X-ray spectromicroscopy with the scanning transmission X-ray microscope at BESSY II

G. Mitrea,[a]* J. Thieme,[a] P. Guttmann,[b] S. Heim[c] and S. Gleber[a]

[a]Institute for X-ray Physics, Georg-August-University Göttingen, Friedrich-Hund-Platz 1, D-37077 Göttingen, Germany, [b]Institute for X-ray Physics, Georg-August-University Göttingen, c/o BESSY GmbH, Albert-Einstein-Strasse 15, 12489 Berlin, Germany, and [c]BESSY GmbH, Albert-Einstein-Strasse 15, 12489 Berlin, Germany. E-mail: gmitrea@gwdg.de

Using the scanning transmission X-ray microscope at BESSY II, colloidal structures from a Chernozem soil have been studied with a spatial resolution around 60 nm and a spectral resolution of 1700 at the K-absorption edge of carbon. Elemental mapping has been used to determine the distribution of organic matter within the colloidal structures. Spectra have been extracted from image stacks to obtain information about the chemical state. For the analysis of the latter, principal component analysis and cluster analysis have been applied. It was possible, for example, to discriminate clay particles against organic components.

Keywords: X-ray spectroscopy; scanning transmission X-ray microscopy; NEXAFS; environmental science.

1. Introduction

1.1. Colloids from the environment

The penetration of soils by the atmosphere, hydrosphere and geosphere is expressed in the pore system of soils, where the distribution of the pore radii covers a wide range from the mm- down to the µm-scale. These pores, and with them the shape of the inner surface of soils, influence important physical and chemical parameters, *e.g.* the transport of substances within the soils. Transport processes are extremely important; examples are the water movement and diffusive transport of nutrients and toxicants. In the range down to 10 µm pore radius the inner structure of soils can be well determined and characterized by porosimetric methods. In the colloidal range with pore radii on the µm- or nm-scale these methods fail. With X-ray microscopy, however, it is possible to image directly the porous inner structure of soils in the colloidal size range and to study it.

The colloid size fraction in soils consists of organic and inorganic particles, which are found almost exclusively clustered into larger associations. Inorganic particles are mainly oxides and clays; organic particles result from the biological activity within the soil. The influence of biological activities is especially prominent in the upper part of soils. Colloidal humic substances, humins, are one result of these activities. These organic particles interact with other colloids, the kind of interaction, of course, depending on the chemical conditions of the environment. Important parameters of soils can be substantially influenced by, for instance, water flow or the transport of matter by diffusion.

Colloidal structures in soils and their morphological changes owing to changes in their environment have been imaged successfully using X-ray microscopy (Thieme *et al.*, 1998). With spectromicroscopy it is possible to add chemical information to this visualization. Elemental mapping performed at the K-absorption edge of carbon results in distribution maps of organic substances within a colloidal structure. Chemical information can be obtained by spectra extracted from stacks of images taken at X-ray energies around that absorption edge. Although the chemistry of the soil colloids is quite complex, it is possible to achieve maps of major constituents within the colloidal structures.

1.2. X-ray spectromicroscopy

The energy range between the carbon and the oxygen K-absorption edges at 284 eV and 543 eV, respectively, the so-called water window, is of particular interest for X-ray spectromicroscopy studies of hydrated samples. While the linear absorption coefficient of water in this energy region is small, it is notably higher for other substances, as shown in Fig. 1 (Thieme *et al.*, 2003). By fine-tuning the energy around X-ray absorption edges in the water window, spatial and spectral information can be obtained from the samples at the same time. When X-rays are directed into an object, some of the photons interact with matter and their energy can be absorbed or scattered, giving rise to an attenuation described by Lambert–Beer's law. The number of photons transmitted through a material depends on the thickness, density and atomic number of the material, and the energy of the indivi-
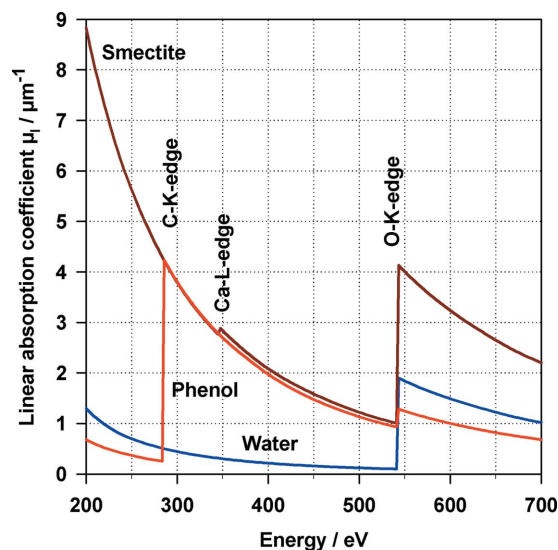
Figure 1
Linear absorption coefficient as a function of energy for water, the organic molecule phenol and smectite ($Si_{12}Al_5MgO_{39}H_{12}$), a clay mineral. Water shows a much lower absorption than other substances in the considered energy region yielding to good contrast for organic and inorganic specimens.
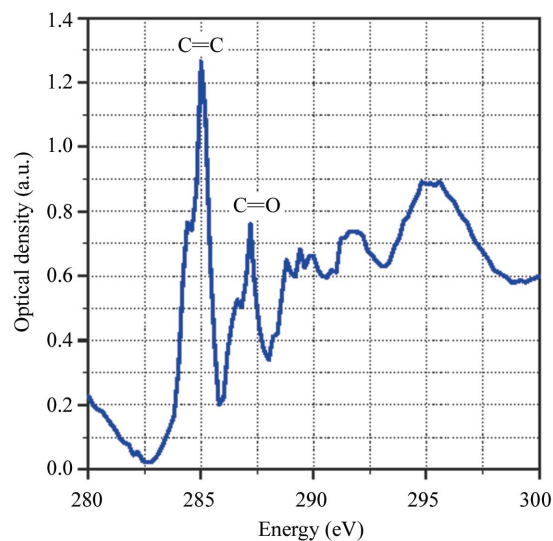

Figure 2
NEXAFS spectrum of a polyimide foil taken with the STXM at U41 at BESSY II. The optical density is plotted as a function of energy. Identification of functional groups is possible: the π^* resonances of the C=C and C=O groups are followed by the σ^* resonances of the C–N, C=C and C=O groups.

dual photons. At certain energies the absorption coefficient increases drastically giving rise to an absorption edge. Each edge occurs when the photon's energy is sufficient to cause excitation of an electron from the inner shells to the continuum state.

The absorption spectrum of an element near the X-ray absorption edge has been shown in the literature to contain much information about the chemical state, the molecular orientation on surfaces, the chemical hybridization and bond length. In particular, the so-called X-ray absorption near-edge structure (XANES) spectroscopy, also known as near-edge X-ray absorption fine-structure (NEXAFS) spectroscopy, is very useful. Around an absorption edge, peaks can be observed representing specific chemical states within the sample. For a uniform sample, the energy of such a peak corresponds to a specific bond of a chemical compound in the sample, while the intensity of the peak is a measure of the chemical compound density (Stöhr, 1992). Using a scanning transmission X-ray microscope (STXM), two methods can be applied to obtain spectroscopical data. The first one is the acquisition of spectra of regions as small as the spatial resolution of the microscope by changing the X-ray energy of the beam passing through the specimen. For studying heterogeneous samples, attention must be paid to move the zone plate along the optical axis to avoid a lateral drift of the focus across the sample. Fig. 2 shows the transmission spectrum of a polyimide foil taken with the STXM at the undulator U41 at the electron storage ring BESSY II in Berlin, Germany. Plotted is the so-called optical density (OD), derived from Lambert–Beer's law as a function of energy,

$$OD(E) = \mu(E)d = -\ln\left[I(E)/I_0(E)\right]. \qquad (1)$$

As the sample is a homogeneous foil, the zone plate has not been moved but is kept at the same focus position corresponding to the starting energy.

The other method of obtaining spectroscopic data, using a STXM, is to obtain spectra extracted from a set of images taken at closely spaced energies throughout the near-edge region, a so-called 'stack of images'. This method has been pioneered by Jacobsen and his group with the STXM at the X-1A beamline at National Synchrotron Light Source, New York, USA (Jacobsen *et al.*, 1991). The images have to be aligned with each other using a correlation method described by Jacobsen *et al.* (1991). After that, a spectrum can be obtained from each selected object point by plotting the optical density of all energies whose images have been taken. Spectra not only from discrete pixels but from larger regions (see Fig. 3) can be obtained by integrating the optical density of all pixels within a chosen area. The necessary I_0 can be derived from regions in the images without object structures. The transmitted intensities through the sample and the sample-free regions as a function of energy may be seen in Fig. 4. This method is more time-consuming than simply recording a spectrum, but it is true spectromicroscopy, having spatially and spectrally resolved data of an object at the same time.

2. Instrumentation: the STXM at BESSY II

The STXM at the electron storage ring BESSY II, in Berlin, Germany, is located at the undulator beamline U41-XM, which serves as a coherent light source (Wiesemann *et al.*, 2000; Wiesemann, 2003). It uses a Fresnel zone plate as a high-resolution optical element to create a diffraction-limited focus
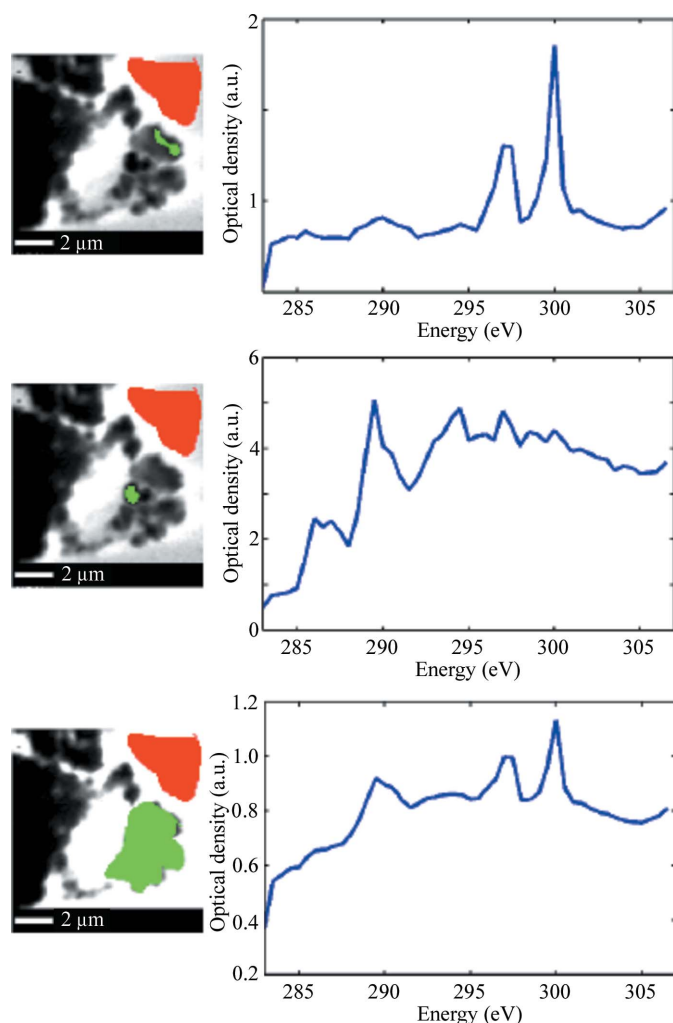
Figure 3
Selected regions for transmitted X-ray intensity I (green), incident X-ray intensity I_0 (red) and the derived spectrum at the C K-edge. The energy step size is 0.5 eV and the sample is dried Chernozem colloids.

Figure 4
Incident X-ray intensity I_0 (red) and transmitted X-ray intensity I (green) as a function of photon energy, corresponding to the spatial regions of the sample depicted in Fig. 3 (top). The energy step size is 0.5 eV and the sample is dried Chernozem colloids.

spot and, by moving the zone plate, scan across the sample. Many STXMs are available worldwide for a large field of experiments, *e.g.* the STXM at the Advanced Light Source (Ade *et al.*, 2003) with particular application to polymers, and a second STXM with applications for molecular environmental science has been installed (Tyliszczak *et al.*, 2003); the STXM at the European Synchrotron Radiation Facility (Barrett *et al.*, 2000) used together with a fluorescence detector. The undulator has been designed to have the maximum brightness of its first harmonic within the energy range of the water window. At E = 283 eV the full width at half-maximum of the first harmonic radiation is $\Delta E \simeq 5$ eV, corresponding to a monochromaticity well below $E/\Delta E \simeq 100$, which is an order of magnitude too wide for spectroscopy. Thus, a monochromator has been designed, built and placed between the undulator and the STXM: it increases the monochromaticity to a value of $E/\Delta E \simeq 1700$, which is sufficient for NEXAFS spectroscopy. The grating has a line density which varies along the grating plane to create parallel radia-
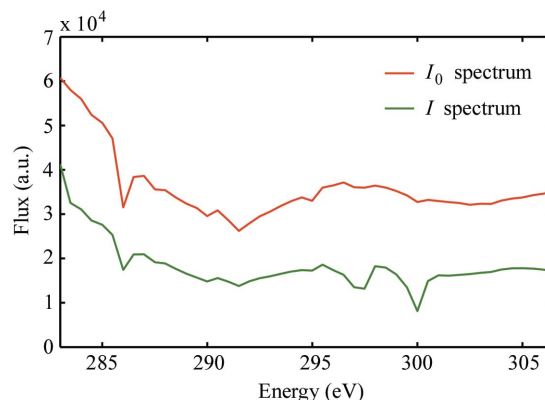
tion out of the divergent beam from the undulator (Früke, 2001). A grating with a mean line density of 1200 lines mm^{-1} has been used for the experiments described here, varying from 1197.32 to 1202.69 lines mm^{-1}. Following calculations, a monochromaticity of $\sim$3500 at the carbon edge should be possible. The diffraction efficiency has been measured to be $\sim$3.0% around the carbon K-edge (Wiesemann, 2003). The measured bandwidth reduction is smaller by a factor of two, which results in part from the fact that the positioning motors have a significant jitter. To avoid this inconvenience and gain in spectral resolution, an exchange of the entire monochromator is under way. With the current monochromator, the resulting maximum energy resolution with the STXM is about 0.18 eV at 300 eV.

The focusing optic is a Fresnel zone plate made by electron beam lithography (Rehbein, 2003). It has an outermost zone width, dr_N, of 50 nm, the number of zones, N, is 1000 and the diameter, $2r_N$, is 200 µm. A central stop of radius r_{CS} = 65 µm has been chosen to obtain large image fields for high-resolution scans. The central stop blocks the zeroth-diffraction-order light from reaching the sample.

The coherently illuminated area from the undulator *via* plane mirror and plane grating at the location of the zone plate has a diameter of 2 mm, thus ensuring that the zone plate with a diameter of 200 µm is illuminated for diffraction-limited resolution of δ = 60 nm in the first diffraction order. The zone plate is mounted on a piezo-electric scanner, permitting its movement in the X, Y directions perpendicular to the X-ray beam direction. Zone plate and piezo scanner are in a high vacuum of 10^{-8} mbar, and there are no windows between the undulator and optical element. An order-sorting aperture (OSA) of radius 40 µm is placed in between the sample stage and zone-plate stage. It selects the light of the first diffraction order to illuminate the sample and blocks the light from higher orders. In addition, the OSA acts as a vacuum window. It can be moved manually in all directions to be aligned with the zone plate. Stepper motors are used for the coarse movement of the sample in the X and Y directions. A

visible-light microscope can be moved in for prefocusing and alignment of the sample. The transmitted intensity through the sample is recorded on a pn-CCD detector (Strüder *et al.*, 1990) of 200 pixels × 64 pixels with 150 μm pixel size, which is used as a configurable detector (Morrison, 1994). A configured detector allows images to be taken in different modes: differential phase contrast (Morrison & Niemann, 1998; Morrison *et al.*, 2003), dark-field contrast (Chapman *et al.*, 1996) and, of course, bright-field contrast. The results presented in this paper were obtained by using the bright-field technique. A single pn-CCD unit of 200 pixels × 64 pixels is used to detect the transmitted X-rays through the sample. The CCD has 64 readout channels. The readout time of one row is 28.5 μs, thus the total readout time for 200 rows is 5.7 ms. Therefore, exposure time is limited by the readout time of the detector. For the characterization of the detector, mono-energetic photons generated by a Fe^{55} source and synchrotron radiation have been used. The noise of the pn-CCD pixels is calculated using the Mn $K\alpha$ emission at $E = 5899$ eV from the Fe^{55} source. A detailed explanation can be found by Wiesemann (2003). The average noise has been found to be 7.0 electrons. To produce one electron–hole pair, a photon with an energy of 3.65 eV is necessary; then an X-ray photon with an energy of 283 eV will generate 76 electrons. Thus, for one photon the average pixel noise is about one order of magnitude lower than its signal. To further decrease the detected noise during experiments, one illuminates the known area on the CCD with low noise levels. During experiments, the detector is kept at liquid-nitrogen temperature (77 K) and vacuum conditions of 3×10^{-8} mbar to reduce the dark current. The maximum photon rate for which the detector keeps its linearity can be calculated. The photon rate which illuminates the zone plate as a function of incident energy has been measured (Wiesemann, 2003) and the maximum has been found to be around 10^{10} photons s^{-1} $(100$ mA$)^{-1}$ at an incident X-ray energy of $E = 350$ eV. The focal length of the zone plate in the first diffraction order is calculated to be 2817 μm at $E = 350$ eV. If the CCD is placed at ∼21 mm from the sample plane, the illuminated area on the detector will have a ring-shaped form of 10 pixels in diameter, corresponding to 1.5 mm. With a continuous readout, 0.285 ms are needed to shift the charge across the 10 pixels. Supposing the transmission efficiency of the zone plate is 10%, and that the two silicon windows (exit and entrance window) used are transmitting around 40% of the incident beam, the maximum photon rate which can reach the detector is 4×10^{8} photons s^{-1} $(100$ mA$)^{-1}$ at $E = 350$ eV. The quantum efficiency of the detector is a function of energy and has been measured to ∼60% at $E = 350$ eV (Hartmann *et al.*, 1995/1996). Following the calculations, one pixel of the CCD will be illuminated by ∼2.4×10^{6} photons s^{-1}, corresponding to 2.3×10^{8} electrons s^{-1}. The full well capacity of one transistor is 10^{5} electrons. Taking the readout time into account, ∼3.51×10^{9} electrons s^{-1} $pixel^{-1}$ can be recorded. This corresponds to ∼3.65×10^{7} photons s^{-1} $pixel^{-1}$, more than one order of magnitude higher than the present maximum illumination, and in this way assuring no absorption saturation of the CCD.

3. Preparation of soil colloids as samples

An otherwise well characterized Chernozem soil has been taken as a first sample from the surrounding area of Göttingen (Ahl *et al.*, 1985) to demonstrate the performance of the STXM. The general characteristics of Chernozem from this area are the homogeneity and an alkaline pH of 8.2–8.4. The chemical composition of this soil is dominated by a high organic content (4.1%) and many inorganic elements (Fe: 1.08%; Mn: 562 p.p.m.; Al: 0.3%; Si: 0.85%) which are present as oxalate and citrates. A 1% dispersion of Chernozem has been made in deionized water. This has been seen already in previous experiments (Thieme & Niemeyer, 1998), using both TXM and STXM. A small quantity of ∼50 μl of this dispersion has been placed on a Si_3N_4 membrane, and allowed to dry. As a result, small aggregations of soil colloids could be found on the membrane. The goal of this paper is to underline the large and complex amount of information obtainable with a STXM. At the same time we want to prove the applicability of the analyzing routines to our set of data recorded with the STXM at BESSY II. For this purpose, thin dried samples from environmental science are a very good approach.

4. Data acquisition using a STXM

4.1. Imaging and elemental mapping

In bright-field X-ray microscopy, the absorption of X-rays as they pass through the specimen is measured. Images of Chernozem soil have been taken at the *K*-absorption edge of carbon. By changing the step size, the resolution in the image can be increased down to the minimum achievable resolution limited by the zone plate used. Zooming is also possible using the STXM, so one can image a large area of the sample and then choose a smaller region for detailed information, increasing also the image resolution owing to the smaller pixel size.

An important feature of the STXM is the capability of mapping chemical and elemental constituents. The dependence of the linear absorption coefficients on energy is element specific. This offers the possibility of mapping chemical elements using the STXM. Fig. 5 shows an example of the Chernozem soil sample. An image of the region of interest below [$E_1 = 280$ eV, (*a*)] and another image above [$E_2 = 310$ eV, (*b*)] the carbon absorption *K*-edge were taken. Away from an absorption edge, the linear absorption coefficient has an E^{-3} dependence. Therefore, a map of thicknesses for the element carbon can be calculated according to (2),

$$d_C = \frac{\ln(I_{02}/I_2) - (E_1/E_2)^3 \ln(I_{01}/I_1)}{\mu_{l2} - \mu_{l1}(E_1/E_2)^3}, \qquad (2)$$

where I_{01} and I_{02} are the intensities of the incident X-ray beams at E_1 and E_2, respectively, while I_1 and I_2 are the intensities of the transmitted X-rays at E_1 and E_2, respectively. The linear absorption coefficients μ_{l1} (at E_1) and μ_{l2} (at E_2) are known from tabulated data (Henke *et al.*, 1993), and the carbon mass density is assumed to be 1.05 g cm^{-3}. The elemental thickness map of carbon for our sample of Cher-
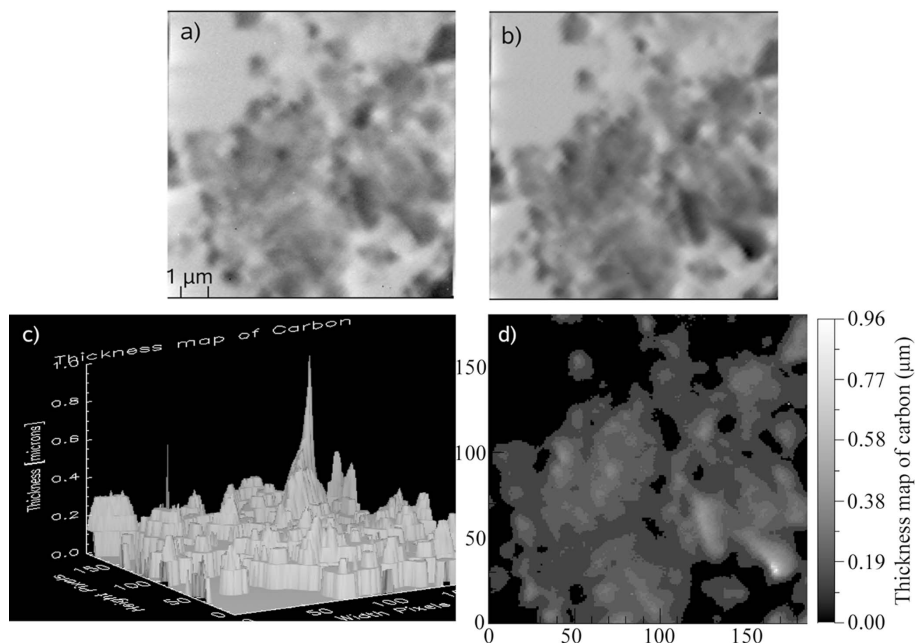
Figure 5
Elemental mapping of carbon in a Chernozem sample using the STXM at BESSY II. 200 pixels × 200 pixels, 20 ms pixel^{-1}, 10 μm × 10 μm, X-ray energy $E = 280$ eV (*a*) and $E = 310$ eV (*b*), in three-dimensions (*c*) and two-dimensions (*d*).

nozem is shown in Figs. 5(*c*) and 5(*d*). An alignment of the two images before calculating the thickness mapping is necessary (see §4.2).

4.2. Spectra and stacks

The STXM can also be used for acquiring spot spectra of samples. This can be done by tuning the undulator energy synchronously with the monochromator energy, and recording the transmitted intensity on the CCD. In Fig. 2, a spectrum of an aromatic polyimide foil taken with the STXM is shown. Comparison with the literature (Stöhr & Anders, 2000; Tanaka *et al.*, 2001) helps to identify in the acquired NEXAFS spectrum the π* resonances of the C═C double bond at ∼285 eV and at ∼286 eV, together with the C═O double bond at ∼287 eV. The broad peaks after the absorption edge, around 289 eV, are the σ* resonances for the C—N single bond and for the above-identified π* resonances. Attempts to record NEXAFS spectra for non-uniformly distributed specimens (Chernozem soil dry sample) have not been successful, yielding noisy spectra, and no fine-structure information could be extracted. The cause is found in the dependence of the focal length of a zone plate on the incident X-ray energy. To obtain spectral information from a non-uniform specimen, the zone plate has to be brought into the correct focus position corresponding to the energy. Thus, not only has the undulator to move synchronously with the monochromator, but at the same time the zone plate has to be translated to its corresponding focus position without additional lateral drifts. Owing to the previously discussed jitter motion of the monochromator, the synchronization of these three elements

is hard to accomplish without losing spectral resolution and it is time-consuming.

Therefore, NEXAFS spectra of non-uniform specimens are not easy to investigate close to the spatial resolution region of a scanning transmission X-ray microscope. An alternative approach, pioneered by Jacobsen *et al.* (1991), has been used. For this approach a stack of images of the sample is taken at many closely spaced energies. The stack of images was acquired manually; after each image the zone plate is moved to its corresponding focus position. A beam shutter is used to block the beam during energy changes. The time to change the energy of the undulator synchronously with the monochromator is added to the images acquisition time, and it is in the range of few seconds per energy step size. In this way, not only is an image of the sample obtained at each energy, but also a spectrum over the entire energy range can be extracted at each pixel. Focusing during the acquisition of a stack will yield to shifts of the images in the X and Y directions, perpendicular to the optical axis. Therefore an alignment of the entire set of images after the data acquisition is necessary and has to be employed. By courtesy of C. Jacobsen, a package of stack analyze routines, developed by his group using the Interactive Data Language (IDL), can be used to analyze the data recorded with the STXM at BESSY II in Berlin. A routine to convert BESSY data into data readable by the stack routines was programmed and used. Now, having the set of data, one has to analyze it. Several ways of analyzing spectroscopic data are used. Gaussian fitting permits the identification of chemical bonds in the sample (see §5.1). Principal component analysis uses a mathematical algorithm which can identify the principal components in the sample and represent them in terms of eigenvalues and eigenvectors (see §5.2). The last analysis method used in this paper is the cluster analysis method (see §5.3). It is a method of grouping the pixels from an entire stack of images, and extracting the maps and spectra of the sample components. To compare the results for all techniques, the same stack of images has been used.

5. Data analysis

5.1. Gaussian fitting

A spectrum, to apply the Gaussian fitting to, was extracted from the stack of images and was obtained by selecting regions of interest, *i.e.* the green zone of Fig. 3 corresponds to $I(E)$ while the red one gives $I_0(E)$. The *Specfit* package, developed by Gleber (2003) in IDL, was used. The spectrum was smoothed using the Golay–Savitzky algorithm, meaning that

Table 1
Functional groups identification of the spectrum extracted from a stack of images using the Gaussian fitting procedure.

Number	Energy (eV)	Group
1	283.8	$\pi^*(C\!=\!C)$
2	285.1	$\pi^*(C\!=\!C)$
3	286.6	$\sigma^*(C\!-\!H)$, $1s(C\!-\!H) \rightarrow \pi^*(C_6H_6)$
4	288.0	$\sigma^*(C\!-\!H)$, $1s(C\!-\!H) \rightarrow \pi^*(C_6H_6)$
5	289.6	$\pi^*(C\!=\!O)$
6	292.2	$\sigma^*(C\!-\!N)$, $\sigma^*(C\!-\!O)$
7	294.4	$\sigma^*(C\!=\!C)$, $\sigma^*(C\!=\!O)$
8	297.2	K L_{III} ($2p_{3/2} \rightarrow 3d$)
9	300.0	K L_{II} ($2p_{1/2} \rightarrow 3d$)
	290.3	Ionization potential
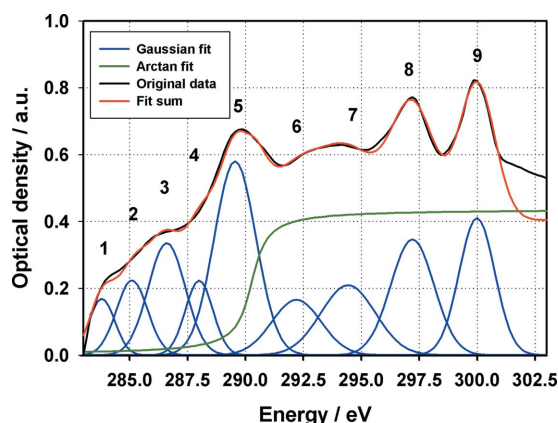
Figure 6
Gaussian fitting of a dry Chernozem spectrum extracted from a stack of images at the carbon K-edge taken using the STXM at BESSY II in Berlin. A Golay–Savitzky smoothing algorithm was applied prior. $\Delta E = 0.5$ eV.

the polynomial that best fits the spectrum is found. Owing to the lack of low energies, no quantitative work was possible, and the Gaussian fitting program was used for peak identification. A constant background was subtracted. When approaching an edge by tuning the energy from the lower end to the edge, the π^* resonances are typically low in energy, followed by Rydberg and/or hydrogen-derived resonances, and, above the ionization potential, σ^* resonances are present. The instrument was calibrated using the well known spectrum of CO_2. Following Smith *et al.* (1998), the $3s$, $3p$, $3d$ and $4p$ absorption peaks were compared, and a shift in energy of -3 eV was found and was included in the succeeding analysis. For the fitting process, the absorption edge was set to 290.3 eV (see Fig. 6 and Table 1). From the fitting process, at the carbon edge, five peaks below the absorption edge and two peaks above it could be derived. Each Gaussian peak corresponds to the electron transitions from the C $1s$ core level to the π^* and σ^* orbitals. At 283.8 eV (1) and 285.1 eV (2), the π^* resonance of the C$=$C double bond might be present. The 286.6 eV (3) peak might correspond to the $1s$ π_1^* resonance of aromatic components in our sample. The 288 eV (4) fitted peak might be the C$-$H bond, but also the π^* resonances of CH_2 and CH_3. The strong peak at 289.6 eV (5) is the π^* resonance of the C$=$O bond. At 292.2 eV the σ^* feature can be assigned to the C$-$O bond, but also to the C$-$N bond (6) and, according

to $1s$ π_1^* of the aromatic group, the second peak is the corresponding σ_1^* resonance of the aromatic group at 294.4 eV (7). At this energy also the σ^* feature of the C$=$O bond might be identified. Compared with the π^* positions, the σ^* energies exhibit larger scatter. The two remaining peaks in the spectrum indicate the potassium transitions $2p_{3/2} \rightarrow 3d$ and $2p_{1/2} \rightarrow 3d$ at 297.2 eV (8) and 300 eV (9), respectively. The shift in absorption energies, compared with the electron binding energies of pure potassium, occurs because it is surrounded by the organic components of the soil.

5.2. Principal component analysis

As shown in Fig. 6, our sample contains organic and inorganic components. Next, we would like to separate these regions with different chemical signatures in the sample. As a first approach, elemental mapping can be used. The thickness map of one chemical element in the sample can be obtained by applying equation (2), as already described in §4.1. In Fig. 6 the L_{III} and L_{II} resonances of potassium are identified. Attempts to chemically map the thickness of potassium were not successful, probably owing to its low concentration compared with that of carbon. One way of finding the distribution of chemical elements in the sample is to apply principal component analysis (PCA). The PCA algorithm has been tested for spectromicroscopical data with a model data set and then applied to known biological samples (Osanna, 2000). A set of significant components is found from our investigated data set, and its characteristic spectra (eigenspectra) (see Fig. 7) and maps (eigenimages) (see Fig. 8). The first component looks like an average of the entire sample, as image as well as spectrum. The second component seems to show the organic content in the sample. The third and fourth components show some characteristic maps, but no clear spectra can be extracted. The fifth and sixth components already look like noise. Nevertheless, to be sure that we do not miss some physical or chemical component in the sample, five significant principal components shall be considered for further analysis. One has to keep in mind that the PCA method is a purely mathematical algorithm and the significant components found have no physical or chemical content. In fact, when taking a close look at the results obtained, except for the first one, all five other significant components (see Fig. 8) show negative density maps (red) with false physical information.

5.3. Cluster analysis

A method of grouping pixels with similar experimentally determined spectra has been applied, and the entire data set has been successively analyzed according to these major groups. Such a method is called cluster analysis, and its applicability to spectromicroscopical data obtained using a STXM has been first tested with a set of simulated data, and then applied to known biological data (Lerotic, 2005). Two ways of grouping pixels have been developed: the Euclidean distance measure and the angle distance measure (Lerotic *et al.*, 2004, 2005; Lerotic, 2005).
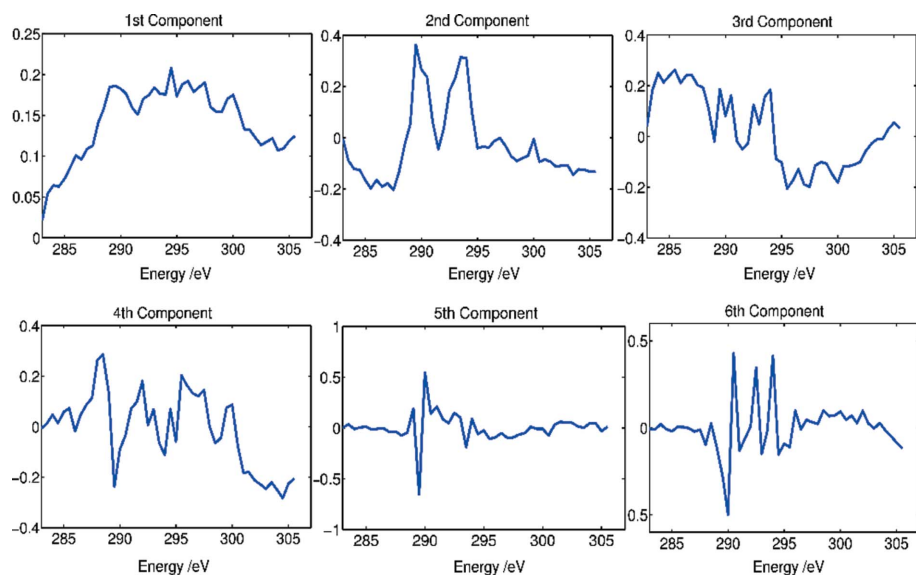
Figure 7
Results of principal component analysis, applied to a stack of images of Chernozem soil taken with the STXM at BESSY II. The eigenspectra of the first six principal components (PC) in the stack are shown. The first PC shows a spectral signature similar to that of the averaged sample, while the second PC shows some peaks around the carbon edge. From the last four PCs, no significant spectral information can be extracted.
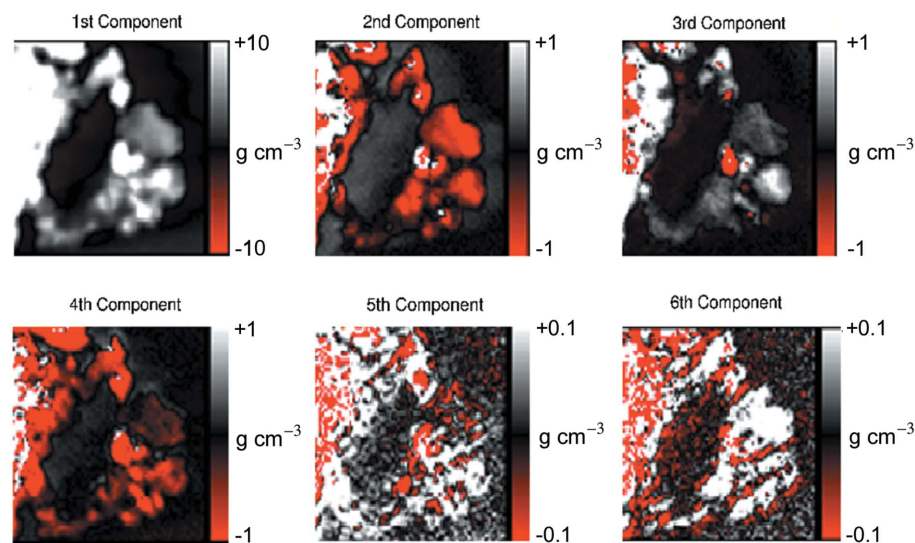

Figure 8
Results of principal component analysis, applied to a stack of images of Chernozem soil taken with the STXM at BESSY II. The first six eigenimages of the stack are shown. The first PC shows a distribution map identical to that of the averaged sample. Negative values in the maps have no physical representations. The fifth and sixth maps show noise, but five PCs will be further considered to avoid losing information.

5.3.1. Euclidean distance measure. The data points are grouped in a simple Euclidean distance measure from the cluster centers. The cluster algorithm starts with an initial guess of the cluster centers uniformly distributed around the origin. Prior, the consideration of five principal components fully describing our sample is assumed.

Fig. 9 represents the spectra of the first five clusters found, and in Fig. 10 the corresponding maps of the clusters are shown. The first cluster shows the presence of potassium and the organic components, the second cluster has a similar spectral representation but with the organic component more present, while the third cluster shows the organic elements present in its spectrum. The fourth cluster has the same organic signature as in the third one, plus an inorganic component, while the fifth represents just noise. Using the cluster analysis Euclidean distance measure, we could obtain spatially delimited cluster thicknesses maps together with their corresponding spectral signatures. We can identify organic and inorganic components in our sample, but we are still not able to separate these components spatially and spectrally. Comparing the first cluster with the original picture, the main part of the cluster area is free of sample. This could be explained as an indication of residuals on the Si_3N_4 foil, but it could also be due to the thickness variation error described by Lerotic (2005). The Euclidean measure method has been proven to find chemical grouping of data and calculate averaged thickness maps and spectra for these groups of data. However, it has been observed that variations in thicknesses in the sample are calculated as distinct clusters.

5.3.2. Angle distance measure. Better results are obtained by clustering the data according to an angle distance measure. As in the case of cluster analysis using Euclidean distance measure, five principal components together with an initial guess of the pixel origin are chosen. We are looking at the angular positions of the pixels in our data according to the chosen centers, and group them together forming the clusters. The number of clusters and the information contained in 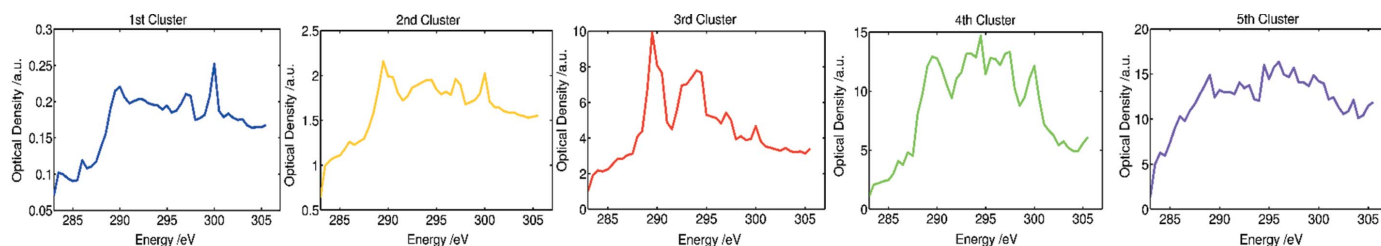each of them, images and spectra, should be the closed representation of chemical distinguished elements in our sample. The results obtained are shown in Figs. 11 and 12. The first cluster represents the sample-free region, while the spectral signature of the second cluster shows the inorganic component. The third cluster has a spectrum with a strong organic signature, but the inorganic component

Figure 9
Results of Euclidean measure, applied to a stack of images of Chernozem soil taken with the STXM at BESSY II. The spectra of the first five clusters are shown. The first cluster contains the spectral signature of potassium edges, the second cluster represents the averaged sample, while organic components can be identified in the third cluster. The fourth cluster looks similar to the third, but the inorganic component is more present. The fifth cluster represents noise.
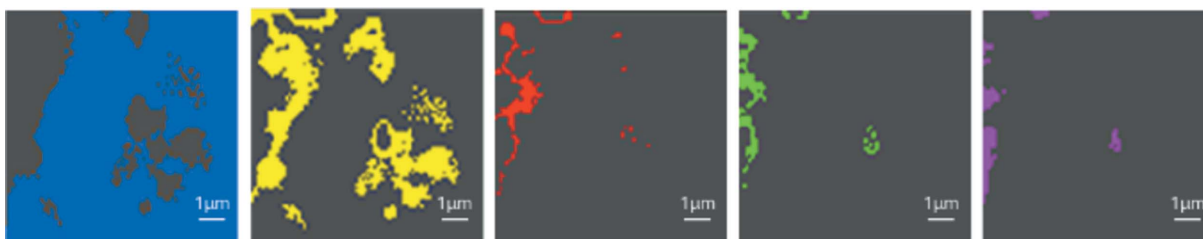

Figure 10
Results of Euclidean distance measure, applied to a stack of images of Chernozem soil taken with the STXM at BESSY II. The distribution maps of the first five clusters are shown. The first cluster maps the sample-free region, the second cluster represents the averaged sample, while the third cluster maps the organic components seen in Fig. 9.
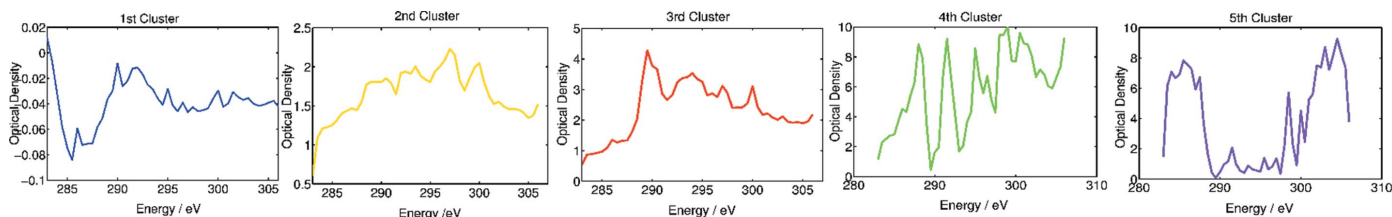

Figure 11
Results of angle distance measure, applied to a stack of images of Chernozem soil taken with the STXM at BESSY II. The spectra of the first five clusters are shown. The first cluster contains the sample-free region, the second cluster's spectrum shows the inorganic component, while in the third cluster the organic component is present; the fourth and fifth spectra contain only noise.

Figure 12
Results of angle distance measure, applied to a stack of images of Chernozem soil taken with the STXM at BESSY II. The distribution maps of the first five clusters are shown. Spatial distribution of the chemical elements identified in Fig. 11.

is also present, while in the spectra of the fourth and the fifth clusters only noise is present.

Some sample parts are perhaps too thick, so for the following analysis the area has been restricted to a smaller subregion. The results are displayed in Figs. 13 and 14. The first resulting cluster with its corresponding spectrum shows clearly the presence of potassium, a very good indication of clay particles present in the soil sample (see, for comparison, Fig. 3, top). The spectrum of the second cluster has four peaks, the two at higher energies are the absorption edges of potassium and the two at lower energies are due to the presence of some organic compounds in the soil. The corresponding image shows a mixture of the clay particle with the soil colloids. The third cluster just gives us an idea of the presence of some organic groups owing to the peak's spectrum around the carbon absorption edge, and from the map
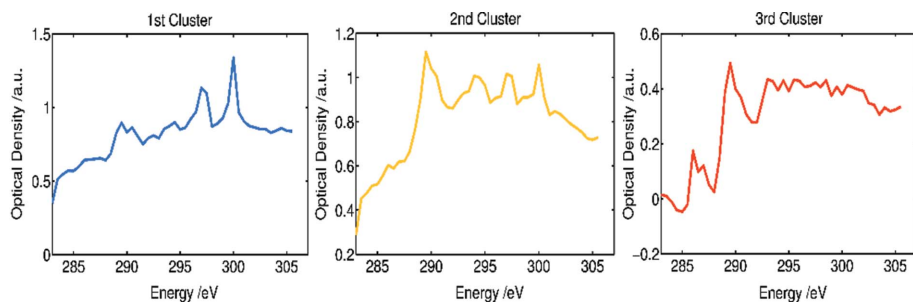
Figure 13
Results of angle distance measure, applied to a restricted region in a stack of images of Chernozem soil taken with the STXM at BESSY II. The spectra of the first three clusters are shown. The first cluster contains potassium, clearly visible in its spectrum (see Fig. 3, top); the second cluster shows the overlapping of organic and inorganic elements in the sample; and the third cluster contains organic components.
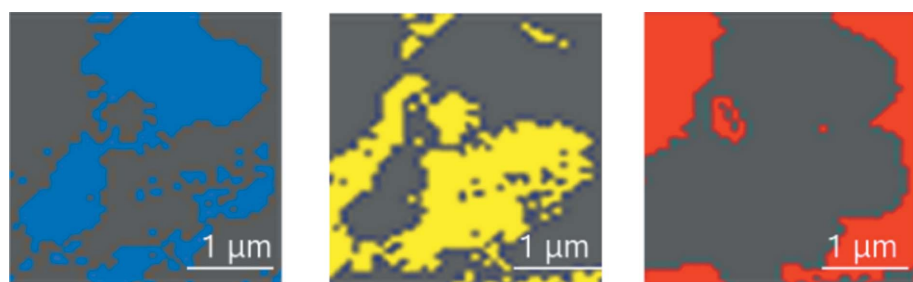

Figure 14
Results of angle distance measure, applied to a restricted region in a stack of images of Chernozem soil taken with the STXM at BESSY II. The distribution maps of the first three clusters are shown (see also Fig. 13).

distribution we could conclude that some residuals are deposited on the Si$_3$N$_4$ foil.

6. Discussion

The spatial resolution is limited by the zone plate used, and is in the range of 60 nm. The measured spectral resolution is ~1700. A Chernozem soil sample was studied, and results are presented in this paper. Analysis of a sample from environmental science, which is very heterogeneous physically as well as chemically, is a real challenge for X-ray spectromicroscopy. Three different analyzing methods were applied to the same data set.

We applied a Gaussian fitting and the results are presented in Fig. 6. The presence of potassium in our specimen is clearly evidenced by its L_{III} and L_{II} absorption edges. The spectrum was fitted using Gaussian line-shapes yielding the identification of organic functional groups in the sample.

The PCA applied to the data helped us to reduce and to noise-filter the spectromicroscopical data obtained from a stack of images at the carbon K-edge. The components obtained are sorted by their degree of significance into the sample. The PCA is a purely mathematical algorithm and does not deliver physical or chemical interpretable results. In Fig. 8, taking into account the last five principal components, the density distributions have been found to have negative values,

obviously with false physical content. Still, the PCA is essentially for the further applied cluster analysis algorithm in that it gives the most significant components and it reduces the data. The PCA is an intermediate result, but it is still unsatisfactory regarding the interpretation of data.

The cluster analysis algorithm was applied to the same spectromicroscopical data set. It has been shown that cluster analysis is able to provide representative spectra for an unknown data set, and map thicknesses corresponding to these spectra. The Euclidean measure method has been shown to find chemical grouping of data and calculate averaged thickness maps and spectra for these groups of data. However, it has been observed that variations in thicknesses in the sample are calculated as distinct clusters. Better results of cluster analysis were obtained using an angle distance measure rather than an Euclidean distance measure. This is in agreement with Lerotic (2005). Cluster analysis Euclidean distance measure made differentiation of organic and inorganic components, *i.e.* spectral information, possible (see Figs. 9 and 10), whereas, using angle distance measure, spatial and spectral identification of organic from inorganic was achieved (see Figs. 13 and 14). The number of clusters and the information contained in each of them, images and spectra, is the closed representation of chemical distinguished elements in our sample. The clay distribution is observed in the first cluster, while some organic residuals are deposited on the Si$_3$N$_4$ foil and clearly seen in the third cluster. The results for the Chernozem sample show that an angle distance measure is more appropriate for spectromicroscopy analysis of real samples with large thickness variations.

The spatial and spectral resolution of the STXM will be improved in the future by using a new and much better monochromator (Guttmann, 2006) and by using new zone plates with smaller outermost zone widths and higher efficiency, thus achieving a higher spatial resolution.

References

Ade, H., Kilcoyne, A. L. D., Tyliszczak, T., Hitchcock, P., Anderson, E., Harteneck, B., Rightor, E. G., Mitchell, G. E., Hitchcock, A. P. & Warwick, T. (2003). *J. Phys. IV*, **104**, 3–8.
Ahl, C., Frede, H. G., Gäth, S. & Meyer, B. (1985). *Mitt. Dtsch. Bodenkundlich. Ges.* **42**, 359–434.

Barrett, R., Kaulich, B., Salom, M. & Susini, J. (2000). *AIP Conf. Proc.* **507**, 458–463.

Chapman, H. C., Jacobsen, C. & Williams, S. (1996). *Ultramicroscopy*, **62**, 191–213.

Früke, R. (2001). Diploma thesis, University of Göttingen, Germany.

Gleber, G. (2003). Diploma thesis, University of Göttingen, Germany.

Guttmann, P. (2006). Private communication.

Hartmann, R., Lechner, P., Strüder, L., Scholze, F. & Ulm, G. (1995/1996). *Metrologia*, **32**, 491–494.

Henke, B. L., Gullikson, E. M. & Davis, J. C. (1993). *Atom. Data Nucl. Data Tables*, **54**, 181–342.

Jacobsen, C., Williams, S., Anderson, E., Browne, M. T., Buckley, C. J., Kern, D., Kirz, J., Rivers, M. & Zhang, X. (1991). *Opt. Commun.* **86**, 351–364.

Lerotic, M. (2005). PhD thesis, Stony Brook University, USA.

Lerotic, M., Jacobsen, C., Gillow, J. B., Francis, A. J., Wirick, S., Vogt, S. & Maser, J. (2005). *J. Electron Spectrosc. Relat. Phenom.* C**144–147**, 1137–1143.

Lerotic, M., Jacobsen, C., Schäfer, T. & Vogt, S. (2004). *Ultramicroscopy*, **100**, 35–57.

Morrison, G. R. (1994). *X-ray Microscopy*, Vol. IV, pp. 479–484. Moscow: Bogorodskii Pechatnik.

Morrison, G. R., Eaton, W. J., Barrett, R. & Charalambous, P. (2003). *J. Phys. IV*, **104**, 547–550.

Morrison, G. R. & Niemann, B. (1998). *X-ray Microscopy and Spectroscopy*, Vol. I, pp. 85–94. Berlin: Springer.

Osanna, A. (2000). PhD thesis, State University of New York at Stony Brook, USA.

Rehbein, S. (2003). *J. Phys. IV*, **104**, 207–210.

Smith, A. P., Coffey, T. & Ade, H. (1998). *X-ray Microscopy and Spectromicroscopy*, Vol. III, pp. 77–82. Berlin: Springer.

Stöhr, J. (1992). *NEXAFS Spectroscopy*, Vol. 25. *Springer Series in Surface Science.* Berlin: Springer.

Stöhr, J. & Anders, S. (2000). *IBM J. Res. Dev.* **44**, 535.

Strüder, L. *et al.* (1990). *Nucl. Instrum. Methods*, A**288**, 227–235.

Tanaka, T., Bando, K. K., Matsubayashi, N., Imamura, M., Shimada, H., Takahashi, K. & Katagiri, G. (2001). *Anal. Sci. Suppl.* **17**, i1077–i1079.

Thieme, J. & Niemeyer, J. (1998). *Prog. Colloid Polym. Sci.* **111**, 193–201.

Thieme, J., Niemeyer, J., Machulla, G. & Schulte-Ebbert, U. (1998). *X-ray Microscopy and Spectroscopy*, Vol. II, pp. 11–19. Berlin: Springer.

Thieme, J., Schneider, G. & Knöchel, C. (2003). *Micron*, **34**, 339–344.

Tyliszczak, T., Warwick, T., Kilcoyne, A. L. D., Fakra, S., Shuh, D. K., Yoon, T. H., Brown, G. E. Jr, Andrews, S., Chembrolu, V., Strachan, J. & Acremann, Y. (2003). *AIP Conf. Proc.* **705**, 1356–1359.

Wiesemann, U. (2003). PhD thesis, Georg-August University, Göttingen, Germany. [http://webdoc.sub.gwdg.de/diss/2004/wiesemann/index.html.]

Wiesemann, U., Thieme, J., Guttmann, P., Niemann, B., Rudolph, D. & Schmahl, G. (2000). *AIP Conf. Proc.* **507**, 430–434.