Journal of Engineering Mathematics **43:** 115–134, 2002.

On the determination of dopant-concentration profiles by grazing emission X-ray fluorescence spectroscopy using the maximum-entropy method

S.M.P. SMOLDERS and H.P. URBACH
Philips Research Laboratory, Professor Holstlaan 4, 5656 AA Eindhoven, The Netherlands

Received 10 August 2001; accepted in revised form 2 May 2002

Abstract. The determination of concentration profiles of impurities in silicon from angle scans of emitted x-ray fluorescence intensities using the maximum-entropy method is studied. Existence and convergence properties of the maxium- entropy method are discussed. The application of the maximum-entropy method to Grazing emission X-Ray Fluorescence Spectromety is compared with an analytical method. It is found that, provided noise levels are sufficiently low, concentration profiles can be reconstructed without using *a priori* knowledge.

Key words: inverse Laplace transform, maximum entropy, semiconductors, spectroscopy, X-ray fluorescence

1. Introduction

The measurement of Group III and Group V dopant-concentration profiles is very important in semiconductor technology. Every measuring technique currently used has some disadvantages. In the new nondestructive analytical technique called Grazing Emission X-ray Fluorescence Spectroscopy (GEXRF), the doped wafer is irradiated by primary x-rays which induce fluorescence at the characteristic wavelengths of the atoms in the wafer. Since the dopant concentrations are low, the fluorescence intensity emitted by the dopant is low also and hence noise due to scattering by the background is generally high. To improve the signal-to-noise ratio, the fluorescence intensity in GEXRF is detected at grazing emission angles (usually < 100 mrad) with respect to the surface of the wafer. This mode of measurement distinguishes GEXRF from convential x-ray fluorescence spectroscopy.

The feasibility of measuring x-ray fluorescence under grazing emission angles was studied for the first time by Becker, Golovchenko and Patel [1]. The determination of concentrations, layer thicknesses or both from GEXRF measurements is very sensitive to noise. Some inverse problems occurring in GEXRF have recently been studied [2]. Inverse problems for a similar technique called TEXRF have been studied in [3] and [4]. TEXRF can be considered as the reciprocal of GEXRF in that the positions of the x-ray tube and the detector are interchanged, *i.e.*, in TEXRF the tube is positioned at grazing angles to the sample, while the fluorescence radiation is detected at a fixed large angle. In all of these papers *a priori* information about the concentrations is used to stabilize the inversion.

In this paper we will apply the maximum-entropy method to the reconstruction of dopant-concentration profiles in silicon. The arsenic-concentration profiles used in the numerical experiments are obtained from simulations of the ion-implantation process. We add Poisson noise of the photon-counting process to the simulated intensities that are used in the reconstruction of the profiles.

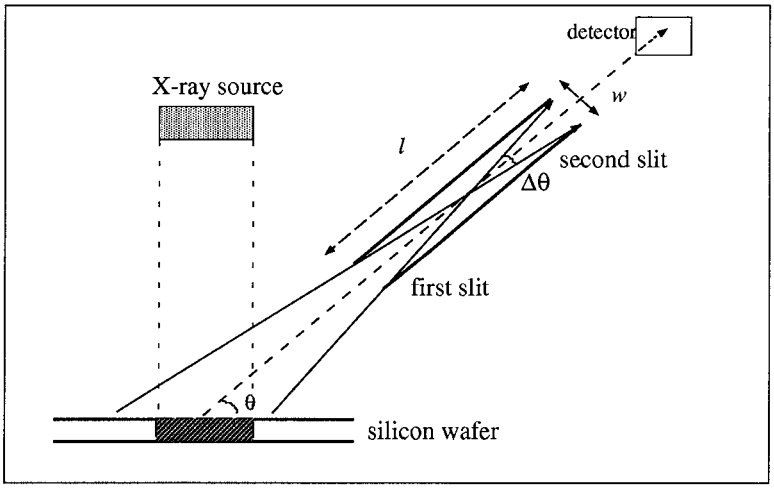

Figure 1. The experimental setup of GEXRF showing the x-ray tube, the sample, the double slit collimator and the detector. The divergence $\Delta\theta$ is also indicated. The detection angle θ is grossly exaggerated and was actually $< 0{\cdot}1$ rad.

In Section 2 we will briefly describe the mathematical model of GEXRF, *i.e.*, the model for computing emitted fluorescence intensities for a given concentration profile. It will be shown that the inverse problem is equivalent to the inversion of the Laplace transform with data on a bounded interval of the real axis. In Section 3 the maximum-entropy method will be formulated as a convex constrained optimisation problem. Some of the mathematical properties of this optimization problem will be discussed in Section 4. In Section 5 we will briefly describe the reconstruction by regularised inversion of the Laplace transform. In this method the measured data are first extrapolated to the entire positive real axis and then the spectral decomposition of the Laplace transform is used to compute the regularised inverse. In the numerical experiments discussed in Section 6, we will compare the results obtained by using the maximum-entropy method to those obtained by the spectral decompostion method (also called Singular-Value Decomposition method) of the Laplace transform. A disadvantage of the SVD for the Laplace transform is that it is not flexible. Another measurement strategy which leads to a different instrumental factor than the one considered in this paper cannot be taken into account in general. Furthermore, the incorporation of *a priori* information is not straightforward. In contrast, the maximum-entropy method is flexible regarding a change of the instrumental factor and the incorporation of *a priori* knowledge. *A priori* information is very important to reduce the sensitivity to noise of the inversion. Some numerical experiments for realistic profiles are described in Section 6. The results suggest that the reconstruction of typical profiles of arsenic in silicon will be possible, provided noise levels are sufficiently low.

We would like to stress here the important role of modelling in the study of the feasability of proposed new measurement techniques such as those considered in this paper. In simulating the reconstructions of freely chosen concentration profiles, we can change signal-to-noise ratios at will, investigate the influence of the shape of the profile on the instability, study the influence of adding more data points, etc. In this way we can gain insight in the limitations of the proposed technique that could not be acquired by experiments alone.

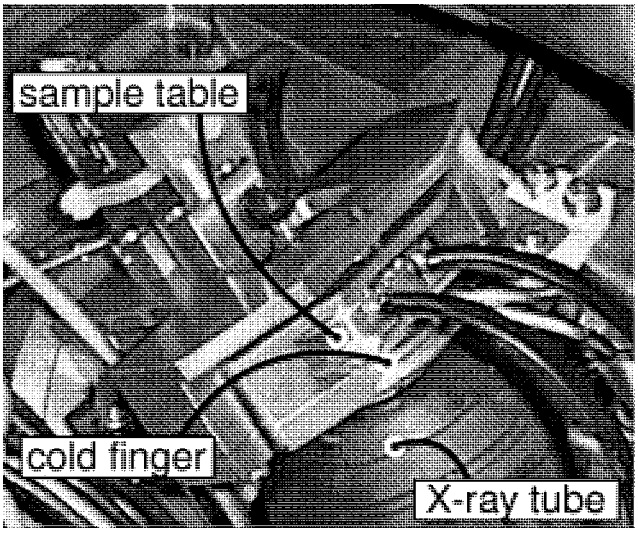

Figure 2. Picture of the interior of the GEXRF spectrometer. The detector is behind the sample table and can therefore not be seen.

2. The mathematical model

Figure 1 shows the setup of the laboratory GEXRF spectrometer. A circular region of the doped silicon wafer with a diameter of approximately 3 cm is irradiated by radiation (power $\approx$ 5 W) from a conventional x-ray tube. Part of this radiation is absorbed and converted into x-ray fluorescence radiation at the characteristic wavelengths of the elements present. A double-slit collimator is used to choose the angle of detection and a crystal monochromator (not shown) is used to select the wavelength of the characteristic radiation. Figure 2 is a picture of part of the GEXRF spectrometer. As x-ray fluorescence is a very inefficient process, virtually all the radiant power incident on the sample is dissipated as heat. Therefore the sample table is air-cooled. In spite of the fact that the region between the sample table and the x-ray tube is kept in a technical vacuum, a so-called cold finger is needed to prevent deposition of carbon from the residual gas on the sample surface. The sample table can be rotated with respect to the direction defined by the double-slit collimator between 0 and 100 mrad with a minimum step size of 0·04 mrad. The double-slit collimator consists of two slits (5–500 μm) 125 mm apart, hence defining angular resolutions between 0·08 and 8 mrad. A more detailed description of the laboratory instrument can be found in [5], [6].

In our model of the instrument we choose a Cartesian coordinate system (x, y, z) of which the z-axis runs perpendicular to the surface of the wafer with z increasing into the sample and with $z = 0$ at the surface of the wafer. Let $f(z)$ be the concentration of the dopant at depth $z > 0$. The concentration of the dopant is thus assumed independent of x and y. An example of a typical concentration profile of arsenic in silicon is shown in Figure 3. It is seen that the profiles are very shallow and that concentration values are low. Therefore, the complex refractive index of the silicon wafer at all the wavelengths of interest can be considered to be the same as that of un-doped silicon. For x-rays the real part of the refractive index n of silicon (and of most other materials) is less than 1 and it is customary to write

$$n = 1 - \delta + \mathrm{i}\beta, \tag{2.1}$$

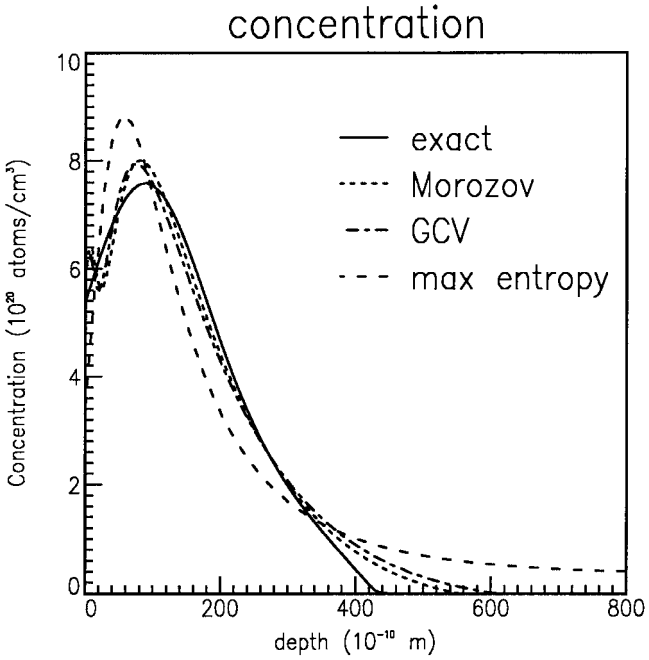

Figure 3. The 'exact' concentration profile (continuous curve) simulated with TSUPREM-4 for an implantation dose of 2×10^{15} As atoms/cm^2 at 8 keV followed by thermal diffusion at 100 °C for 20 s. The dashed profiles are obtained by regularised inversions of noisy fluorescence data at 120 equidistant angles using the maximum-entropy method and Tikhonov's inversion with both Morozov's principle and GCV to determine the regularization parameter. The divergence admitted by the two slits is $\Delta\theta = 0{\cdot}4$ mrad. The measurement time was 180 s per angle, which corresponds to a mean relative error in the data of 0·74%.

where δ and β are small positive numbers of an order of 10^{-3} or less. Because the region $z < 0$ between the sample table and the x-ray tube is in a technical vacuum, the refractive index is 1 for $z < 0$. Let λ be the wavelength of a characteristic line of the dopant and let n be the (known) refractive index of the wafer at this wavelength. The fluorescence intensity per steradian emitted at an angle θ with respect to the surface of the wafer is given by (see [6], [7]

$$I(\theta) = CT(\theta) \int_0^\infty \exp[-2kz\, \Im\mathfrak{m}\sqrt{n^2 - \cos^2\theta}]\, f(z)\, \mathrm{d}z, \tag{2.2}$$

where $\Im\mathfrak{m}$ means the imaginary part, $k = 2\pi/\lambda$ is the wave number in vacuum, C is a known constant which depends on the fluorescence wavelength λ and on the primary radiation, but is independent of the emission angle θ, and $T(\theta)$ is the squared absolute value of the transmission coefficient for the surface $z = 0$:

$$T(\theta) = \left| \frac{2\sin\theta}{\sin\theta + \sqrt{n^2 - \cos^2\theta}} \right|^2. \tag{2.3}$$

The exponential factor in the integrand accounts for absorption of the fluorescence radiation by the sample. Because the real part of the refractive index is smaller than 1, there exists a so-called critical angle defined by

$$\cos\theta_{\mathrm{crit}} = 1 - \delta. \tag{2.4}$$

At detection angles smaller than θ_{crit} the wave inside the wafer is evanescent in the z-direction. The imaginary part of the square root at the right-hand side of (2.2) is then relatively large;

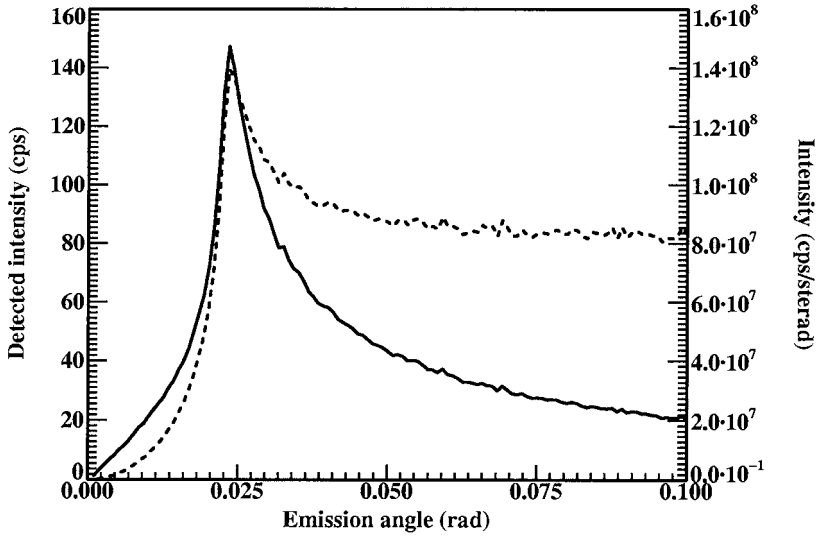

Figure 4. The fluorescence intensity $I(\theta)$ (dashed curve) emitted per second and per solid angle and the number of counts per second $P_{\text{detector}}(\theta)$ (continuous curve) measured by the detector as functions of the emission angle θ for the the Lα line of the continuous dopant profile of arsenic in silicon shown in Figure 3. The slit widts corresponds to the divergence $\Delta\theta = 0{\cdot}4$ mrad. The Poisson noise that was added corresponds to a measurement time of 180 s per angle and a mean relative error of 0·74%.

therefore only atoms very close to the surface contribute appreciably to the measured intensity, which is consequently relatively low. This is illustrated by Figure 4 in which the emitted and (simulated) detected intensities are shown of the Lα line of As corresponding to the profile of Figure 3.

For small widths of the slits of the collimator, the intensity actually measured by the instrument is, to a good approximation, given by ([6])

$$P_{\text{detector}}(\theta) = S(\theta)\, I(\theta), \tag{2.5}$$

where $S(\theta)$ is a monotonically decreasing instrumental factor (see Figure 5).

It is assumed that the main source of noise in the data is Poisson noise caused by the photon counting. This implies that, when the measurement time for angle θ is t_θ, the relative error in the data at that angle is $1/[P_{\text{detector}}(t_\theta)\, t_\theta]^{1/2}$, *i.e.*, the reciprocal of the square root of the total counts $P_{\text{detector}}(t_\theta)\, t_\theta$. To study the influence of noise on the reconstruction, we will add several levels of noise to the simulated data.

Define

$$p(\theta) = 2k\, \mathfrak{Im}\, (n^2 - \cos^2\theta)^{1/2}, \tag{2.6}$$

and

$$F(p) = I(\theta)/CT(\theta). \tag{2.7}$$

Then (2.2) implies

$$F(p) = \int_0^\infty \exp(-pz)\, f(z)\, \mathrm{d}z. \tag{2.8}$$

The interval $(p_{\min}, p_{\max})$ corresponding to the largest possible range of detection angles $0 < \theta < \pi/2$ is given by

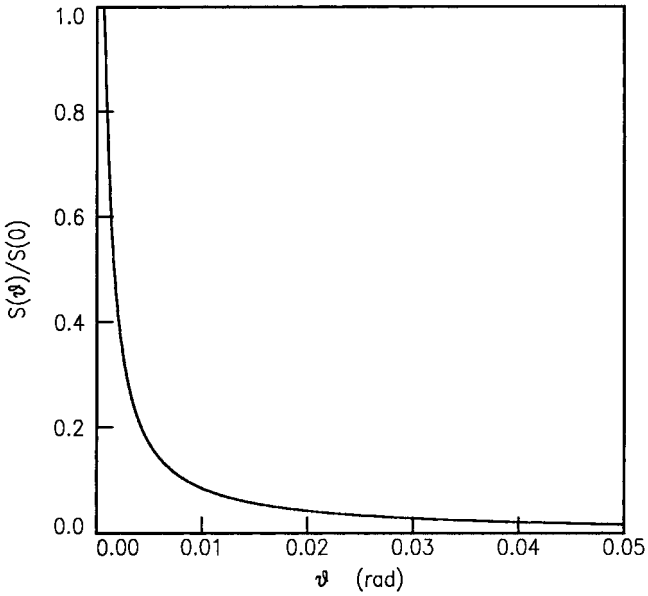

Figure 5. The normalised instrumental factor $S(\theta)/S(0)$ as a function of the detection angle θ. The slit width $w = 24{\cdot}7\ \mu$m and the slit distance $l = 123{\cdot}6$ mm, which implies that the divergence of the detected beam is $\Delta\theta = 0{\cdot}4$ mrad.

$$p_{\min} = p(\pi/2) = 2\,k\,\beta, \tag{2.9}$$

$$p_{\max} = p(0) = 2\,k\,\Im m\,(n^2 - 1)^{1/2} \approx 2\sqrt{2}\,k\,\Im m(-\delta + \mathrm{i}\beta)^{1/2}. \tag{2.10}$$

In Figure 6 the function $p(\theta)$ is shown for the Lα line of As in silicon. Because the detection angles are limited to grazing angles ($\theta \ll \pi/2$), $p_{\min}$ is in practice larger than $2\,k\,\beta$. However, as is confirmed by the simulations, for angles larger than approximately 6°, the data do not contain additional information about the profile. Hence, the bulk of the information is contained in the data corresponding to the smallest angles. We conclude that the concentration profile has to be determined by an inversion of the Laplace transform using (noisy) data restricted to a small interval of the positive real axis.

To understand the instability of the inversion of the truncated Laplace transform, it is helpful to consider the inversion of the Laplace transform itself. The Laplace transform of a square integral function $f : (0, \infty) \to \mathcal{R}$ can be shown to be square integrable. Hence, the Laplace transform can be regarded as an operator $\mathcal{L} : L^2(0, \infty) \to L^2(0, \infty)$. The following properties of this operator are proved in, *e.g.*, [8] and [9]. The operator $\mathcal{L}$ is symmetric and bounded, but it is not compact. Its spectrum is given by the interval $[-\pi^{1/2}, \pi^{1/2}]$. For $-\pi^{1/2} \leq \lambda \leq \pi^{1/2}$ with $\lambda \neq 0$, we define τ by $\tau = (1/\pi)\mathrm{arccosh}(\pi/\lambda^2)$. Then the eigenfunction corresponding to λ is

$$e_\lambda(z) = \frac{1}{|\lambda|(\pi^2 - \lambda^4)^{1/4}} \left\{ \Gamma(1/2 + \mathrm{i}\tau)^{1/2}\, \frac{z^{-1/2-\mathrm{i}\tau}}{\sqrt{2\pi}} + \Gamma(1/2 - \mathrm{i}\tau)^{1/2}\, \frac{z^{-1/2+\mathrm{i}\tau}}{\sqrt{2\pi}} \right\}, \tag{2.11}$$

when $\lambda > 0$ and

$$e_\lambda(z) = \frac{1}{|\lambda|(\pi^2 - \lambda^4)^{1/4}} \left\{ \Gamma(1/2 + \mathrm{i}\tau)^{1/2}\, \frac{z^{-1/2-\mathrm{i}\tau}}{\sqrt{2\pi}} - \Gamma(1/2 - \mathrm{i}\tau)^{1/2}\, \frac{z^{-1/2+\mathrm{i}\tau}}{\sqrt{2\pi}} \right\}, \tag{2.12}$$

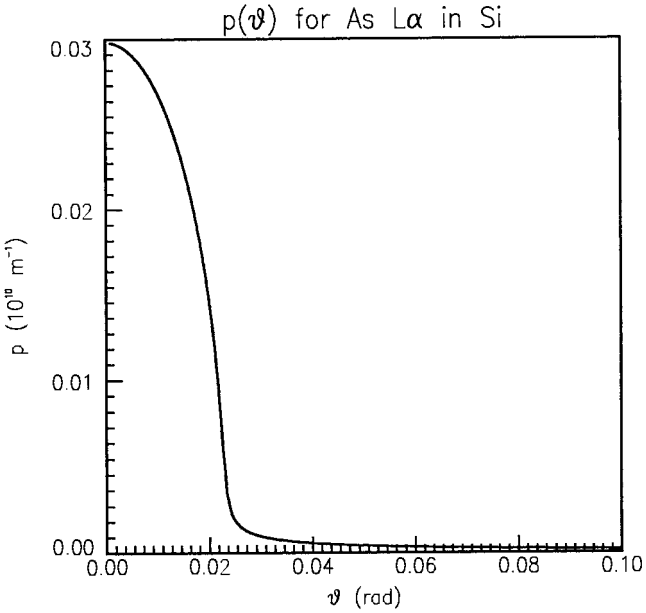

Figure 6. The function p(θ) for the As Lα line in silicon. We have $p_{\min} = 1{\cdot}88 \times 10^{-5}$ Å^{-1} and $p_{\max} = 2{\cdot}98 \times 10^{-2}$ Å^{-1}.

when $\lambda < 0$, where Γ is the Gamma function. The spectral decomposition of $\mathcal{L}$ is thus

$$\mathcal{L}(f) = \int_{-\sqrt{\pi}}^{\sqrt{\pi}} \lambda(f, e_\lambda) e_\lambda \, \mathrm{d}\lambda, \tag{2.13}$$

where $(.,.)$ is the scalar product of $L^2(0, \infty)$. It is seen that, when $\lambda \to 0$, the eigenfunctions become increasingly oscillatory. We remark that the eigenfunctions are not elements of the space $L^2(0, \infty)$. Furthermore, (2.13) should in general be interpreted as the principle-value integral:

$$\int_{-\sqrt{\pi}}^{\sqrt{\pi}} = \lim_{\delta \to 0} \int_{-\sqrt{\pi}}^{-\delta} + \int_{\delta}^{\sqrt{\pi}}, \tag{2.14}$$

where the limit converges in general only in the sense of $L^2(0, \infty)$.

It can be shown that, when f in $L^2(0, \infty)$ satisfies $\mathcal{L}(f) = 0$, then f vanishes. Hence, the solution of

$$\mathcal{L}(f) = F, \tag{2.15}$$

provided it exists, is unique and given by

$$f = \mathcal{L}^{-1}(F) = \int_{-\sqrt{\pi}}^{\sqrt{\pi}} \frac{1}{\lambda}(F, e_\lambda) e_\lambda \mathrm{d}\lambda. \tag{2.16}$$

However, this expression is extremely unstable due to the eigenvalue in the numerator of the integrand and it is therefore useless when F is obtained from noisy measurements.

It is not difficult to see that the inverse of the Laplace transform is also unique when the data are restricted to a bounded interval $(p_{\min}, p_{\max})$ of the positive real axis. In fact, for f

in $L^2(0,\infty)$, the function $F(p) = \mathcal{L}(f)(p)$ is analytic in the positive half space $\mathfrak{Re}\, p > 0$ and hence, if F is known on a countable set inside the interval $(p_{\min}, p_{\max})$, it has a unique extension to a function in $L^2(0,\infty)$. Hence, the concentration profile is uniquely determined by a countable set of data inside the interval $(p_{\min}, p_{\max})$.

It may be expected that the inversion of the Laplace transform with data restricted to $(p_{\min}, p_{\max})$ is even more unstable than the inversion of the Laplace transform with data on the half line. It is thus clear that the inversion must be regularised.

The inversion of the (truncated) Laplace transform is a problem that is encountered frequently in applied physics. In their recent review paper, Istratov *et al.* [10] mentioned many different regularisation methods that have been tried in many different applications. One method of regularisation is to replace the inverse problem by a constrained optimization problem in which a certain object functional is optimised under the constraint that, for the optimum profile f, $\mathcal{L}(f)$ shall differ from the measured data by no more than the noise level. In the next section we will consider an optimization problem of this type that was introduced by Jaynes [11] and is called the maximum-entropy method.

3. The maximum-entropy method

In the maximum-entropy method it is necessary to assume a finite depth of the concentration profile. Let d therefore be so large that for $z > d$ the amount of the dopant is negligible. Usually the total amount of the dopant is known from the ion-implantation process and we therefore normalise the concentrations so that $\int_0^d f(z)\,\mathrm{d}z = 1$. Suppose that the intensities are measured at angles θ_i, $i = 1, \cdots, M$. Let p_i correspond to angle θ_i and let F_i^δ correspond to the noisy data at θ_i with the noise level described by the parameter δ. The F_i^δ are derived from detected counts $P_{\text{detector}}(\theta_i)$ by making use of (2.5) and (2.7).

The maximum-entropy method is formulated as a constrained optimisation problem $\mathcal{P}$ in which the negative of the entropy functional is minimised over the set of normalised non-negative functions in $L^1(0,d)$:

$$(\mathcal{P}) \qquad \text{minimize } f \mapsto \int_0^d f(z)\log f(z)\,\mathrm{d}z, \tag{3.1}$$

subject to:

$$f \in L^1(0,d), \quad f(z) \geq 0 \quad \text{for almost all } z \in [0,d],$$

and

$$\int_0^d f(z)\mathrm{d}z = 1,$$

and

$$\frac{1}{M}\sum_{i=1}^{M} w_i |\mathcal{L}(f)(p_i) - F_i^\delta|^2 \leq \epsilon,$$

where the w_i are positive weigths with total sum 1:

$$\frac{1}{m}\sum_{i=1}^{m} w_i = 1, \tag{3.2}$$

and $\epsilon > 0$ is the regularisation parameter. By choosing the w_i appropriately, we can give additional weights to those angles for which the measurements are more accurate, *i.e.*, for which the number of counts is relatively large. We choose

$$w_i \propto P_{\text{detector}}(\theta_i)\, t_{\theta_i}/(F_i^{\delta})^2. \tag{3.3}$$

Since $1/[P_{\text{detector}}(\theta_i)\, t_{\theta_i}]^{1/2}$ is the relative error of the measurement at angle θ_i, w_i as defined by (3.3) is proportional to the reciprocal square of the absolute error at that angle. When exact data are used in simulations (*i.e.*, no noise is added), we choose all weigths w_i to be equal and ϵ to be very small. For stability the regularisation parameter ϵ must satisfy

$$\epsilon \geq \frac{1}{m} \sum_{i}^{m} w_i |F(p_i) - F_i^{\delta}|^2, \tag{3.4}$$

where $F(p_i)$ corresponds to the exact data at angle θ_i.

After Jaynes [11], [12] had stressed the importance of information-theoretic methods for ill-posed problems in statistical mechanics, the maximum-entropy method was applied as a general method for solving ill-posed problems in many different fields [13]. In our problem, the differential form $f(z)\,\mathrm{d}z$ in (3.2) can be interpreted as the probability of finding a fluorescence atom in the layer of infinitesimal thickness $\mathrm{d}z$ at depth z. Given the measurements, the solution to (3.2) is the most likely distribution of fluorescence atoms when *a priori* all depths are considered equally probable. Accordingly, in the absence of measurements, the solution of problem (3.2) is the constant profile.

When the *a priori* probability of a fluorescence atom being at a certain depth is a known function of depth, the entropy functional can be modified to incorporate this knowledge. In fact, when $\phi(z)$ is the *a priori* probability function, one should take as the entropy functional

$$\int_0^d \left\{ f(z) \log \frac{f(z)}{\phi(z)} + \phi(z) - f(z) \right\} \mathrm{d}z. \tag{3.5}$$

The solution without measurements is then equal to $\phi(z)$. One could apply this more general entropy functional, for instance, when GEXRF is used to improve a concentration profile that has been estimated initially using another measurent technique. Alternatively, *a priori* knowledge about the concentration profile can be incorporated into the problem by adding more contraints. For instance, when the widths of the slits would be increased to increase the signal-to-noise ratio, it may be necessary to replace the multiplicative instrumental factor (2.5) by a convolution integral. This kind of modification of the measurement procedure means that the Laplace transform is replaced by some other linear operator. For the maximum-entropy method this means that the expression for $\mathcal{L}(f)$ in the constraint in problem (3.2) must be replaced by the modified expression. However, for the SVD method described in Section 5, the consequences are more far-reaching because the eigenfunctions of the new operator in general are not known analytically. This would make this method much less actractive.

In the present paper we will not consider the incorporation of *a priori* knowledge, however important that may be for the stabilisation of the inversion, but study instead the possibility of reconstructing concentration profiles without using any *a priori* knowledge.

4. Some mathematical properties of the maximum-entropy method

Optimisation problem (3.2) has some nice properties. The function

$$s \mapsto s \log s, \tag{4.1}$$

is strictly convex for $s \geq 0$ and has a minimum of $-1/e$ which is attained for $s = 1/e$. For convenience we define $s \log s = +\infty$ for $s < 0$. It is then clear that for every measurable function f (positive or not) the entropy functional

$$\mathcal{F}(f) = \int_0^d f(z) \log f(z) \, \mathrm{d}z, \tag{4.2}$$

has a meaning, finite or infinite. It was shown in [14] that $f \mapsto \mathcal{F}(f)$ is a strictly convex functional on the domain $\{f \in L^1(0, d); \quad \mathcal{F}(f) < \infty\}$. Hence $\mathcal{F}$ is lower-semicontinuous with respect to the weak topology on $L^1(0, d)$ inherited from the dual $L^\infty(0, d)$. The functional $\mathcal{F}$ has the following property:

Proposition 1 *Suppose that $f_n \to f$ in the weak topology of $L^1(0, d)$ and suppose that $\lim_{n\to\infty} \mathcal{F}(f_n) = \mathcal{F}(f) < \infty$. Then $\lim_{n\to\infty} \| f_n - f \|_1 = 0$.*
Here $\|.\|_1$ is the norm on $L^1(0, d)$. This important property was proved in [15].

The following existence result applies to optimisation problem (3.2).

Theorem 1 *If the feasible set of problem (3.2) is not empty (i.e. if ϵ is sufficiently large), the optimisation problem (3.2) has a solution and the solution is unique.*

This theorem was proved in [15] and [16] using the weak topology on $L^1(0, d)$. Since the functional $\mathcal{F}$ is lower semicontinuous with respect to the weak topology on $L^1(0, d)$, the theorem would follow immediately if it could be proved that the feasible set of (3.2) or some equivalent set, were sequentially compact with respect to the weak topology. In fact, in that case a minimising sequence would have a weakly converging subsequence which converges to the solution of the optimisation problem. The feasible set in (3.2) is convex and closed, and therefore it is weakly closed, but unfortunately it is *not* weakly compact. However, by using a theorem of Dunford-Pettis-De La Vallée-Poussin [17, pp. 239] it was proved in [15], [16] that, for every constant $C > 0$, the set defined by

$$\mathbf{A}_C = \{f \in L^1(0, d); \quad \mathcal{F}(f) \leq C\}, \tag{4.3}$$

is weakly sequentially compact in $L^1(0, d)$. Hence, it follows that the intersection of $\mathbf{A}_C$ and the feasible set of (3.2) is weakly sequentially compact. Since one may clearly replace the feasible set of (3.2) by its intersection with $\mathbf{A}_C$ for C large enough, the lower semi-continuity of the object functional implies the existence of a solution of the optimisation problem. The fact that this functional is strictly convex implies uniqueness of the solution.

In order to clarify the connection between the solutions of the maximum-entropy method and the actual profile, we will prove below that, when the number of angles is increased, the errors in the measurements are decreased and the regularisation parameter is decreased in an appropriate way; the solutions of the optimisation problems converge in the $L^1(0, d)$-norm to the actual concentration. To be more precise, assume that superscript m refers to the maximum-entropy method when m angles θ_i^m $i = 1, \cdots, m$ are used. Let $p_1^m < p_2^m < \cdots < p_m^m$ be the values of p corresponding to the angles θ_i^m at which noisy data F_i^m, $i = 1, \cdots, m$ have been measured. Furthermore, let w_i^m be the weights used (not necessarily given by (3.4)) and let ϵ^m be the regularisation parameter. Let f^m be the solution of optimisation problem

(3.2) corresponding to these choices. Finally, let $F^\infty(p)$ be the exact data and let f^∞ be the exact concentration profile so that

$$\mathcal{L}(f^\infty) = F^\infty. \tag{4.4}$$

We define the noise level δ^m in the mth data set F^m by

$$\delta^m = \frac{1}{m}\sum_{i=1}^{m} w_i^m |F_i^m - F^\infty(p_i^m)|^2. \tag{4.5}$$

We assume furthermore that for some $p_{\min} \leq \tilde{p}_{\min} < \tilde{p}_{\max} \leq p_{\max}$ we have

$$p_1^m = \tilde{p}_{\min} \text{ and } p_m^m = \tilde{p}_{\max}, \quad \text{for all } m, \tag{4.6}$$

and that there exist constants $C_1 > 1$ and $C_2 > 0$ such that for all m we have

$$\frac{1}{C_1 m} \leq p_{i+1}^m - p_i^m \leq \frac{C_1}{m}, \text{ for all } i = 1, \cdots, m-1, \tag{4.7}$$

and

$$w_i^m \geq C_2, \text{ for all } i = 1, \cdots, m. \tag{4.8}$$

Note that property (4.7) means that, in the limit $m \to \infty$, the angles for which measured data are available become equidistributed. Furthermore, if we would choose the w_i^m according to (3.4), property (4.8) means that no angle exists for which the absolute error in the measurement can increase beyond all bounds.

Theorem 2 *When*

$$\lim_{m\to\infty} \delta^m = 0 \text{ and } \lim_{m\to\infty} \epsilon^m = 0, \tag{4.9}$$

such that $\epsilon^m \geq \delta^m$ for all m, the solutions of the maximum-entropy method converge to the exact concentration profile in the L^1 norm:

$$\lim_{m\to\infty} \int_0^d |f^m - f^\infty| \, \mathrm{d}z = 0. \tag{4.10}$$

The theorem is proved in the Appendix.

Convergence in the L^1 norm is, of course, not a very strong statement. In practice one would like to have a stronger result, *e.g.*, convergence in the maximum norm when the real profile f^∞ is continuous, but it seems that this has not been proved so far.

In the numerical computations that will be described in Section 6 we will compare the profiles that we reconstruct using the maximum-entropy method with the chosen real profiles f^∞. In addition, we compare the performance of the maximum-entropy method with another regularisation method based on the spectral decomposition of the Laplace transform. For completeness we briefly describe this regularisation method in the following section.

5. Reconstruction using the regularised inversion of the Laplace transform

Let, as before, f^∞ be the real profile and F^∞ be the Laplace transform of f^∞:

$$\mathcal{L}(f^{\infty}) = F^{\infty}. \tag{5.1}$$

Let F^{δ} denote a noisy finite data set, with δ a value for the level of the noise. The set of points p_i corresponding to the data is contained in the interval $[\tilde{p}_{\min}, \tilde{p}_{\max}]$ and includes $\tilde{p}_{\min}$ and $\tilde{p}_{\max}$. In the reconstruction method based on the regularised inversion of the Laplace transform the data set is first interpolated linearly inside the interval $(\tilde{p}_{\min}, \tilde{p}_{\max})$ and extrapolated to the entire half line $p \geq 0$. For the extrapolation to $0 \leq p \leq \tilde{p}_{\min}$ the coefficients a_j in

$$F^{\delta}(p) = \sum_{j=0}^{N} a_j p^j, \tag{5.2}$$

are determined by fitting the given values of F^{δ} in the interval $[\tilde{p}_{\min}, \tilde{p}_{\max}]$. For $p > \tilde{p}_{\max}$ use is made of the formula

$$F^{\delta} = \sum_{j=1}^{M} \frac{b_j}{p^j}, \tag{5.3}$$

where the b_j are again determined by fitting the data in the interval $[\tilde{p}_{\min}, \tilde{p}_{\max}]$. The choices for N and in particular for M are crucial here. Too small values give inaccurate extrapolations; too large values cause instabilities due to the noise. In practice $N, M \approx 4$ is appropriate.

Once the extrapolations have been performed, the profile is reconstructed by use of a regularised version of formula (2.16) for the inverse Laplace transform:

$$f_{\rho}^{\delta} = \int_{-\sqrt{\pi}}^{\sqrt{\pi}} \frac{\lambda}{\lambda^2 + \rho} (F^{\delta}, e_{\lambda})\, e_{\lambda}\, \mathrm{d}\lambda, \tag{5.4}$$

where $\rho > 0$ is the regularisation parameter. The function f_{ρ}^{δ} is an approximation of the profile f^{∞}, obtained by the reconstruction (*i.e.*, the extrapolation of the data and the application of formula (5.4)). The regularisation parameter must not be chosen too large to ensure that the reconstructed profile does not differ too much from the actual profile and it must not be chosen too small to prevent instabilities due to noise in the data. Morozov's discrepancy method [18, pp. 87–94] and generalised cross validation (GCV) [19, Chapter 4] are two methods by which ρ can be determined automatically from the noisy data. Both methods have been implemented for application in GEXRF and it was found that they give similar results.

Because the eigenfunctions $e_{\lambda}(z)$ become increasingly oscillatory for $\lambda \to 0$, the numerical integrations to compute the scalar products $(F^{\delta}, e_{\lambda})$ and the integrals (5.4) should be done such that the error made is uniformly small, irrespective of the period of the oscillations (for details we refer to [20]). Provided the numerical integrations are done in this way, the described method is accurate and very fast because the extrapolations require only small computational efforts and because the inversion fromula (5.4) does not require a matrix inversion. The main disadvantage of the method is that it can only be applied to the unconstrained inversion problem. For more details about the application of this method to GEXRF we refer to [20].

6. Numerical results

In the numerical computation of the solution of the maximum-entropy method the concentration profiles are approximated by piecewise linear functions. The grid that we used consisted typically of 60 equidistant points z_i. The optimisation problem was then solved numerically in

the 60-dimensional space of piecewise linear functions. This requires a lot of computing power ($\approx$40 m. of CPU on a HP N4000/540) and the convergence of the numerical optimisation is sometimes rather slow. But eventually the solution of the optimisation problem is always obtained, irrespective of the initial guess for the profile. The reason is that the object functional is strictly convex and therefore local minima do not exist.

We assume that the laboratory GEXRF instrument described in [5] is used in the simulated experiments. However, the x-ray tube is assumed to have a rotating Rh anode by which the emitted power is increased to 20 kW. The two slits of the collimator are 20 mm wide in the direction perpendicular to the plane of drawing in Figure 1. Their distance is 123·6 mm and the minimum width of the slit in the plane of drawing is 10 μm, which corresponds to a minimum divergence of $\Delta\theta_{\min} = 0{\cdot}16$ mrad. The minimum step size in the detection angle is 0·05 mrad, which is consequently well below the minimum divergence.

In the examples that we will discuss below the dopant is arsenic (As). We simulated the dopant concentration profiles of As using the package TSUPREM-4 (Avant! Corporation, Fremont, Ca, U.S.A). With this package the ion implantation process and subsequent nonlinear thermal diffusion can be simulated under various experimental conditions. The fluorescence intensities used as data correspond to the Lα line with a wavelength of 9·671 Å. At this wavelength we have in silicon $\beta = 1{\cdot}4475 \times 10^{-5}$ and $\delta = 2{\cdot}63 \times 10^{-4}$. Using (2.4) we conclude that the critical angle is 22·9 mrad. It was found to be sufficient to restrict the detection angles to 0 mrad $< \theta <$50 mrad (larger angles do not contain additional information of the profile). For equidistant detection angles the maximum number of angles corresponding to a given divergence $\Delta\theta$ is thus given by

$$N_\theta = 0 \cdot 05/\Delta\theta. \tag{6.1}$$

Choosing more detection angles would amount to oversampling the data. For the minimum slit width of 10 μm, Equation (6.1) implies $N_\theta = 312$.

The concentration profile obtained with TSUPREM-4 when 2×10^{15} As atoms/cm^2 of energy 8 keV are implanted into silicon and the implantation is followed by thermal diffusion at 1000 °C for 20 s is given by the continuous curve shown in Figure 7. The maximum concentration is only $7{\cdot}6 \times 10^{20}$ atoms/cm^3 at a depth of 100 Å. The simulated fluorescence intensities at the Lα line are correspondingly low. They are shown in Figure 4. The slit width was assumed to be $w = 24{\cdot}7$ μm, which corresponds to a divergence of $\Delta\theta = 0{\cdot}4$ mrad. The maximum number of equidistant angles of detection given by Equation (6.1) is then 125.

Also shown in Figure 7 are reconstructed profiles that we obtained using exact data $P_{\text{detector}}(\theta)$ (*i.e.*, without noise) at 120 equidistant angles between $\theta_{\min} = 1$ mrad and $\theta_{\max} = 50$ mrad. The results of both the maximum-entropy method and Tikhonov's regularised inversion (GCV) are shown. The GCV method yields for the regularisation parameter $\rho_{\text{GCV}} = 9{\cdot}3 \times 10^{-5}$ (the regularisation parameter is non-zero due to small numerical and rounding errors and due to errors caused by the extrapolations). The order of the extrapolation was 3 for both large and small angles. The weights $\tilde{w}_i$ in the inequality constraint in (3.2) of the maximum-entropy method were all chosen to be equal to 1 and ϵ was chosen to be very small, namely

$$\epsilon = 0{\cdot}5 \times 10^{-7} \times \frac{1}{m} \sum_{i=1}^{m} |F_i^\delta|^2 \tag{6.2}$$

It can be seen in Figure 7 that the profile we obtained using the maximum-entropy method differs slightly from the profile found by means of Tikhonov's method. The former is better at

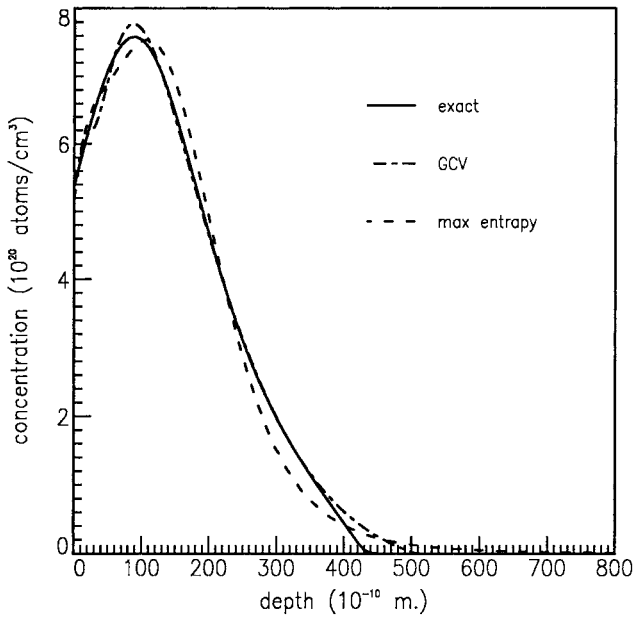

Figure 7. For the same concentration profile as in Figure 3, the dashed curves shown are obtained by inversion of the fluorescence data *without noise being adding* at 120 equidistant angles using the maximum-entropy method and Tikhonov's method with the regularization parameter determined by GCV. The divergence admitted by the two slits is again $\Delta\theta = 0{\cdot}4$ mrad.

depths above 100 Å where the latter slowly oscillates. The maxium value of the solution obtained by the maximum-entropy method is slightly too high. The result obtained with Tikhonov's method can be improved further by using $M = 6$ in extrapolation (5.3). The extrapolation of the data to small angles is much more critical than that to large angles. The reconstructed profile found with the maximum-entropy method can be improved somewhat by using more data points (angles).

Next we will consider the inversion of data to which Poisson noise was added. The concentration profile of Figure 3 was reconstructed using the the maximum-entropy method and using Tikhonov's method with data to which Poisson noise was added corresponding to a measurement time of 180 s per angle. This means that the mean relative error (computed over all angles) in the data was 0·74%. In this case the weights w_i were chosen to be proportional to the reciprocal of the square of the absolute error in the data (see (3.4)). Regularisation parameter ϵ was chosen as follows

$$\epsilon = \frac{1}{m}\sum_{i=1}^{m} w_i |\Delta F_i|^2, \tag{6.3}$$

where ΔF_i is the estimated absolute error in the F_i. In Tikhonov's inversion, Morozov's discrepancy method and the GCV method gave similar values for regularisation parameter ρ ($0{\cdot}68 \times 10^{-3}$ and $0{\cdot}18 \times 10^{-2}$, respectively) and the reconstructed concentration profiles are similar. The reconstructed profile shown in Figure 3 obtained using the maximum-entropy method shows a maximum concentration that is too high. The solutions obtained by Tikhonov's method show an oscillation in the first 50 Å. For this profile, the maximum-entropy method turned out to be more sensitive to noise than Tikhonov's method. Note that, although

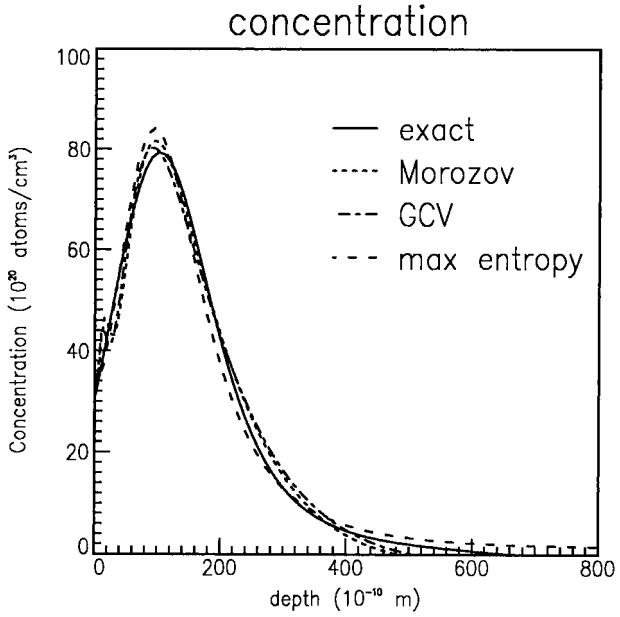

Figure 8. The concentration profile computed by TSUPREM-4 (continuous curve) and the profiles obtained through regularised inversions of noisy fluorescence data at 120 equidistant angles using the maximum-entropy method and Tikhonov's inversion with both Morozov's principle and GCV (dashed curves) to determine the regularization parameter. The divergence was again $\Delta\theta = 0{\cdot}4$ mrad. The concentration profile was obtained using an implantation dose of 2×10^{16} atoms/cm^2 at 8 keV, followed by thermal diffusion at 1000 °C for 1 minute. The measurement time was 900 s. per angle, which corresponds to a mean relative error in the data of 0·25%.

in Tikhonov's method it is not required explicitly that the concentration must be non-negative; negative values were never obtained.

For a dose of 2×10^{16} atoms/cm^2, and an energy that is again 8 keV and annealing at 1000 °C for 1 minute, TSUPREM-4 predicts the profile shown as the continuous curve in Figure 8. Again Poisson noise was added to the simulated fluorescence intensities at the 120 equidistant angles corresponding to a measurement time of 900 s per angle. Since the concentrations are higher than in the previous cases, the fluorescence intensities are higher, and therefore the relative errors for the same measurement time are smaller. The mean of the relative errors in the data is now 0·25%.

The reconstructed profile of the maximum-entropy method in particular is now much better than for the profile of Figure 3. The values for the regularisation parameter for Tikhonov's method were $\rho = 0{\cdot}10 \times 10^{-3}$ using Morozov's method and $\rho = 0{\cdot}27 \times 10^{-3}$ using GCV. The primary reason why the reconstruction was better in this case than in the case of Figure 3 is, of course, that the concentration is higher, and hence the noise is lower for the same measurement time. But also for the same mean relative error in the data, the profile shown in Figure 8 is easier to reconstruct than that of Figure 3. It is less steep because of the longer annealing time, and therefore the coefficients corresponding to small eigenvalues in the expansion of the profile on the basis of eigenfunctions of the Laplace-transform operator have less weight than in the case of Figure 3.

Since for Poisson noise the relative error in the data decreases in proportion to only the square root of the total counts, increasing measurement times and/or the source strength gives only a small improvement in the reconstructed profiles. One can use simulations to compute

the strength of the x-ray source that would be necessary to achieve a prescribed accuracy in the reconstruction of a given profile using a desired measurement time per angle. For example, when it is required that the measurement time be 30 s for all angles, so that for 120 angles the total measurement time is 1 hour, the required intensity should be approximately 200 times larger than that yielded by a conventional x-ray tube with a rotating anode. Therefore, in actual experiments a conventional x-ray tube can in general not be used and a high-power source, such as a synchrotron, is needed. To what extent *a priori* knowledge about the profile could be traded off against a lower exposure dose remains a subject for future investigation.

The accurate reconstruction of the profiles for sufficiently high signal-to-noise ratios is reason for the cautious conjecture that in problems of chemical analysis, where concentration levels are not low, shallow concentration profiles can be determined accurately using GEXRF. Of course, when the concentrations are not low, the refractive index will in general be a function of the unknown concentration and the emitted fluorescence intensity will consequently depend nonlinearly on the concentration. However, to obtain the emitted intensity, it will in general not suffice merely to substitute the relationship between the refractive index n and the unknown concentration f in formula (2.2). In fact, a more drastic modification of (2.2) will be necessary, because the radiation is scattered by the inhomogeneous refractive index. It seems nevertheless conceivable that the ill-posedness of the problem will still be caused mainly by the absorption of radiation in the sample, *i.e.*, by the exponential factor occuring in (2.2). The conclusion drawn for the case of low concentrations, namely that the reconstruction of concentration profiles is possible for sufficiently high signal-to-noise ratio, therefore leads us to expect that shallow profiles with normal concentration levels can be reconstructed from GEXRF data obtained with a conventional laboratory x-ray tube as source.

7. Conclusions

The determination of concentration profiles of a dopant in silicon using an angle scan of noisy x-ray intensities obtained by Grazing Emission X-Ray Fluorescence Spectroscopy is an ill-posed problem. It can be formulated as a problem of the inversion of the Laplace transform for incomplete and noisy data. We have formulated and applied the maximum-entropy method to this inversion. The maximum-entropy method is a flexible method which allows us to change the constraints for the concentration profiles easily. Hence, *a priori* knowledge about the profile can be incorporated.

Some properties of the maximum-entropy method were discussed. The maximum-entropy method leads to a convex constrained optimization problem which has a unique solution. It was shown that the solutions of the discretised optimization problems always converge in the sense of L^1 to the actual concentration profile, provided that the noise is reduced and the number of data is increased.

Numerical experiments using simulated data to which realistic values of Poisson noise were added showed that the reconstruction of the profile is possible without *a priori* knowledge when the signal-to-noise level is sufficiently high. The noise levels are required to be so small and the measurement times, when using a conventional x-ray tube are consequently so long, that the necessary stability of the spectrometer could not be maintained. Hence, for the low concentration values considered in this paper, synchroton radiation will have to be used. On the other hand, for problems with higher concentration values, the signal-to-noise ratio may be sufficiently high, so that a conventional x-ray source might well be sufficient. But in

that case the unknown concentration profile influences the refractive index of the sample and hence the inverse problem becomes nonlinear. It is reasonable to expect, however, that the sensitivity to noise will not be worse in the nonlinear case than in the linear one.

Acknowledgments

The authors would like to thank Mr. J.J. Rusch of Philips Research Laboratory for his extensive help with the optimisation programme for the maximum-entropy method. Furthermore, we are grateful to Dr. P.K. de Bokx and Dr. P.A. Stolk of the same laboratory for discussions on experimental aspects of GEXRF and for running the simulations with TSUPREM-4, respectively.

Appendix A: Proof of Theorem 2

In the proof Proposition 1 will be essential. We define the quadratic forms

$$\mathcal{G}^m(F) = \frac{1}{m}\sum_{i=1}^{m} w_i^m \, |F(p_i^m)|^2, \tag{A1}$$

and denote the feasible set of the mth optimisation problem by $\mathbf{A}^m$:

$$\begin{aligned}\mathbf{A}^m = \{f \in L^1(0,d); \quad f(z) \geq 0 \quad a.e., \quad \mathcal{F}(f) < \infty, \quad \textstyle\int_0^d f(z)\,\mathrm{d}z = 1,\\ \mathcal{G}^m(\mathcal{L}(f) - F^m) \leq \epsilon^m\}.\end{aligned} \tag{A2}$$

Definition (4.5) of δ^m implies that

$$\mathcal{G}^m(\mathcal{L}(f^\infty) - F^m) = \delta^m \leq \epsilon^m, \tag{A3}$$

hence the exact profile is in the feasible sets of all optimisation problems:

$$f^\infty \in \mathbf{A}^m, \quad \text{for all } m. \tag{A4}$$

Therefore

$$\mathcal{F}(f^m) \leq \mathcal{F}(f^\infty), \tag{A5}$$

which implies that the sequence $\{\mathcal{F}(f^m)\}$ is bounded. By the relative compactness of the set $\mathbf{A}_C$ mentioned above, there exists a subsequence of $\{f^m\}$, which we will continue to write as $\{f^m\}$, which convergences weakly in $L^1(0,d)$, *i.e.*, there exists $g \in L^1(0,d)$ such that

$$\lim_{m\to\infty} f^m = g, \quad \text{weakly in } L^1(0,d). \tag{A6}$$

We will show that $g = f^\infty$. The weak convergence (A6) implies that

$$\lim_{m\to\infty} \mathcal{L}(f^m)(p) = \mathcal{L}(g)(p), \text{ for all } p \geq 0. \tag{A7}$$

Because the weakly converging sequence $\{f^m\}$ is bounded in the norm of $L^1(0,d)$, the corresponding sequence of Laplace transforms $\{\mathcal{L}(f^m)\}$ is an equicontinuous uniformly bounded sequence of functions on $[0,\infty)$. Hence, by the Ascoli-Arzelá theorem, there exists a subsequence, which we again will write as $\{f^m\}$, such that on a compact set $\{\mathcal{L}(f^m)\}$ converges uniformly to $\mathcal{L}(g)$. Now for $F, \tilde{F} \in \mathcal{C}[\tilde{p}_{\min}, \tilde{p}_{\max}]$ we have

$$\begin{aligned}
&\left|\mathcal{G}^m(F - F^\infty) - \mathcal{G}^m(\tilde{F} - F^\infty)\right| \\
&\quad \le \frac{1}{m}\sum_{i=1}^{m} w_i^m \left|F(p_i^m) + \tilde{F}(p_i^m) - 2F^\infty(p_i^m)\right| \left|F(p_i^m) - \tilde{F}(p_i^m)\right| \\
&\quad \le \left(\|F - F^\infty\|_\infty + \|\tilde{F} - F^\infty\|_\infty\right) \|F - \tilde{F}\|_\infty.
\end{aligned} \tag{A8}$$

Here and in the following the norm $\|F\|_\infty$ is the maximum norm on the space $\mathcal{C}[\tilde{p}_{\min}, \tilde{p}_{\max}]$. By substituting $F = \mathcal{L}(f^m)$ and $\tilde{F} = \mathcal{L}(g)$ we obtain

$$\begin{aligned}
&\left|\mathcal{G}^m(\mathcal{L}(f^m) - F^\infty) - \mathcal{G}^m(\mathcal{L}(g) - F^\infty)\right| \\
&\quad \le (\|f^m\|_1 + \|g\|_1 + 2\|F^\infty\|_\infty) \, \|\mathcal{L}(f^m) - \mathcal{L}(g)\|_\infty,
\end{aligned} \tag{A9}$$

where we used $\|\mathcal{L}(f)\|_\infty \le \|f\|_1$ for $f \in L^1(0, d)$. As a weakly converging sequence, $\{f^m\}$ is bounded with respect to the L^1-norm and because $\{\mathcal{L}(f^m)\}$ converges uniformly to $\mathcal{L}(g)$ on the interval $[\tilde{p}_{\min}, \tilde{p}_{\max}]$, it follows from (A9) that

$$\lim_{m\to\infty} \left[\mathcal{G}^m(\mathcal{L}(f^m) - F^\infty) - \mathcal{G}^m(\mathcal{L}(g) - F^\infty)\right] = 0. \tag{A10}$$

But

$$\begin{aligned}
\lim_{m\to\infty} \mathcal{G}^m(\mathcal{L}(f^m) - F^\infty) &\le 2 \lim_{m\to\infty} \mathcal{G}^m(\mathcal{L}(f^m) - F^m) + 2 \lim_{m\to\infty} \mathcal{G}^m(F^m - F^\infty) \\
&\le 2 \lim_{m\to\infty} \epsilon^m + 2 \lim_{m\to\infty} \delta^m = 0,
\end{aligned} \tag{A11}$$

and hence also

$$\lim_{m\to\infty} \mathcal{G}^m(\mathcal{L}(g) - F^\infty) = 0. \tag{A12}$$

Then, using (4.7), (4.8) and (A12), we have

$$\begin{aligned}
&\lim_{m\to\infty} \sum_{i=1}^{m-1} [\mathcal{L}(g)(p_i^m) - F^\infty(p_i^m)]^2 (p_{i+1}^m - p_i^m) = \\
&\qquad \le \frac{C_1}{C_2} \lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} w_i^m [\mathcal{L}(g)(p_i^m) - F^\infty(p_i^m)]^2 \\
&\qquad = \frac{C_1}{C_2} \lim_{m\to\infty} \mathcal{G}^m(\mathcal{L}(g) - F^\infty) = 0.
\end{aligned} \tag{A13}$$

Now property (4.7) implies for every function F which is continuously differentiable on $[\tilde{p}_{\min}, \tilde{p}_{\max}]$:

$$\begin{aligned}
\left|\int_{\tilde{p}_{\min}}^{\tilde{p}_{\max}} F(p)\,\mathrm{d}p - \sum_{i=1}^{m-1} F(p_i^m)\,(p_{i+1}^m - p_i^m)\right| &\le \sum_{i=1}^{m-1} \int_{p_i^m}^{p_{i+1}^m} \left|F(p) - F(p_i^m)\right| \mathrm{d}p \\
&\le \|F'\|_\infty \sum_{i=1}^{m-1} \frac{1}{2} \left(p_{i+1}^m - p_i^m\right)^2 \le \frac{C_1^2}{2m} \|F'\|_\infty.
\end{aligned} \tag{A14}$$

Hence for continuously differentiable F:

$$\int_{\tilde{p}_{\min}}^{\tilde{p}_{\max}} F(p)\,\mathrm{d}p = \lim_{m\to\infty} F(p_i^m)(p_{i+1}^m - p_i^m). \tag{A15}$$

Apply this to $F = [\mathcal{L}(g) - F^\infty]^2$, and use (A13)) to conclude that

$$\int_{\tilde{p}_{\min}}^{\tilde{p}_{\max}} \left[\mathcal{L}(g)(p) - F^\infty(p)\right]^2\,\mathrm{d}p = 0. \tag{A16}$$

Hence $\mathcal{L}(g) = F^\infty$ on $[\tilde{p}_{\min}, \tilde{p}_{\max}]$ and by the uniqueness of the Laplace transform it follows that

$$g = f^\infty. \tag{A17}$$

Hence, $\lim_{m\to\infty} f^m = f^\infty$ weakly in $L^1(0, d)$ and because $f \mapsto \mathcal{F}(f)$ is lower semicontinuous with respect to the weak topology, we get

$$\mathcal{F}(f^\infty) \le \liminf_{m\to\infty} \mathcal{F}(f^m) \le \limsup_{m\to\infty} \mathcal{F}(f^m) \le \mathcal{F}(f^\infty), \tag{A18}$$

where the last inequality follows from (A5). Hence

$$\lim_{m\to\infty} \mathcal{F}(f^m) = \mathcal{F}(f^\infty). \tag{A19}$$

Then Proposition 1 implies that

$$\lim_{m\to\infty} \|f^m - f^\infty\|_1 = 0. \tag{A20}$$

We conclude therefore that the sequence $\{f^m\}$, and also every subsequence of $\{f^m\}$, has a subsequence which converges in $L^1(0, d)$. Hence, the sequence $\{f^m\}$ itself converges in $L^1(0, d)$. This completes the proof.

References

1. R.S. Becker, J.A. Golovchenko and J.R. Patel, X-ray evanescent-wave absorption and emission. *Phys. Rev. Lett.* 50 (1983) 153–156.
2. Y.C. Sasaki and K. Hirokawa, Depth profile measurement by using a refracted x-ray fluorescence method. *Appl. Phys. Lett.* 58 (1991) 1384–1386.
3. S. Hasegawa, Y. Ino, and H. Daimon, Chemical analysis of surfaces by total-reflection-angle x-ray spectroscopy in RHEED experiments, *Jpn. J. Appl. Phys.* 24 (1985) L387–L398.
4. H. Schwenke, J. Knoth, L. Fabry, S. Pahlke, R. Scholz and L. Frey, Measurement of shallow arsenic impurity profiles in semiconductor silicon using time-of-flight secondary ion mass spectrometry and total reflection x-ray fluorescence spectrometry. *J. Electrochem. Soc.* 144 (1997) 3979–3983.
5. P.K. de Bokx and H.P. Urbach, Laboratory grazing-emission x-ray fluorescence spectrometer. *Rev. Sci. Instrum.* 66 (1995) 15–19.
6. H.P. Urbach and P.K. de Bokx, Calculation of intensities in grazing-emission x-ray fluorescence, *Phys. Rev. B* 53 (1996) 3752–3763.
7. H.P. Urbach and P.K. de Bokx, Grazing Emission x-ray fluorescence from multilayers. *Phys. Rev. B* 63 (2001) 1–17.
8. P.R. Chernoff, Spectral representation of the Laplace transform. *Am. Math. Monthly* 101 (1994) 366–367.
9. D.S. Gilliam, J.R. Schulenberger, and J.L. Lund, Spectral representation of the Laplace and Stieltjes transform. *Mat. Applic.* 7 (1998) 101–107.
10. A.O. Istratov and O.F. Vyvenko, Exponential analysis in physical phenomena. *Rev. Sci. Instr.* 70 (1999) 1233–1257.
11. E.T. Jaynes, Information theory and statistical mechanics. *Phys. Rev.* 106 (1957) 620–630.

12. E.T. Jaynes, Information theory and statistical mechanics II. *Phys. Rev.* 108 (1957) 171–190.
13. E.T. Jaynes, On the rationale of maximum-entropy methods. *Proc. IEEE* 70 (1982) 939–952.
14. R.T. Rockafellar, Integrals which are convex functionals, II. *Pacific J. Math.* 39 (1971) 439–469.
15. J.M. Borwein and A.S. Lewis, Convergence of best entropy estimates. *SIAM J. Optim.* 1 (1991) 191–205.
16. P.P.B. Eggermont, Maximum entropy regularization for Fredholm integral equations of the first kind. *SIAM J. Math. Anal.* 24 (1995) 1557–1576.
17. I. Ekeland and R. Teman, *Convex Analysis and Variational Problems.* Amsterdam: North-Holland (1976) 401 pp.
18. A.N. Tikhonov and V.Y. Arsenin, *Solutions of Ill-Posed Problems*. Washington: Winston and Sons (1977) 258 pp.
19. G. Wahba, *Spline Models for Observational Data*. Philadelphia: SIAM (1990) 169 pp.
20. C. Kok and H.P. Urbach, On the regularization of the inverse Laplace transform in grazing-emission x-ray fluorescence spectroscopy. *Inv. Probl. Eng.* 7 (1999) 433–470.