1. HRTEM AND NANOTECHNOLOGY

With the rapidly escalating attention recently being given to nanoscale science and technology, techniques capable of structural characterization on the nanometer scale have central importance. The high-resolution transmission electron microscope (HRTEM) has evolved over many years to such an extent that resolving powers at or close to the one Ångstrom (0.1 nm) level can nowadays be attained almost on a routine basis. Using correct operating conditions and well-prepared samples, high-resolution image characteristics are interpretable directly in terms of projections of individual atomic-column positions. With quantitative recording and suitable image processing, atomic arrangements at defects and other inhomogeneities can be reliably and accurately determined. Since nanoscale irregularities have a marked influence on bulk behavior, the HRTEM has become a powerful and indispensable tool for characterizing nanostructured materials. This chapter first outlines some of the basic theoretical principles and practical aspects of the HRTEM. Representative applications to problems involving nanostructured materials are then described. Ongoing trends are identified and some underlying problems associated with use of the HRTEM are briefly discussed. The interested reader is referred to review articles [1–4], conference proceedings [5–10] and research monographs [11–13] for further information about HRTEM and additional applications.
2. PRINCIPLES AND PRACTICE OF HRTEM

2.1. Basis of Image Formation

The process of image formation in the HRTEM can be considered as occurring in two stages. Incoming or incident electrons undergo interactions with atoms of the specimen, involving both elastic and inelastic scattering processes. The electron wavefunction which leaves the exit surface of the specimen is then transmitted through the objective lens and subsequent magnifying lenses of the electron microscope to form the final enlarged image. Our attention here will be focused on those electrons that are elastically scattered since these contribute to formation of the high-resolution bright-field image. Note, however, that the inelastically scattered electrons can also be used to provide invaluable information about the sample composition using the technique of electron-energy-loss spectroscopy (EELS). Moreover, electrons scattered to very large angles are utilized for Z-contrast annular-dark-field imaging in the scanning transmission electron microscope (STEM). These possibilities are explored elsewhere in other chapters.

Electron scattering is a strongly dynamical process, unlike X-ray or neutron scattering, so that a simple kinematical scattering approximation is insufficient for understanding image formation for all but the thinnest of samples. Multiple electron scattering and large phase changes are typical for most samples, which means that the relative heights of the different atoms in the specimen must be taken into account when attempting quantitative image interpretation. Image simulations, which must also take into account additional effects caused by the objective lens, are essential before detailed information about atomic arrangements at crystal defects such as dislocations and interfaces can be confirmed. Several alternative approaches to image simulation have been developed [14, 15]. The most widespread is based on n-beam dynamical theory [16, 17], and is termed the multislice approach. In this theory, atoms in the specimen are considered as located on narrowly-separated planes (or slices), normal to the beam direction. The electron wavefunction is then propagated slice-by-slice through the sample to eventually form the exit-surface wavefunction. This iterative process lends itself to convenient computer algorithms which enable rapid computations to be carried out, which are especially useful during the refinement of unknown defect structures [18, 19]. Further information about other equivalent approaches to electron scattering can be found in the monograph by Cowley [20].

The electron wavefunction at the exit surface of the specimen must still be transferred to the final viewing screen or recording medium. This process is dominated by the transfer characteristics of the objective lens of the microscope. The effect of this lens on image formation is conveniently understood by reference to what is termed the phase contrast transfer function (TF), as described by Hanszen [21]. The TF is both specimen- and microscope-independent so that a single set of universal curves enables the transfer characteristics of all objective lenses to be described. Electron microscopes that have different objective lenses, or operate at different accelerating voltages, are easily compared by using suitable scaling factors. Figure 1 shows TFs computed for the optimum defocus of the objective lens for a typical 400-kiloVolt HRTEM. Two

Figure 1. Transfer functions for high-resolution phase-contrast imaging at optimum defocus for objective lens \( C_s = 1.0 \) mm of 400-kV HREM: (a) coherent illumination; and (b) partially coherent illumination, with focal spread of 8 nm and incident beam convergence half-angle of 0.5 mrad. Arrow indicates interpretable resolution limit.

Cases are shown, corresponding to: (a) coherent, and (b) partially coherent, incident electron illumination. Notice that the TF has an oscillatory nature, meaning that electrons scattered to different angles undergo relative phase reversals, which would lead directly to artefactual detail in the final image. Note also that the shape of the TF is further affected by defocus changes, which result in additional phase changes that will alter the image appearance. It is thus essential to have an accurate knowledge of the lens defocus, since much of the detail contained in the recorded high-resolution image will otherwise be uninterpretable. In the ideal case, the incident electron beam will be a coherent, monochromatic plane wave. In practice, some inevitable loss of coherence will result from focal spread (temporal coherence) and finite beam divergence (spatial coherence) [22, 23]. Mathematically, the effects of partial coherence can be represented by envelope functions. These will cause dampening of the TF for spatial frequencies corresponding to larger scattering angles, as illustrated, for example, by curve (b) in Fig. 1. Note also that the positions of the TF crossovers, or “zeroes”, are not affected by the envelope functions. However, specimen information that is scattered to higher spatial frequencies, which would be equivalent to higher image resolution, is liable to be lost during image formation. Lanthanum hexaboride has been used as the electron source for calculation of the curve (b) with partial coherence. As discussed in the following section, additional specimen information can possibly become available by using a high-coherence electron source such as the field-emission gun (FEG).

2.2. Definitions of Resolution

Resolution is traditionally discussed in terms of the ability of an imaging system to discriminate between two discrete objects, and is usually closely coupled to the wavelength of the incident illumination. Since high-energy electrons have picometer wavelengths, resolution limits on the same scale might be anticipated. In practice, electron lenses
have unavoidable aberrations, and perfectly coherent electron sources are unattainable. It then becomes necessary to make a compromise between the traditional diffraction limit, which varies inversely as the aperture angle, and spherical aberration, which varies rapidly with the cube power of the angle. The end result is an expression of the form

\[ d = AC_s^{-1/4} \lambda^{3/4} \]  

(1)

where \( C_s \) is the spherical aberration coefficient of the objective lens, \( \lambda \) is the electron wavelength, and the value of the constant \( A \) depends on certain assumptions that are made about the imaging conditions.

In practice, there are several resolution limits that are applicable to any specific HRTEM [24]. These are conveniently understood by reference to the TF of the objective lens. The interpretable resolution, which is sometimes known as the structural or point resolution, is defined by the position of the first zero crossover of the TF at the optimum defocus. This resolution is indicated by the arrow in Fig. 1. This specific defocus gives the widest possible band of spatial frequencies without a phase reversal, and the corresponding interpretable resolution is then given by \( \delta \sim 0.66(C_s \lambda^{3})^{1/4} \). Values of \( C_s \) inevitably increase as the electron energy is increased due to limits on the saturation of the pole-piece material. However, overall improvements in \( \delta \) should be obtained due to reductions in \( \lambda \) as the accelerating voltage is increased. Typical resolution figures using realistic \( C_s \) values (which range from 0.3 mm at 100 keV to about 1.5 mm at 1.0 MeV) are in the range of 0.25 to 0.12 nm. Other practical factors such as the size and cost of higher-voltage electron microscopes, as well as the increasing likelihood of electron irradiation damage due to the higher energy electron beam, become important considerations. Intermediate-voltage HRTEMs which operate in the range of 200 to 400 kV have become widespread due to the impact of these factors.

The instrumental resolution or information limit of the HRTEM is usually defined in terms of the damping produced by the envelope functions. A value of approximately 15% [i.e., \( \exp(-2) \)] is commonly taken as the resolution cutoff since this level is regarded as the minimum acceptable for image processing requirements [25]. In recent 200- or 300-kV HRTEMs equipped with an FEG electron source, this resolution limit extends well beyond the interpretable resolution. Because of the oscillatory nature of the TF, the very fine detail present in the image at such resolution levels will, however, not be easily related to specimen features. Higher-order diffracted beams with inverted phase can be prevented from contributing to the image by using an objective aperture of suitable diameter located in the back focal plane of the objective lens. Conversely, provided that the defocus and \( C_s \) values are known with sufficient accuracy, then the phase-modulating effects of the TF can be removed by \( a \) posteriori image processing. Image interpretability with improved resolution can be achieved, as demonstrated by the pioneering work of Coene et al. [26] who used a focal-series reconstruction approach to resolve oxygen atoms for the first time in a high-temperature superconductor.

The term lattice-fringe resolution refers to the finest spacings of the lattice fringes that are visible in high-resolution images from a particular HRTEM. These lattice fringes
result from the interference between two or more diffracted beams from a crystalline material. This resolution is determined by the overall instrumental stability of the HRTEM, and is often a direct reflection of freedom of the microscope environment from adverse external factors such as acoustic noise, mechanical vibrations, and stray magnetic fields. It must be noted, however, that lattice fringes with very fine spacings do not usually provide any useful information about local atomic arrangements. The interfering diffracted beams typically originate from comparatively large specimen areas, and the lattice fringe images may be recorded at significant underfocus conditions when there are many TF oscillations. For many years, the lattice-fringe resolution was widely regarded as the ultimate figure of merit for an HRTEM. It should be appreciated that the interpretable and instrumental resolution limits are nowadays considered far more useful for comparison purposes.

2.3. Lattice Imaging or Atomic Imaging

Most elemental and compound materials have unit cells that are relatively large in comparison with the resolution limits of commercially available HRTEMs. High-resolution lattice-fringe images are thus relatively straightforward to obtain when these materials are viewed in major low-index projections. However, it is important to appreciate that only a very small subset of these lattice images can be interpreted in terms of atomic arrangements. The basic problem is that the phase reversals that result from TF oscillations when the defocus of the objective lens is changed make it difficult to recognize or select the appropriate optimum defocus. The characteristic image of a crystal defect or the Fresnel fringe along the edge of the sample is often indispensable for recognizing a particular defocus setting. Some prior knowledge or calibration of the focal step size corresponding to the finest focus control is thus essential. The situation is even more complicated for materials with small unit cells which have comparatively few diffracted beams contributing to the final image. For these materials, identical images, referred to as Fourier or self-images, recur periodically with changes in defocus, with the period given by $2d^2/\lambda$, where $d$ is the corresponding lattice spacing [27, 28]. With a half-period change in defocus, all image features reverse in contrast, with black spots becoming white and vice versa. As an example, a series of image simulations for [001] tin dioxide is shown in Fig. 2, where the major reversals in contrast of the image features from white to black to white to black are readily apparent [29].

Further misleading complications are likely to arise in thicker crystals because of dynamical multiple scattering. As the sample thickness is increased, intensity will be progressively lost from the directly transmitted beam, and it will instead build up in the diffracted beams. Eventually, a thickness is reached where the direct beam will be relatively low in intensity (in the vicinity of what is termed the thickness extinction contour), and the resulting image will be dominated by (second-order) interference processes between the various diffracted beams. In some small-unit-cell materials, these interferences can often lead to pairs of white spots, graphically referred to as dumbbells, which appear to portray faithfully all atom positions in the unit cell [30]. Except under highly specific thickness and defocus values, it has been found that the separation
II. Electron Microscopy

Figure 2. Through-focal series of image simulations for crystal of tin dioxide in [100] projection: 500 keV, $C_s = 3.5$ mm, crystal thickness $-2.8$ nm. Focus extends in 10 nm steps from $-140$ nm (top left) to $+10$ nm (bottom right) [29].

between these spots will invariably be slightly different from the correct projected atomic separations for the specific material [31]. Moreover, erroneous image features can also be anticipated to occur at structural discontinuities such as interfaces [32]. Such interference lattice-fringe images from thicker specimen regions should thus be considered as nothing more than interesting coincidental effects and they should never ever be interpreted in terms of atomic positions.

2.4. Instrumental Parameters

In common with other sophisticated pieces of equipment, the HRTEM has an array of adjustable parameters that must often seem almost overwhelming to the newcomer. However, only a small subset must be known with any degree of accuracy. Experimental determination of these instrumental parameters is well documented elsewhere [24], so these details are not reproduced here. For atomic-resolution imaging, it is essential that the accelerating voltage be fixed and highly stable, preferably to within one part per million (ppm), but the exact value is relatively unimportant. A similar situation holds for the current of the objective lens, which will determine its focal length and several other imaging parameters, including $C_s$, the spherical aberration coefficient. In practice, it is highly desirable to operate with a fixed objective lens current, due to the extreme sensitivity of several adjustable parameters such as the incident-beam tilt alignment and the objective lens astigmatism to the exact current setting. The objective lens current should thus be monitored continuously, and when the sample is tilted or another field of view is selected, the sample height should be adjusted rather than altering the objective lens current. Knowledge of the objective lens defocus is critical to image interpretation because of the extreme sensitivity of the image appearance to the defocus value. The highly characteristic image appearance
of complicated structures at the optimum defocus can be recognized in some special cases, but for unknown aperiodic features such as a dislocation or grain boundary, some other means for selecting defocus is required. It has become increasingly common to generate a through-thickness-through-focus tableau of images, at least of the perfect crystal structure, before commencing microscopy [33]. Some electron microscopists have recommended recording a focal series of images with pre-calibrated focal step sizes, which thereby reduces somewhat the subjective element of the matching process [34]. Methods based on cross-correlation [35], and non-linear least squares [36] have been developed for matching experimental and simulated images, and these can provide determination of both defocus and local specimen thickness on a local scale.

At extreme resolution limits, several higher-order objective-lens aberrations have a dominant role in determining the overall image integrity. Two-fold image astigmatism is well-known to result in focal length differences in orthogonal image directions, with a corresponding effect on the image appearance, but its presence can usually be detected by reference to the image of a thin amorphous material, such as carbon or germanium: Fourier transformation reveals a set of concentric rings, that are more or less elliptical depending on the amount of astigmatism present [37]. The successful implementation of image reconstruction schemes involving either focal series [26] or off-axis electron holography [38] requires knowledge of the spherical aberration coefficient with an error of no greater than 1%. The graphical method of optical diffractogram analysis [37] is no longer sufficiently accurate, and alternative methods are required in order to reach the desired level of accuracy [39]. The third-order aberrations of axial coma and three-fold astigmatism are more difficult to detect and quantify because their influence is only clearly visible in images of amorphous materials recorded with tilted incident illumination [40, 41]. Detection and correction of third-order aberrations is discussed further in Section 3.4.

2.5. Further Requirements

There are additional requirements that must be satisfied before the atomic-scale resolution of the HRTEM can be fully utilized for characterizing nanostructured materials. The sample region being imaged must be sufficiently thin, typically meaning that its thickness should be on the order of 15 nm or less depending on its composition. Otherwise, multiple electron scattering effects become significant, and the very fine details present in the image are not likely to be interpretable in terms of atomic arrangements, even with the assistance of image simulations. Since the final recorded image represents a two-dimensional projection of the crystal structure, the incident electron beam must be aligned closely with a major zone axis of the crystal. The tolerance for adjustment of the crystal alignment to avoid overlap of projected atomic columns obviously becomes more demanding for thicker crystals and smaller column separations [42]. The examination of crystalline defects must be confined to planar faults, such as twin boundaries and stacking faults, and linear defects, such as dislocations, which are aligned so that they are parallel with the incident beam direction. The defects should also be periodic through the entire projected structure. Inclined faults and curved
grain boundaries will not in general be amenable to fruitful study by high-resolution imaging (nor any other imaging technique). Despite these various constraints, there have been many, many successful studies where the HREM has proven indispensable in evaluating nanostructured materials. The highly selective group of examples briefly described below is chosen to be illustrative of the myriad possibilities. The need for complementary information from macroscopic measurements, including electrical and optical properties, as well as other types of microscopy, should not be overlooked.

2.6. Milestones

The performance of the “high” resolution electron microscope first exceeded that of the optical microscope in the mid-1930s [43], but direct correspondence between lattice images and projected crystal structure of large-unit-cell block oxides was not achieved for many years [44, 45]. Improved instrumentation eventually led to directly interpretable information about atomic arrangements for metals, ceramics and semiconductors. Individual atomic columns in a small gold particle were resolved [46], and several high-voltage HRTEMs surpassed the 0.2 nm interpretable resolution limit on a routine basis in the early 1980s [47–50]. Intermediate-voltage HRTEMs that could attain this performance level regularly also became available commercially [31, 51, 52]. The highly coherent illumination available with 300-kV FEG TEMs facilitated the achievement of instrumental resolutions closely approaching 0.1 nm [53], and image interpretation to the same level was achieved by through-focal series reconstruction [26] and off-axis electron holography [38]. The latest generation of high-voltage (1–1.5 MV) HRTEMs have succeeded in closely approaching the long-sought-after goal of 0.1 nm without reversals in the contrast transfer function [54–56], enabling direct image interpretation without the need for a posteriori image processing. Significant information transfer beyond the first zero crossover of 0.105 nm was clearly demonstrated by the 1.25-MeV HRTEM in Stuttgart [57]. Sub-Ångstrom electron microscopy to resolutions of better than 0.09 nm has since been achieved using exit-wave retrieval [58, 59] and also using aberration-corrected annular-dark-field imaging with a scanning transmission electron microscope [60].

3. APPLICATIONS OF HRTEM

Imaging with the HRTEM enables individual atomic columns to be resolved in most inorganic materials, making it possible to determine the atomic-scale microstructure of lattice defects and other inhomogeneities. Structural features of interest include planar faults such as grain boundaries, interfaces and crystallographic shear planes, linear faults such as dislocations and nanowires, as well as point defects, nanosized particles and local surface morphology. Additional information can be extracted from high-resolution studies, including unique insights into the controlling influence of structural discontinuities on a range of physical and chemical processes such as phase transformations, oxidation reactions, epitaxial growth and catalysis. The HRTEM has impacted many scientific disciplines, and the technique has generated a vast scientific literature, far too extensive to be reviewed in any detail. Our intent here is to describe,
albeit very briefly, some representative examples that illustrate the range of applications, while additional information and many more examples can be found elsewhere [1–13].

3.1. Semiconductors

The characterization of semiconductors has been a highly active and fruitful area of research for HRTEM. Because of the relatively large unit-cells of elemental and compound semiconductors, ranging from 0.54 nm (Si) to 0.65 nm (CdTe), high-resolution lattice-fringe images are easily obtained in the $<110>$ orientation. However, individual atomic columns are not separately resolved unless the microscope resolution allows $\{004\}$-type reflections to contribute to the imaging process [61]. True atomic imaging for elemental Si and Ge was demonstrated in $<100>$, $<111>$ and $<013>$ orientations under carefully chosen imaging conditions using a 400-keV HRTEM [62]. Discrimination between atomic species on the basis of image spots of different intensity can be achieved for compound semiconductors with sufficiently different atomic numbers, which is particularly useful for analysing interfaces between dissimilar materials.

**Dislocations:** Individual atomic columns have not actually been resolved in almost all HRTEM structural studies so far reported. Extensive comparisons with simulated images for various models of Lomer edge dislocation cores in Si and Ge enabled the locations of atomic columns to be deduced to within $\sim0.025$ nm [63]. HRTEM analysis of dissociated $60^\circ$ dislocations in CdTe showed that alternative structural models (glide set and shuffle set) could be clearly differentiated, but again without resolution of individual atomic columns [64]. By taking advantage of the improved microscope resolution that has recently become available, more accurate modelling of dislocation core structures in semiconductors should nowadays become possible. Figure 3 shows an example, recorded with the 1.25-MeV HRTEM in Stuttgart, of a symmetrical Lomer edge dislocation at a Ge/Si(001) heterointerface, together with the structural model derived directly from the HRTEM image appearance [65].

![Figure 3.](image)
Interfaces: Many semiconductors have grain boundaries (GBs) and heteroepitaxial interfaces that are almost atomically flat and can be aligned edge-on to the incident beam direction, with the crystals on both sides of the boundary oriented to a low-index zone axis. The Ge (130) $\Sigma = 5$ GB was studied in [001] and [013] projections, so that a complete three-dimensional crystallographic analysis could be performed [66]. The Ge (112) $\Sigma = 3$ GB was shown to have a $c(2 \times 2)$ periodic supercell of the geometrical coincidence lattice of the boundary [67], and the rotation-twin nature of a GaP $\Sigma = 3 \{111\}$ twin boundary was unambiguously identified from the characteristic image features [57].

Heterointerfaces are of immense practical importance. Interface roughness can be made apparent by choosing appropriate thickness and defocus values to emphasize contrast differences between materials, and composition gradients can be assessed from the abruptness with which the two characteristic contrast motifs terminate at the interface. For materials with large lattice mismatch, significant image features can usually be interpreted without resorting to special defocus/thickness combinations. Examples of large misfit systems where the atomic structure of interfacial misfit dislocations have been successfully determined include the GaAs/Si(001) interface [68], the CdTe(001)/GaAs(001) interface [69], and the CdTe(111)/GaAs(001) interface [70]. For silicide-silicon interfaces, the atomic facetting of asymmetrical twins and asymmetrical “hetero”-twins was investigated [71], and structural studies have been reported for various CoSi$_2$ (A, B)/Si(111) interfaces [72, 73]. Figure 4 shows an interesting example of a CoSi$_2$ nanowire formed by self-assembled epitaxial growth on Si(100) substrate at 750°C [74].

The Group III-nitrides of AlN, GaN and InN have potential applications in short wavelength optoelectronic devices, based on their wide bandgaps which range from 1.9 eV (InN), to 3.4 eV (GaN), to 6.2 eV (AlN), and their excellent thermal properties make them ideal candidates for high-temperature and high-power devices [75]. However, the lack of substrate materials that are both lattice- and thermally-matched...

Figure 5. High-resolution electron micrograph showing cross-section of GaN/SiC interface. Analysis of lattice spacings along planes labeled $a$, $b$, and $c$ confirmed interface abruptness [76].

represents a serious obstacle to ongoing research. Thus, there is much interest in understanding the atomic structure at the substrate-nitride interface. Figure 5 shows a high-resolution image of an GaN/SiC interface, recorded at the optimum defocus so that atomic-column projections have black contrast. Image analysis established that the GaN/SiC interface was atomically abrupt (between the planes labelled $a$ and $b$), and that atomic arrangements across the interface primarily consisted of N bonded with Si, but with some Ga bonded with C in order to maintain charge balance [76].

3.2. Metals

Defects in metals pose greater challenges for characterization using HRTEM because of reduced unit-cell dimensions. Atomic imaging is restricted to comparatively few low-index zone axes even with the latest generation of instruments. High-resolution imaging requires very thin foils but substantial atomic rearrangements, especially relaxation in the vicinity of lattice defects and crystal surfaces, are liable to alter the defect structure so that it becomes atypical.

**Grain boundaries and interfaces**: High-resolution observations of symmetrical Au $<110>$ tilt GBs established the presence of different recurring structural units [77]. Structural characterization of an Al $[100]$ 45° twist plus 17.5° tilt GB revealed that this asymmetrical GB was composed of a mixture of two basic structural units [78]. Observations of a Mo bicrystal revealed the presence of both $\Sigma = 25$ and $\Sigma = 41$ structural units along the GB [79]. Bi segregation was reported to modify the structure of grain boundaries in Cu [80], and GBs became faceted and an ordered Bi layer was observed at a $\Sigma = 3$ twin boundary [81]. Figure 6 shows an example of an Al 6° [001]
symmetric tilt GB with prominent 1/2<110> edge dislocations that accommodate the tilt misalignment [82].

Structural modelling based upon atomistic simulations has accompanied many HRTEM GB studies. The incoherent Al {112} twin GB was observed [83], and comparisons were made with structures calculated using the Embedded Atom Method (EAM). For the Al Σ = 9 (221)[110] GB, glide-plane and mirror-plane symmetric structures were observed to alternate periodically along the boundary; the predictions of the various atomistic modelling approaches could only be distinguished for one of these two structures but not the other [84]. The atomic structure of the Nb (310) twin boundary was generated using interatomic potentials derived from several approaches but only one, the so-called model-generalized pseudopotential theory, successfully predicted the mirror symmetry later observed in experimental images [85]. For the Nb Σ = 25 (710)/[001] twin, a multiplicity of stable, low-energy structures were predicted by EAM: four possibilities out of 13 remained after careful comparison with experimental micrographs [86]. Molecular dynamics simulations of the Ag Σ = 3<110> (211) twin boundary predicted a thin boundary phase having the rhombohedral 9R structure, and this prediction was confirmed by experimental HRTEM micrographs [87]. Figure 7 is an experimental image showing the existence of the same rhombohedral phase at an incoherent Cu Σ = 3 (211) boundary [88].
Atomic-level structural investigations of intermetallic alloys and several metal/metal systems have been reported. For example, the NiAl $\Sigma = 5 \{001\}(310)$ GB was imaged by high-resolution electron microscopy and analyzed with the assistance of image simulations [89]: changes in local stoichiometry and a rigid body translation along but not normal to the boundary were inferred. A quantitative study of the NiAl $\Sigma = 3 \{111\}$ GB enabled atomic positions at the boundary core to be determined with an accuracy of $\sim 0.015$ nm [90].

Dislocations: Determination of the atomic structure of dislocation cores is simplified by the presence of only one atomic species but the core may not be stable in metal foils thin enough for atomic-resolution viewing due to the possibility of core spreading or even defect motion in the form of glide to the thin foil edge [91]. Moreover, structural rearrangements during high-resolution observations of thin metal foils have been observed by many workers (see, for example, [92]). The $1/3<1120>\{1010\}$ edge dislocation in $\alpha$-Ti was determined to have a planar elongated core structure [93]. Atomic modelling of the core structure of $a<100>$ and $a<110>$ dislocations in the intermetallic alloy NiAl revealed that the former had large elastic strain fields but were usually undissociated, whereas the latter either decomposed into other types of dislocations or climb-dissociated into two partial dislocations [94]. An investigation of dislocation core structures in the ordered intermetallic alloy TiAl has also been reported [95].

3.3. Oxides and Ceramics

Most ceramic materials consist of two or more atomic species and only rarely can the separate elements be discriminated in HRTEM micrographs. The image of the 6H polytype of SiC in the (1120) orientation shown in Fig. 8 is a particularly noteworthy...
example since the closest separation of the clearly resolved Si and C columns is only 0.108 nm [55]. Structure imaging in the electron microscope originated with large-unit-cell block oxides, although the image interpretation was initially based on contrast features corresponding to the location of tunnels [45]. Improvements in HRTEM resolution to better than 0.2 nm led to oxide images in which the black spots visible at optimum defocus were interpretable in terms of atomic column positions. It is then possible to deduce directly the detailed atomic structure of complicated shear defects and precipitation phenomena observed in doped and slightly reduced oxides, as represented by the example of a pentagonal bipyramidal defect as shown in Fig. 9 [96].

Grain boundaries. Symmetric tilt GBs in NiO bicrystals had a multiplicity of distinct structural units [97], and asymmetric structural models based on experimental micrographs, differed from symmetric models derived from theoretical considerations. Observations of the $\Sigma = 5 \langle 210 \rangle / \langle 001 \rangle$ symmetric tilt GB in yttrium aluminum garnet were compared with model atomic structures [98]. Structural investigations of a near-$\Sigma = 5 \langle 210 \rangle$ GB in TiO$_2$, rutile, revealed a stepped boundary with well-defined lattice dislocations at the steps, and extended, flat terraces that consisted of $\Sigma = 5 \langle 210 \rangle$ segments with mirror glide symmetry [99]. Observation of an undoped SrTiO$_3$ $\Sigma = 5 \langle 130 \rangle$ symmetrical tilt GB revealed that it was composed of repeating structural units [100]. Examination of a $25^\circ \langle 001 \rangle$ tilt boundary in SrTiO$_3$ provided an initial structural model, and bond-valence sum calculations based on electron-energy-loss spectroscopy at the boundary were then used to refine the O atom positions [101]. A combined structural

Figure 9. (a) Atomic-resolution electron micrograph of nonstoichiometric (W, Nb)O$_{2.93}$ showing pairs of pentagonal bipyramidal columnar defects. (b) Corresponding structural model. Occupied tunnel sites are located by direct visual inspection [96].

and spectroscopic investigation of coherent (111) twins in BaTiO$_3$ enabled a modified structural model of the boundary to be proposed [102]. The atomistic structure of 90° domain walls in ferroelectric PbTiO$_3$ thin films was investigated by using digital processing to determine lattice parameter variations across the walls [103].

Interfaces. Knowledge of metal–ceramic interfaces promotes a better understanding of bulk mechanical properties. For thin layers of Nb and Mo deposited on R-plane sapphire, misfit dislocations were offset from the interface for Nb films whereas they were localized very close to the interface for Mo films [104]. Interactive digital image matching was used in a comprehensive study of the Nb/sapphire interface: and translation vectors were determined with a precision of $\sim 0.01$ nm [105]. The atomic core structure of the dislocations was determined in another study of the same interface [106].

Ceramics. Most ceramics are close-packed materials with tetrahedral bonding so that atomic resolution is very difficult to obtain. Careful examination of {100} nitrogen platelets in diamond led to the development of a novel “nitrogen-fretwork” model [107], while later observations led to a more refined model [108]. Interfacial structures and the defects that occurred during heteroepitaxial growth of β-SiC films on TiC substrates have been characterized [109]. The atomic structure of Ti(C,N)-TiB$_2$ interfaces was investigated using a combined microscopy-simulation study [110].

3.4. Surfaces

Several different TEM configurations provide atomic-scale information about surfaces [111]. In surface profile imaging [112], the electron microscope is operated in the normal HRTEM mode and the optimum defocus image displays the surface profile at the structural resolution limit. Gold received much early attention, mainly because its surface was relatively inert and electron irradiation readily caused desorption of surface...
II. Electron Microscopy

Figure 10. Surface profile images showing reconstructed CdTe (001) surface at different temperatures: (a) 2×1 at 140°C; (b) 3×1 at 240°C [127].

contamination overlayers, and also because its large interatomic spacings and high atomic-column visibility made it easy to characterize atomic rearrangements [113]. Thus, profile imaging was first applied to observations of a 2×1 reconstruction of the gold (110) surface [114]. Adsorbed Bi atoms have been identified on reconstructed Si(111) surfaces using an ultrahigh-vacuum (UHV) transmission electron microscope [115], and the overgrowth of Au on ZnTe has been investigated [116]. In later experiments involving in situ Au evaporation under UHV conditions, the ×5 superperiod associated with the corrugated (5×28) reconstructed Au surface was observed [117].

Oxide surfaces are generally more straightforward to prepare for profile imaging [118], but surface modification may occur under intense electron irradiation due to electron-stimulated desorption of oxygen from near-surface regions [119, 120]. A complex spinel catalyst developed surface rafts, identified as ZnO, following prolonged use as an oxidation catalyst [121]. Surface profile images were central to studies of terbium oxide [122], Eu₂O₃ [123], and β−PbO₂ [124]. A direct correlation was made between the exposed surface structure of V₂O₅ oxide catalysts and the catalyst selectivity [125].

Profile images of semiconductor surfaces can only really be considered as valid when the surface has been cleaned inside the microscope. A novel 1×1 dimer reconstruction of a Si (111) surface was reported after the sample had been heated in situ to 1000°C [126]. As shown by the surface profile images in Fig. 10, the CdTe (001) surface undergoes a reversible phase transformation from a 2×1 structure at temperatures below about 200°C to a 3×1 structure at higher temperatures [127]. The atomic columns at the surface were located to within ~0.01 nm, and structural models were
developed: the $2 \times 1$ was determined as being Cd-stabilized whereas the $3 \times 1$ was found to be Te-stabilized.

3.5. Dynamic Events

Dynamic events such as phase transitions, defect motion and interface dynamics can be documented directly at the atomic-scale by means of a low-light-level TV camera attached to the base of the microscope without any loss of resolution. However, image recording is not useful for quantitative image analysis due to reduced dynamic response of the camera, as well as image distortions. Further information about dynamic in situ studies can be found in the chapter by Sharma and Crozier.

*Surfaces.* Surface profile imaging revealed movement of Au atomic columns across extended Au (110) surfaces [113], and rapid structural changes and the existence of “atom clouds” extending out from Au surfaces were reported [128]. Hopping of Au columns between surface sites on small Au particles was recorded at TV rates [129, 130], and similar hopping of Pt columns was also studied [131].

*Small particles.* Under strong electron irradiation, small metal particles (<8–10 nm) rapidly change their shape and orientation. Structural rearrangements from single crystal twinned to multiply-twinned were observed in small Au particles [130, 132]. Structural rearrangements as well as surface hopping were also documented in small particles of Pt [131] and Rh [133]. In the case of small Ru particles (~2.5 nm), the internal stacking changed between cubic-close-packed (ccp) and hexagonal-close-packed (hcp) which is the stable bulk form of Ru [134]. The term “quasi-melting” has been used to describe the structural fluctuations which have been reported for Au clusters supported on pillars of MgO [135].

4. CURRENT TRENDS

4.1. Image Viewing and Recording

Image viewing or recording should not normally be expected to affect resolution limits of the HRTEM but the recording media must still be properly optimized to ensure efficient operation. Ideally, every incident electron should be detected but readout noise and shot noise could affect the overall recording efficiency. The low-light-level TV camera has steadily evolved to the point where it has replaced the fluorescent screen for most image viewing. A camera is easily attached beneath the viewing chamber without affecting microscope performance, and high brightness images can be obtained even for very low exposure levels. Enlarged specimen detail is easily visible on a TV monitor, and permits image focussing and rapid astigmatism correction, while dynamic events within the sample can be viewed and recorded at TV rates. Nevertheless, high dark current and amplifier noise limit dynamic range, the input-output linearity is poor, and the number of resolvable picture elements is restrictive. The intensified TV camera is seriously inadequate as a recording medium for quantitative HRTEM studies.

The development of the slow-scan CCD camera has led to a revolution across the entire field of electron microscopy, and it has particular value for quantitative
HRTEM applications. The potential of the CCD camera had long been recognized [136], especially its sensitivity, wide dynamic range and overall usefulness for the electron microscopist [137]. However, dynamic viewing at TV rates seems unlikely to be achieved without sacrificing image quality. The imaging properties of the slow-scan CCD camera, the intensified TV camera and the photographic plate have been compared [138], and a comprehensive overview of characterization methods and important detection parameters, such as modulation transfer function and input-output linearity, for CCD cameras has been published [139]. The fixed location of the CCD camera enables geometric distortions of the imaging system to be accurately compensated [140], which is advantageous for extracting quantitative phase information during off-axis electron holography. On-line acquisition of a digital signal also enables automated microscope control or “autotuning” [137], as described in the following section.

4.2. On-Line Microscope Control

As resolution limits improve, it becomes progressively more difficult for an operator to adjust focus, to correct the objective lens astigmatism and to align the incident beam direction (coma-free alignment) with the accuracy that is required to ensure that image interpretation is not compromised. Accordingly, attention has been directed towards on-line computer control or “autotuning” of the microscope [137, 141]. The desired end-product of routine, top-quality micrographs should then leave the microscopist free to concentrate on solving the particular materials problem at hand.

Several criteria have been proposed as a basis for autotuning, including diffractograms, beam-tilt-induced image displacement, and contrast analysis for amorphous materials [24]. These methods rely upon signals that are fed into a computer, which analyses the data and then makes appropriate on-line adjustments to the microscope controls. Contrast analysis, the first autotuning method to be implemented successfully [141], locates a global minimum of the image variance as the computer successively iterates through the focus, beam tilt controls (in two orthogonal directions) and two objective lens stigmator controls. High accuracy for beam alignment and astigmatism can be achieved but the method is inefficient in terms of dose and inapplicable in the absence of amorphous material.

The latest variant of autotuning to be implemented [137], which is based on automated diffractogram analysis (ADA), utilizes diffractograms computed from slow-scan CCD images of amorphous material. The method works best at high image magnification, but is again inapplicable in the absence of an amorphous specimen region. The set of diffractograms in Figure 11, recorded (a) before, (b) after one cycle, and (c) after two cycles, of autotuning, illustrate the improvements that can be obtained with a 200-keV FEG-TEM. Astigmatism correction and focus adjustment to within 1 nm can be achieved, and the beam-tilt alignment is better than 0.1 mrad, which is beyond the adjustment/detection limit of the available instrumentation. Autotuning should be regarded as an essential procedure during preparation of the microscope for HRTEM observation.
4.3. Detection and Correction of Third-Order Aberrations

All electron lenses suffer from performance-limiting aberrations that must be corrected when possible, or at least quantified and accounted for during image interpretation. Two-fold astigmatism and third-order axial coma can be corrected during autotuning as discussed in the previous section. Detection and correction of spherical and chromatic aberration, as well as three-fold objective-lens astigmatism become much more critical as microscope resolution limits extend towards and beyond the 0.1-nm barrier [142–144].

Spherical aberration is unavoidable in rotationally symmetric electron lenses [145, 146]. Elimination of $C_s$ (and $C_c$) by a suitable combination of multipole elements attracted much attention over many years but all early correction attempts failed, due primarily to insufficient electrical stability and lack of alignment precision [146]. In recent years, a double-hexapole $C_s$-corrector system attached to a 200-keV FEG-TEM has enabled the normal 0.23 nm interpretable resolution limit of the instrument to be surpassed, with a level of $\sim$0.13 nm eventually being achieved after adventitious instabilities were removed [147]. Concurrently, a corrector system, incorporating multiple quadropole-octopole elements, has been applied to the probe-forming lens of a 100-keV scanning transmission electron microscope, and probe sizes of less than 0.1 nm can be achieved [148]. Note that both approaches to aberration correction are completely dependent on computer analysis of the imaging conditions and high-precision feedback to the numerous deflector and corrector power supplies. Some initial experiences with aberration-corrected HRTEM are described later in Section #4.5.

Knowledge of the chromatic aberration coefficient of the objective lens is not critical for high-resolution imaging since $C_c$ does not affect the interpretable resolution, and the temporal coherence envelope is determined by an effective focal spread, which can be estimated empirically if required [23]. Nevertheless, $C_c$ does impact temporal coherence, which is performance-limiting for an LaB$_6$ electron source, so that reduction...
or correction of chromatic aberration is desirable and worth pursuing. Correction of $C_c$ in a low-voltage scanning electron microscope was achieved [149] but no substantial success in the energy range applicable for HREM (upward from about 100 kV) has so far been reported. An alternative approach to reducing chromatic effects is to locate a monochromator immediately following the electron gun, which could serve to reduce the energy spread to about 0.1 eV [150], but this improvement can only be achieved at the expense of reduced beam current [151].

Three-fold astigmatism was quantified during the first comprehensive study of coma-free alignment [40]. However, it was not until the implementation of the ADA method for autotuning that the disturbing implications of three-fold astigmatism for high-resolution imaging at the 0.10 nm level came to be fully appreciated [41]. Like axial coma, three-fold astigmatism is effectively invisible in an axial bright-field image or the corresponding diffractogram since its basic effect is an asymmetrical shift of phase information. However, image simulations have shown that the impact on very-high-resolution images of crystalline materials can be highly detrimental, depending on the relative orientation of the three-fold astigmatism and any crystal symmetry directions [143, 144]. The magnitude of three-fold astigmatism can be estimated using diffractograms from four mutually orthogonal beam-tilt directions, and correction can be achieved using a pair of sextupole stigmator coils located near the back focal plane of the objective lens: Reasonable adjustments can be reached using fixed correction currents in existing objective stigmator coils provided that the coils are energized separately rather than being wired in pairs [152]. Microscopists purchasing new microscopes should insist that this highly desirable correction of three-fold astigmatism be done during installation and commissioning of their instrument.

4.4. Quantitative HRTEM

A major attraction of HRTEM is the possibility that atomic arrangements at local irregularities such as dislocations and interfaces can be determined to very high accuracy, some times closely approaching 0.01 nm. However, the refinement process is heavily demanding both of the microscopist and the instrument, as well as being computing-intensive and often extremely time-consuming with much trial-and-error parameter fitting. Prior knowledge (or elimination) of essential experimental parameters facilitates the goal of interactive structure refinement. Correction of the three-fold astigmatism at the time of microscope installation should reduce its effect to the level where it has almost negligible influence on the image. Routine application of autotuning, which implies the availability of an online CCD camera as well as computer control of relevant power supplies, would remove two-fold astigmatism (two parameters) and coma/beam tilt (two more parameters) from consideration. Thus, additional resources must be committed in order to make progress towards the ultimate goal of real-time, interactive structure refinement. For aperiodic structural features, it appears that this goal will not be easily attained since some parameters can only be optimized iteratively [105], although it is possible to automate considerably parts of the iteration process [153].
Historically, the determination of defect structures has relied upon qualitative comparisons between experimental micrographs and image simulations that were based on various alternative structural models. The acceptability of a specific structural model was generally considered as being enhanced when an image match was achieved for multiple members of a focal series (see, for example, refs. 85, 107). An alternative approach has been to overlay projected atomic column positions on the experimental and/or simulated images [66], and other studies have utilized superposition [79] or subtraction [154] of simulated and experimental images during refinement. A non-linear least-squares optimization approach was used to refine atomic positions at Nb GBs [155].

Attention has been given to the issue of quantifying the “goodness of fit” between experimental micrographs and the corresponding simulated images, as derived from postulated structural model(s). The so-called reliability or R-factors of X-ray or neutron diffraction studies refer to the contents of the entire unit cell, unlike aperiodic defects such as grain boundaries or dislocation cores studied by HRTEM where such well-defined discrete entities do not generally exist. Thus, alternative image agreement factors (IAFs) have been proposed and utilized at different times, and a useful summary of those most commonly used can be found in the Appendix of ref. [105]. However, it should be noted that slightly different results can be obtained depending on which IAF is used, possibly because different IAFs weight bright and dark contrast areas differently. It would be helpful if defocus and other microscope parameters were removed from the actual refinement process by instead using the complex exit-surface wavefunction of the specimen as the basis for comparison between simulation and experiment. At that stage, a $\chi^2$ goodness-of-fit criterion would be an appropriate test of overall convergence [156]. An additional benefit of determining the exit-surface wavefunction would be the availability of both phase and amplitude information about electron scattering by the sample. The possibilities for determining unknown object structures would be enhanced because this wavefunction directly reflects electron scattering by the object.

4.5. Aberration-Corrected HRTEM

Compensation of the spherical aberration of the objective lens offers the exciting prospect of directly interpretable image detail extending out to the HRTEM information limit without the need to unscramble the artefactual detail normally caused by TF oscillations. One additional benefit of the aberration-corrected HRTEM is that image delocalization, which is a major source of imaging artefacts at discontinuities such as interfaces and surfaces, is markedly reduced [147, 157]. Another benefit is that other imaging aberrations are also substantially reduced, which should simplify the process of exit-wave retrieval using through-focal reconstruction [158], and also alleviate the accuracy needed for sample tilting. However, it is relevant to reiterate here that conventional TF theory for the HRTEM refers to phase contrast imaging, so that when the $C_s$ value is reduced exactly to zero it would be necessary to use a projected charge density approach for image interpretation [159]. In practice, having an adjustable spherical aberration can provide additional flexibility to the microscopist.
interested in solving particular materials problems. For example, a small negative $C_s$ with a slightly overfocus condition enabled imaging of oxygen atom columns in a perovskite ceramic [160]. It is clear that much experimentation is still required to examine the full range of possibilities for aberration-corrected HRTEM, and possibly reach some consensus about standard imaging conditions.

5. ONGOING PROBLEMS

5.1. The Stobbs’ Factor

It has become increasingly obvious, and highly disconcerting, that there are substantial, seemingly inexplicable, discrepancies between the contrast levels of experimental and simulated images as well as diffracted beam intensities [161]. These differences were not apparent in earlier qualitative studies using photographic film when there was no simple measure of absolute intensity, and the contrast range in image simulations could easily be scaled to match that of the experimental micrographs. Initial quantitative studies revealed major differences in image contrast, sometimes by factors as large as 6 to 8 [161], although factors of about three are reported to be more typical [162].

Obvious sources of error, such as contributions from inelastic scattering and surface contamination overlayers, have not adequately accounted for these contrast differences [163], stimulating further concerted efforts to identify the origin(s) of what has come to be called the Stobbs’ factor [164]. Thermal diffuse scattering has been at least partly implicated by recent experiments combining energy-filtered imaging with off-axis electron holography [165], leading to suggestions for further experiments that might finally be definitive [164].

5.2. Radiation Damage

Interactions between the highly energetic electron beam and the sample within the electron microscope are always likely to result in permanent structural modification. There are two basic types of electron beam damage [166]: radiolytic processes (sometimes known as “ionization damage”) involve electron–electron interactions and affect most covalent and ionic solids; and direct atomic “knock-on” displacements, which occur above characteristic energy thresholds. An electron energy of 400 keV is sufficient to cause bulk displacements in elements as heavy as copper ($Z = 29$). Moreover, because of reduced binding energy, the energy needed for the activation of surface sputtering or for diffusion at defects or interfaces is considerably less than bulk values. For example, under continuous electron irradiation (400-keV, 5–15 $\text{A/cm}^2$), electron-stimulated desorption causes depletion of oxygen from the near-surface region of maximally valent, transition-metal oxides, leaving thin layers of reduced oxide covering the surface [119]. In a quantitative comparison of difference images, preferential damage was documented to occur at a Cu/sapphire interface [167], which limited the useful viewing time to a total of $\sim 10$ mins ($1250$ keV, magnification of 600,000×, specimen current density of $\sim 1.6$ $\text{A/cm}^2$).

Thus, the electron microscopist must be continually alert to the likelihood that the sample morphology has altered, probably irreversibly, during HRTEM imaging.
and especially microanalysis. Higher viewing magnifications as a means to improve signal statistics result in higher current densities, which will mean higher damage rates since the current density at the sample increases with the square of the magnification when the screen brightness is maintained at a constant level. For quantitative studies, the image magnification and the beam current density should be limited whenever possible, and the region of sample studied must be periodically checked for signs of structural change. A cautious microscopist will monitor the sample appearance during observation and thus erroneous results can be discounted once changes start to become apparent [168]. Digital micrographs of an NiAl $\Sigma = 3 (111)$ twin boundary, recorded with a slow-scan CCD camera in a high-voltage HREM, were compared at regular intervals as a means of establishing the useful observation time [90]. Finally, it should be apparent that structural change is even more likely to occur when the highly intense, focussed probe of the STEM is used for nanoscale microanalysis, and this possibility should always be monitored.

5.3. Inversion of Crystal Scattering

Inversion of crystal scattering to retrieve the crystal potential in the presence of dynamical scattering is a major unresolved problem. Several methods exist for retrieving the exit-surface wavefunction, and Fourier inversion can be used to extract the crystal potential for very thin samples when the kinematical or weak phase object approximations are valid. However, ab initio inversion of crystal scattering to retrieve the crystal potential has received comparatively scant attention over the years. An iterative method based on inversion of the multislice algorithm has been proposed [169], but further work is still needed to explore the applicability of the method, and for extending thickness limits in the case of non-periodic wavefields [170]. An alternative approach based on a simulated annealing algorithm has been explored for a perfect GaAs crystal with zincblende structure [171]. Successful reconstruction was achieved for a thickness of 5.6 nm but not for a thickness of 11.2 nm. Overall, it still holds true that because of the likelihood of multiple solutions to the inverse scattering problem for almost any sample of reasonable thickness, the uniqueness of the inversion process for unknown structures remains an unresolved issue.

6. SUMMARY AND FUTURE PERSPECTIVE

This chapter has provided an overview of HRTEM, with the objective of highlighting some of its applications and achievements, as well as identifying areas of ongoing research and development. The HRTEM enables the atomic structure of interfaces and defects to be determined routinely, reliably and with very high positional accuracy, thus providing better insights into the physical behavior of many nanostructured materials. Experts in the HRTEM field should find ways to ensure that their colleagues who are interested in the properties of these types of materials are given the opportunity and assistance required to capitalize on its attractive possibilities. Meanwhile, further developments in quantification and many novel applications can be anticipated. Online microscope control (“autotuning”), digital recording and computer...
II. Electron Microscopy

processing, (almost) real-time structure refinement, and in situ environmental electron microscopy are likely areas of concentrated activity. Many challenges remain. The differences between simulated and experimental contrast levels need to be fully explained. Better approaches to inversion of crystal scattering are needed. Operating conditions for aberration-corrected imaging need to be further explored. These issues will surely receive much attention over the next several years.

REFERENCES