional N) both randomly and in incipient clusters (2N cluster), and also in the form of the above interstitial N complexes (a schematic of these complexes is shown in Fig. 2.13).

The equilibrium lattice positions of the atoms in fully relaxed Ga(AsN) complexes used as input in the HAADF-STEM simulations were calculated using density functional theory within the local density approximation, as implemented in
Fig. 2.13  (a) Plot of the contrast vs. nitrogen content from the simulated images of the different N-complexes. Open symbols correspond to the valley between the columns and solid symbols to the atomic columns; a schematic picture of each complex is included where the dark atoms represent N, the dark gray represent Ga and the light gray represent As; (b) averaged dumbbell spacing ratio measured from the simulated images for the different complexes vs. N content (reproduced from [74]).

VASP [84]. All the calculations were performed using 64-atom cubic cells. All the atoms were relaxed to minimize the Feynman–Hellmann forces to below 0.02 eV/Å.

HAADF-STEM images were computed using the simulation software SICSTEM [85]. Elastic scattering was implemented following Kirkland’s approach, while TDS was included using a local TDS absorptive potential approach. Images corresponding to 70-nm-thick structures (which is the thickness of the sample foil where the experimental images were acquired) for the different complexes were simulated, where the N complexes were located in random positions in three dimensions. The dumbbells’ spacing and intensity valley/peak have been measured individually for each dumbbell in each image, and then averaged. It is worth noting that a mismatch of 2–4 times between STEM images and simulations [86] has been reported previously, associated with an overestimation of the degree of coherence in the electron source [87]. Because of this, the comparison of the experimental and simulated images will be based on the evolution of contrast with composition, not aiming to compare absolute values of intensity between theory and experiment.

The simulated images for substitutional N have shown an increased intensity coming from the N containing area, presumably due to the local distortion of the lattice. In order to separate the effect of the Z number from that due to the deformation of the structure, the simulation of a GaAs cell mimicking the local distortion of substitutional N was performed, where the N atom of a relaxed
Ga(AsN) structure was substituted by As without changing the position of the atoms. The simulation of such a structure exhibited exactly the same intensity as the simulated substitutional N in Ga(AsN). This result shows the strong effect of the static atomic displacements in HAADF-STEM images. This distortion of the lattice has an effect similar to freezing in lattice vibrations, enhancing the effective Debye–Waller factor of the host lattice [88] and, in some cases as shown in this work, it can dominate over the Rutherford-scattering component.

Figure 2.13 shows the plots of the contrast Ga(AsN)–GaAs (a) and the average dumbbell spacing ratio (b) vs. N content derived from the simulated images. As can be observed, the contrast for substitutional N shows an increasing slope for higher N contents, whereas the experimental study showed saturation for increasing N. The dumbbell spacing ratio also follows the opposite tendency with respect to the experimental study, increasing with N content. This different behavior reveals the existence of alternative N configurations in the alloy besides random substitutional N.

According to Fig. 2.13a, both the N–As split interstitial and the (AsGa–NAs)_{nn} pair complexes show a pronounced increase in contrast for the higher N compositions, whereas the experimental results show the opposite. Moreover, the evolution of dumbbell spacing ratio with the N content does not agree with the experimental observations for any of these complexes. For the N–N split interstitial, both the Ga(AsN)–GaAs contrast and the dumbbell spacing ratio stay very stable with increasing N content because of the small local distortion when substituting As for two N. Consequently, the existence of this complex in the alloy cannot explain the experimental results.

The 2N substitutional complex is the only one for which the simulations agree with the experimental results. Thus, as can be observed, the evolution of contrast with composition increases with N content, with a reduction in the slope for higher N contents. The dumbbell spacing ratio, on the contrary, decreases for higher N contents with the same tendency as in the experimental curve in Fig. 2.12e. These results clearly point to the absence of interstitials in the alloy and to the onset of N clustering. Substitutional N pairing up to four neighbor position in GaAs_{1-x}N_{x} alloys with x < 0.025 has been reported previously [89], in good agreement with our experimental observations.

### 2.7 Review on InN Nanostructures

Despite being traditionally a material discriminated against the rest of the family of III-nitrides, the interest in InN started growing after the re-establishment in 2002 of its band gap to a value close to 0.7 eV [90–93] from the 1.9 eV that was assumed before [94]. This observation opened a new field of possibilities, since the combination of InN with GaN and/or AlN to form ternary or quaternary alloys could lead to high-efficiency optoelectronic devices covering a wide range of wavelengths, from the infrared to the ultraviolet including the whole visible spectral range [95]. In the same way, the gradual improvement of the different growth techniques led to
the fabrication of high-quality crystalline InN layers that presented largely improved electrical properties compared with the past [96,97]. The attempt to join the promising properties of InN with those derived from quantum confinement effects made the fabrication of InN-based nanostructures the next challenge. Unlike some other systems such as InAs, (InGa)As, or GaN, results about the development of InN nanostructures have emerged only during the last few years. It was not until 2003 when Briot et al. [98] first achieved the growth of InN QDs grown on GaN by metalorganic vapor phase epitaxy, and later in the same group, also on AlN and Si substrates [99]. In the same year, Cao et al. reported the growth of InN QDs and their shape and size dependence on the MBE deposition conditions [100]. After this, important improvements were achieved regarding smaller sizes and higher densities [101–107], and recently the emission, even very poor, of InN QDs [108,109] and growth and optical properties of cubic InN dots [110] have been reported. Other types of InN nanostructures fabricated so far include single [111] and multiple [112–115] quantum wells, nanocolumns [116,117], and nanowires [118–120]. However, even if the number of publications related to these nanomaterials has increased considerably during the last years, the number of them including characterization by HRTEM still remains scarce. Some of the results that will be described later in this section correspond to the first ones reported in the literature about structural and chemical characterization by (HR)TEM of InN QDs. This technique was also used to analyze the crystalline structure and quality [117], or the Si/InN and Si/AlN/InN interfaces in InN nanocolumns [116]; to confirm the cubic structure of the QDs [110] and the existence of oxide precipitates in InN nanostructures [121]; or to investigate the relaxation mechanisms of InN nanowires on GaN [120]. Our main contribution to the field is related to the characterization by means of electron beam-related techniques of uncapped and GaN-capped InN QD heterostructures grown on GaN/Sapphire substrates by metalorganic vapor phase epitaxy [98]. Even though a wider number of techniques were employed and an extensive work has been carried out, here we will mainly concentrate on the results derived from the HRTEM analysis.

The investigated InN/GaN quantum dots present a well-defined flat hexagonal shape with an average height of 12 nm and diameters of 72 and 120 nm for uncapped and capped QDs, respectively [122]. When observed in plan-view TEM (PVTEM) geometry, the presence of three-directional moire fringes patterns is revealed in all the uncapped QDs. These patterns were used to estimate the average degree of plastic relaxation of InN in the heterostructure, showing that it is close to a full relaxation, which was confirmed later by HRTEM [123]. As in other cases of highly mismatched heterostructures, this high degree of relaxation is related to the presence of a dense misfit dislocation (MD) network at the interface between the two materials. The application of conventional diffraction contrast TEM in plain view orientation to study the MD network is not feasible due to the large number of dislocations and their associated strain fields. However, these problems were overcome after the application of the GPA method [31] to high-resolution PVTEM micrographs, which allowed us to obtain strain maps at the atomic scale of the heterostructure and successfully perform a complete characterization of this MD network [124]. Figure 2.14a displays a HR-PVTEM micrograph of an uncapped
InN QD, and Fig. 2.14b the associated strain map. In the latter, three families of MDs (red lines) are distinguished, without interaction between them or node generation, but forming a sort of “stars of David” network. Moreover, these maps also revealed the behavior of the MDs when they are close to the edge of the QD. From the strain maps and the corresponding elastic constants, stress maps were built in areas containing the boundary of the QD, as the one marked with a white square in Fig. 2.14a. The results, displayed in Fig. 2.14c–e, turned out to be in perfect agreement with the stress distribution around an edge threading dislocation. This means that not only a dense network of MDs exists at the interface InN/GaN, but there is also another network of threading dislocations surrounding the system, as a consequence of the bending of the MDs when they are close to the edge of the QDs due to the free surface forces.

Additionally, the above-mentioned Moiré fringes revealed extra information about the system: the interruption of the Moiré lines in the region close to the center of the QD (see arrow in Fig. 2.15a) is a consequence of the presence of a threading dislocation in that area. To assess if that TD is located in the GaN, the QD, or both, inverse fast Fourier transform of the area marked with an orange square in Fig. 2.14a was built considering the reflections that correspond to GaN and InN separately, and drew for both cases a Burgers circuit around the area of interest.
Figure 2.15 (a) PVTEM micrograph of an uncapped InN QD showing one family of Moiré fringes. The arrow indicates a discontinuity of one of these lines. (b and d) Inverse fast Fourier transforms of the area containing the discontinuity marked with an orange square in Fig. 2.14a for InN and GaN, respectively.

Figure 2.15b, c shows the results: for InN the circuit closes perfectly, whereas for GaN the same closure fails. This means that the TD is located in the GaN, does not propagate into the InN QD, and has at least an edge component of the Burgers vector. Finally, applying the invisibility criterion to conventional diffraction contrast micrographs confirms that the dislocation is pure edge type. This was not an isolated observation, but a general rule for every dot observed in large areas of the sample, so these pure edge-type TDs in GaN are assumed to be related to the nucleation mechanism of the InN QDs [125]. None of the classical nucleation models found in the bibliography (Volmer–Weber, Stranski–Krastanow, and Burton–Cabrera–Frank) fulfilled our observations, so even though the basic interactions that lead to the nucleation of the InN QDs on top of pure edge-type dislocations in GaN are still not clear, our observations suggest a new mechanism never reported before.

Finally, HRTEM was also used to investigate the aging process of the uncapped InN QDs after 36 months of exposure to environmental conditions [126, 127]. While the GaN capping layer prevented the capped QDs from interacting with air and then no structural changes are observed, the uncapped dots suffered a transformation from the hcp structure to cubic phases in the more external layers, while the core of the dot remained as wurtzite InN. These cubic phases turned out to be composed of indium oxide, and highly oxygen-doped cubic InN was proposed as an intermediate phase between the original wurtzite InN and the final product, In$_2$O$_3$.

2.8 Crystalline, Compositional, and Strain TEM Assessments of High-Quality Epilayers of Ternary and Quaternary III–N Alloys

Semiconductor stoichiometric solid solutions composed of half of N atoms and half of a single element or a random mix of specific elements of the group III of the periodic table (i.e., Al, Ga, and/or In) energetically tend to stabilize
in the hexagonal wurtzite-type phase (2H-polytype) and constitute the base of numerous recent and foreseen advantageous technological applications. These new materials have mainly been developed through the last 15 years and were commonly fabricated as thick deposits on different substrates or pseudosubstrates such as bare or modified sapphire, silicon carbide, gallium arsenide, or silicon wafer surfaces since no technique for obtaining single crystals existed. Due to many deleterious effects for their growth, such as lattice mismatch-induced stress and other thermodynamical restrictions, these materials develop and suffer from the formation of extended defects and crystalline or compositional phase separations. Nevertheless, the use of MBE and other techniques is starting to allow epitaxial depositions of III-N binary compounds (AlN, GaN, and InN) and their ternary and quaternary alloys (AlGaN, InAlN, InGaN, and InAlGaN) in the form of single crystals. The current trend on semiconductors downscaling to get device miniaturization—and hence improvements by quantum confinement and/or electron accumulation effects—compels to obtain nano-sized motifs with high quality and controlled size and shape of layers, wires, or dots depending on the desired performance. In the described context, any heteroepitaxial volume of III-N under a critical thickness, even unfaulted, is subjected to a tetragonal-like distortion. The amount of this deformation can be calculated through the elastic theory, and its determination depends on the fine fourfold knowledge of material features, namely, lattice mismatch, surface orientation, chemical composition, and directional anisotropic stiffness constants. TEM-related techniques permit to collect important structural and chemical information to solve these problems. This section focuses on the importance of these aspects and describes some examples of how TEM can contribute to characterize heterosystems consisting of mismatched and nearly lattice-matched thin films of hexagonal InGaN, InAlN, AlGaN, and InAlGaN grown on the basal plane of unstrained GaN buffer thick layers grown on sapphire.

2.8.1 Previous Considerations

Films based on a hexagonal III-N compound or alloy, In$_x$Al$_y$Ga$_z$N ($x + y + z = 1$), with a definite composition ($x$, $y$, $z$), are often subjected to biaxial stresses. On polar surfaces, these stresses promote a strain parallel to the basal plane ($\epsilon_a$) which is opposite and related to the perpendicular strain to this (0001) plane ($\epsilon_c$). Provided that the quality of the layer is high enough (single phased with crystalline and compositional homogeneity and stoichiometry), the anisotropic elastic theory assumes that the experimental lateral strain coefficient $R_S(x, y, z) = \epsilon_c / \epsilon_a$ should be equal to the biaxial strain relaxation coefficient $R_B(x, y, z)$, which is dependent on the ratio of the crystal elastic constants, $C_{ij}(x, y, z)$. Lattice and elastic constants of ternaries and quaternaries can be calculated for any composition within contour conditions which correspond to the different binary III-N, using Vegard’s law ($d_0$In$_x$Al$_y$Ga$_z$ = $x$ $d_0$InN + $y$ $d_0$AlN + $z$ $d_0$GaN). Table 2.1 lists reliable elastic constants statistically calculated as far as recently claimed accurate lattice constants.
Table 2.1 Reference material parameters used in this work

<table>
<thead>
<tr>
<th></th>
<th>InN</th>
<th>AlN</th>
<th>GaN</th>
</tr>
</thead>
<tbody>
<tr>
<td>2H- $a_0$ (Å)</td>
<td>3.53774</td>
<td>3.11197</td>
<td>3.18840</td>
</tr>
<tr>
<td>2H- $c_0$ (Å)</td>
<td>5.70374</td>
<td>4.98089</td>
<td>5.18500</td>
</tr>
<tr>
<td>$C_{11}$ (GPa)</td>
<td>237 ± 7</td>
<td>395 ± 5</td>
<td>374 ± 4</td>
</tr>
<tr>
<td>$C_{12}$ (GPa)</td>
<td>106 ± 4</td>
<td>137 ± 3</td>
<td>138 ± 4</td>
</tr>
<tr>
<td>$C_{13}$ (GPa)</td>
<td>85 ± 3</td>
<td>107 ± 3</td>
<td>101 ± 4</td>
</tr>
<tr>
<td>$C_{33}$ (GPa)</td>
<td>236 ± 6</td>
<td>404 ± 6</td>
<td>395 ± 5</td>
</tr>
<tr>
<td>$C_{44}$ (GPa)</td>
<td>53 ± 3</td>
<td>117 ± 2</td>
<td>98 ± 3</td>
</tr>
</tbody>
</table>

Reference [44] [128] [44]

Fig. 2.16 Values of $R^B(x, y, z)$ for the whole In$_x$Al$_y$Ga$_z$N compositional range. These values should theoretically be coincident with the experimental ratio of lateral strains, $R^S(x, y, z) = \varepsilon_c / \varepsilon_a$

for hexagonal binaries: InN ($y = 0$ and $z = 0$), AlN ($x = 0$ and $z = 0$), and GaN ($x = 0$ and $y = 0$).

Figure 2.16 shows the expected values of $R^B$ for any perpendicularly strained polar wurtzite In$_x$Al$_y$Ga$_z$N. This plot and further calculations have been obtained by using the data in Table 2.1 and that $R^B(x, y, z) = -2C_{13}(x, y, z)/C_{33}(x, y, z)$. It is very important to note that the assumption of a linear trend for $C_{ij}(x, y, z)$ among binaries implies a bowing behavior for $R^B(x, y, z)$. From the assumption that $R^B(x, y, z) = R^N(x, y, z)$, compositions can be indirectly calculated by the use of the lattice parameters of the films. However, the erroneous applications of Vegard’s rule have led to big underestimations and overestimations of sets of $x, y$ (InAlN); $x, z$ (InGaN); $y, z$ (AlGaN); or $x, y, z$ (InAlGaN) values in the past.

2.8.2 Briefly, a Complete (S)TEM Study

In a first stage, careful analyses of conventional diffraction contrast micrographs (DCTEM) and selected area electron diffraction (SAED) patterns might be enough to know if the analyzed III-N materials are of a sufficient crystalline quality and worth to deserve a deeper insight [44, 128–130]. In this case, being the nitride layers of good quality, sharp diffraction spots associated with the substrate materials together with a unique set coming from a single crystalline ternary or quaternary
III-N layer are only expected to be present in an electron diffraction pattern. Figure 2.17 presents a cross-section (XTEM) panoramic image representative of an InAlGaN/GaN/sapphire heterostructure, and its respective SAED pattern with reflections indexed into the inset.

It is well known that the low miscibility of binary compounds, especially when using InN, often leads to phase separation effects. Contrasts associated with these effects have been reported by means of DCTEM and SAED for ternary and quaternary alloys, for example, in extensive studies made by Prof. Mahajan’s group in the last decade [131–137].

The only defects that are still acceptable in III-Ns for their adequate performance are the threading dislocations (TDs). It is not easy to get rid of them in film fabrications mainly due to the involved misfit strains among layers. Although some scientists found that they are electronically beneficial, the main goal is to have the lower possible densities; and a study of their null, occasional or common presence is feasible by DCTEM. If the semiconductor material is properly oriented in the electron microscope, the reflections 000N or 011N could be used on bright-field (BF), dark-field (DF), or weak-beam (WB) modes, and imaging is acquired for screw component (their Burger vectors are “c” or “a + c”) or for edge plus mixed (“a” or “a + c”) TDs, respectively. Figure 2.18 shows DC-XTEM images (g = 0002 near the [1120] zone axis) representative of non-defective alloys in InAlN/GaN and InAlGaN/GaN heterostructures, and of a horizontally phase-separated quaternary alloy (top right image). Screw component TDs presenting different behaviors, when stopping at or crossing the interfaces, are also visible.
HRTEM and STEM techniques are applied to demonstrate the quality of the material. While HRTEM images permit a structural comprehension of layers’ sequence and their interfaces at an atomic level, to analyze the chemical homogeneity in layers, a combination of nano-probe energy dispersive X-ray (EDX) spectra and HAADF images, both in STEM mode, is useful and needed. EELS or energy-filtered TEM can also be helpful for this purpose. The 2D arranged atomic columns in the HRTEM detail of Fig. 2.19 illustrate an example of sharp interfaces among single crystals. The nano-interlayers labeled as “spacer” served as a buffer between the epilayer and the GaN in these cases. See in the same figure HAADF micrographs of various alloys with constant compositions [deduced from smooth and invariable atomic mass (Z)-contrasts], except for the one horizontally phase-separated InAlGaN bilayer. Note that for any of the epilayers claimed as homogeneous, for 10–20 EDX-STEM spectra taken for nano-sized regions of every ternary and quaternary alloy in different positions, the measured compositions spread less that 2% of the average for each film. This spread is on the level of the absolute error of the technique, giving this fact another hint on the uniformity of the composition.

Nevertheless, these examples of high homogeneity are still unusual in thin III-N epilayers and similar good qualities as those presented here have been rarely claimed, as, for example, in InAlN [138] or in InGaN [139], but for a very rough surface in the latter case. Some representative examples in the recent literature can be found for the state-of-the-art, but defective, ternary and quaternary III-Ns studied by some combination of HRTEM, HAADF, EDX, and EELS analyses. In this way, for InAlN nearly lattice matched to GaN, there are reports on horizontal composition separation [128,129], or on In segregation to vertical domain boundaries [140,141] or to surface V-defects [142, 143]. For AlGaN, composition modulation [136], phase-separated domains [144] sometimes having tunnel defects and associated dislocations with erratic behaviors [145] have also been measured. On the other hand, surface-grown InGaN thin films showed bilayered structures of In- and Ga-rich InGaN parallel to the growth plane [146]. For InGaN/GaN and
AlGaN/GaN, most of these kinds of studies focused on the presence of V-defects, atomic fluctuations or clustering, and electron beam damage in multi-quantum wells [147, 148]. Only a few advanced TEM studies begin to consider InAlGaN alloys, a system where the control of composition is less well understood during the growth. Nano-scale In clustering in InAlGaN was observed by HRTEM [149], and a high-quality InAlGaN layer was elsewhere claimed [150]. However, the contents of both In and Al in the latter cases are in practice very low, with a lower probability of defects promoted by the lack of affinities among In, Ga, and Al atoms.

### 2.8.3 Analyses of Lateral Strains

Once the high compositional homogeneity is confirmed, the combination of chemical and lattice measurements can facilitate a complete analysis of the perpendicular deformations in the layers, assuming they behaved biaxially strained as expected by the elastic theory. Accurate measurements of lattice parameters are often supplied by high-resolution XRD analyses, but they can also be estimated by TEM with precisions and accuracies below 1% [151]. In this way, the comparison of the positions of sharp spots corresponding to nitrides with respect to those of a present unstrained internal standard in the same SAED pattern is needed for claiming such accuracies. In the example shown in Fig. 2.20 for an InGaN/GaN/AlN/sapphire heterostructure, having one pixel resolution for InGaN- and Al2O3-related far reflections, lattice measurements of the ternary alloy with an error of ±0.001 Å are possible [44]. In the case of phase-separated films or for those having small solid solution inhomogeneities, the variation in the composition can create a distribution of d-spacing for a crystallographic plane and this approach becomes complex, although multi-peak-finding retrieval methods can be applied as it is well accepted in XRD theory.

![Fig. 2.20](image)

**Fig. 2.20** Brightness profile for specific diffraction spots from a SAED pattern taken for a InGaN/GaN/AlN/Al2O3 heterostructure (adapted from [44])
To determine the compositions $x$, $y$, $z$, the general (2.4) can be built from the considerations given in Sect. 2.8.1. The only required parameters to solve (2.4) are those listed in Table 2.1 together with measured $c$ and $a$ lattice parameters. Moreover, a second input is needed for solving a system of two equations, which is straightforward for the ternary alloys being $x = 0$ if AlGaN; $y = 0$ if InGaN; or $z = 0$ if InAlN. In the case of a quaternary alloy, the value of any $x/y$, $x/z$, or $y/z$ experimental ratio can be used

$$
\frac{c - (x c_{0}^{\text{InN}} + y c_{0}^{\text{AlN}} + z c_{0}^{\text{GaN}})}{x c_{0}^{\text{InN}} + y c_{0}^{\text{AlN}} + z c_{0}^{\text{GaN}}} \times \frac{x a_{0}^{\text{InN}} + y a_{0}^{\text{AlN}} + z a_{0}^{\text{GaN}}}{a - (x a_{0}^{\text{InN}} + y a_{0}^{\text{AlN}} + z a_{0}^{\text{GaN}})} = -2 \frac{x C_{13}^{\text{InN}} + y C_{13}^{\text{AlN}} + z C_{13}^{\text{GaN}}}{x C_{33}^{\text{InN}} + y C_{33}^{\text{AlN}} + z C_{33}^{\text{GaN}}}. \tag{2.4}
$$

When an agreement is found among calculated and experimental compositions, the high quality of the layers are confirmed since the expected elastic response is fulfilled. Then, a derivation of lateral strains is further comprehensible. Note that the lack of agreement between experimental and calculated compositions, even for single crystals, may indicate an absence of perfect stoichiometry of the layers [128].

A collection of the EDX-measured compositions of ternary and quaternary alloys, together with their thicknesses and strains are listed in Table 2.2. The utilized measured lattice parameters to make the composition calculations come from the best quality samples of our own research, and except for the micrographs related to the bilayers (inhomogeneous layers), the images shown in this section come from any of these listed samples [44, 128–130]. Note the agreement inside the EDX error as a proof of their good quality (i.e., expected elastic response).

Let us notice to finish that the solutions of compositions calculated with other well-accepted but simpler approaches for quaternaries [152] lead to unrealistic values, sometimes very far away from those demonstrated in this work. In conclusion, the application of Vegard’s rule among binaries without any modification

<table>
<thead>
<tr>
<th>Composition</th>
<th>Thick (nm)</th>
<th>$\varepsilon_c$ (%)</th>
<th>$\varepsilon_a$ (%)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>In$<em>{0.42}$Ga$</em>{0.58}$N</td>
<td>25</td>
<td>0.81</td>
<td>-1.75</td>
<td>[44]</td>
</tr>
<tr>
<td>In$<em>{0.51}$Ga$</em>{0.49}$N</td>
<td>25</td>
<td>0.21</td>
<td>-0.65</td>
<td>[44]</td>
</tr>
<tr>
<td>In$<em>{0.62}$Ga$</em>{0.38}$N</td>
<td>25</td>
<td>0.75</td>
<td>-1.07</td>
<td>[44]</td>
</tr>
<tr>
<td>In$<em>{0.70}$Ga$</em>{0.30}$N</td>
<td>25</td>
<td>0.33</td>
<td>-0.55</td>
<td>[44]</td>
</tr>
<tr>
<td>In$<em>{0.81}$Ga$</em>{0.19}$N</td>
<td>25</td>
<td>0.08</td>
<td>-0.14</td>
<td>[44]</td>
</tr>
<tr>
<td>In$<em>{0.18}$Al$</em>{0.82}$N</td>
<td>48</td>
<td>-0.01</td>
<td>0.01</td>
<td>[128]</td>
</tr>
<tr>
<td>Al$<em>{0.42}$Ga$</em>{0.58}$N</td>
<td>20</td>
<td>-0.43</td>
<td>0.81</td>
<td>[129]</td>
</tr>
<tr>
<td>In$<em>{0.10}$Al$</em>{0.67}$Ga$_{0.23}$N</td>
<td>50</td>
<td>-0.01</td>
<td>0.36</td>
<td>[129]</td>
</tr>
<tr>
<td>In$<em>{0.01}$Al$</em>{0.42}$Ga$_{0.57}$N</td>
<td>45</td>
<td>-0.36</td>
<td>0.67</td>
<td>[129]</td>
</tr>
<tr>
<td>In$<em>{0.04}$Al$</em>{0.39}$Ga$_{0.57}$N</td>
<td>45</td>
<td>-0.22</td>
<td>0.57</td>
<td>[129]</td>
</tr>
<tr>
<td>In$<em>{0.08}$Al$</em>{0.37}$Ga$_{0.55}$N</td>
<td>20</td>
<td>0.006</td>
<td>-0.04</td>
<td>[129]</td>
</tr>
</tbody>
</table>
should be the only method for calculating ternary and quaternary lattice and elastic constants.

Acknowledgements  This work was supported by the European Commission (SANDiE European Network of Excellence NMP4-CT-2004-500101 and COST MP0805), the Spanish MICINN (TEC2005-05781-C03-02/MIC, TEC2008-06756-C03-02/TEC, and CONSOLIDER CSD2009-00013), and the Junta de Andalucía (Ref. P08-TEP-03516 and P09-TEP-5403) with EU-FEDER cofinancing. The authors would like to thank their collaborators and other coauthors of the articles reviewed here (A.G. Taboada, A.J. Papworth, A. Yáñez, B. Alén, D. Fuster, D.G. Morgan, F. de la Pena, J.M. Ripalda, L. González, L. Kirste, M.F. Chisholm, M.H. Du, M.H. Gass, M. Hopkinson, M. Varela, M. Walls, N.D. Browning, O. Ambacher, P.J. Goodhew, Q.M. Ramasse, R. Aidam, S.B. Zhang, S. Hauguth-Frank, S.J. Pennycook, T. Lim, V. Cimalla, V. Lebedev, and Y. González).

References

34. M.J. Hýtch, J.L. Plamann, Ultramicroscopy 87, 199 (2001)
38. M.J. Hýtch, T. Plamann, Ultramicroscopy 87, 199 (2001)
42. P.D. Robb, A.J. Craven, Ultramicroscopy 109, 61 (2008)
68. S.I. Molina et al., Nanotechnology 17, 5652 (2006)
High-Resolution Electron Microscopy of Semiconductor

118. Y.L. Chang, F. Li, Z. Mi
143. Z.L. Mião et al., Appl. Phys. Lett. 95, 231909 (2009)
149. J. Wu et al., Nanotechnology 17, 1251 (2006)
Amalia Patanè
Naci Balkan
Editors

Semiconductor Research

Experimental Techniques

With 192 Figures
Preface

This book intends to provide its readers with the fundamentals and applications of experimental techniques commonly used in semiconductor research. Each chapter describes the physics concepts underlying a specific technique and its latest developments in the investigation of novel semiconductor materials and heterostructures, including InN, dilute nitride III–V–N–V alloys, InAs/InSb heterostructures and self-assembled quantum dots.

Chapter 1 focuses on the investigation of semiconductor surfaces by reflection high-energy electron diffraction (RHEED) and low-energy electron diffraction (LEED). These diffraction-based techniques give access to the structural properties of crystalline layers and junctions, which represent the core of modern devices. Several examples of LEED and RHEED patterns are described and provide the reader with the basic tools for interpreting RHEED and LEED data. Chapter 2 describes transmission electron microscopy (TEM) and high resolution electron microscopy techniques. These are based on the analysis of a transmitted electron beam through an electron-transparent sample. Operation principles of TEM and examples of spatial mapping of composition and strain at the nanoscale and atomic resolution are the focus of this chapter. Chapter 3 reviews common techniques used to investigate the energy and momentum relaxation rates of hot carriers in semiconductors where the carrier heating is achieved by either the application of an electrical field or by an optical excitation. This condition is frequently met in optical and electronic device. Also, this chapter reviews steady-state spectral and transient measurement techniques. Chapter 4 describes the principles, experimental setups, and theoretical approaches used in optical modulation spectroscopy studies. Special attention is dedicated to contactless electroreflectance (CER) and photoreflectance (PR). These are non-destructive techniques, which are widely applied to study the band structure properties of semiconductor materials and devices. The complementary technique of photoluminescence (PL) and its relation to absorption and photoluminescence excitation (PLE) spectroscopy are described in Chap. 5. Typical experimental setups for optical studies, with and without an applied magnetic field, are discussed and examples of their application to study electronic properties, disorder effects and carrier thermalization in III–V semiconductor alloys and heterostructures.