# Enhancement of the energy resolution in ion-beam experiments with the maximum-entropy method

R. Fischer, M. Mayer, W. von der Linden, and V. Dose
*Max-Planck-Institut für Plasmaphysik, EURATOM Association, Boltzmannstr. 2, D-85748 Garching, Germany*
(Phys. Rev. E **55**, 6667 (1997), received 22 January 1997)

In ion beam experiments with MeV ions, the energy resolution of the detector and the energy spread of the incident ion beam restrict the energy and depth resolution. It is possible to deconvolve the measured spectra if the energy transfer function of the apparatus is known. However, this leads to an ill-posed problem. The Bayesian probability theory allows one to tackle consistently such types of problems. Using the maximum-entropy prior, it is possible to obtain the probability distribution for the desired energy and depth profiles. An enhancement of the energy resolution by a factor of 8 in Rutherford backscattering analysis of thin films, resulting in an energy resolution of about 2.5 keV with semiconductor detectors, is obtained.

02.50.-r, 07.05.Kf, 79.20.Rf

## I. INTRODUCTION

In most ion beam experiments, semiconductor detectors such as silicon surface barrier detectors or particle-implanted and passivated silicon (PIPS) detectors are used for the detection and energy analysis of backscattered or recoiled particles. Semiconductor detectors are small, relatively cheap, and easy to use. The introduction of semiconductor detectors has made ion-beam analysis a powerful and widely used method in material analysis. However, the major disadvantage of semiconductor detectors is their limited energy resolution, typically of about 12–15 keV full width at half maximum (FWHM), thus limiting the mass and depth resolution in ion-beam experiments.

The apparatus-induced energy resolution in ion-beam experiments is determined by (a) the energy resolution of the detector; (b) the electronic noise of the detector-preamplifier system; (c) the energy spread of the incident beam; and (d) the kinematic energy spread due to geometry (solid angle of detection, beam spot size). The electronic noise contribution of the detector-preamplifier system is typically about 5 keV for a well-designed system, but can be lowered to about 0.5 keV with a cryogenic detector-preamplifier system [1]. The kinematic energy spread can be reduced by a small beam spot size and a small detector solid angle. A small solid angle, however, increases the fluence and time necessary for a measurement with sufficient statistics, so that some compromise has to be accepted. The energy spread of the incident beam is hard to determine, but can be assumed to be of the order of about $\Delta E/E \leq 5 \times 10^{-3}$.

The major contribution to the energy broadening visible in experimentally determined spectra is due to the semiconductor detector. For an ideal detector, the statistical distribution of the output pulses would have zero variance for incident monoenergetic particles with energy $E_0$. Due to electronic energy-loss straggling in the dead layer of the detector, thickness variations of the dead layer, and the statistics of electron-hole pair creation, the response function of a semiconductor detector always has nonzero variance [2]. The particle energy is statistically distributed between electronic excitation and ionization and nonelectronic energy losses by energy transfer to silicon atoms. Due to this statistical nature of electron-hole pair creation, the physical limit for the resolution of PIPS detectors without a dead layer would be about 6 keV FWHM for 3-MeV $^4$He particles [2]. For detectors with a dead layer, an energy resolution of about 8 keV FWHM was achieved [2], which is quite close to the physical limits.

A much better energy resolution of the order of $\Delta E/E = 5 \times 10^{-4}$ can be obtained with a magnetic spectrograph, such as the Munich Q3D magnetic spectrograph [3]. However, a magnetic spectrograph is large (in the range of meters) and therefore expensive.

The different contributions to the energy broadening can be summarized in a transfer function of the whole system: the apparatus function. The measured spectrum $\tilde{f}(E)$ is given by the convolution

$$\tilde{f}(E) = \int_{-\infty}^{\infty} f(E')\, A(E-E')\, dE', \qquad (1)$$

where $f(E)$ is the spectrum that would be measured with an ideal system with no intrinsic energy broadening and $A(E)$ is the apparatus function. If $f(E)$ and $A(E)$ are known, it is simple to calculate the measured spectrum $\tilde{f}(E)$. This convolution is performed by all programs for the simulation of backscattering spectra, such as RUMP [4]. Mostly, a Gaussian apparatus function is used. The variance of the Gaussian is adjusted for best fit to the measured spectrum.

The reverse and more interesting way, however, i.e., the determination of the deconvolved spectrum $f(E)$ from the measured spectrum $\tilde{f}(E)$ and the apparatus function $A(E)$, is not trivial at all. This problem is ill-posed, and some form of regularization is needed to recover $f(E)$. A recently published paper for resolution correction of Rutherford backscattering spectroscopy (RBS) spectra reviews the quality of a few methods for dealing with inversion problems [5]. A consistent probabilistic theory to obtain unbiased results from incomplete or noisy data is provided by the Bayesian probability theory. Often, it is called the maximum-entropy (ME) method, though the Bayesian calculus is the correct theory for logical inference and ME is important for assigning prior probabilities. The Bayesian probability theory, with the ME method, has become a powerful and widely used tool in image processing and data analysis during the past ten years [6–10]. We have used the ME method for the deconvolution of measured RBS spectra to improve the energy resolution.

## II. BAYESIAN PROBABILITY THEORY AND MAXIMUM ENTROPY

The direct inversion of the intrinsic spectrum $f(E)$ is an ill-posed problem if the eigenvalue spectrum of the apparatus transfer function $A$ varies over orders of magnitude. A commonly used approximation is a Gaussian transfer function, which illustrates the ill-posed nature of the problem. The direct inversion leads to meaningless results, since the statistical error of the experimental data is amplified by the inverse of the small eigenvalues, which results in artificial structures and even negative RBS intensities in the deconvolved spectrum. To overcome this problem and to separate the signal from the noise, the statistical nature of the error has to be taken into account properly. We want to find the most noncommittal reconstruction compatible with the data.

A self-consistent mathematical tool for this task is given by the Bayesian probability theory, which provides a general and consistent frame for logical inference. The Bayesian probability theory allows one to exploit any type of testable information, such as noisy experimental data, expectation values, positivity constraints, or other forms of prior knowledge. An important class of problems comprises the determination of positive and additive distribution (PAD) functions, such as the RBS spectrum. The most uninformative prior for a PAD is the entropic prior [11]. The Bayesian probability theory combined with the entropic prior is referred to as quantified maximum entropy (QME), which has been applied successfully to various data-analysis problems [6–10]. The mathematical and numerical details of the QME procedure are explained in great detail in [9]. Here, we will merely outline the key ideas of the QME procedure and introduce a recently developed advantageous extension of QME, namely, the adaptive kernel method.

The goal is to determine the *posterior probability density* $P(f|d,\sigma,I)$ for the RBS spectrum $f_j$ at the $N$ energies $E_j$, given $N_d$ experimental data $d_i$, the respective errors $\sigma_i$, and further prior knowledge summarized in $I$. The notation with the vertical bar denotes conditional probabilities, based on either empirical or theoretical information and further background information $I$. The posterior probability density represents all the information necessary to decide how reasonable a solution $f_j$ is. It allows one to determine many quantities of interest, such as the posterior mode resulting in the most probable solution, the mean $< f_j >= \int f_j P(f|d,\sigma,I)\, d^N f$, confidence intervals, etc. The Bayes theorem relates the as yet unknown $P(f|d,\sigma,I)$ to quantities that are known, namely, the *likelihood probability density* $P(d|f,\sigma,I)$ and the *prior probability density* $P(f|I)$, via

$$P(f|d,\sigma,I) = \frac{P(d|f,\sigma,I)\,P(f|I)}{P(d|I)}. \tag{2}$$

The historical terms "posterior" and "prior" have a logical, rather than a temporal, meaning. They simply mean "with" and "without" the new data taken into account. The Bayes theorem is a consequence of the two forms of the product rule $P(AB|C) = P(A|BC)P(B|C) = P(B|AC)P(A|C)$, where $A$, $B$, and $C$ are propositions. The theorem is fundamental to all scientific work, as it provides a formal rule for updating knowledge in the light of new data or learning from observations. In the Bayesian probability theory, *probabilities are not frequencies*, although frequency arguments are often important for assigning priors and frequency estimates can be derived from Bayesian probabilities.

The probability density $P(d|I)$ is a normalization constant because the problem is formulated for given data. The likelihood function describes the error statistics of the experiment. In the present case of a counting experiment with a large number of counts, we are dealing with a Gaussian likelihood function,

$$P(d|f,\sigma,I) = \frac{1}{\prod_{i=1}^{N_d}\sqrt{2\pi}\sigma_i} exp\left(-\frac{1}{2}\chi^2\right), \tag{3}$$

where (4)

$$\chi^2 = \sum_{i=1}^{N_d}\left(\frac{d_i - \sum_{j=1}^{N} A_{ij} f_j}{\sigma_i}\right)^2, \quad \sigma_i = \sqrt{d_i}. \tag{5}$$

The uninformative prior $P(f|I)$ for a PAD is the entropic prior

$$P(f|\alpha,I) = \frac{1}{Z}\exp\left(\alpha S\right) \tag{6}$$

$$S = \sum_{j=1}^{N} f_j - m_j - f_j \ln\left(\frac{f_j}{m_j}\right) \quad , \tag{7}$$

where $S$ is the information-theory entropy or information divergence relative to the default model $m_j$. The prior has to be normalized $Z = \int d^N f\, P(f|\alpha,I)$. We use an uninformative flat default model, $m_j = c$, where the most probable value for $c$ is given by minimizing $\chi^2$. The regularization parameter $\alpha$ is a nuisance parameter that has to be marginalized $P(f|I) = \int d\alpha\, P(f|\alpha,I)\,P(\alpha)$, where the uninformative prior for $\alpha$ is given by Jeffreys' prior $P(\alpha) \propto 1/\alpha$. A commonly used method of handling this improper prior is the evidence approximation with the steepest-descent method [11,12]. Recently we showed that for some inversion problems the correct marginalization of $\alpha$ in some sensible range, $[\alpha_{min}, \alpha_{max}]$, is compelling [10].

In addition to the properties of a PAD, the background knowledge $I$ summarizes model assumptions, such as the discretization scheme of the spectrum. The discretization scheme accounts for the smallest local resolution possible. The number of grid cells defines the number of degrees of freedom (DOF) of the image parameterization. To be sure not to lose information, a sufficiently fine grid has to be chosen. A disadvantage of a fine grid is the overfitting of the data noise. Accordingly, the grid has to be incorporated into the Bayesian analysis, either by choosing the best grid via model selection or by marginalization over all possible grids. It is expected that the solution has local smoothness and structures only if they are supported by the data. A smooth solution is expected when arbitrary permutation of the cells $i$ results in loss of information. This is definitely the case for RBS spectra or, in general, for any image. Alternatively, instead of searching for the optimal cell size distribution we may achieve smoothness on a fine grid by introducing correlations between neighboring cells. Though the DOF remain unchanged, the effective DOF (eDOF) are reduced. A measure of the eDOF is given below. Spatial smoothing as a regularizing technique is often applied to ill-posed inversion problems by adding a functional containing some derivative of the solution multiplied by a Lagrange parameter to the misfit $\chi^2$, e.g., $\chi^2 + \lambda \int f''(x)^2\, dx$. One has to find a criterion for choosing the strength of the regularization $\lambda$. But the existence of exactly one "optimal" $\lambda$ is not rigorous. In addition, the optimal $\lambda$ depends on the subdivision of the spectrum. For example, the strength of regularization $\lambda$ of a spectrum with an intense narrow peak and a weak broader peak differs from the result obtained if we join the regularized two spectra, each with one peak. In the Bayesian probability theory, such a regularization parameter is a nuisance parameter that has to be marginalized. Furthermore, instead of using a constant overall smoothing property, we need local smoothness, which allows a high smoothness level in unstructured regions of the spectrum where only background occurs, and a low smoothness level where structures arise. Local smoothness of the image $f$ is not included in standard ME methods because the entropic prior contains no correlation between the cells. An arbitrary permutation of the cells results in the same entropy, whereas the spectrum *looks quite different.*

We impose smoothness on $f$ through a convolution of a hidden density $h$ with a smoothing kernel functional $B$:

$$f(x) = \int dy\, B\left(\frac{x-y}{b(y)}\right) h(y). \tag{8}$$

Note that the local kernel width $b(y)$ varies with $y$. The shape and width distribution $b(y)$ of the kernel $B$ have to be determined by Bayesian methods. Knowing the normalization, the mean, and the standard deviation of $B$, the kernel functional form has to be Gaussian according to the ME principle [13]:

$$B\left(\frac{x-y}{b(y)}\right) = \frac{1}{\sqrt{2\pi}b(y)} \exp\left[-\frac{1}{2}\left(\frac{x-y}{b(y)}\right)^2\right]. \tag{9}$$

We do not expect the shape of the kernel functional to matter greatly, but in some applications the infinite range of a Gaussian may not apply. For example, if we expect sharp edges in an image, the prior information would favor a compact kernel, such as a truncated parabola [14].

In the Bayesian approach, the nuisance parameters $h$ and $b$ have to be marginalized:

$$P(f|d,\sigma,I) = \int d^N h\, d^N b\, P(f,h,b|d,\sigma,I) \tag{10}$$

$$\propto \int d^N h\, d^N b\, \delta(f - B*h) P(d|h,b,\sigma,I) \times P(h|b)\, P(b). \tag{11}$$

For the prior probability density $P(h|b)$, we use the entropic prior as was previously done for $f$ [15],

$$P(h|b,\alpha) = P(h|\alpha) = \left(\prod_i^N (\frac{\alpha}{2\pi h_i})^{\frac{1}{2}}\right) \exp(\alpha S(h)), \tag{12}$$

with the entropy

$$S(h) = \sum_i^N h_i - m_i - h_i \ln\left(\frac{h_i}{m_i}\right). \tag{13}$$

Note the measure $d^N h / \prod_i h_i^{1/2}$, which has to be applied to integrals of the probability over $h$ space [15]. The prior is independent of $b$ and ensures positivity in $f$. As shown above, the regularization parameter $\alpha$ has to be marginalized. For simplicity, we use the historic ME approach, where $\alpha$ is chosen such that $\chi^2/N_d \approx 1$ [16]. In our earlier papers, it is shown that historic ME provides quite the same regularization compared with the marginalization over $\alpha$, whereas the evidence approximation tends to over-fit the data [17]. At any rate, we observed that the importance of the prior is strongly reduced when using the adaptive kernel approach. The range over which $\alpha$ has to be varied to observe significant changes in the image exceeds by far the differences of $\alpha$ of the various approaches.

The prior probability $P(b)$ has to reflect all the information we have about the kernel widths $b$. We know the scale of $b$: the upper limit $b_u$ of each kernel width $b_i$ is given by the total image range (overall smoothing), and the lower limit $b_l$ is given by the value of $b_i$, which gives negligible contributions to neighboring image cells (no smoothing). In addition, we favor a blur width distribution where blur widths from neighboring cells differ by no more than one cell width. Otherwise the data would be artificially separated into background and signals, although in some cases this may be advantageous. This also enforces smooth variations in $b$. We chose the Gaussian prior for the derivative of $b$, measured in cell units, with a standard deviation of one for $b \in [b_l, b_u]$:

$$P(b) \propto \begin{cases} \exp(-0.5 \sum_i (b_i - b_{i-1})^2) & : \quad b \in [b_l, b_u] \\ 0 & : \quad \text{elsewhere.} \end{cases} \quad (14)$$

The Bayesian analysis yields the entire probability distribution (11), which is, however, somewhat complicated due to the presence of the $\delta$ function. This is of minor importance, since our interest usually focuses on expectation values of some functional of $f$, for example, the posterior mean $< f >$ and the posterior variance $< (\Delta f)^2 >$. Though the evaluation of the multidimensional integrals is cumbersome, this can be tackled by Monte Carlo importance sampling (see, for example, [18]). Alternatively, the different natures of $h$ and $b$ can be exploited and the integration over $h$ and $b$ can be split. The hidden image $h$ represents the signal intensity, whereas the set of kernel widths $b$ characterizes the complexity of the model. This can easily be seen in the limits $b \to \infty$ and $b \to 0$. For $b \to \infty$, the image is described solely by the mean of the density $h$ with an eDOF of one, whereas, for $b \to 0$, the kernel goes into the $\delta$ functional, which results in $f(x) = h(x)$. We obtain the conventional result where the image is reconstructed pointwise with an eDOF of $N$. Blur widths between the two limits result in local smoothing with locally varying eDOF.

A measure for the eDOF of the model parameters $b$ is given by

$$eDOF = \sum_i \sqrt{eigenvalue_i(B^T B)}. \quad (15)$$

The largest eigenvalue is one, due to normalization of $B$. For $b \to 0$ the eDOF is $N$ and for $b \to \infty$ the eDOF is one. The large eigenvalues define eigenvectors that are essential for describing the data, and the small eigenvalues define eigenvectors describing insignificant contributions due to noise.

The splitting of the integration is readily done with the evidence approximation [11,12], which can be applied if the probability for the kernel widths, given the data $P(b|d)$, is strongly peaked at some value $\hat{b}$, $P(b|d) \approx \delta(b - \hat{b})$:

$$P(f|d,\sigma,I) \propto \int d^N h \, \delta(f - \hat{B} * h) \, P(d|h,\hat{b},\sigma,I) \, P(h|I). \quad (16)$$

The optimal kernel widths $\hat{b}$ are determined by maximizing the marginal $P(b|d)$:

$$P(b|d) \propto \int d^N h \, P(d|h,b) \, P(h) \, P(b). \quad (17)$$

The multidimensional integrals are routinely determined by the steepest-descent approximation, where $\ln(P(d|h,b)\,P(h))$ is expanded up to second order about its maximum $\hat{h}$. The resulting Gaussian integral yields

$$P(b|d) \approx P(b) \, P(d|\hat{h},b) \, P(\hat{h}) \, det^{-\frac{1}{2}} H, \quad (18)$$

where $H$ is the Hessian of $\ln(P(d|h,b)\,P(h))$. In the present case of a linear model, $d = A * f$, the Hessian can easily be calculated:

$$H = B^T A^T diag(1/\sigma^2) A B + diag(\alpha/h). \quad (19)$$

Given the most probable kernel widths $\hat{b}$ and most probable hidden image $\hat{h}$ we calculate the *maximum posterior* (MAP) solution from $\hat{f} = f(\hat{h}, \hat{b})$.

The factor that favors the simplest model is called "Occam's razor". The Bayes theorem does recognize simplicity as one component of the inference. The quantity that governs Occam's razor is the volume of the prior probability covered by the high-likelihood region [19]. Two cases have to be distinguished: The models differ in the number of parameters (DOF), e.g., the number of grid cells, or the models have a constant DOF but differ in the values of additional model parameters determining the eDOF, as in the present case of the adaptive kernels with the additional parameters $b$. In the former case, the prior probability density of the parameters depends on the manifold of the parameter space, where the prior decreases as the DOF increases. In the latter case, the volume of the hypothesis space is constant. Neglecting the entropic contribution, Occam's razor is determined by the volume of the likelihood-probability density, which depends implicitly on the model parameters. In the adaptive kernel approach, the DOF is given by the number of image cells $N$ and the model parameters $b$ define the complexity of the model (eDOF). The volume of the prior probability covered by the high-likelihood region is already calculated in eq. (18). The $B^T B$-term in the Hessian reflects the smoothing power of the adaptive kernels. The determinant of $H$ decreases as the kernel widths increase.

We are mainly interested in the posterior expectation $< f_l >$ and point estimates of the variance of the posterior probability density $var(f_l) = < f_l^2 > - < f_l >^2$. The variance describing the uncertainty of the image $f$ is correlated with the kernel width. To illustrate this, consider a mock data set, consisting of $N$ samples of a constant, distorted by a constant noise level $\sigma$. In the limit $b \to 0$, where no smoothing occurs, the variance of the uncorrelated image points is $\sigma^2$. In the limit $b \to \infty$, where overall smoothing occurs, the data is described by the mean value, and the variance is $\sigma^2/N$ (the variance of the mean). Between the two limiting cases, the variance is approximately given by $\sigma^2/b$. Thus, unstructured data regions that can be described with large kernel widths result in reconstructions with a low error level.

To summarize the ingredients of the reconstructing procedure in the Bayesian framework, we have to start with the Bayes theorem, assigning the likelihood probability from the error statistics of the experiment and the prior probability for a positive additive distribution. In addition, we have to introduce the adaptive kernels, to take into account our background information about the correlated cells. The nuisance parameters have to be marginalized. The numerical calculations are rather cumbersome, but do not exceed the experimental efforts.

## III. EXPERIMENT

The following measurements were performed at the 2.6 MeV Van-de-Graff accelerator at the Max-Planck-Institut für Plasmaphysik in Garching. The incident beam of 1.5 MeV or 2.6 MeV $^4$He was collimated to $0.5 \times 0.5$ mm$^2$. Backscattered particles were detected at a scattering angle of 165°. The solid angle of the detector is $1.08 \times 10^{-3}$ sr. A standard PIPS detector (EG&G Ortec BE-012-025-100) with an active area of 25 mm$^2$ and a nominal resolution of 12 keV FWHM for 5.486 MeV $\alpha$ particles was used. The actual noise width of the detector-preamplifier-amplifier chain was determined with a test pulser to be about 13 keV FWHM. No efforts to achieve a better resolution were performed. The thickness of the entrance Au electrode of the detector is 40.0 $\mu$g/cm$^2$ ($1.2 \times 10^{17}$ atoms/cm$^2$).

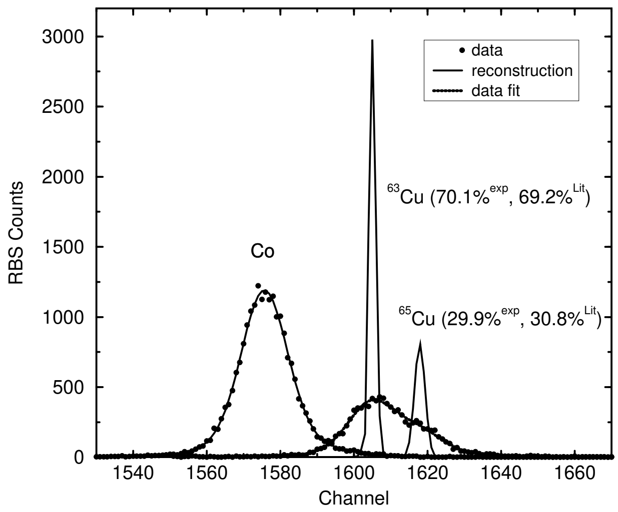


FIG. 1. RBS spectra of thin Co and Cu films on a Si substrate, measured with 2.6 MeV $^4$He at 165°. The Cu spectrum is deconvolved with the apparatus broadening function obtained from the Co spectrum. The two Cu isotopes are clearly resolved with a measured abundance ($^{exp}$) very close to the natural abundance ($^{Lit}$). Energy calibration: 1.19 keV/channel.

## IV. RESULTS AND DISCUSSION

The apparatus-energy-broadening function has to be determined simultaneously with the spectrum we are interested in. This is done by measuring the RBS spectra of a thin cobalt layer on a Si substrate. Co is chosen, since it is isotopically pure. The thickness of the Co layer was chosen to be $1.5 \times 10^{16}$ atoms/cm$^2$ (0.75 nm) to obtain sufficient intensity as well as negligible multiple scattering.

The energy loss of the incoming and backscattered $^4$He ions when penetrating the Co layer is about 1 keV for each direction. The energy-loss straggling of the $^4$He ions is about 1 keV in Bohr approximation. For an ideal apparatus and detector with no intrinsic energy broadening, the Co backscattering spectrum should therefore have a width of about 3 keV, which is much lower than the actual detector resolution.

The measured backscattering spectrum of the Co layer is shown in Fig. 1. The energy calibration gives 1.19 keV/channel. The width of the Co peak is about 19 keV FWHM due to the different energy-broadening contributions: 13 keV due to electronic noise, about 4 keV due to energy-loss straggling in the Au entrance electrode, about 6 keV due to energy-loss straggling in the dead layer of the detector [2], about 6 keV due to the statistics of electron-hole pair creation [2], about 3 keV due to energy loss and energy-loss straggling in the Co layer, and about 10 keV due to the initial beam-energy straggling and geometrical-energy straggling. The peak is nonsymmetric and non-Gaussian mainly in the wings of the distribution.

In the first step, we have to reconstruct the apparatus function from the measured data of the Co layer. This reconstruction is necessary because the measured data suffer from statistical fluctuations due to the number of counts. The solid line in Fig. 1 is the reconstruction of the apparatus-broadening function and was obtained with the described Bayesian analysis, using the adaptive kernel approach. For the matrix $A$ we have to choose the unity matrix because we want to exploit only the smoothing property of the adaptive kernel approach.

With the best estimate for the apparatus-broadening function, we are now able to deconvolve the Cu spectrum. This is also shown in Fig. 1. The thickness of the Cu layer was $1.5 \times 10^{16}$ atoms/cm$^2$. The energy of the backscattered $^4$He-ions from Cu is about 2029 keV, which is very close to the energy of 1990 keV of $^4$He backscattered from Co, so we can use the same apparatus function. The apparatus function is energy dependent, but only varies slowly with energy (see below). After deconvolution, the two isotopes $^{63}$Cu and $^{65}$Cu are clearly resolved. The FWHM of the dominant $^{63}$Cu peak after deconvolution is 2.6 keV, which is about 8 times better than the achieved experimental resolution and far beyond any conceivable experimental resolution with semiconductor detectors. The smaller peak of $^{65}$Cu is slightly broader because the signal-to-noise ratio is lower, resulting in less structure. The measured abundances of the isotopes are 70.1% $^{63}$Cu and 29.9% $^{65}$Cu. This has to be compared with the natural abundance of 69.2% $^{63}$Cu and 30.8% $^{65}$Cu.

It should be noted that only the measured and reconstructed apparatus function and no additional prior information, such as knowledge of the existence of two peaks, peak positions, peak area ratios or whatever, was used for the deconvolution. The eDOF as calculated from Eq. (15) is 14.7. This value results mainly from the 14 cells spanning the two peaks.

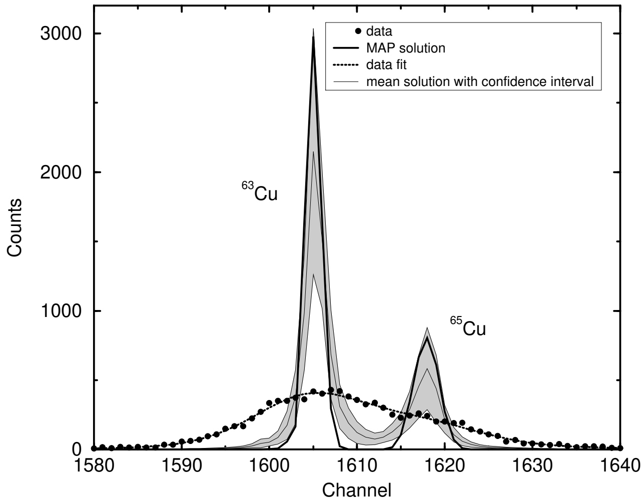


FIG. 2. The MAP solution and the mean solution with the confidence interval as two presentations of the most important properties of the posterior probability density.

A comparison of the MAP solution $f(\hat{h}, \hat{b})$ and the mean solution $< f >$ with the confidence interval $(\Delta f)^2$, is depicted in Fig. 2. The mean solution is slightly smoother compared to the MAP solution, reflecting the skewness of the posterior probability density. Notice that the two solutions are only different ways to represent the most important properties of the posterior probability density. The two presentations coincide if the signal-to-noise ratio goes to infinity. The confidence interval represents $\pm 1$ standard deviation of the posterior probability density.

The dashed line in Fig. 2 is the convolution of the deconvolved spectrum with the apparatus function.

A more challenging example, which also shows the resolution limitations for a given signal-to-noise ratio, is a thin Fe film on Si, which is depicted in Fig. 3. The Fe film thickness was $1.7 \times 10^{16}$ atoms/cm$^2$. Whereas the two prominent isotopes, $^{54}$Fe and $^{56}$Fe, can clearly be resolved, the third isotope, $^{57}$Fe, is only indicated in the high-energy tail of the most prominent peak. The measured abundances of the isotopes are 8.1% $^{54}$Fe and 91.9% for the joined $^{56}$Fe and $^{57}$Fe peaks. This has to be compared with the natural abundance of 5.9% $^{54}$Fe and (91.6% + 2.2%) for $^{56}$Fe and $^{57}$Fe. The isotope $^{58}$Fe, with a natural abundance of 0.3%, is not visible in this measurement and would require an improved signal-to-noise

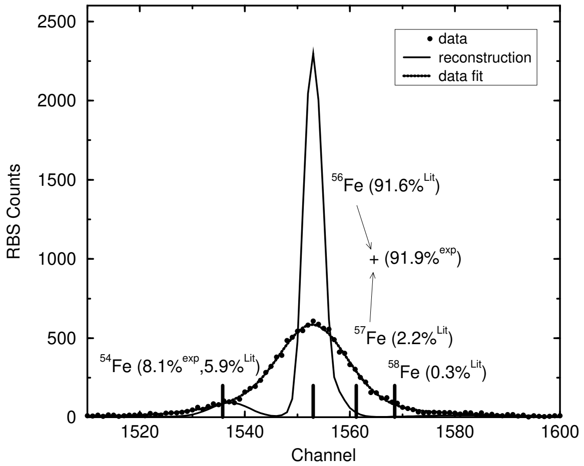


FIG. 3. RBS spectrum of a thin Fe film on a Si substrate, measured with 2.6 MeV $^4$He at 165°. The Fe spectrum is deconvolved with the apparatus broadening function obtained from the Co spectrum. The black lines indicate the theoretical positions of the 4 Fe isotopes. The isotopes $^{54}$Fe and $^{56}$Fe are clearly resolved with a measured abundance ($^{exp}$) very close to the natural abundance ($^{Lit}$), the isotope $^{57}$Fe is hardly resolved.

ratio.

Besides the separation and quantification of different masses, a common task in RBS analysis is the determination of depth profiles. To demonstrate the achievable increase in depth resolution, a Co-Au multilayer on a Si substrate, consisting of six layers of Au and six layers of Co with a layer thickness of about $8.3 \times 10^{16}$ atoms/cm$^2$ each, is deconvolved with the Co (Au) apparatus-broadening functions, which are measured individually with thin films, as described above. The measured and deconvolved spectra are shown in Fig. 4. The resolution of the six single layers of each constituent is significantly enhanced, and the layers can be separated. The width of the peaks increases for larger target depths due to energy-loss straggling in the target. Whereas at the target surface the Au layers can be clearly separated, they overlap more and more in larger target depths. For thick targets, the effect of energy-loss straggling becomes the dominant factor that determines and limits the depth resolution.

The inset in Fig. 4 shows the two nearly identical apparatus-broadening functions from Co and Au. Note the high-energy tail appearing in both reconstructions. This is most likely due to the energy distribution of the incident ions. The small deviation at the low-energy side is an effect of the different stopping powers in Co and Au. The Co apparatus-broadening function was shifted by 213 channels to coincide with the Au function.

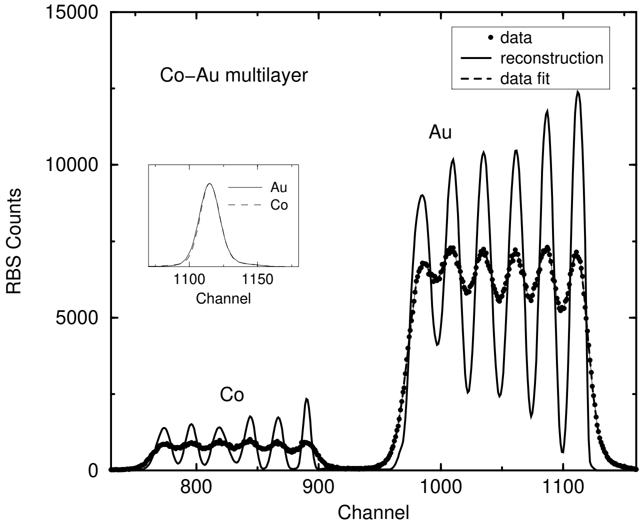


FIG. 4. RBS spectra of a Co-Au multilayer film on a Si substrate, measured with 1.5 MeV $^4$He at 165°. The Co (Au) part of the spectrum is deconvolved with the apparatus broadening function obtained from a spectrum of a thin Co (Au) layer, respectively. The inset shows the two nearly identical apparatus broadening functions from Co (Au), respectively. The Co apparatus broadening functions is shifted by 213 channels to coincide with the Au function.

## V. CONCLUSIONS

The energy resolution in ion-beam experiments with MeV ions, such as Rutherford backscattering, is limited by the intrinsic energy resolution of the detector and the energy-broadening function of the whole apparatus, thus limiting the energy and depth resolution in ion-beam experiments. The apparatus-energy-broadening function can be measured at a specific energy by backscattering from thin films with film thicknesses in the range of $10^{16}$ atoms/cm$^2$ of monoisotopic elements, such as cobalt or gold. The knowledge of the apparatus-energy-broadening function allows the deconvolution of measured backscattering spectra by means of Bayesian probability theory and the entropic prior to obtain the most probable solution. An enhancement of the energy resolution by a factor of 8 in Rutherford backscattering analysis of thin films, resulting in an energy resolution of about 2.5 keV with semiconductor detectors, is obtained. The energy resolution before deconvolution was about 20 keV. The deconvolution allows the separation of the isotopes $^{63}$Cu and $^{65}$Cu or $^{54}$Fe and $^{56}$Fe in thin films with 2.6 MeV $^4$He ions at a scattering angle of 165°.

For thick films, the energy dependence of the apparatus function, due to the response function of a semiconductor detector, may become a limiting factor for deconvolution. For a proper deconvolution, the energy-dependent apparatus function must be known with great accuracy. Fortunately, the energy dependence of the energy broadening is relatively small, so the described method also works for thicker films as long as the apparatus function does not change significantly. For larger energy intervals, the apparatus function has to be measured at different energies to allow deconvolution.

The described method allows the deconvolution of the energy-broadening contributions of the detector, electronic noise, geometrical-energy straggling, and energy straggling of the incident beam. In larger depths, the broadening contribution due to energy-loss straggling and multiple scattering becomes dominant and results in a decrease of the depth resolution. The deconvolution of the effects of energy-loss straggling and multiple scattering seems possible; however, an accurate theoretical description of these energy-broadening contributions with an accuracy of the order of percent is necessary.


[1] M. Geretschläger. Nucl. Instr. Meth. 204 (1983) 479.
[2] E. Steinbauer, P. Bauer, M. Geretschläger, G. Bortels, J.P. Biersack, and P. Burger. Nucl. Instr. Meth. B 85 (1994) 642.
[3] G. Dollinger, T. Faestermann, and P. Maier-Komor. Nucl. Instr. Meth. B64 (1992) 422.
[4] R. Doolittle. Nucl. Instr. Meth. B9 (1985) 344.
[5] H. Ellmer and D. Semrad. Phys. Rev. E 54 (1996) 3569.
[6] S.F. Gull and J. Skilling. IEE Proc. 131 F (1984) 646.
[7] P.C. Gregory and T.J. Loredo. Astrophys. J. 398 (1992) 146.
[8] W. von der Linden, M. Donath, and V. Dose. Phys. Rev. Lett. 71 (1993) 899.
[9] W. von der Linden. Appl. Phys. A 60 (1995) 155.
[10] R. Fischer, Th. Fauster, W. von der Linden, and V. Dose. Surf. Rev. & Lett. 3 (1996) 1393.
[11] J. Skilling. Quantified maximum entropy. In *Maximum Entropy and Bayesian Methods*, P. Fougère, Ed. Kluwer Academic Publishers, Dordrecht, 1990, p. 341.
[12] S.F. Gull. Developments in maximum entropy data analysis. In *Maximum Entropy and Bayesian Methods*, J. Skilling, Ed. Kluwer Academic Publishers, Dordrecht, 1989, p. 53.
[13] L.R. Mead and N. Papanicolaou. J. Math. Phys. 25 (8) (1984) 2404.
[14] R.C. Puetter. Int. J. Image Sys. & Tech. 6 (4) (1995) 314.
[15] J. Skilling. Fundamentals of maxent in data analysis. In *Maximum Entropy in Action*, B. Buck and V. Macaulay, Eds. Clarendon Press, Oxford, 1991, p. 19.
[16] J. Skilling. Classic maximum entropy. In *Maximum Entropy and Bayesian Methods*, J. Skilling, Ed. Kluwer Academic Publishers, Dordrecht, 1989, p. 45.
[17] R. Fischer, W. von der Linden, and V. Dose. On the importance of $\alpha$ marginalization in maximum entropy. In *Maximum Entropy and Bayesian Methods*, R. Silver and K. Hanson, Eds. Kluwer Academic Publishers, Dordrecht, 1996, p. 229.
[18] J. Ruanaidh and W. Fitzgerald. *Numerical Bayesian*

*Methods Applied to Signal Processing*. Springer, New York, 1996.

[19] E.T. Jaynes. *Probability Theory: The Logic of Science*. unpublished.