Nuclear Instruments and Methods in Physics Research B 93 (1994) 334–339
North-Holland

NIM B
Beam Interactions with Materials & Atoms

Deconvolution of noisy data with strong discontinuity and uncertainty evaluation

F. Schiettekatte *, R. Marchand, G.G. Ross
I.N.R.S.-Énergie et Matériaux, 1650 Montée Ste-Julie, Varennes, Québec J3X 1S2, Canada

Received 25 June 1993; revised form received 28 March 1994

A new method for deconvoluting data with statistical noise and sharp spike in the actual profile has been developed. It is applicable to a kernel of the general form and it allows to reveal some structure information which was not visible in the original data signal. The unknown function is assumed to be a superposition of a delta function and cubic splines. It is well adapted to the measurement of depth profiling of elements in matter by means of nuclear microanalysis.

1. Introduction

One technique frequently used to remove noise from digital measurements consists of an optimal (Wiener) filtering following a fast Fourier transform (FFT) [1]. This method has been used successfully in a number of experiments [2,3]. FFT can equally be used to perform deconvolutions. The convoluted signal would first be filtered. A division by the FFT of the resolution function and an inverse FFT would then be used to produce the deconvoluted signal. Unfortunately, this approach is limited by two serious constraints. The first has to do with the form of the kernel involved in the convolution. The second is related to the presence of spikes in the actual signal function.

A general expression for the convolution of a function of one variable $f(x)$ is

$$g(x) = \int_0^1 f(x')K(x, x')\,\mathrm{d}x' \qquad (1)$$

where K is the kernel of the convolution and g is the convoluted function. In Eq. (1), f and K are assumed to be defined for $0 \le x' \le 1$ and for $x \in \mathbb{R}$. In order for a one-dimensional FFT to be readily applicable, the kernel K must be of the form

$$K(x, x') = k(x - x'). \qquad (2)$$

The Fourier transform could be applied to the general form of K but this would require a two-dimensional FFT. For many problems of interest, the amount of data is such that a 2D FFT would be impractical. This is particularly true when real time data analysis is required in a PC environment. An additional difficulty in some measurements comes from the presence of physical sharp spikes in the actual signal $f(x')$. These spikes have high frequency components that would be lost in the filtered signal.

In this paper we present an efficient technique, based on finite elements, which is applicable to the deconvolution problem of noisy data with a kernel of the general form (such as Eq. (1)) and a signal function with a δ-function-like spike. This technique will then be applied to deconvolute hydrogen implantation profiles in materials by elastic recoil detection analysis [4]. In Section 2, we briefly describe the experimental conditions and the nature of the measurements which led us to develop this technique. In Section 3 we describe the deconvolution method in detail. The method is applied to computer simulated measurements in Section 4 and its accuracy is assessed. Our deconvolution method is applied to experimental measurements in Section 5. Finally, Section 6 gives a summary and some concluding remarks.

2. Experimental situation

In this section we describe the nature of the experimental condition which has motivated the development of a new deconvolution technique. One method used to measure quantitative depth profiles of light ions (all

* Corresponding author, tel. +1 514 449 8130, fax +1 514 449 8102.

SSDI 0168-583X(94)00194-Z

H and He isotopes) implanted in various substrates is elastic recoil detection analysis with an electromagnetic particle filter (ERD-E × B) [4]. The measurements obtained from this technique are convolutions of an actual depth profile with a kernel function which is also a function of depth [5]. As a result, K is not of the simple form given in Eq. (2). For example, in Sections 4 and 5 a deconvolution problem with a kernel of the form

$$K(x, x') = \frac{1}{\sqrt{2\pi}\sigma(x')} \exp\left[\frac{-(x-x')^2}{2\sigma^2(x')}\right] \tag{3}$$

is considered. In Eq. (3), the parameter σ is given by

$$\sigma^2(x') = \sigma_d^2 + \delta_{c1}x' + \delta_{c2}x'^2. \tag{4}$$

and so depends on x'. In this expression, the coefficients σ_d, δ_{c1}, δ_{c2} have to be determined for each combination of incident particle, recoiled particle and substrate. The coefficient σ_d is associated with the detector resolution. Coefficients δ_{c1} and δ_{c2} account for multiple scattering and straggling of the incident and of the recoiled particles as they move through the substrate. As an example, for H implanted in Be and depth profiled by means of a 350 keV He beam, these coefficients are: $\sigma_d = 4.95$ nm, $\delta_{c1} = 0.794$ nm, $\delta_{c2} = 3.96 \times 10^{-3}$, with x' given in nm [5]. In this particular case, expressions (3) and (4) for the standard deviation are valid for $x < 100$ nm.

One important feature of our measurements is the presence of a monatomic layer of H adsorbed at the surface of the substrate. This layer is due to the presence of contaminants such as water vapour (even when a trapping cryostat is used) and the migration of deeper hydrogen retained by binding energy. This spike in the hydrogen density at the surface appears as a Gaussian density profile in the ERD-E × B measurement.

Another difficulty has to do with the presence of statistical noise in the measurements. In certain substrates, significant desorption processes have been observed during measurements [8]. Hence, the most reliable profiles are obtained for the lower fluences of incident particle. This procedure unfortunately leads to higher statistical noise. Better statistics can be obtained by repeating the measurement at several locations on the substrate (when the implantation is uniform) and by bringing the detector closer to the substrate. These procedures are unfortunately not always practical. Hence, it is imperative to construct a method which can perform reliable deconvolutions on data with relatively high statistical noise.

3. Deconvolution technique

We now describe the method used to determine the actual deposition profile $f(x)$ from a measured signal $h(x)$. The method is not limited to a particular kind of measurement technique. It is presented here in general terms and is applicable to the deconvolution of any function h with statistical noise. Let $f(x)$ be the profile to be determined and $g(x)$ its convolution with respect to kernel K, as defined in Eq. (1). If $n(x)$ represents statistical noise ($n(x) < g(x)$), then the actual measurement from which f must be determined is $h(x) = g(x) + n(x)$. The solution to the deconvolution problem consists first of expressing the unknown function f as a sum of known basis functions,

$$\bar{f}(x) = \sum_{i=0}^{n} \alpha_i \psi_i(x) \approx f(x), \tag{5}$$

for $0 \le x < L$, the interval of definition of $f(x)$. In Eq. (5), the bar indicates that the function represents an approximation to the actual profile $f(x)$. If FFTs were used, the basis functions ψ_i would represent sine and cosine functions with arguments being multiples of $2\pi x/L$. For reasons explained in the previous section, it is more convenient to use a combination of a cubic spline ($i = 1, \ldots, n$) and a delta function ($i = 0$). The n cubic splines are defined on an $n-2$ point partition of the interval. Substituting this expression in Eq. (1) yields

$$\bar{h}(x) = \sum_{i=0}^{n} \alpha_i \int_0^L \mathrm{d}x'\, K(x, x')\psi_i(x') \approx h(x). \tag{6}$$

The coefficients α_i are then determined by minimizing the mean square deviation between the approximate convolution and the measured signal h. It is important, however, to impose two constraints in the minimization process. One is that the total number of counts (area under the curve) must be the same in both h and $\bar{h}$. The second is that the approximate function $\bar{f}$ must be non-negative. The problem then reduces to that of minimizing a function H of the form

$$H = S + \lambda \int \mathrm{d}x \left(h(x) - \sum_{i=0}^{n} \alpha_i \int_0^L \mathrm{d}x'\, K(x, x')\psi_i(x') \right), \tag{7}$$

with

$$S = \int \mathrm{d}x \left(h(x) - \sum_{i=0}^{n} \alpha_i \int_0^L \mathrm{d}x'\, K(x, x')\psi_i(x') \right)^2. \tag{8}$$

The minimisation must be done with respect to the parameters α_i and λ. The parameter λ is a Lagrange multiplier. It is introduced to enforce the first constraint. With the second constraint, this problem is a classical quadratic programming problem. It is solved numerically with a commercial routine [6].

Methods have been proposed to determine the optimal distribution of nodes when representing a function

with cubic splines [7]. A number of trials have been made for the problem at hand, and a uniform distribution of the nodes has been found to be quite satisfactory in all cases considered. In the following, we therefore limit our attention to partitions with a uniform distribution of nodes. Another parameter which needs to be determined is the number of nodes in the partition. This point is addressed in the next section.

4. Numerical experiments

In this section we present deconvolution results from signals $h(x)$ which have been generated by an artificial function $f(x)$. The comparison between the approximate deconvoluted profile $\bar{f}(x)$ and the known exact profile $f(x)$ serves to assess the accuracy of the method. It is also useful in providing a simple rule for the number of nodes $n-2$ needed in the partition. To be specific, these numerical experiments consist of the following steps:

1) An arbitrary distribution $f(x)$ is constructed.

2) A convolution $g(x)$ is calculated from Eq. (1).

3) A histogram is constructed from N random numbers, consistent with distribution $g(x)$. That is, for N random numbers $0 \le r \le 1$, $h(x)$ is defined as the number of random numbers which satisfies

$$\int_{-\infty}^{x_i} \mathrm{d}x\, g(x) \le r \int_{-\infty}^{\infty} \mathrm{d}x\, g(x) < \int_{-\infty}^{x_i+\Delta} \mathrm{d}x\, g(x) \qquad (9)$$

divided by Δ, where Δ is the value between two distinct channels.

4) A deconvolution is then applied to $h(x)$, as described in the previous section.

Two example results are illustrated in Fig. 1. In example (a), the histogram $h(x)$ is defined on the interval $[0, \infty]$, with $\Delta = 0.785$ keV, and $N = 144$. The deconvolution is performed using $n = 7$ cubic splines. Despite the relatively large statistical fluctuations in $h(x)$, the agreement between f and $\bar{f}$ is seen to be quite good. Also, one interesting feature of this deconvolution technique is its ability to reproduce the sharp density spike at the surface, while efficiently filtering undesirable noise. In example (b), we see how our method can reveal structures hidden by the experimental resolution. This result has been obtained with $N =$ 6667 and 12 cubic splines.

Similar numerical experiments have been made to determine a simple relation for the number of cubic splines to use as a function of the number N of detected particles. A small number of splines will tend to sample more of the convoluted function and, thus, smooth out statistical noise more effectively. A larger number of splines, however, is required in order to reproduce any structure present in the function f. If too many splines are used for a given count, non-physical structures associated with noise will result. If too few are used, the deconvoluted profile may miss some important structures. In order to determine the best number of splines, a series of numerical experiments have been conducted. We have calculated the mean square deviation

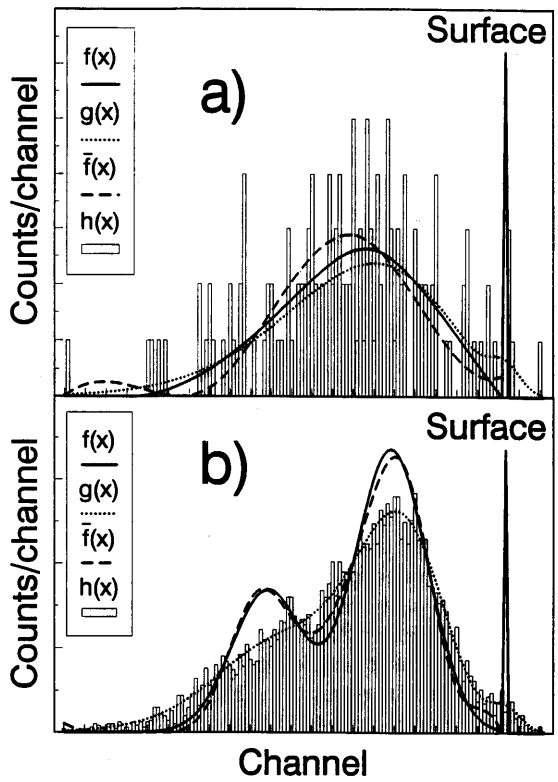

Fig. 1. Deconvolution of artificially constructed signals. In (a), $h(x)$ is constructed from 144 random numbers distributed according to $g(x)$, and six cubic splines are used in the deconvolution. In (b), 12 cubic splines are used to deconvolute a signal $h(x)$ constructed from 6667 random numbers distributed according to a convoluted two-peak $f(x)$.

$$\frac{1}{M}\sum_{i=1}^{M}\int_{-\infty}^{\infty}\left(f(x)-\bar{f}_i(x)\right)^2 \mathrm{d}x,$$

where $M = 50$ and $\bar{f}_i(x)$ is obtained each time from a new generated $h(x)$. We have made the test for various values of n and N, for some selected functions $f(x)$. Empirically, the smallest error is obtained with

$$n \simeq 5.2 + 0.083N^{1/2}, \qquad (10)$$

for $N \ge 100$. This expression for n is used in all the calculations presented here. This expression has been found empirically for distributions over approximately 150 channels. A different number of channels may require a different expression for n.

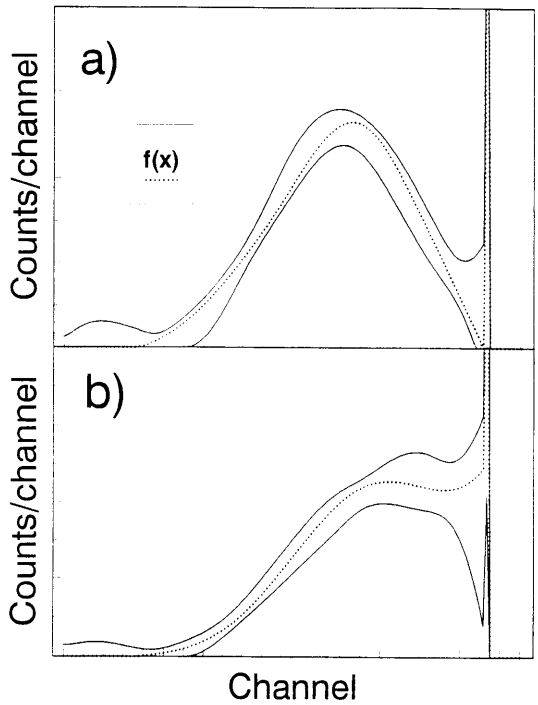

Fig. 2. Standard deviation (———) calculated at each channel of $\tilde{f}(x)$ obtained from 50 different $h(x)$ generated in each instance with 144 random numbers. Two distributions are considered. In (a), the bulk concentration vanishes at the surface. In (b), it is finite. The distribution $f(x)$ assumed in both cases is shown with dotted lines.

The uncertainty which results from the deconvolution is illustrated in Fig. 2. The dotted curve shows the actual profile $f(x)$. The upper and lower curves are obtained respectively by adding and subtracting the mean deviation of the deconvoluted signals obtained from a large set of different $h(x)$, each one being generated with 144 random numbers. When subtracting the mean square, if a negative result is found, the result is assumed to be zero. The distance between the dashed and the solid curves represents the uncertainty in the deconvolution. We note that the maxima of the distribution for the deconvoluted signal in Fig. 1 may be shifted compared with those of the actual profile $f(x)$. This shift is small, however, and it is mainly due to statistical errors.

Finally, we have studied the behaviour of the mean of those standard deviations over the deconvoluted profiles as a function of the number of detected particles. The results are reported in Fig. 3 for the distribution function $f(x)$ of Figs. 2a and 2b. The relative error defined as $e = \sqrt{S/N}$ is found empirically to scale as $e = \pm 0.083 N^{-1/3}$ (S is defined in Eq. (8)).

5. Application to experimental results

Fig. 4 shows an example application to the measurement of an actual implantation profile. This measurement is for hydrogen implanted at 1.5 keV near saturation in a beryllium substrate. The analyzing beam con-

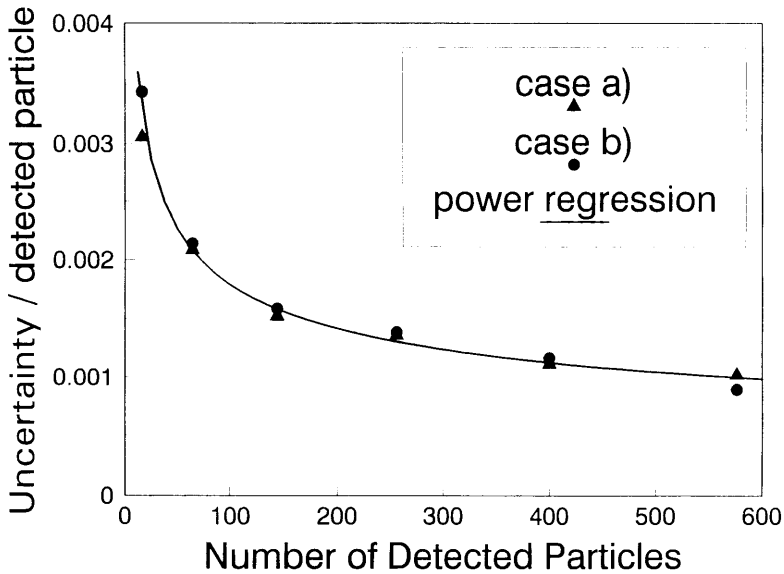

Fig. 3. Mean of the standard deviations (excluding surface spike) for different numbers of detected particles. The profiles considered are those of Fig. 2a (▲) and Fig. 2b (●).

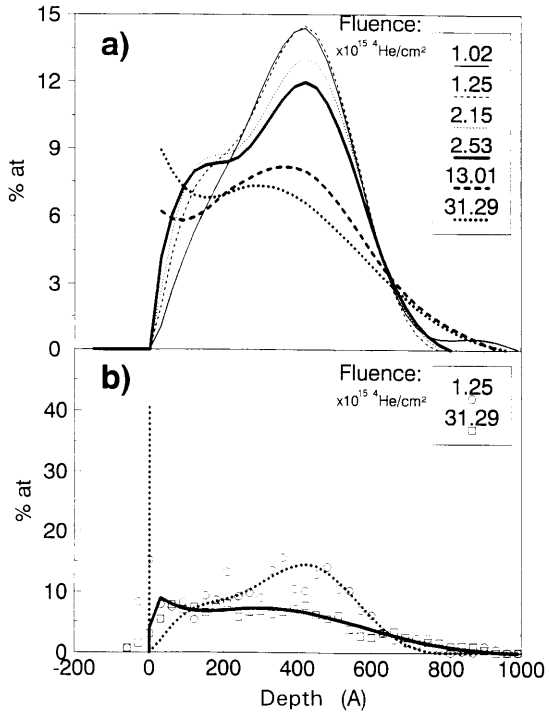

Fig. 4. Profile modification and desorption of hydrogen (1.3×10^{17} at./cm^2) implanted in beryllium at 1.5 keV, induced by the 350 keV ^{4}He probing beam. (a) Deconvoluted profiles computed with various fluences of the beam. Surface spikes have been subtracted for clarity. (b) The raw signal and the deconvoluted profiles for two selected fluences considered in (a).

sists of a 350 keV He beam incident on the surface at 25°. In Fig. 4, each profile is obtained with incident He fluences varying from 1.02×10^{15} to 31.3×10^{15} ^{4}He/cm^2. The deconvolution results are plotted in Fig. 4a. For clarity, the surface spike in hydrogen density has been subtracted from the deconvoluted profiles. This figure clearly shows the modifications to the implanted hydrogen depth profile as the fluence of the He beam increases. The initial implantation profile has a maximum near 42 nm, and a full width at half maximum of approximately 40 nm. After being exposed to a small fluence of the probing beam, the depth profile begins to diffuse partly in the defect profile created during the implantation, thus forming a protuberance closer to the surface (in agreement with corrected TRIM [9] runs). After a large fluence, we observe a surface density with an absolute maximum at the surface. The profile also diffuses significantly deeper into the material. Fig. 4b shows experimental data together with the deconvoluted signal computed with two fluences. For the lower fluence, the surface contamination concentration amounts to a relatively low value of 15 at.%. When deconvoluted, this fraction increases considerably to approximately 40 at.%. This spike is desorbed relatively rapidly and it is essentially absent at the higher fluence.

Fig. 5 shows a hydrogen profile implanted to saturation in carbon with an energy of 1 keV. The use of our deconvolution procedure reveals the presence of two maxima in the actual profile; a feature which is not readily visible from the raw data. This structure would

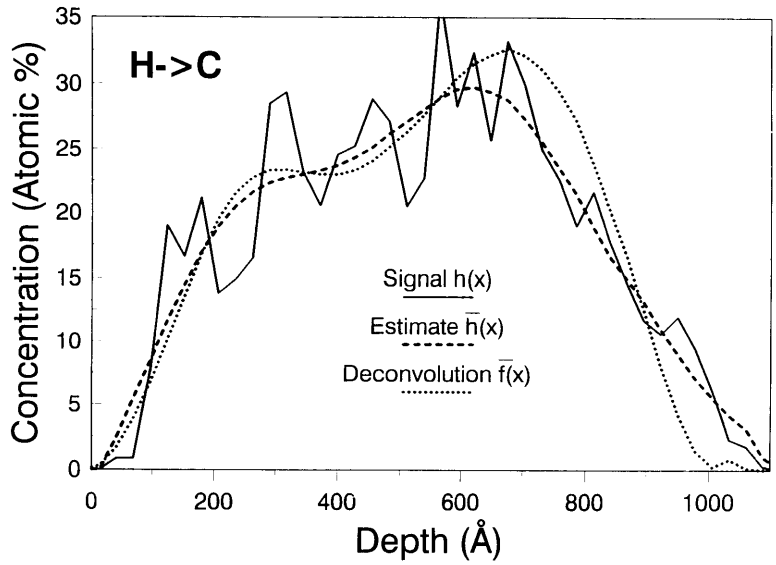

Fig. 5. Comparison between deconvoluted profile (· · · · · ·), the original signal (———) and our estimate of the convoluted profile $\bar{h}(x)$ (- - - - - -) (see Eq. (6)), for H implanted to saturation in carbon.

be due to a fast release of hydrogen from a depth near to the mean range, i.e. where the hydrogen concentration was at its highest value [10].

6. Summary and conclusion

A new technique for deconvoluting data with statistical noise and sharp physical spikes has been developed. The unknown function is represented as a superposition of a delta function and cubic splines. While a non-uniform distribution of nodes is possible, a definition of splines on a uniform partition of the interval is found to be satisfactory in all the cases considered. This method does not completely remove statistical error, but it considerably enhances the signal-to-noise ratio and produces smooth deconvoluted profiles. Also, compared with more usual methods which use FFT filtering and inverse FFT, the present technique has the advantage of preserving physical spikes in the function to be deconvoluted. It is also applicable to a more general form of the kernel involved in the convolution.

The method has been applied to the measurement of hydrogen density profiles in beryllium and carbon substrates. It has served to demonstrate the modification in the implanted hydrogen profile caused by exposure to an energetic helium probing beam. It shows also that one can get structure information from ERD profiles.

Acknowledgement

This work has been supported by the Natural Sciences and Engineering Research Council of Canada.

References

[1] W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, Numerical Recipes (Cambridge University Press, New York, 1986) chap. 12.6.
[2] K. Piyakis and E. Sacher, Appl. Surf. Sci. 55 (1992) 159.
[3] P.J. Chen, M.L. Colaianni and J.T. Yates, J. Vac. Sci. Technol. A 8 (1990) 764.
[4] G.G. Ross, B. Terreault, G. Gobeil, G. Abel, C. Boucher and G. Veilleux, J. Nucl. Mater. 128/129 (1984) 730.
[5] G.G. Ross and L. Leblanc, Nucl. Instr. and Meth. B 62 (1992) 484.
[6] The NAG Fortran Workstation Library, Release 1 (NAG, Oxford, 1991) subroutine E04NAF.
[7] P.D. Loach and A.J. Wathen, IAM J. Numer. Anal. 11 (1991) 393.
[8] G.G. Ross and I. Richard, Nucl. Instr. and Meth. B 64 (1992) 603.
[9] J.F. Ziegler and J.P. Bierzack, Nucl. Instr. and Meth. 174 (1990) 190.
[10] B. Terreault, R.G. Saint-Jacques, G. Veilleux, J.G. Martel, J. L'Ecuyer, C. Brassard and C. Cardinal, J. Nucl. Mater. 68 (1977) 334.