Deconvolution of Rutherford backscattering spectra: An inverse problem

Heinz Ellmer and Dieter Semrad
Institut für Experimentalphysik, Johannes-Kepler Universität Linz, A-4040 Linz-Auhof, Austria

(Received 23 April 1996)

We review the problems associated with resolution correction and discuss some of the most promising procedures to solve them. We select four methods to deconvolute heavy particle backscattering spectra: (i) parametrization of the theoretical function, (ii) using histograms with variable bin width, (iii) modified Landweber iterations, and (iv) using mollifiers. To judge the quality of the methods, we treat simulated backscattering spectra as obtained in Rutherford backscattering measurements with solid-state detectors. Our results are as follows: When *a priori* information on the shape of the spectra is available, parametrization of the problem is superior to all other methods. When information on high-frequency components of the spectrum (e.g., on sharp edges) is of primary interest, the use of histograms with variable-bin width might provide good results. In all other cases, the choice of the best procedure depends on the specific problem and our ability to optimize the adjustable parameter of the specific method. [S1063-651X(96)11010-2]

PACS number(s): 02.50.−r

I. INTRODUCTION

A basic problem in counting experiments is this: when signals are processed, they are blurred by stochastic processes inherent in the corresponding transfer elements, i.e., the detection system. To get the undistorted information on the original distribution of signals, one has to apply some sort of resolution correction. This is a typical inverse problem [1], i.e., inverse in causality, if one is interested in the causes of an observed effect.

This review has been stimulated on the one hand by some recent papers in physics [2–7] and in mathematics journals [8,9], on the other hand by our attempts to obtain the undistorted shape of the yield enhancement found in Rutherford backscattering (RBS) spectra taken at exactly 180° [10]. We have tested a great number of methods that can be applied to deconvolute light ion backscattering spectra, and we will present the results of the four most promising procedures. To judge the quality of the deconvolution procedures, we have to know the undistorted spectra. Therefore, all ''measured'' spectra in this contribution have been obtained by simulation.

When a projectile with well defined energy E_0 enters a surface-barrier semiconductor detector (SBD) or a particle-implanted and passivated-silicon (PIPS) detector, it loses an energy δE_d by random processes in the entrance window. This part of the initial energy does not contribute to the detector signal. The remaining energy $E_0-\delta E_d$ will be partitioned in a stochastic way into electronic excitation or ionization of the detector atoms δE_e and into nonelectronic processes $\delta E_{\rm ne}$ (e.g., production of phonons). The energy δE_e is then available for the creation of electron-hole pairs. Finally, the charge of these pairs is converted into pulse height by standard electronics. Due to stochastic processes involved in this energy-to-pulse-height conversion, we measure a density distribution in pulse height.

Using many different energies E_0 and identifying the centers of gravity of the corresponding distributions with the primary energies E_0, we obtain the energy calibration of the detector system. Without loss of generality, we assume this calibration to be linear, which is not strictly true, due to the energy dependence of δE_d. With this calibration in mind, we may consider the distribution in pulse height (for a particular E_0) a distribution in energy E. We call it the resolution function of the detector system, $k(E-E_0)$. For all methods discussed here, $k(E-E_0)$ must be known; it can be obtained, e.g., from the spectra of projectiles backscattered from a very thin layer. This function turns out to be approximately Gaussian, but asymmetric.

Due to energy-loss straggling on the way into and out of the target, projectiles scattered from thin layers at a larger depth will show a broader distribution. It might be useful to include this energy-loss straggling in the resolution function, thus making the function depend explicitly on energy E_0, i.e., leading to $k(E-E_0;E_0)$. In the following, we will restrict our considerations to measurements of near-surface layers where the shape of k can be assumed to be independent of E_0. Nevertheless, all procedures discussed will work equally well with an energy-dependent shape of the resolution function, provided this dependence is known.

II. FORMULATION OF THE PROBLEM

When the detector is exposed to a *true* spectrum of energies E_0 described by a density distribution $f(E_0)$, the *ideal measured* spectrum $h(E)$ is given by

$$h(E)=\int_{-\infty}^{\infty}k(E-E_0)f(E_0)dE_0=(k*f\ )(E), \qquad (1)$$

where $k*f$ means a convolution. Although it is not necessary, we have here assumed that k and f are independent. The integral kernel k is the resolution function discussed above. Normally, the spectra consist of a number of values at discrete energies E^i, obtained by means of a multichannel analyzer (MCA), where i is the channel number. So we can replace the convolution integral by a sum. To take counting statistics in the individual channels into account, we add some noise $r_\delta(E^i)$ to the convoluted spectrum. We know that $r_\delta(E^i)$ is governed by Poisson statistics dependent on the number of counts per channel and that it may be characterized by a standard deviation δ. Thus we obtain the *actually measured* spectrum $h_\delta(E^i)$:

$$h_\delta(E^i)=\sum_{j=0}^{j_{\max}} k(E^i-E_0^j)f(E_0^j)+r_\delta(E^i)$$
$$=(k*f\ )(E^i)+r_\delta(E^i). \tag{2}$$

As in Eq. (1), the index 0 refers to the true energy. In this discrete formulation, the convolution reduces to the multiplication of the matrix $\underline{\mathbf{k}}$ with elements $k_{ij}=k(E^i-E_0^j)$, by the vector $\mathbf{f}$ with components $f_j=f(E_0^j)$ $(i,j=1,2,3,\ldots,N)$, where N is the number of data points, e.g., the number of channels of the MCA. The noise $\mathbf{r}_\delta$ and the measured spectrum $\mathbf{h}_\delta$ are also vectors with the same dimension as $\mathbf{f}$. So we can simply write

$$\mathbf{h}_\delta=\underline{\mathbf{k}}\mathbf{f}+\mathbf{r}_\delta. \tag{3}$$

Convolution in energy domain transforms into a product in Fourier domain; therefore, the Fourier transform of Eq. (2) reads

$$\mathcal{F}(\mathbf{h}_\delta)=\mathcal{F}(\mathbf{k})\mathcal{F}(\mathbf{f})+\mathcal{F}(\mathbf{r}_\delta), \tag{4}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ are the operators of Fourier transform and of inverse Fourier transform, respectively, given by

$$\mathcal{F}(q)(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}q(y)e^{ixy}dy, \tag{5}$$

$$q(x)=\mathcal{F}^{-1}[\mathcal{F}(q)](x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathcal{F}(q)(y)e^{-ixy}dy. \tag{6}$$

Formally, one gets the true spectrum $f(E_0)$ from Eq. (4) by

$$\mathbf{f}=\mathcal{F}^{-1}\left(\frac{\mathcal{F}(\mathbf{h}_\delta)}{\mathcal{F}(\mathbf{k})}\right)-\mathcal{F}^{-1}\left(\frac{\mathcal{F}(\mathbf{r}_\delta)}{\mathcal{F}(\mathbf{k})}\right). \tag{7}$$

We focus on the first term of Eq. (7), which is available from experiment. The asymmetric detector resolution function $k(E-E_0)$ may be well described by a sum of two Gaussian distributions shifted with respect to each other [11–13]. Its Fourier transform is again a sum of two Gaussians, which decreases at high frequencies ω proportional to $\exp(-\omega^2\sigma^2/2)$, where σ is the standard deviation of the smaller of the two Gaussian peaks. In contrast, for the transform of the noisy spectrum h_δ at high frequencies we have to assume a fairly uniform distribution at least up to values corresponding to the channel width of the MCA. Hence the argument of the first term in Eq. (7) might become arbitrarily large for high Fourier components:

$$\frac{\mathcal{F}(h_\delta)}{\mathcal{F}(k)}\approx e^{\omega\sigma^2}\xrightarrow{\omega\to\infty}\infty. \tag{8}$$

A forced cutoff or damping (filtering) at high frequencies ω will result in loss of information about $\mathbf{f}$ and will destroy structures formed by high Fourier components, e.g., edges. We want to emphasize that Eq. (8) suggests using a measured and hence noisy resolution function $\mathbf{k}$ with high Fourier components rather than an analytical form with less high frequencies.

These considerations show that the inverse of the problem formulated by Eq. (2) is not well posed in Hadamard's sense [14,15]: (i) the solution does not exist in the strict sense, (ii) the solutions might not be unique, and (iii) solutions might not depend continuously on the data, i.e., small changes in the data might cause arbitrarily large changes in the result if the ill-posedness of the problem is not carefully taken into account. Unfortunately, even if the problem is well posed it still can be ill-conditioned, which will result in numerical instabilities. This will be discussed in more detail later on. We want to point out that our problem is much more unstable than, e.g., the inversion of radon transform used in computerized tomography [16]. This is due to the very smooth kernel in the integral equation of the first kind, Eq. (1). As shown by Eq. (8), it is that feature of the kernel which makes even small errors with high Fourier components give rise to large oscillations in the solution of the inverse problem [15].

Strictly speaking, Eq. (3) is still incomplete: we should have taken into account so-called ''ghosts.'' Ghosts are functions g, which do not vanish identically but which fulfill $\underline{\mathbf{k}}\mathbf{g}=\mathbf{0}$. So, Eq. (3) can also be written as

$$\mathbf{h}_\delta=\underline{\mathbf{k}}\mathbf{f}+\mathbf{r}_\delta=\underline{\mathbf{k}}(\mathbf{f}+\mathbf{g})+\mathbf{r}_\delta. \tag{3a}$$

These functions g are invisible to the inversion (see, e.g., [16]). As a matter of principle, they can not be reconstructed from the data. One can only find the best approximate solution to the problem by means of ''normal equations''; see Eq. (13) below.
To further investigate the error resulting from deconvolution, we use the discrete presentation, Eq. (3). One formally gets $\mathbf{f}$ by multiplying Eq. (3) by $\underline{\mathbf{k}}^{\mathbf{-1}}$ from the left-hand side:

$$\mathbf{f}=\underline{\mathbf{k}}^{-1}(\mathbf{h}_\delta-\mathbf{r}_\delta). \tag{9}$$

Unfortunately, we cannot subtract the (unknown) noise $\mathbf{r}_\delta$. The obvious consequence would be to consider the noise of the measured spectrum to be due to a scatter of the original spectrum $\mathbf{f}_\delta$ (which we will indicate by the index δ) and to solve

$$\mathbf{h}_\delta=\underline{\mathbf{k}}\mathbf{f}_\delta \tag{10}$$

by inversion of the matrix $\underline{\mathbf{k}}$:

$$\overline{\mathbf{f}}_\delta=\underline{\mathbf{k}}^{-1}\mathbf{h}_\delta. \tag{11}$$

Due to errors in matrix inversion, both numerical and systematic, we get a deconvoluted spectrum $\mathbf{f}_\delta$ different from the true spectrum $\mathbf{f}$. Assuming that $\|\Delta\underline{\mathbf{k}}\|\leqslant\|\underline{\mathbf{k}}^{-1}\|^{-1}$, where $\underline{\mathbf{k}}\|\Delta\|$ is the norm of $\Delta\mathbf{k}$ [see Eq. (13)], the total error $\Delta\mathbf{f}=\mathbf{f}-\mathbf{f}_\delta$ can be estimated using Eq. (12) [17]:

$$\frac{\|\Delta\mathbf{f}\|}{\|\mathbf{f}\|}\leqslant\frac{A_{\mathrm{cond}}(\mathbf{k})}{1-A_{\mathrm{cond}}(\mathbf{k})\ \dfrac{\|\Delta\underline{\mathbf{k}}\|}{\|\underline{\mathbf{k}}\|}}\left(\frac{\|\Delta\underline{\mathbf{k}}\|}{\|\underline{\mathbf{k}}|}+\frac{\|\mathbf{h}-\mathbf{h}_\delta\|}{\|\mathbf{h}\|}\right). \tag{12}$$

We see that $\Delta\mathbf{f}$ depends on the relative data error $(\|\mathbf{h}-\mathbf{h}_\delta\|/\|\mathbf{h}\|)$, as expected. But it also depends on the quality of the numeric algorithm used to invert $\underline{\mathbf{k}}$, specified by the condition of $\underline{\mathbf{k}}$, $A_{\mathrm{cond}}(\underline{\mathbf{k}})$, and by the relative error $\|\Delta\underline{\mathbf{k}}\|/\|\underline{\mathbf{k}}\|$. $\|\Delta\underline{\mathbf{k}}\|$ includes all errors due to the experimental determination of $\underline{\mathbf{k}}$, deviation of the calculated inverse from the true one, etc. In Eq. (13) we give one practicable set of definitions of the norm of $\|\mathbf{f}\|$ and of $\|\underline{\mathbf{k}}\|$:

$$\|\mathbf{f}\| = \sqrt{\sum_{i=1}^{n} f_i^2}, \quad \|\underline{\mathbf{k}}\| = \max_i \sum_j |k_{ij}|,$$

$$A_{\mathrm{cond}}(\underline{\mathbf{k}}) = \|\underline{\mathbf{k}}\| \|\overline{\underline{\mathbf{k}}}^{-1}\| \quad (\text{always} \geqslant 1). \tag{13}$$

The matrix $\overline{\underline{\mathbf{k}}}^{-1}$ is the approximate inverse of $\underline{\mathbf{k}}$ obtained by some inversion method. To stress how inaccurate a matrix inversion may be, we mention that for a Cholesky algorithm [17], i.e., straightforward triangularization of the matrix, the condition of $\underline{\mathbf{k}}$ can be as large as 2×10^5. Fortunately, there are other algorithms that provide a more stable inversion, e.g., the Householder algorithm [18], with $A_{\mathrm{cond}}(\underline{\mathbf{k}})=1$. But even so, $\|\Delta\underline{\mathbf{k}}\|/\|\underline{\mathbf{k}}\|$ could still be much larger than the relative error of the data.

We arrive at a slightly more stable problem if instead we look for a vector $\overline{\mathbf{f}}_\delta$ which—under certain constraints—minimizes the norm of the differences:

$$\|\mathbf{h}_\delta - \underline{\mathbf{k}}\mathbf{f}_\delta\| \rightarrow \min. \tag{14}$$

This will lead to the so-called normal equation:

$$\underline{\mathbf{k}}^H \mathbf{k} \mathbf{f}_\delta = \underline{\mathbf{k}}^H h_\delta. \tag{15}$$

The superscript H marks the adjoint (or Hermitian conjugate). The best approximate solution of Eq. (14) is then given by

$$\overline{\mathbf{f}}_\delta = (\underline{\mathbf{k}}^H \underline{\mathbf{k}})^{-1} \underline{\mathbf{k}}^H \mathbf{h}_\delta. \tag{16}$$

Notice the difference between Eq. (11) and Eq. (16). The quantity $(\underline{\mathbf{k}}^H\underline{\mathbf{k}})^{-1}\underline{\mathbf{k}}^H$ represents the so-called pseudoinverse (Moore-Penrose inverse) [17] of $\underline{\mathbf{k}}$, which always exists and which is unique. But this does not yet take into account the ill-posedness of the problem explicitly.

So we still need a method that can treat noisy data. If the theoretical form of $\mathbf{f}$ is known as a function of a small number of parameters, we can fit a parameterized function to $\mathbf{h}_\delta$ using any optimization technique, such as line (descent-direction)-searching algorithms guided, e.g., by a steepest-descent criterion [19,20] (Sec. III A), least-squares fit, Newton techniques [19,20], Newton techniques combined with Tikhonov regularization [7,21], or a maximum-entropy method [6,22]. If one is interested in characteristics of the spectrum characterized by high-frequency Fourier components (e.g., sharp edges), the representation of the theoretical spectrum by a histogram with adjustable bin width might give good results (Sec. III B).

If no information about the true spectrum is available, we have to use regularization (stabilization) techniques. Regularization, e.g., Tikhonov regularization, is the approximation of an ill-posed problem by a family of neighboring well-posed problems. This family is characterized by a stabilization parameter that has to be chosen judiciously. Alternative procedures are Landweber iterations [23,24]—another regularization method (Sec. III C)—or using mollifiers to construct an approximate inverse of $\underline{\mathbf{k}}$ [8,9] (Sec. III D).

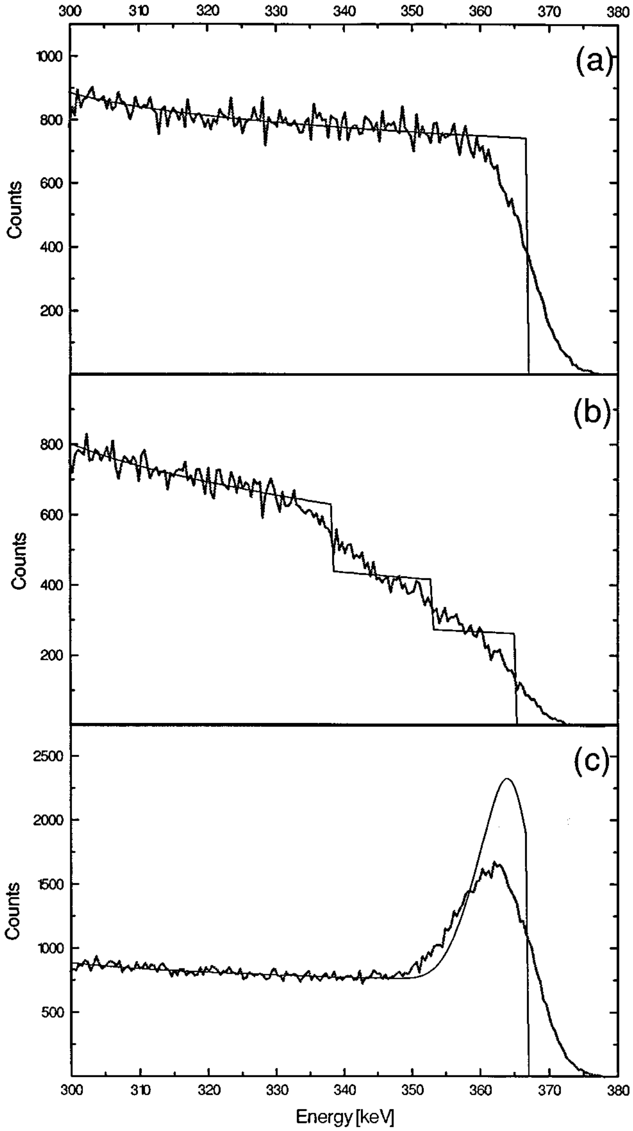

FIG. 1. Theoretical spectrum of 400 keV helium projectiles backscattered (a) from a one-component target, (b) from a three-component target, and (c) from a one-component target showing the 180° yield enhancement (see text) (thin solid line). Also shown are the simulated spectra that result from measurements with a detector of 9.1 keV energy resolution (thick solid line).

III. DECONVOLUTION TECHNIQUES

A. Parametrization of the spectrum

To obtain measured spectra, we take theoretical spectra, convolute them with the resolution function and add stochastic variables to represent counting statistics (see Fig. 1). The theoretical spectra are given as a function of energy, and they depend on a small number of parameters.

In the deconvolution process, these parameters are determined from the measured spectra by minimizing the norm given by Eq. (14). When we apply the parametrization method, we know the exact functional dependence of our theoretical spectra on the parameters, and this is indeed a crucial requirement. Hence our quantitative interpretations are correct provided we have found a function that describes the spectrum properly.

To test the method, we consider three cases shown in Figs. 1(a), 1(b), and 1(c). Case 1 represents an RBS spectrum that might have been obtained by backscattering 400 keV He^+ ions from a thick one-component target. The corresponding trial function f_1 depends on four parameters: the position of the high-energy edge a_1 and three parameters a_2, a_3, and a_4 giving the shape of the spectrum; in all these considerations we neglect a possible high-energy background due to, e.g., pulse pileup.

$$f_1(E_0)=\begin{cases} a_2-a_3E_0+\dfrac{a_4}{E_0}, & E_0\leqslant a_1 \\ 0 \quad E_0>a_1 \end{cases}. \tag{17}$$

Case 2 corresponds to a three-component target [Fig. 1(b)]. We assume the partial spectra to have the same functional dependence, but with different parameters. So the theoretical spectrum $f_2=f_{2,1}+f_{2,2}+f_{2,3}$ depends on 12 parameters:

$$f_{2,1}(E_0)=a_2-a_3E_0+\frac{a_4}{E_0}, \quad E_0\leqslant a_1,$$

$$f_{2,2}(E_0)=a_6-a_7E_0+\frac{a_8}{E_0}, \quad E_0\leqslant a_5, \tag{18}$$

$$f_{2,3}(E_0)=a_{10}-a_{11}E_0+\frac{a_{12}}{E_0}, \quad E_0\leqslant a_9.$$

Case 3 represents backscattering at 180° showing a near-surface yield enhancement [10,25]. As a trial function, we use here a standard backscattering spectrum plus two Gaussians, all cut off at the high-energy edge a_1 of the spectrum [Fig. 1(c)]. This function [Eq. (19)] has been found to reproduce measured spectra fairly well. It depends on eight parameters:

$$f_3(E_0)=a_2-a_3E_0+\frac{a_4}{E_0}+\frac{a_5}{\sqrt{2\pi}a_7}\exp\left(-\frac{(E_0-a_6)^2}{2a_7^2}\right)$$
$$+\frac{a_g}{\sqrt{2\pi}1.3a_7}\exp\left(-\frac{[E_0-(a_6-a_7)]^2}{2(1.3a_7)^2}\right),$$
$$E_0\leqslant a_1. \tag{19}$$

We have found [11] that asymmetric peaks such as that typical of 180° enhancement may be described by two Gaussians with one of them smaller in height, broader by a factor of 1.3, and shifted by one standard deviation towards lower energies. We use a similar function to describe the asymmetric detector resolution (in all three cases):

$$k(E-E_0)=\frac{1}{1+q}\left[\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(E-\overline{E_0})^2}{2\sigma^2}\right)\right.$$
$$\left.+\frac{q}{\sqrt{2\pi}1.3\sigma}\exp\left(-\frac{[E-(\overline{E_0}-\sigma)]^2}{2(1.3\sigma)^2}\right)\right]. \tag{20}$$

Here, we have chosen $q=0.3$ for the ratio of areas and $\sigma=3.57$ keV for the standard deviation of the principal Gaussian. This gives a full width at half maximum of 9.1 keV, in agreement with experimental data of our cooled 300 mm^2 PIPS detector for 400 keV He projectiles. By introducing the energy E_0, we take into account the difference between the mean of the principal Gaussian peak (E_0) and the mean of the total resolution function (E_0) to avoid having the spectrum shift through convolution:

$$\overline{E_0}=E_0+\frac{\sigma q}{1+q}. \tag{21}$$

To get the deconvoluted spectrum $\overline{\mathbf{f}_\delta}$, we optimize the parameters a_i with a line-searching algorithm with steepest-descent criterion to minimize the norm of differences [Eq. (14)].

B. Histogram with adjustable binning

We again start with Eq. (14). The spectrum $h_\delta(E^i)$ ($i=1,\ldots,N$) is displayed on a MCA with channels of constant width Δ, N being typically 1024 or 2048. To prevent loss of information, Δ should be small compared to the width of the detector resolution function [Eq. (20)]. A very simple, but numerically expensive way would now be the variation of the content of each channel $f_\delta(E_0^j)$ until $\|\mathbf{h}_\delta-\mathbf{k}\mathbf{f}_\delta\|$ reaches a minimum. However, in backscattering spectra there might be regions where the spectrum height $f(E_0^i)$ stays constant within the spread given by counting statistics; there we might choose the width Δ^m of the mth bin to be a multiple of Δ, thus reducing the number of bins from N to M. By lumping together the content of adjacent channels, some information may be lost. But this may be outweighed by the reduction of the relative uncertainty of the bin content and by the reduction of the number M of contents $\mathbf{F}_\delta(E_0^m)$, $m=1,\ldots,M$; in most cases M can be chosen to be more than one order of magnitude smaller than N. As the bin width of the measured and of the theoretical spectrum do not coincide, we have to calculate $f_\delta(E_0^j)$ in Eq. (14) by means of characteristic functions $\Theta(E_0-E_0^m,\Delta^m)$ representing the mth bin. They are defined as

$$\Theta(E_0-E_0^m,\Delta^m)=\begin{cases} 1, & E_0^m-\dfrac{\Delta^m}{2}\leqslant E_0\leqslant E_0^m+\dfrac{\Delta^m}{2} \\ 0, & \text{elsewhere.} \end{cases} \tag{22}$$

The spectrum $\overline{\mathbf{f}_\delta}(E_0)$ now follows as

$$\overline{f_\delta}(E_0)=\sum_{m=0}^{M-1}\overline{F_\delta}(E_0^m)\Theta(E_0-E_0^m,\Delta^m), \tag{23}$$

so that we finally get the discrete spectrum

$$\overline{f_\delta}(E_0^j)=\sum_{m=0}^{M-1}\overline{F_\delta}(E_0^m)\Theta(E_0^j-E_0^m,\Delta^m). \tag{24}$$

By iteration, we optimize the following parameters by minimizing the norm of the difference $\|\mathbf{h}_\delta - \underline{\mathbf{k}}\overline{\mathbf{f}}_\delta\|$: the number of bins M, their widths Δ^m, and their contents $F_\delta(E_0^m)$. We use the same kind of optimization algorithm as in Sec. III A to find a set of best parameters. We start with a small number of bins, e.g., four; then we iteratively determine those two adjacent bins where a decrease in the width of one at the expense of the other will make the norm $\|\mathbf{h}_\delta - \underline{\mathbf{k}}\overline{\mathbf{f}}_\delta\|$ decrease most, with Δ as the lower bound to Δ^m. When this procedure has made the norm converge to a relative minimum, we double the number of bins and repeat the procedure. This will, in general, give a smaller relative minimum for the norm. We terminate the iteration when this reduction of the norm becomes insignificant. In general, 16–32 bins lead to a sufficiently accurate result.

C. Tikhonov regularization, Landweber iterations

We want to describe two regularization methods in a very simplified way and to apply one of them to our problem. For detailed information, see, e.g., Refs. [1, 14, 21, 23, 24]. We try to damp the influence of the noise $\mathbf{r}_\delta$ given by Eq. (2) by replacing the ill-posed problem [Eq. (14)] with a neighboring well-posed problem depending on an adjustable parameter α:

$$\|\mathbf{h}_\delta - \underline{\mathbf{k}}\mathbf{f}_\delta\| + \alpha\|\mathbf{f}_\delta\| \rightarrow \text{min}. \tag{25}$$

In this way, the solution will be stabilized by some *a priori* information on the result. In our case this information is that the norm of $\mathbf{f}_\delta$ should be as small as possible; this means that we are looking for a solution as smooth as possible. We could also have looked for a solution that minimizes curvature or, more exactly, that minimizes the norm of the second derivative ($\|\mathbf{h}_\delta - \underline{\mathbf{k}}\mathbf{f}_\delta\| + \alpha\|\mathbf{f}''_\delta\| \rightarrow \text{min}$). It can be shown [21], that Eq. (25) has a unique minimizer $\mathbf{f}_\delta^\alpha$:

$$\overline{\mathbf{f}}_\delta^\alpha = (\alpha\underline{\mathbf{U}} + \underline{\mathbf{k}}^H\underline{\mathbf{k}})^{-1}\underline{\mathbf{k}}^H\mathbf{h}_\delta. \tag{26}$$

Here, $\underline{\mathbf{U}}$ is the unit matrix. Note that for $\alpha=0$ we get Eq. (16), which characterizes the best approximate solution of our original problem. Hence, Eq. (26) is a stabilized version of Eq. (16). The regularization parameter α has to be chosen according to the scatter of the data, quantified by δ. If one simply takes a power of δ for α, i.e., $\alpha \propto \delta^n$, one finds [14] that the best choice would be $\alpha(\delta) \propto \delta^{2/3}$. However, the choice of parameter α becomes very difficult if one wants to obtain the optimum convergence rate, α must then be determined from the following nonlinear equation (Morozov's principle of discrepancy; for further discussion the reader is referred to [26–28]):

$$\|\mathbf{k}\mathbf{f}_\delta^{\alpha(\delta)} - \mathbf{h}_\delta\| = \delta. \tag{27}$$

This is the so-called Tikhonov regularization, the most prominent regularization method for ill-posed problems. It has been successfully used, e.g., for the deconvolution of spectra in fluorescence spectroscopy [7]. But it requires an exceedingly large numerical effort to determine the appropriate regularization parameter α and to invert the matrix in Eq. (26).

Instead of solving Eq. (26), it appears better to solve Eq. (15) in a stable manner by Landweber iterations. Again, we present the result without proof; for further details see Refs. [23, 24]. Following the Banach fixpoint theorem [29], one has to determine $\mathbf{f}_\delta$ from

$$\overline{\mathbf{f}}_\delta = (\underline{\mathbf{U}} - \beta\underline{\mathbf{k}}^H\underline{\mathbf{k}})\overline{\mathbf{f}}_\delta + \beta\underline{\mathbf{k}}^H\mathbf{h}_\delta \tag{28}$$

by successive approximations (note that $\underline{\mathbf{U}}\overline{\mathbf{f}}_\delta = \overline{\mathbf{f}}_\delta$):

$$\begin{aligned} \mathbf{f}_\delta^0 &= \beta\underline{\mathbf{k}}^H\mathbf{h}_\delta, \\ \mathbf{f}_\delta^n &= \mathbf{f}_\delta^{n-1} + \beta(\underline{\mathbf{k}}^H\mathbf{h}_\delta - \underline{\mathbf{k}}^H\underline{\mathbf{k}}\mathbf{f}_\delta^{n-1}). \end{aligned} \tag{29}$$

The limits for the parameter β to make this procedure converge are $0 < \beta \leqslant 2/\|\underline{\mathbf{k}}\|^2$. In our case $\underline{\mathbf{k}}$ is given by the resolution function, which is normalized to 1, so that $\|\underline{\mathbf{k}}\|=1$ [see Eq. (13)]. Convergence is achieved when Eq. (14) is approximately fulfilled, i.e., when $\|\mathbf{h}_\delta - \underline{\mathbf{k}}\mathbf{f}_\delta^n\| \leqslant \varepsilon$, with ε being sufficiently small. Instead of minimizing the difference, we make the ratio converge to 1. So we try to modify Landweber iterations described above in the following way:

$$\begin{aligned} \overline{\mathbf{f}}_\delta^0(E_0^i) &= (\underline{\mathbf{k}}^H\mathbf{h}_\delta)(E_0^i), \\ \overline{\mathbf{f}}_\delta^n(E_0^i) &= \overline{\mathbf{f}}_\delta^{n-1}(E_0^i)\left(\frac{(\underline{\mathbf{k}}^H\mathbf{h}_\delta)(E_0^i)}{(\underline{\mathbf{k}}^H\underline{\mathbf{k}}\mathbf{f}_\delta^{n-1})(E_0^i)}\right)^\beta. \end{aligned} \tag{30}$$

This modification gives a steeper edge of the spectrum and avoids undershooting at the bottom of this edge. Contrary to Eq. (29), Eq. (30) does not tolerate local negative values of $\underline{\mathbf{k}}^H\underline{\mathbf{k}}\mathrm{f}_\delta^{n-1}$. The convergence rate and the error in the result depend strongly on the appropriate choice of β and on the breakoff criterion for the iteration. We want to emphasize that the solutions of Eqs. (26), (29), and (30) are—in a mathematical sense—as smooth as possible. This means that they have to be superpositions of sinusoidal functions leading to the well known oscillations in the result (see Sec. IV). In the present case, their amplitudes depend on both the regularization parameter and the breakoff criterion. Hence, both have to be dealt with carefully. We found that too many iterations may give a result as poor as too few. A breakoff criterion might be to keep the oscillations smaller than the noise δ of the data.

An alternative exists if one has some *a priori* information on the theoretical spectrum, e.g., that far from the edge, the shape of the theoretical spectrum is hardly changed by convolution. So one can this take as the breakoff criterion, if the norm of the difference of the calculated spectrum $\mathbf{f}_\delta^n$ and the measured spectrum $\mathbf{h}_\delta$,$\|\mathbf{h}_\delta - \mathbf{f}_\delta^h\|$,is sufficiently small. This breakoff criterion for a selected low-energy interval, together with $\beta=0.5$, has been applied in our calculations. We found that this criterion seems to be superior to others commonly used.

D. Mollifier method

We want to compute the inverse of the kernel in Eq. (2) by means of mollifiers [8,9] $u_c(E_0^i - E_0^j)$, which may depend on a parameter or on a set of parameters c. Mollifiers are approximate solutions of the following equation:

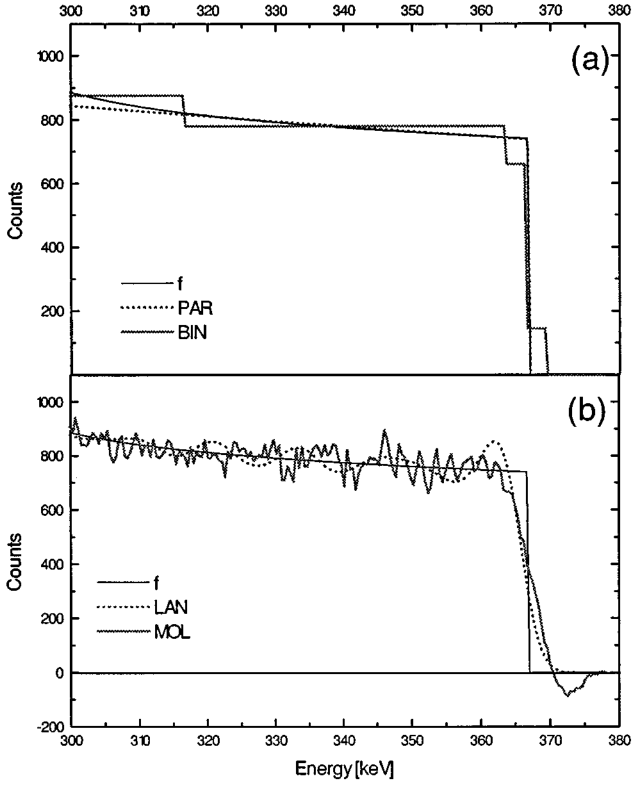

FIG. 2. The theoretical backscattering spectrum from a one-component target (thin solid line, symbol f) of Fig. 1(a) is shown together with the deconvoluted spectra (a) using parametrization (broken line, PAR) or using histograms with adjustable binning (thick solid line, BIN) and (b) using Landweber iterations (broken line, LAN) or using mollifiers (thick solid line, MOL).

$$f(E_0^i)=\sum_j u_c(E_0^i-E_0^j)f(E_0^j). \tag{31}$$

Obviously, $u_c(E_0^i-E_0^j)=\delta(E_0^i-E_0^j)$ would be the exact solution. Here, the δ function is defined by $\delta=1$ for $i=j$ and $\delta=0$ for $i\neq j$. The goal of this procedure is to find a so-called reconstruction kernel $v_c(E_0^i-E^j)$ so that an approximate solution of our problem can be determined from

$$\overline{f_\delta}(E_0^i)=\sum_j v_c(E_0^i-E^j)h_\delta(E^j). \tag{32}$$

We can here apply Eq. (10), since these general considerations are not restricted to noisy data. So we replace $\mathbf{h}_\delta$ by $\underline{\mathbf{k}}\overline{\mathbf{f}_\delta}$. Using the definition of the adjoint matrix $\mathbf{k}^H$, we can write

$$\begin{aligned}\sum_j v_c(E_0^i-E^j)h_\delta(E^j)\\ &=\sum_j v_c(E_0^i-E^j)\sum_n k(E^j-E_0^n)\overline{f_\delta}(E_0^n)\\ &=\sum_n\left[\sum_j v_c(E_0^i-E^j)k^H(E^j-E_0^n)\right]\overline{f_\delta}(E_0^n).\end{aligned} \tag{33}$$

It follows from Eqs. (31)–(33) that the matrix $\underline{\mathbf{v}}_c$ $[v_{c(ij)}=v_c(E_0^i-E^j)]$ has to be an approximate solution of

$$\sum_j v_c(E_0^i-E^j)k^H(E^j-E_0^n)=u_c(E_0^i-E_0^n). \tag{34}$$

The mollifier $\mathbf{u}_c$ allows us to construct an approximate inverse $\underline{\mathbf{v}}_c$ of $\underline{\mathbf{k}}$ that maps the data $\mathbf{h}_\delta$ to a regularized solution $\overline{\mathbf{f}_\delta}$ [Eq. (32)]. It should fulfill two criteria: (i) $\mathbf{f}$ has to be approximated as well as possible, and (ii) the influence of the noise $\mathbf{r}_\delta$ has to be reduced. These two criteria contradict each other, so one has to find a compromise between reproducing $\mathbf{f}$ and smoothing the data.

Of course, the exact reconstruction kernel would be the deconvoluted δ function, i.e., a function that yields the δ function when convoluted with the detector resolution function; unfortunately, such a function does not exist. Instead of the δ function, one might use a narrow Lorentz distribution as a mollifier:

$$u_c(E_0^i-E_0^j)\propto\frac{1}{1+[(E_0^i-E_0^j)/c]^2}. \tag{35}$$

As usual in this field, the parameter c gives the width of the mollifier. From Fourier and inverse Fourier transform, we get from Eq. (34) the kernel by

$$v_c=\mathcal{F}^{-1}\left(\frac{\mathcal{F}(u_c)}{\mathcal{F}(k)}\right). \tag{36}$$

If we look at the real part of the reconstruction kernel [Eq. (36)] we learn that a feasible ansatz might be a product of a cosine and a Gaussian function. To obtain a more appropriate result, we replace the parameter c with two adjustable parameters a and b:

$$v_{a,b}(E_0^i-E^j)=\cos\left(\frac{(E_0^i-E^j)}{a}\right)\exp\left(-\frac{(E_0^i-E^j)^2}{b^2}\right). \tag{37}$$

The parameters were determined by an optimization algorithm for the special case of a rectangular function: we have calculated the convolution of a rectangular function with our resolution function [Eq. (20)] and tried to reproduce the original function by using Eq. (32), with the kernel taken from Eq. (37). The best results were obtained with $a=0.0997$ keV and $b=2.42$ keV. By choosing parameter b, emphasis can be put on either one of the two criteria given above. Choosing $b=2.72$ keV with $a=0.0997$ keV results in a better deconvolution of the rectangular spectrum but gives a standard deviation five times as large. With negligible noise $\mathbf{r}_\delta$, a much better deconvolution would be possible ($a=0.0595$ keV, $b=2.31$ keV), but this $\underline{\mathbf{v}}_c$ completely fails for noisy data.

Using Eq. (37), the shape of the reconstruction kernel can be varied only within a limited range. We found that by modifying the kernel locally, we can reproduce the rectangular function in a better way. To do this, we start with the optimum parameters a and b and change the reconstruction kernel bin by bin. Our optimization process does not change the symmetry of the function, although it would be evident in view of the asymmetry of the resolution function; but this does not lead to a significantly better deconvolution. The

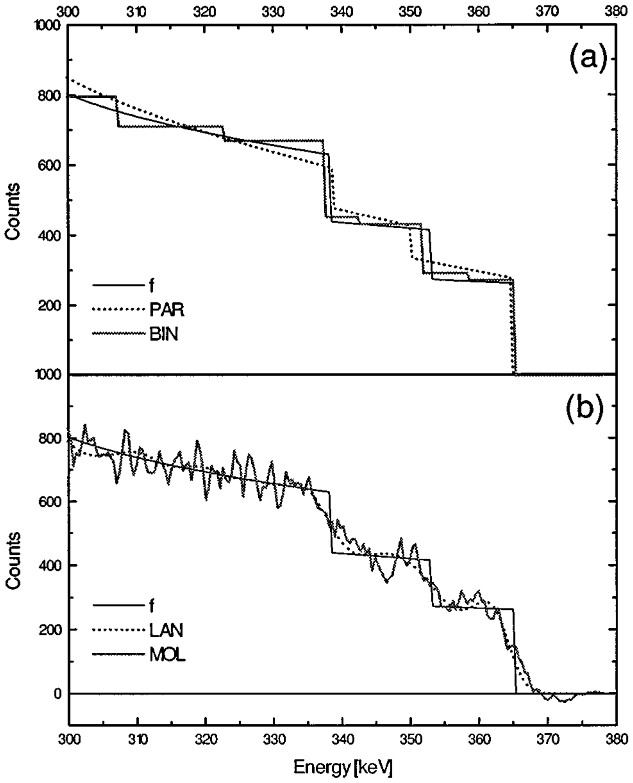

FIG. 3. Same as Fig. 2 but for a backscattering spectrum from a three-component target as in Fig. 1(b).

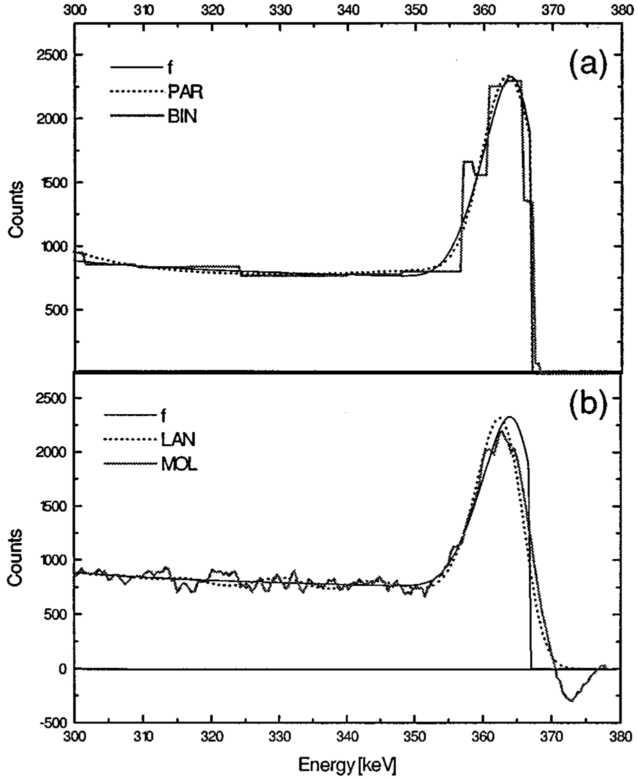

FIG. 4. Same as Fig. 2 but for a backscattering spectrum showing the 180° yield enhancement [Fig. 1(c)].

result is a function with less regular oscillations than those produced by Eq. (37). We call this method ''mollifier with postoptimization.''

IV. RESULTS

In Fig. 1 we show both the theoretical spectra and the corresponding measured spectra obtained by convolution with the resolution function and by adding some noise. In mathematical language, the convolution operator maps a theoretical spectrum from its domain of definition into its image space or range. Instead of ''domain of definition,'' we use here the term ''object space,'' which is more familiar to physicists. The crucial point is that in actual practice, the quality of deconvolution has to be assessed in image space. So any disagreement in object space between theoretical and deconvoluted spectra, which becomes apparent in Figs. 2, 3, and 4, does not appear in image space. Due to the smoothing properties of the resolution function, it is lost when both spectra are convoluted. This fact also reflects the ill-posedness of this inverse problem (see Sec. II).

First we discuss the results for a simple one-component spectrum. One usually is interested in the spectrum height and in the position of the high-energy edge. With only the measured spectrum at hand, as in Fig. 1(a), one would extrapolate the plateau of the spectrum towards the edge and one would define the position of the edge where the measured spectrum has half the height of the extrapolated plateau. In the case of asymmetric resolution functions, this introduces a systematic shift due to the difference between the median and the center of gravity of the resolution function. In Fig. 2(a) we show the theoretical spectrum f together with the results when methods using parametrization (PAR) and variable width binning (BIN) are applied to the measured spectrum. The results using modified Landweber iterations (LAN) and mollifiers with post-optimization (MOL) are shown in Fig. 2(b). Clearly, PAR yields the best agreement; the original and the deconvoluted spectra are almost indistinguishable. For BIN, the rather coarse binning results in an incorrect height of the edge, but the position is well reproduced. Using more bins would have made agreement in object space better, but without any discernible improvement in image space. Both LAN and MOL give spectra with steeper edges than the measured spectrum, but with an overshoot on top or at bottom, respectively. The essential difference is that the intersection of the deconvoluted spectrum with half of its extrapolated plateau should now give an unbiased estimate of the position of the edge, irrespective of the asymmetry of the resolution function. However, we do not believe that deconvolution of simple backscattering spectra based upon LAN or MOL will essentially improve evaluation.

In Fig. 3 we show the corresponding spectra from a three-component target. We want to point out that although the shape of the high-energy edge of the measured spectrum [Fig. 1(b)] is completely smeared, all methods detect its triple structure. Only PAR benefits from the information that the spectrum is composed of three partial spectra [Eq. (18)]. From this point of view, the quality of the result using PAR is rather poor. The best quantitative agreement is obtained with BIN [Fig. 3(a)]. The oscillations in the result of LAN [Fig. 3(b)] fit snugly into the steps of the theoretical spectrum, but this could also be fortuitous. MOL [Fig. 3(b)] yields a spectrum that does not allow any quantitative evaluation.

The goal of deconvolution of 180° backscattering spectra is to determine both height and position of maximum enhancement as accurately as possible. These quantities provide insight into small-angle scattering cross sections in solids. Here it is most evident that the knowledge of the exact function used to generate the theoretical spectrum favors PAR [Fig. 4(a)] and, in fact, the height of the maximum is perfectly reproduced and its position is only slightly shifted towards smaller energies. BIN [Fig. 4(a)] gives excellent data for the height, but there is no way to determine accurately the position of the maximum. Both LAN and MOL [Fig. 4(b)] result in a maximum at too low energies. In view of this fact, it is of no significance that LAN reproduces the height of the peak.

From these calculations we draw the following conclusions. The essential drawback of all deconvolution methods is that the quality assessment has to be performed in image space where the high-frequency components are damped by convolution with the Gaussian-shaped resolution function. In addition, (i) the more *a priori* information on the theoretical spectrum is available, the better deconvolution will work; hence, parametrization of the theoretical spectrum, if possible, is mostly superior to all other methods; (ii) the simple method of using histograms with variable bin width works unexpectedly well in the case of backscattering spectra from fairly homogeneous targets; (iii) all other methods that need no *a priori* information will work better when the spectrum does not contain high-frequency elements such as sharp edges.

ACKNOWLEDGMENT

We thank Professor H. Engl and Professor H. Paul for valuable discussions during this work. This work was supported by the Jubiläumsfonds der Oesterreichischen Nationalbank under Contract No. 4680.

[1] J. Keller, Am. Math. Monthly **83**, 107 (1976).

[2] Q. Xie and N. X. Chen, Phys. Rev. E **52**, 6055 (1995).

[3] H. Busse, K. Wandelt, and G. R. Castro, J. Electron. Spectrosc. Relat. Phenom. **72**, 311 (1995).

[4] R. Serimaa, J. Non-Cryst. Solids **193**, 372 (1995).

[5] H. Schafer, J. Magn. Res. A **116**, 145 (1995).

[6] P. J. Cumpson, J. Electron. Spectrosc. Relat. Phenom. **73**, 25 (1995).

[7] G. Landl, T. Langthaler, H. W. Engl, and H. F. Kauffmann, J. Comput. Phys. **95**, 1 (1991).

[8] A. K. Louis and P. Maaβ, Inverse Probl. **11**, 1211 (1995).

[9] A. K. Louis and P. Maaβ, Inverse Probl. **6**, 427 (1990).

[10] P. P. Pronko, B. R. Appleton, O. W. Holland, and S. R. Wilson, Phys. Rev. Lett. **43**, 779 (1979).

[11] D. Semrad and P. Bauer, Nucl. Instrum. Methods **149**, 159 (1978).

[12] A. L'Hoir, Nucl. Instrum. Methods **223**, 336 (1984).

[13] G. Bortels and P. Collaers, Appl. Radiat. Isot. **38**, 981 (1987).

[14] C. W. Groetsch, *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind* (Pitman, Boston, 1984).

[15] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems* (Kluwer, Dordrecht, 1996).

[16] F. Natterer, *The Mathematics of Computerized Tomography* (Teubner-Wiley, Stuttgart, 1986).

[17] J. Stoer and R. Burlisch, *Einführung in die Numerische Mathematik unter Berücksichtigung von Vorlesungen von F. L. Bauer* (Springer, Berlin, 1972).

[18] A. S. Householder, *The Theory of Matrices in Numerical Analysis* (Blaisdell, New York, 1969).

[19] R. Fletcher, *Practical Methods of Optimization* (Wiley, Chichester, 1980), Vol. 1.

[20] R. Fletcher, *Practical Methods of Optimization* (Wiley, Chichester, 1981), Vol. 2.

[21] A. N. Tikhonov and V. Y. Arsenin, *Solution of Ill-Posed Problems* (Winston, Washington, DC, 1977).

[22] *Maximum-Entropy and Bayesian Methods in Inverse Problems*, edited by C. R. Smith and W. T. Grandy, Jr. (Reidel, Dordrecht, 1985).

[23] L. Landweber, Am. J. Math. **73**, 615 (1951).

[24] B. Hofmann, *Regularization for Applied Inverse and Ill-Posed Problems* (Teubner, Leipzig, 1986).

[25] H. Ellmer, W. Fischer, A. Klose, and D. Semrad, Rev. Sci. Instrum. **67**, 1794 (1996).

[26] V. A. Morozov, *Methods for Solving Incorrectly Posed Problems* (Springer, New York, 1984).

[27] H. W. Engl and H. Gfrerer, Appl. Numer. Math. **4**, 395 (1988).

[28] H. Gfrerer, Math. Comput. **49**, 507 (1987); **49**, S5 (1987).

[29] M. A. Krasnosel'skii, *Approximate Solutions of Operator-Equations* (Wolters-Noordhoff, Groningen, 1972).