Improving the energy resolution of X-ray and electron energy-loss spectra

R.F. Egerton a,b,*, H. Qian a,b, M. Malac a,b

a Physics Department, University of Alberta, Edmonton, Alta., Canada T6G 2J1
b National Institute for Nanotechnology, Edmonton, Alta., Canada

Abstract

We discuss some practical problems of improving the resolution of X-ray and electron spectra. Iterative Bayesian methods promise greater resolution enhancement than Fourier techniques but they also give rise to spectral artifacts. Satellite peaks are generated adjacent to strong peaks in the original spectrum and oscillatory artifacts become prominent after a large number of iterations, particularly when the original data contain high noise content. In the case of valence-electron energy-loss spectra, satellite peaks are reduced by removing the zero-loss peak prior to spectral sharpening. Even so, care should be exercised in interpreting low intensity at low energy loss (after sharpening) as evidence for a bandgap in the electronic density of states.

q 2006 Elsevier Ltd. All rights reserved.

Keywords: EELS; Energy resolution; Maximum entropy; Maximum likelihood; Monochromator

1. Introduction

Resolution is nearly always of concern to spectroscopists. X-ray energy-dispersive spectroscopy (XEDS) in an electron microscope provides a convenient method of elemental analysis with high spatial resolution (down to 1 nm) but provides an energy resolution of the order of 100 eV, which can give rise to problems of peak overlap particularly at low photon energies. Wavelength dispersive and bolometric detectors bring the energy resolution closer to 1 eV but at the expense of reduced collection efficiency and increased complexity. X-ray absorption spectroscopy, using an electron synchrotron and monochromator as the radiation source, can achieve a resolution below 0.1 eV, allowing the study of absorption-edge fine structure. In this case, the spatial resolution is limited to about 30 nm by practical problems of zone-plate fabrication.

Electron energy-loss spectroscopy (EELS) in the transmission electron microscope (TEM) is capable of very high spatial resolution (below 1 nm) but its spectral resolution is limited by the energy width of the electron gun, of the order of 1 eV for a thermionic source and possibly as low as 0.23 eV for a field-emission source (Kimoto et al., 2005). Electron monochromators are now commercially available for the TEM and can reduce this energy spread to below 0.2 eV (Brink et al., 2003). Computer processing of energy-loss spectra provides a less expensive alternative and has been successfully employed to improve the energy resolution of core-loss (inner-shell) spectra (Overwijk and Reefman, 2003; Gloter et al., 2003; Kimoto et al., 2003). However, the energy resolution obtainable in the core-loss region is limited by the core-level width and by final-state broadening, typically to a few electron volt (Egerton, 2003). Therefore, improvements in energy resolution are more feasible in the case of valence-loss spectra (energy loss E < 50 eV) and are most urgently required in the case of bandgap studies at energy losses below 5 eV (Batson et al., 1986; Lazar et al., 2003).

In this short communication, we explore some of the practical problems encountered when applying software methods of peak sharpening. Although we concentrate on valence-loss spectra, the methods and conclusions should apply equally to other forms of spectroscopy.

2. Fourier deconvolution

Fourier-transform methods provide a simple method of controlling the resolution of spectral data. If Z(E) is the resolution function of the spectrometer system (assumed to have a linear response) and P(E) represents the physical data, as recorded by an ideal spectrometer, the measured spectrum J(E) is a convolution of these two functions and can be represented by:

* Corresponding author. Physics Department, University of Alberta, Edmonton, Canada T6G 2J1. Tel.: +1 780 492 5095; fax: +1 780 492 0714.
E-mail address: regerton@ualberta.ca (R.F. Egerton).
Taking the Fourier transform of both sides of Eq. (1) gives
\[ f(f) = p(f)z(f), \]
where \( f \) is a Fourier frequency (in units of eV\(^{-1}\)), so that
\[ p(f) = \frac{j(f)}{z(f)} \]
In practice, the energy scale is digitized into a limited number \( n \) of channels and each computed Fourier transform occupies an equal number of channels (\( i = 1 \) to \( n \)), so Eq. (2) becomes:
\[ p_i = \frac{j_i}{z_i}. \]
Therefore, the instrumental broadening should be removable entirely by dividing Fourier coefficients, followed by an inverse transform to yield \( P(E) \). However, spectra are usually oversampled (recorded with energy increments less than the width of the resolution function) so that the coefficients \( z_i \) become very small for values of \( i \) that are close to \( n \). In this case, the high-frequency components in \( J(E) \) consist predominantly of noise; because they are greatly amplified by Fourier division, the inverse transform yields \( P(E) \) submerged in high-frequency noise, such that it is often unrecognizable.

A partial solution is to attenuate the high-frequency Fourier coefficients, in which case the energy broadening of the spectrometer system is only partially corrected. The factor by which the spectral resolution can be improved then depends on the signal/noise ratio (SNR) of the recorded data and its degree of oversampling, as illustrated in Fig. 1. If the zero-loss peak contains four data points within its full width at half maximum (FWHM), the resolution improvement is limited to a factor of about 1.5 for practical values of SNR.

Usually, Fourier processing of electron energy-loss spectra includes the removal of plural scattering, which involves the introduction of a logarithm function into Eq. (2) (Johnson and Spence, 1974; Egerton, 1996). But for reasonably thin specimens, this complication does not significantly change the feasibility of resolution improvement.

### 3. Bayesian techniques

The main problem with the Fourier method of deconvolution is that it treats the spectral signal and its noise content on an equal basis. To distinguish between these two components, the measured spectrum should be represented as:
\[ J(E) = P(E)^*Z(E) + N(E) \]
At high energy loss \( E \), the noise component \( N(E) \) is largely electron-beam shot noise, whose amplitude is proportional to the square root of \( J(E) \), whereas in the low-loss region there may be a substantial contribution from detector noise, which is independent of \( J(E) \).

It is possible to envisage an iterative procedure, in which an estimate \([P(E)]\) of the unBroadened spectrum \( P(E) \) is obtained and gradually refined by successive approximation, for example, by minimizing the parameter:
\[ \chi^2 = \frac{1}{N} \sum_i (J_i - [P_i]Z_i) \]
In the absence of noise, \( \chi^2 \) could in principle approach zero, but in practice this will not happen because of the noise term \( N(E) \). One can also maximize the entropy, defined as
\[ S = \frac{1}{N} \sum_i [P_i] \log([P_i]/B_i) - 1 \]
where \( B_i \) is termed a Bayesian prior, representing the starting point of the iteration. Maximizing entropy is equivalent to minimizing the information content of the derived spectrum \([P(E)]\), thereby avoiding the violent oscillations which result when high-frequency components of the noise \( N(E) \) are treated as real data. Further discrimination against noise comes from insisting that all of the derived data points \([P_i]\) must be positive, if they are to represent physically real electron (or photon) intensities. Maximum-entropy deconvolution has been used to substantially improve the visibility of K-edge energy-loss fine structure (Overwijk & Reefman, 2003). Maximum likelihood (ML) deconvolution is based on maximizing the joint probability of observing the measured data set, which for uncorrelated noise is given by
\[ p(J_1...J_n) = \prod_i p(J_i) \]
where the product on the right-hand side includes terms from \( i = 1 \) to \( n \). Assuming Poisson noise distributed about a mean value \((PZ)_i\) that represents the noiseless data
\[ p(J_i) = [(P^*Z)_i]^i \exp\left(-[(P^*Z)_i]/(J_i)\right) \]
The maximum in Eq. (7) is obtained by applying the criterion: \( d(\ln p(J_i))/dp = 0 \), subject to the requirement that the solution is positive. This iterative procedure was first described by Richardson (1972) and shown to be equivalent to maximum likelihood by Lucy (1974), so it is commonly known as the Richardson–Lucy (RL) algorithm. It has been widely used for increasing the visibility of two-dimensional data, including astronomical images, and has been adapted to processing the output of a CCD camera attached to a parallel-recording energy-loss spectrometer (Gloter et al., 2003). The RL method

Fig. 1. Factor by which the resolution can be enhanced (while retaining SNR > 3) by Fourier-transform deconvolution, based on results from Egerton and Crozier (1988), and by maximum-entropy processing, based on Eq. (9). This factor is plotted as a function of the signal/noise ratio in the original spectrum.
has also been used to improve the energy resolution of XEDS data (Watanabe and Williams, 2003).

As a result of applying prior knowledge about the data-acquisition process (e.g. positive intensities) and explicitly recognizing the presence of noise, the effect of noise on the process of resolution improvement is reduced, as shown in Fig. 1. Here, the straight line represents the ‘superresolution coefficient’

$$SR = \frac{1}{3} \log_2\left[1 + (\text{SNR})^2\right]$$

derived from Bayesian statistical theory (Kosarev, 1989; Overwijk and Reefman, 2000). Based on Fig. 1, it appears that Bayesian methods offer substantially greater resolution enhancement compared to simple Fourier techniques.

4. Tests on valence-loss spectra

We have applied a commercial version of the Richardson–Lucy maximum-likelihood method (De Noyer, 2003) to electron energy-loss data. Fig. 2 shows a valence-loss spectrum recorded from a thin film of the organic semiconductor pentacene, before and after RL peak sharpening. As the number of iterations increases, the energy resolution improves, as can be seen from the width of the zero-loss peak, which in Fig. 2 has fallen from 1.3 to 0.8 eV after 15 iterations. Since the algorithm typically starts with a smoothed version of the original data, spectral noise is not apparent. However, the procedure does introduce artifacts in the form of satellite peaks that appear on either side of the zero-loss peak, see Fig. 2.

These so-called wing peaks have been discussed by Reefman (2000). They are visible in the RL-processed data of Watanabe and Williams (2003) and might be mistaken for the presence of minor elements. In the case of valence-loss spectra, however, the wing peaks occur mainly in the vicinity of the zero-loss peak, which has the smallest width and (usually) by far the highest intensity. If the zero-loss peak is removed prior to RL sharpening (e.g. by Fourier-log deconvolution, which also removes plural scattering), the intensity of the wing satellites is greatly reduced, as shown in Fig. 3.

With increasing number of iterations, the wing-peak intensity increases and a second kind of artifact appears, in the form of a rapid oscillation of intensity throughout the spectrum, see Fig. 4. This illustrates another undesirable feature of the algorithm, namely that it often does not converge towards a stable result. Instead, the operator must terminate the
program after an empirically determined number of iterations, based on the required degree of sharpening or on the tolerable amplitude of the oscillatory artifacts. In the case of noisy two-dimensional images, attenuation of the high-frequency components (prior to iterative resolution enhancement) has been found helpful in obtaining convergence (Kawata and Ichioka, 1980).

One practical question in both Fourier-log and Bayesian deconvolution is how to obtain the instrumental resolution function $Z(E)$. In the case of energy-loss spectroscopy, a zero-loss peak (ZLP) can usually be recorded separately, with the specimen shifted out of the electron beam. However, there is evidence that this peak is narrower than the ZLP measured in the presence of a specimen (Ishizuka, 2004, personal communication). The greater width of the specimen’s zero-loss peak could be an indication of vibrational-mode losses, which are included within the peak because of the limited energy resolution. However, it could also be an instrumental effect, arising from the fact that elastic and phonon scattering in the specimen increase the spectrometer-object size (in TEM image mode) or introduce spectrometer-aberration broadening of the spectrum (in TEM diffraction mode).

Whatever the cause, it appears preferable to obtain $Z(E)$ from the spectral data. One technique for doing this is to fit the data-ZLP to a simple function, such as a Gaussian or...
Lorentzian, provided the no-specimen ZLP appears symmetric (Kimoto et al., 2005). Alternatively, the no-specimen ZLP could be fitted to an appropriate function and then broadened so that its width (at half height) matches the ZLP of the spectral data. In the case of the pentacene data processed here, $Z(E)$ was taken as a Gaussian of width equal to that of the recorded spectrum.

5. Tests on simulated low-loss spectra

To test whether our observations might depend on exactly how the Richardson–Lucy algorithm is implemented, we compared the behaviour of two commercially available versions (De Noyer, 2003; Ishizuka, 2004) and one from Brookhaven National Laboratory (Wu, 2004). These computer routines were applied to the same data set, namely a simulation of a plasmon-loss spectrum (including plural scattering and noise) generated by the SPECGEN program (Egerton, 1996). The result of RL processing was found to be almost identical over most of the spectral range, except in the immediate vicinity of the zero loss peak where the satellite wings appeared slightly different; see Fig. 5. In each case, it was necessary to terminate the process after a certain number of iterations (in the range 10–50) to avoid the appearance of repetitive peak-like artifacts.

By varying the noise content of the simulated spectra, we found that the oscillatory artifacts appear sooner (after fewer iterations) as the noise increases. In fact, these artifacts appear to be a smoothed and (eventually) amplified version of the original noise; see Fig. 6a and b. With a broad plasmon peak and no spectral noise (Fig. 6c), we observed pronounced wing peaks (adjacent to the zero-loss component) but no oscillatory artifacts after several thousand iterations, see Fig. 6d.

We also tested the degree to which the sharpened spectrum depends on the choice of the Bayesian prior $B_i$. After 50 iterations, the result was almost the same, regardless of whether $B_i$ was taken as the raw data, or as a constant equal to the average value of the raw data.

6. Relevance to bandgap studies

It is apparent from Figs. 2b, 4 and 6d that RL deconvolution can give rise to a region of low or zero intensity (between the ZLP and the wing peak) that could be mistaken as representing the conduction-valence bandgap of an insulating or semi-conducting specimen. Therefore, great care must be taken in interpreting these deconvolved spectra. A spectrometer system fitted with a monochromator may represent a safer alternative for the measurement of bandgaps (or states within the gap) by electron energy-loss spectroscopy.

7. Conclusions

Bayesian deconvolution methods can be useful for improving the resolution of spectral data, by a factor that depends on the recorded signal/noise ratio. However, they are liable to introduce spectral artifacts and must therefore be used with caution. One type of artifact is the emergence of satellite peaks adjacent to any strong peak in the raw data. In the case of valence-loss spectra, these wing peaks are significantly reduced by removing the zero-loss peak prior to spectral sharpening. Associated with the wing peaks is a region of low or zero intensity adjacent to the zero-loss peak, which could be mistakenly interpreted as representing the bandgap of an insulator.

After a large number of iterations, spurious peaks appear throughout the spectrum, apparently generated by the noise content of the original spectrum. As a result, the iterative procedure must be monitored and terminated, based on visual inspection.

Acknowledgements

The authors wish to thank Lin de Noyer, Kazuo Ishizuka and Lijun Wu for use of their deconvolution programs, and the Natural Sciences and Engineering Research Council for financial support.

References


