

Statistically sound evaluation of trace element depth profiles by ion beam analysis

K. Schmid *, U. von Toussaint

Max-Planck-Institut für Plasmaphysik, Boltzmannstraße 2, D-85748 Garching b. München Germany

Abstract

This paper presents the underlying physics and statistical models that are used in the newly developed program NRADC for fully automated deconvolution of trace level impurity depth profiles from ion beam data. The program applies Bayesian statistics to find the most probable depth profile given ion beam data measured at different energies and angles for a single sample. Limiting the analysis to % level amounts of material allows one to linearize the forward calculation of ion beam data which greatly improves the computation speed. This allows for the first time to apply the maximum likelihood approach to both the fitting of the experimental data and the determination of confidence intervals of the depth profiles for real world applications. The different steps during the automated deconvolution will be exemplified by applying the program to artificial and real experimental data.

1 Introduction

The determination of trace element depth profiles (concentrations in the order of 1%) from ion beam (IBA) methods is of importance in different areas of material science. For instance in the field of plasma wall interaction in nuclear fusion devices, measuring deuterium depth profiles in metals like tungsten is one of the fundamental methods to investigate the retention of hydrogen in plasma exposed first wall materials. Since the solubility of hydrogen in W is close to zero, hydrogen is only located at natural defects and radiation damage sites at concentrations ranging from ≈ 0.1 at. frac. near the surface to 10^{-7} at. frac. throughout the bulk. We will show that at such low concentrations the forward

*Corresponding Author
Tel: +49 89 3299 2228
Fax: +49 89 3299 2279
E-mail: Klaus.Schmid@ipp.mpg.de

calculation of the IBA data can be linearized with respect to the concentration dependence, since the stopping power is essentially not modified by variations in the concentrations of the trace element. This linearization greatly improves the computation speed and allows one to perform an extensive Markov-Chain Monte Carlo (MCMC) based search for the most probable (maximizing the likelihood) depth profile given all experimental data. This ability to use multiple data sets makes it ideally suited for deconvoluting D depth profiles. Typically these are determined from nuclear reaction analysis (NRA) data acquired at different energies of the incident ion beam. In this approach each energy predominantly probes a certain depth region due to the peaked nature of the ${}^3\text{He}(p, \alpha)\text{D}$ nuclear reaction used.

The determination of depth profiles by simultaneously fitting multiple data sets determined from different IBA experiments is an ideal application of Bayesian statistics. This approach has been applied to the deconvolution of IBA data [1, 2, 3] and is summarized in the review by Jeynes [4]. There have been many approaches to determining depth profiles by fitting experimental scattering data starting with the early work by Doolittle [5] evolving into complete code packages like DataFurnace [6] or SimNRA [7]. A review and detailed comparison of current codes can be found in [8]. The codes differ mainly in the figure of merit function (e.g. minimizing χ^2 or maximizing the likelihood) and the type of optimization algorithm used (e.g. Simplex Methods[7] or variations of MCMC schemes[6, 2, 3]). Another important difference are the confidence intervals of the determined depth profiles. Only very few codes yield statistically sound confidence intervals. In [9] Bayesian statistics are used in combination with an MCMC scheme to estimate the confidence intervals of depth profiles determined from IBA measurements. Similarly DataFurnace uses an MCMC scheme to assign confidence intervals using Bayesian statistics to the depth profiles previously determined by minimizing χ^2 . An example of a χ^2 based fit to experimental data using simulated annealing can be found in e.g. [10, 11].

Independently on how the confidence intervals are assigned, using a minimum χ^2 as figure of merit to fit the depth profile introduces the ambiguity that χ^2 always (must) decrease with the number of free parameters i.e. layers. So increasing the number of layers will yield a better fit which in the end can result in fitting the noise of the experimental data. While this can be avoided in an "ad-hoc fashion" by penalizing the addition of new layers in the calculation of the χ^2 (e.g. [11]), this approach will not necessarily lead to the most probable number of layers given the data. To avoid this ambiguity NRADC (Nuclear Re-

action Analysis DeConvolution) uses Marginalization to find the most probable number of layers given the experimental data. This procedure also uses a χ^2 base likelihood estimate but penalizes solutions with a large number of layers following the concept of "Ockham's Razor" (see section 4.1). It is followed by a maximum likelihood [2, 3] search for the concentrations in each layer. So the important difference to previous approaches is that it chooses the number of free parameters to fit based on a probabilistic argument. Both searches are based on MCMC chains and require $\approx 10^6$ forward calculations. In comparison only 3000 MCMC steps were performed in [9] to infer the confidence intervals. This shows why the maximum likelihood approach has been generally considered too computationally expensive due to the large number of forward calculations required [4]. But due to the linearization of the forward calculation in NRADC, such a large number of forward calculations become manageable and a full deconvolution including confidence intervals both based on the maximum likelihood approach takes less than ≈ 10 minutes on a normal office PC. In contrast a Bayesian inference calculation is an over night task for codes that include the full non linear forward calculation.

In the approach utilized in this paper generality is sacrificed for computational efficiency. Despite the linearization our approach still is applicable to a large number of problems where trace amounts of material are to be detected.

One reason why NRADC has been developed was that often NRA experiments are evaluated "manually", only considering the proton spectrum peak integrals, thereby completely neglecting the information in the individual proton spectra. This issue of ambiguity during manual deconvolution of IBA data is of course not new and has been discussed previously (e.g. [4] in 2003). However since then still a large number (e.g. [12, 13, 14, 15, 16]) of publications based on a manual fitting of the experimental data have been published. The depth profiles therein have no statistically sound confidence intervals and thus numerous discussions have arisen on whether or not certain features in the depth profiles are real or just artifacts from the deconvolution. Due to the fact that manual evaluation of data appears still to be common today, we will revisit the topic based on our linearized forward calculation which allows to show the ambiguities based on mathematical arguments.

The paper will present the underlying physics and statistical models that are used in the newly developed program NRADC. The deconvolution of depth profiles using NRADC will be exemplified by applying it to real world and artificially generated data sets from D depth profiling by ion beam analysis. However

it must be stressed that the program is not limited to the deconvolution of D depth profiles, but can generally be applied to the evaluation of ion beam data, as long as the conditions allowing one to linearize the forward calculation are fulfilled.

The paper first outlines how D depth profiles are measured using NRA and how the forward calculation of such IBA energy spectra can be linearized. Also the limits of applicability of this approximation are discussed. Next based, on the linear forward calculation, the deconvolution of the depth profile from the measured data is described. In this section different approaches ranging from direct inversion of the linear system to a Markov-Chain Monte Carlo search for the most probably depth profile are compared. Also potential pitfalls of the conventional, manual evaluation are pointed out in this section. Finally the application of the NRADC program to artificial and real world examples is shown, discussing the potential for improved depth profiling by applying a statistically sound evaluation of the data.

2 NRA energy scan depth profiling of D in W

To determine the depth profile of D in W or other bulk materials typically the ${}^3\text{He}(p, \alpha)\text{D}$ nuclear reaction is used. The energy dependence of this nuclear reaction is depicted in Fig. 1 with data taken from [17]. It exhibits a peak at 600keV ${}^3\text{He}$ energy. Therefore varying the primary energy amounts to varying the depth being predominantly probed for D, since the ${}^3\text{He}$ ions lose energy as they penetrate the target. Therefore for a given energy of the incident ${}^3\text{He}$, the energy axis in Fig. 1 can also qualitatively be seen as a depth axis. This allows one to compare the probing depth with a typical D depth profile also shown in Fig. 1 whose depth axis (upper x-axis) is matched to the energy axis of the cross section data. This D profile exhibits a sharp peak at the surface and a long tail into the bulk. Comparing the two curves in Fig. 1 it becomes evident that a certain ${}^3\text{He}$ energy does not exclusively probe a certain depth: Despite the low cross section at the surface, the high concentration of D there, contributes essentially the same amount of counts to the NRA spectrum, as the low concentration region at the depth where the cross section has its maximum. This is best visualized by the plot of the product of the cross section with D concentration shown in Fig. 1. Therefore the contributions of the D amount at individual depths to an NRA spectrum are heavily convoluted, making a manual

evaluation of the data measured at different energies by eye almost impossible. Therefore typically the manual evaluation is limited to only matching the total integral over the NRA energy spectra measured at different ^3He energies [17]. This approach ignores the entire information available in the energy spectra of the produced protons and alpha particles. Merely looking at the integrals is prone to artifacts since it is an ill posed inversion problem (see section 4 and 5). Also, without further assumptions about the principal shape of the D depth profile the thus determined profiles are not in general unique.

To determine the most probable depth profile given the NRA spectra measured for a single sample under different experimental conditions (e.g. energies, angle) a fast forward calculation model is needed. This forward calculation yields the NRA spectra corresponding to an assumed input depth profile which can then be compared to the experimental data.

3 Fast forward calculation of IBA data by linearization

The general idea behind a forward calculation is that it yields a functional relationship between the parameters to be estimated (e.g. the trace element depth profile) and the ideal (noiseless) data which can then be compared to the real noisy data resulting from an experiment. From this probabilistic comparison between the ideal forward calculated data and the experimental data the most probable choice of the trace element depth profile can be determined.

The forward calculation of IBA data is described in much detail for example in [4, 18, 7]. Here only the general idea behind the forward calculation and how it can be linearized for trace level concentrations will be described. The approach is similar to that in [19] but finally uses SimNRA to determine the design matrix (see below) for the linear system. However this is only performed once during the entire deconvolution procedure. All the forward calculations are done based on design matrices derived from this initial one! This is why NRADC is so much faster than other optimization codes since it only performs the full nonlinear calculation once. A complete deconvolution including confidence intervals takes at most 10 minutes including the initial calculation of the design matrix by SimNRA.

An IBA spectrum consists of M-channels each of which contain the number particles entering the detector from the target in an energy window $E_j + \Delta E_j$

with ($j \in 1 \dots M$). The energy E_j of the the j-th channel and its width ΔE_j are given by the energy calibration which has to be determined experimentally. The input depth profile into the forward calculation is given by N-layers of width Δx_i and concentration of the trace element c_i^T with ($i \in 1 \dots N$). In addition to the trace element depth profile concentrations, also the background target element concentrations $c_{i,k}^B$ are needed since they mainly define the stopping power. The index k thereby runs over all background target elements. Given the depth profiles, the primary ^3He energy E , the differential reaction cross section $\frac{d\sigma}{d\Omega}$, the analysis dose Φ (i.e. # of ^3He ions) and the solid angle $\Delta\Omega$ of the detector the number of counts Q_j in spectrum channel j is given by Eq. 1.

$$\begin{aligned}
Q_j &= \sum_{i=1}^N c_i^T * f_j \left(S(E, c_i^T, c_{i,k}^B), \frac{d\sigma}{d\Omega}(E), \Delta\Omega, \Phi \right) & (1) \\
S(E, c_i^T, c_{i,k}^B) &= c_i^T * S_i^T(E) + \sum_k c_{i,k}^B * S_k^B(E) \\
c_i^T &= \text{Concentration of trace element in layer i} \\
c_{i,k}^B &= \text{Concentration of background element k in layer i} \\
S_i^T(E), S_j^B(E) &= \text{Stopping power in pure elements} \\
f_j() &= \text{Complex non linear function}
\end{aligned}$$

Basically Eq. 1 states that the counts in a spectrum channel is given by the sum of the contributions from each layer in the input depth profile into the forward calculation.

Following Bragg's rule [20], the stopping power $S(E, c_i^T, c_{i,k}^B)$ in Eq. 1 is a linear combination of the energy loss in the individual elements weighed by the concentrations of the elements. Therefore it becomes obvious that low concentrations elements only contribute a negligible amount to the total stopping power. This fact allows to neglect the variations in Q_j due to variations in $S(E, c_i^T, c_{i,k}^B)$ due to variations c_i^T in Eq. 1. In this approximation Eq. 1 becomes a linear equation since the c_i^T dependence is removed from f_j which can be written as Eq. 2

$$\begin{aligned}
Q_j &= \sum_{i=1}^N c_i^T * f_j \left(S(E, c_{i,k}^B), \frac{d\sigma}{d\Omega}(E), \Delta\Omega, \Phi \right) & (2) \\
&= A_{j,i} c_i^T
\end{aligned}$$

Thus the forward calculation becomes a matrix multiplication which is very fast compared to the full non linear forward calculation. The design matrix $A_{j,i}$ in Eq. 2 is constant under variations in c_i^T and can be precalculated for different experimental conditions. Its elements $A_{j,i}$ describe the contribution of the trace element concentration in the i-th layer to counts detected in j-th IBA spectrum channel. In NRADC the program SimNRA [7] is used to produce $A_{j,i}$. SimNRA performs the full non linear forward calculation similar to DataFurnace, including proper handling of energy loss straggling and finite energy resolution of the detector. This fully automated procedure only needs the depth discretization of the trace element (given by a vector of layer thicknesses Δx_i) and background target depth profile. The details on how to choose this depth discretization will be discussed later. To determine the contribution of layer i to each energy channel Q_j a small amount of D is placed into layer i and the resulting spectra are calculated using SimNRA for each spectrum type (e.g. protons, alphas) and for each experimental condition EC_k . Repeating this for each layer, yields the contribution of each trace element depth profile layer i to each forward calculation data channel j for each experimental condition (e.g. Energy & Angle of incidence of the primary ^3He beam). This results in a design

matrix layout as in Eq. 3

$$\mathbf{A} = \begin{pmatrix} P_{Spec}(L_0, EC_0, Chan_0) & \dots & P_{Spec}(L_N, EC_0, Chan_0) \\ \vdots & \ddots & \vdots \\ P_{Spec}(L_0, EC_0, Chan_{N_C}) & \dots & P_{Spec}(L_N, EC_0, Chan_{N_C}) \\ \vdots & \ddots & \vdots \\ P_{Spec}(L_0, EC_{N_{EC}}, Chan_0) & \dots & P_{Spec}(L_N, EC_{N_{EC}}, Chan_0) \\ \vdots & \ddots & \vdots \\ P_{Spec}(L_0, EC_{N_{EC}}, Chan_{N_C}) & \dots & P_{Spec}(L_N, EC_{N_{EC}}, Chan_{N_C}) \\ \alpha_{Spec}(L_0, EC_0, Chan_0) & \dots & \alpha_{Spec}(L_N, EC_0, Chan_0) \\ \vdots & \ddots & \vdots \\ \alpha_{Spec}(L_0, EC_0, Chan_{N_C}) & \dots & \alpha_{Spec}(L_N, EC_0, Chan_{N_C}) \\ \vdots & \ddots & \vdots \\ \alpha_{Spec}(L_0, EC_{N_{EC}}, Chan_0) & \dots & \alpha_{Spec}(L_N, EC_{N_{EC}}, Chan_0) \\ \vdots & \ddots & \vdots \\ \alpha_{Spec}(L_0, EC_{N_{EC}}, Chan_{N_C}) & \dots & \alpha_{Spec}(L_N, EC_{N_{EC}}, Chan_{N_C}) \end{pmatrix} \quad (3)$$

L_i = Contribution of layer i

EC_j = Experimental condition j

N = Number of trace element depth profile layers

N_C = Number of channels in spectrum

N_{EC} = Number of experimental conditions

P_{Integ} = Contribution to proton peak integrals

P_{Spec} = Contribution to proton spectrum channels

α_{Spec} = Contribution to α spectrum channels

Multiplying the matrix in Eq. 3 with the trace element depth profile concentration vector c_i^T yields the complete forward calculation of all experimental data types (Proton integrals, Proton energy spectra and alpha spectra) for all experimental condition EC_j and can be directly compared to an experimental data vector D_{Type, EC_j}

Of course this linearization of the forward calculation is only applicable for cases where $c_i^T * S_i^T(E) \ll \sum_k c_{i,k}^B * S_k^B(E)$ is valid (see Eq. 1). The stopping power for the pure elements scales as the nuclear charge squared. This means that what can be considered to be a trace level amount depends on the type of ele-

ment under consideration. Since the stopping power scales with nuclear charge (Z) squared, this approximation is valid for hydrogen isotopes up the several % while it is hardly applicable for determining depth profiles of high Z impurities like W.

4 Deconvolution of IBA data

The steps in deconvoluting IBA data will be shown based on a mock data set (proton and alpha spectra) created from a typical D in W depth profile. This depth profile consists of a composition vector c_i^T and layer thickness vector Δx_i . The mock data set is generated by multiplying c_i^T with a design matrix $A_{j,i}$ generated for the following "experimental" conditions: Perpendicular angle of incidence and ^3He energies of 600keV, 1200keV, 2400keV, 3200keV, 4500keV and 6000keV. This yields the proton and alpha spectra to which Poisson noise is added to mimic the counting statistics of an IBA measurement. The typical depth profile is shown as open square symbols in Fig 2 and the corresponding artificial data set is shown as open symbols in Fig 3. To deconvolute this artificial data set i.e to reproduce the underlying c_i^T different approaches are possible: Since the forward calculation is linear, the simplest approach is to invert the design matrix using singular value decomposition (SVD). To include the experimental uncertainties i.e the Poisson noise in the counting statistics, the inversion of the linear system can be written as a weighted least squares problem as in Eq. 4. From now on all bold font variables in equations are vectors or matrices while normal variables with indices correspond to vector or matrix elements.

$$\chi^2 = \sum_{j=1}^M \left(\frac{D_j - \sum_{i=1}^N A_{j,i} c_i^T}{\sigma_j} \right)^2 \rightarrow \text{Min} \quad (4)$$

$$\sigma_j = \sqrt{D_j} = \text{Experimental variance (Poisson noise)}$$

To solve Eq. 4 the SVD of the design matrix weighted with the experimental variance is calculated first. This yields the left and right singular vectors stored as column orthonormal matrices \mathbf{U} and \mathbf{V} and the singular values stored as the diagonal elements of matrix \mathbf{W} . From this SVD result the least squares

composition vector result \mathbf{c}^{LSQ} of Eq. 4 is determined as shown in Eq. 5.

$$B_{j,i} = \frac{A_{j,i}}{\sqrt{D_j}} \quad (5)$$

$$\mathbf{B} = \mathbf{U} \cdot \mathbf{W} \cdot \mathbf{V}^T = \text{SVD result}$$

$$E_j = \frac{D_j}{\sqrt{D_j}}$$

$$\mathbf{c}^{\text{LSQ}} = \mathbf{V}^T \cdot \left(\frac{\mathbf{U}^T \cdot \mathbf{E}}{\text{Diag}(\mathbf{W})} \right) \quad (6)$$

The division by the singular values $\text{Diag}(\mathbf{W})$ in Eq. 5 is to occur per vector component i.e $(\mathbf{U}^T \cdot \mathbf{E})_i / \text{Diag}(\mathbf{W})_i$.

In addition to the solution for the composition vector SVD also allows to calculate the condition number of the design matrix which qualitatively states how sensitive \mathbf{c}^{LSQ} is to variations (e.g. noise) in the data vector \mathbf{D} . The condition number C is given by the ratios of the largest to the smallest singular value of the design matrix i.e $C = \max(\text{Diag}(\mathbf{W})) / \min(\text{Diag}(\mathbf{W}))$. For values of C larger than ≈ 1000 the system is considered ill conditioned and the result for \mathbf{c}^{LSQ} is affected by the counting statistics and is generally meaningless.

To assign confidence intervals $\Delta \mathbf{c}^{\text{LSQ}}$ to the components of the least squares solution vector \mathbf{c}^{LSQ} the approach outlined in chapter 3.5 in [21] is used. According to the Bayes's theorem the probability of the estimated parameters \mathbf{c}^{LSQ} to be true given the experimental data \mathbf{D} and further prior knowledge \mathbf{I} about \mathbf{c}^{LSQ} is given by Eq. 7

$$P(\mathbf{c}^{\text{LSQ}}|\mathbf{D},\mathbf{I}) \propto P(\mathbf{D}|\mathbf{c}^{\text{LSQ}},\mathbf{I}) * P(\mathbf{c}^{\text{LSQ}}|\mathbf{I}) \quad (7)$$

$$P(\mathbf{D}|\mathbf{c}^{\text{LSQ}},\mathbf{I}) = \text{Likelihood function}$$

$$P(\mathbf{c}^{\text{LSQ}}|\mathbf{I}) = \text{Prior function}$$

The likelihood function thereby states how probable it would be to measure the experimental data \mathbf{D} given that the depth profile \mathbf{c}^{LSQ} and our linear forward calculation model were true. In the case of forward calculation of IBA counting spectra the likelihood function is given by a product of Poisson distributions in

Eq. 8

$$\begin{aligned}
\mathbf{f} &= \mathbf{A} \cdot \mathbf{c}^{\text{LSQ}} & (8) \\
\mathbf{A} &= \text{Design matrix} \\
\mathbf{f} &= \text{forward calculation result vector} = \\
&\quad \text{mean value of expected counts} \\
P(\mathbf{D}|\mathbf{c}^{\text{LSQ}}, \mathbf{I}) &= \prod_{j=1}^M \frac{f_j^{D_j}}{D_j!} \exp(-f)
\end{aligned}$$

The product in Eq. 8 runs over all proton and alpha spectrum channels and compares the experimental data D_j to expected the data f_j from the forward calculation result.

The prior function contains additional information about the parameters to be estimated (e.g that the components of \mathbf{c}^{LSQ} should lie between 0 and 1.) As is shown in [21] the confidence intervals for the estimated parameters can be calculated from the diagonal terms of the covariance matrix \mathbf{cov} . The covariance matrix describes correlations between estimated parameters (e.g. concentrations in different layers of the trace element depth profile). The covariance matrix is calculated according to Eq. 9.

$$\begin{aligned}
L &= \ln(P(\mathbf{c}^{\text{LSQ}}|\mathbf{D}, \mathbf{I})) & (9) \\
G_{j,i} &= \frac{\partial^2 L}{\partial c_j^{\text{LSQ}} \partial c_i^{\text{LSQ}}} \\
\mathbf{cov} &= -\text{Inverse}[G_{j,i}] & (10)
\end{aligned}$$

The "Inverse" in Eq. 9 stands for the matrix inverse of matrix elements enclosed by the square brackets.

To calculate L in Eq. 9 a flat (i.e constant) prior function $P(\mathbf{c}^{\text{LSQ}}|\mathbf{I})$ is used in Eq. 7 and the Likelihood function is approximated by $P(\mathbf{D}|\mathbf{c}^{\text{LSQ}}, \mathbf{I}) \propto \exp(-\chi^2/2)$. This essentially amounts to approximating the Poisson noise of the experimental IBA counting data with gaussian noise which is justified for counts > 10 . This finally yields the following expressions for L , \mathbf{cov} and the

least squares solution confidence intervals Δc_i^{LSQ} .

$$\begin{aligned}
L &= const. - \frac{\chi^2}{2} \\
G_{k,s} &= \frac{\partial^2 L}{\partial c_k^{LSQ} \partial c_s^{LSQ}} = \sum_{j=1}^M \frac{A_{j,k} A_{j,s}}{\sigma_j^2} \\
\mathbf{cov} &= -\text{Inverse}[\mathbf{G}] \\
\Delta c_i^{LSQ} &= \sqrt{cov_{i,i}}
\end{aligned} \tag{11}$$

In order to apply the least squares solution technique to the artificial data set in Fig. 3 the design matrix \mathbf{A} has to be calculated. Apart from the experimental parameters used to measure the alpha and proton spectra one has to specify a certain discretization Δx_i of the depth profile. This choice of the depth discretization is crucial for finding the most probable depth profile as shall be demonstrated using by deconvoluting the mock data in Fig. 3 using different assumptions about the depth discretization.

Since the depth discretization is known for the artificial data set it can be used directly to produce the appropriate design matrix $\mathbf{A}^{\text{Perfect}}$ which of course is then identical to that used to produce the artificial data. Using $\mathbf{A}^{\text{Perfect}}$ in the least squares solution technique yields the depth profile $\mathbf{c}^{\text{LSQ,Perfect}}$ (open circles) overlaid to the true depth profile c_i^T in Fig. 2. Despite the Poisson noise on the artificial data set the agreement between the two profiles is good as was to be expected from the condition number of ≈ 10 of $\mathbf{A}^{\text{Perfect}}$. In reality of course the depth discretization is essentially unknown, only the maximum information depth is given based on the stopping power of the ^3He ions in the target. For 6MeV ^3He the maximum information depth in W is of the order of 10 μm . Therefore one has to choose a depth discretization for these first 10 μm . For peaked cross sections as is the case for the $^3\text{He}(p, \alpha)\text{D}$ nuclear reaction a "reasonable" guess for the depth discretization can found using the following scheme: In a first step for each ^3He energy E_j^{3He} used in the measurement find the depth x_j^{center} where the ^3He has reduced its energy to where the cross section has its peak. In the second step find the depth x^{Max} where the ^3He with the highest E_j^{3He} has reduced its energy to a minimum value e.g 100keV.

Then the initial depth discretization is given by Eq. 12

$$\begin{aligned}
\Delta x_1 &= x_1^{center} & (12) \\
\Delta x_i &= x_i^{center} - x_{i-1}^{center} \text{ with } i \in (2, m - 1) \\
\Delta x_m &= x^{Max} - x_{m-1}^{center} \\
m &= \text{Number of different energies during measurements}
\end{aligned}$$

This type of depth discretization centers the layers of the depth profile around the regions where for a given E_j^{3He} the measurement is most sensitive (i.e has the highest cross section) for D. Based on this depth discretization and the rest of the experimental parameters a new design matrix \mathbf{A}^{Guess} can be calculated which then yields the new least squares solution $\mathbf{c}^{LSQ,Guess}$. $\mathbf{c}^{LSQ,Guess}$ is compared to the true depth profile c_i^T in Fig. 4. The first apparent difference is that the number of layers in $\mathbf{c}^{LSQ,Guess}$ is less than in c_i^T and while the average amount of D in both profiles is similar the shape of both profiles is not very similar. In particular the appearance of deep minima occurs. What this example shows is that during the deconvolution of IBA data to obtain depth profiles it is not sufficient to just optimize the composition \mathbf{c} but also the depth discretization $\Delta \mathbf{x}$ must be part of the optimization i.e the most probable composition and layer thickness vector must be found.

4.1 Finding the most probable depth discretization

To find the optimum $\Delta \mathbf{x}$ a large number of different $\Delta \mathbf{x}$ have to be compared, which requires a large number of forward calculations. Since the design matrix depends on $\Delta \mathbf{x}$ one would have to recompute it for each choice of $\Delta \mathbf{x}$. Since this is a very expensive calculation a method of varying $\Delta \mathbf{x}$ without having to recompute the whole matrix was developed: Based on an initial choice for $\Delta \mathbf{x}$ from Eq. 12 a depth discretization was produced by splitting each of the layers in $\Delta \mathbf{x}$ into n sub layers yielding $\Delta \mathbf{x}^{OS}$. The OS thereby stands for "over sampling". Based on $\Delta \mathbf{x}^{OS}$ a design matrix \mathbf{A}^{OS} is calculated. By restricting the possible choices of $\Delta \mathbf{x}$ to those that can be generated by "binning" together sub layers in $\Delta \mathbf{x}^{OS}$ the full re-calculation of the design matrix can be avoided: Binning two sub layers together in the current depth discretization corresponds to adding the corresponding columns in \mathbf{A}^{OS} . Therefore for a certain binning layout the appropriate design matrix can be derived quickly from \mathbf{A}^{OS} by adding the appropriate columns.

Finding the most probable $\Delta \mathbf{x}$ actually consists of two problems: First find the number N of layers in $\Delta \mathbf{x}$ and second their individual thicknesses Δx_i . To find the most probable N^{OPT} number of layers one has to assign a probability to a certain choice of N . Following the approach in [21] chapter 4.2 the probability for a choice of N is given in Eq. 13 using a similar approach as in Eq. 7

$$P(N|\mathbf{D}, I) \propto P(\mathbf{D}|N, I) P(N|I) \quad (13)$$

\mathbf{D} = Experimental data

$P(N|I)$ = Flat prior i.e is constant

Since a flat, constant prior probability is used in Eq. 13, $P(N|\mathbf{D}, I)$ only depends on $P(\mathbf{D}|N, I)$ i.e the probability of the data \mathbf{D} given the number of layers N in the depth profile used in the forward calculation model. Since the model in addition to N also takes $\Delta \mathbf{x}$ and \mathbf{c} as input parameters $P(\mathbf{D}|N, I)$ has to be determined for any arbitrary choice of $\Delta \mathbf{x}$ and \mathbf{c} . Therefore $P(\mathbf{D}|N, I)$ has to be determined by marginalizing (see e.g. [21] p. 4) over $\Delta \mathbf{x}$ and \mathbf{c} as in Eq. 14.

$$P(\mathbf{D}|N, I) \propto \int \cdots \int P(\mathbf{D}, \mathbf{c}, \Delta \mathbf{x}|N, I) dC_i dx_i \quad (14)$$

applying Bayes theorem

$$P(\mathbf{D}, \mathbf{c}, \Delta \mathbf{x}|N, I) \propto P(\mathbf{D}|\mathbf{c}, \Delta \mathbf{x}, N, I) P(\mathbf{c}|N, I) P(\Delta \mathbf{x}|N, I)$$

assuming constant priors for $\Delta \mathbf{x}$ and \mathbf{c} yields

$$P(\mathbf{c}|N, I) P(\Delta \mathbf{x}|N, I) = \frac{1}{(c_{max} - c_{min})^N (\Delta x_{max} - \Delta x_{min})^N}$$

and applying least squares likelihood function

$$P(\mathbf{D}|\mathbf{c}, \Delta \mathbf{x}, N, I) \propto \exp\left(-\frac{\chi^2}{2}\right)$$

finally yields

$$P(\mathbf{D}|N, I) \propto \frac{1}{(c_{max} - c_{min})^N (\Delta x_{max} - \Delta x_{min})^N} \int \cdots \int \exp\left(-\frac{\chi^2}{2}\right) dC_i dx_i$$

Marginalization means that the probability for a layer number is given by integrating over all possible choices, weighted by the respective probability. The multidimensional integral in Eq. 14 can be solved by monte carlo integration but this is rather time consuming. Therefore still following the approach in [21] chapter 4.2, χ^2 is replaced by a taylor expansion around the χ_{OPT}^2 corresponding to the best least squares fit for parameter values \mathbf{c}^{BEST} and $\Delta \mathbf{x}^{BEST}$. This reduces the integral to a multivariate gauss integral and finally $P(N|\mathbf{D}, I)$ is

given by Eq. 15

$$\begin{aligned}
P(N|\mathbf{D}, I) &\propto P(\mathbf{D}|N, I) \\
&\propto \frac{N!(4\pi)^N}{(c_{max} - c_{min})^N (\Delta x_{max} - \Delta x_{min})^N \sqrt{Det(\nabla\nabla\chi^2)}} \exp\left(-\frac{\chi_{BEST}^2}{2}\right) \\
Det(\nabla\nabla\chi_{BEST}^2) &= \text{Determinant of the hessian of } \chi^2 \text{ evaluated at} \\
&\quad \mathbf{c}_{BEST} \text{ and } \Delta\mathbf{x}_{BEST}
\end{aligned} \tag{15}$$

The expression for $P(N|\mathbf{D}, I)$ contains two counteracting contributions: While the $\exp\left(-\frac{\chi^2}{2}\right)$ term penalizes large discrepancies between the data and the forward calculation, the fraction in front of it (Ockham Factor) penalizes having a large number of free parameters i.e it favors simple models.

While for a given $\Delta\mathbf{x}$ the best least squares \mathbf{c}_{BEST} can be found directly by applying SVD as described above, finding the optimum $\Delta\mathbf{x}$ requires more effort. For a given number N the search for $\Delta\mathbf{x}_{BEST}$ occurs in two steps: First different $\Delta\mathbf{x}$ are generated by randomly binning sub layers (see above) and the the corresponding least squares result for \mathbf{c}^{LSQ} is calculated by SVD and from it the corresponding χ^2 value is derived according to Eq. 4. As part of this first random search a lot of $\Delta\mathbf{x}$ configurations are encountered that yield a \mathbf{c}^{LSQ} which is outside the prior range ($c_{Min} = 0 \leq c_i \leq c_{Max} = 1$). In this case the $\Delta\mathbf{x}$ configurations is considered invalid and is discarded. If a $\Delta\mathbf{x}$ configurations with a valid \mathbf{c}^{LSQ} is found it is stored together with its χ^2 value. Based on the $\Delta\mathbf{x}$ with the lowest χ^2 value resulting from this first step, a Markov Chain Monte Carlo (MCMC) optimization is started as the second optimization step. In each MCMC step a layer k in $\Delta\mathbf{x}$ is chosen randomly and its thickness is either increased or decreased by adding or removing dl sub layers from the selected Δx_k . For the so generated new configuration the least squares \mathbf{c}^{LSQ} and $\chi^2 \equiv \chi_{BEST}^2$ is determined. From χ_{BEST}^2 the probability $P(N|\mathbf{D}, I)$ of the current configuration is calculated according to Eq. 15. If $P(N|\mathbf{D}, I)$ is smaller or equal than the previous probability or if $P(N|\mathbf{D}, I)$ is larger than a some random number, the step is accepted otherwise the step is rejected and the previous $\Delta\mathbf{x}$ configuration is restored. Accepting a given configuration with a probability determined by some random is required to maintain the detailed balance criterion for the MCMC chain. This entails that also steps have to be taken which result in configurations with a lower probability.

For a given N this MCMC iteration converges to the optimal configuration $\Delta\mathbf{x}^{BEST,N}$ with a corresponding $\mathbf{c}^{BEST,N}$ and the lowest $\chi_{BEST,N}^2$.

This optimization procedure is repeated for different layer numbers N and finally N^{OPT} is selected by comparing the probability $P(N^{OPT}|\mathbf{D}, I)$ calculated according to Eq. 15 from the individual results for $\mathbf{c}^{BEST, N}$ and $\chi_{BEST, N}^2$. Typically N^{OPT} is of the order of the number of different experimental conditions (e.g. ^3He energies) that entered the evaluation.

Applying the search for N^{OPT} to the artificial data in Fig. 3 yields parameter estimates for \mathbf{c}^{LSQ} that can be compared to the true depth profile c_i^T in Fig. 5. Compared to the initial approach with non optimized depth discretization $\Delta\mathbf{x}$ the optimized version yields a better resemblance with the true profile. A perfect fit is not possible since the depth resolution of the NRA method does not allow to resolve the near surface features without the detailed information about the depth discretization which in reality is not available.

Once N^{OPT} and the corresponding $\Delta\mathbf{x}^{BEST}$ are found, the result for \mathbf{c} can further refined by dropping the least squares likelihood approximation in its determination via SVD. As already shown in Eq. 7 and 8 the probability for a certain \mathbf{c} is essentially given by the likelihood function when one assumes flat constant prior functions. For NRA data the likelihood function is given by the Poisson statistics as in Eq. 8. Using Eq. 8 to assign a probability to a composition vector \mathbf{c} and MCMC optimization can be applied to find the most probable composition vector. The figure of merit in this MCMC optimization is $P(\mathbf{c}|\mathbf{D}, \mathbf{I})$ from eqs. 7 and 8. Again as in the search for N^{OPT} a new configuration \mathbf{c} from a single MCMC step is accepted if it improves the figure of merit or is rejected otherwise and the configuration is restored to its previous state. During the MCMC optimization of \mathbf{c} three stages are distinguished "Burn in", "Sample" and "Measure" differing in the step width \mathbf{dc} used to generate the next trial configuration. During the "Burn in" stage the components of the \mathbf{dc} vector are varied such as to achieve a step acceptance fraction of $\approx 50\%$. Next during the "Sample" stage \mathbf{dc} is constant and a certain number of steps are performed after which the chain is assumed to have converged. In a converged MCMC run the composition configurations \mathbf{c} are accepted with a probability equal to $P(\mathbf{c}|\mathbf{D}, \mathbf{I})$. Therefore during the "Measurement" stage the accepted \mathbf{c} configuration vectors are collected and stored in a histogram which finally contains the probability densities for each component c_i of \mathbf{c} . These probability densities are often well represented by normal distributions. In these cases confidence intervals can be deduced which are then equal to the least squares confidence intervals introduced above (see Eq. 11). However in some cases multi modal probability densities occur for which no simple confidence intervals

can be derived. For these cases no corresponding least squares approximation characterized by a single number can be given.

Applying the MCMC optimization of \mathbf{c} to the artificial data in Fig. 3 using the result from a previous N^{OPT} search, yields parameter estimates for \mathbf{c}^{MCMC} that can be compared to the true depth profile c_i^T in Fig. 6. Compared to the result for \mathbf{c} from the N^{OPT} search there is further improvement in the resemblance of the true depth profile c_i^T . But the main advantage of the MCMC optimization is determination of the probability densities for each component c_i which can detect cases where the least squares confidence intervals are not applicable.

5 Possible pitfalls during manual deconvolution

When performing the deconvolution manually, directly following the approach in [17] i.e. only taking into account the proton peak integrals the system becomes badly conditioned as shall be shown here. Qualitatively the problem lies in the highly convoluted nature of the proton peak integrals as was already explained in section 2 based on Fig. 1: For high ^3He energies the sensitivity for D (i.e the cross section) is low in the near surface region and has its maximum at some depth where the ^3He has been slowed down to an energy where its sensitivity (i.e its cross section) for D is at its maximum (see Fig. 1). Still the contribution ($\propto \text{concentration} * \text{crosssection}$) of the high near surface concentration of D to the proton peak integral can be equal to that of the low concentration region at the depth where the cross section has its maximum. Therefore the contributions of the D amount at individual depths to the proton peak integrals are heavily convoluted, making a manual evaluation very difficult and prone to artifacts. This fact can also be illustrated in a more mathematical sense by investigating the confidence intervals and condition numbers of a design matrix $\mathbf{A}^{Partial}$ corresponding to the typical information content used in a manual evaluation: the proton peak integrals. Based on a reasonable guess for the depth discretization according to Eq. 12 a design matrix $\mathbf{A}^{Partial}$ was generated for the experimental conditions (i.e. ^3He energies) used for the artificial data shown in Fig. 3. This procedure yielded a 10×10 $\mathbf{A}^{Partial}$ matrix (10 energies and 10 layers) with a condition number of $\approx 4 \times 10^4$ and error bars of the order of 10^{-2} . So for a typical depth profile with near surface concentrations of $\approx 10^{-2}$ and $\approx 10^{-4}$ in the bulk this corresponds to a relative error of 100% near the surface and 10000% in

the bulk. This shows that a manual evaluation can lead to statistically totally insignificant results.

6 Real world example

In Fig. 7 the experimental data for an NRA measurement of a D implanted W sample is shown. NRA proton spectra were collected at 600, 1200, 2400 and 4500keV, alpha spectra only at 600keV because the alpha spectra were overshadowed by backscattered ^3He from the W substrate at higher energies. The data was evaluated both manually and by applying NRADC. The manual evaluation was limited to matching the proton peak and alpha spectra while NRADC took all the available information into account. The NRADC GUI greatly facilitates energy calibration and entering all the experimental details and is usually quite a bit faster than the manual evaluation. The results are compared in Fig. 7 and Fig. 8. In Fig. 7 the experimental data is compared to the result of the forward calculation resulting from a full (N^{OPT} search + MCMC optimization of \mathbf{c}) NRADC run. The match with the experimental data is perfect. In Fig. 8 the depth profiles obtained from NRADC are compared to the result from the manual deconvolution. The first striking feature is that the manual deconvolution used \approx two times more layers despite the fact that it only took a fraction of the available data into account. Based on Ockham's razor the manual evaluation result is much less likely than the NRADC result which with fewer free parameters produces an equally good (maybe even better) fit to the data. While both results have a similar shape they differ at some locations by more than an order of magnitude. Also the manual evaluation can not produce sound confidence intervals for the composition. The confidence intervals attached to the manual evaluation result have been calculated by NRADC according to Eq. 11 from a design matrix which was based on the same depth discretization as the manual evaluation and included all available experimental data.

Generally this kind of error estimation is usually omitted during a manual evaluation and it is therefore not possible to judge whether or not certain result features (like the second maximum in the depth profile) are actually significant or are just within the experimental uncertainties.

NRADC comes with a graphic user interface that simplifies the input of the experimental conditions and data. It further facilitates the re-use of input data (energy calibration, regions of interest etc.) for cases where measurements of

different samples were performed under the same experimental conditions. The typical run time for the actual deconvolution is in the order of 400 seconds on a quad core PC. Most of the computational time is spent searching for the most probably depth discretization which can be efficiently parallelized by probing a certain layer number on each thread. This makes NRADC much faster than a manual evaluation.

7 Conclusions

A program NRADC has been developed that utilizes the concepts of Bayesian statistics to perform statistically sound deconvolution of trace element depth profiles from IBA data. Based on the assumption that trace amounts of elements do not vary the stopping power significantly, the forward calculation of IBA data, given the elemental depth profiles, was linearized. This results in a very fast forward calculation allowing one to perform Markov Chain Monte Carlo (MCMC) searches for the most probable (maximum likelihood) depth profile, defined by its depth discretization and composition. This sets NRADC apart from existing automated data evaluation tools in that it not just finds the optimal values for the fit parameters (compositions) but also chooses their numbers based on a statistically sound probabilistic argument.

The high speed and statistically sound confidence intervals make NRADC a powerful tool for IBA data analysis.

References

- [1] N. Barradas, R. Mateus, M. Fonseca, M. Reis, K. Lorenz, I. Vickridge, Nucl. Inst. Meth. B 268 (2010) 1829.
- [2] U. von Toussaint, R. Fischer, K. Krieger, V. Dose, New Journal of Physics 1 (1998) 11.1.
- [3] U. von Toussaint, R. Fischer, V. Dose, AIP Conference Proceedings 568 (2001) 615.
- [4] C. Jeynes, N. P. Barradas, P. K. Marriott, G. Boudreault, M. Jenkin, E. Wendler, R. P. Webb, J. Phys. D. Appl. Phys. 36 (2003) R97.
- [5] T. Dittmar, P. Roubin, E. Tsitrone, E. Gauthier, A. Hakola, J. Likonen, et Al., Nucl. Inst. Meth. B B15 (1986) 227.
- [6] N.P.Barradas, C.Jeynes, R.P.Webb, Appl. Phys. Lett. 71 (1997) 291.
- [7] J. L. Duggan, I. Morgan (Eds.), Proceedings of the 15th International Conference on the Application of Accelerators in Research and Industry, Vol. 475, American Institute of Physics Conference Proceedings, 1999.
- [8] N. Barradas, K. Arstila, G. Battistig, M. Bianconi, N. Dytlewski, C. Jeynes, E. K. tai, G. Lulli, M. Mayer, E. Rauhala, E. Szilagyi, M. Thompson, Nucl. Inst. Meth. B 262 (2007) 281.
- [9] N. Barradas, C. Jeynes, M. Jenkin, P. K. Marriot, Thin Solid Films 343-344 (1999) 31.
- [10] N. Barradas, C. Jeynes, Nucl. Inst. Meth. B 266 (2008) 1875.
- [11] N. P. Barradas, P. K. Marriott, C. Jeynes, R. P. Webb, Nucl. Inst. Meth. B 136-138 (1998) 1157.
- [12] V. K. Alimov, W. M. Shu, J. Roth, et Al, J. Nucl. Mat 417 (2011) 572.
- [13] V. K. Alimov, H. Nakamura, B. Tyburska-Pueschel, O. V. Ogorodnikova, J. Nucl. Mat 414 (2011) 479.
- [14] O. V. Ogorodnikova, T. Schwarz-Selinger, S. K., J. Appl. Phys 109 (2011) 013309.
- [15] K. Sugiyama, V. K. Alimov, J. Roth, Physical Scripta T138 (2009) 014026.
- [16] T. Dittmar, P. Roubin, E. Tsitrone, E. Gauthier, A. Hakola, J. Likonen, et Al., Physical Scripta T138 (2009) 014027.
- [17] V. K. Alimovr, M. M. und J. Roth, Nucl. Instrum. Meth. B 234 (2005) 169.
- [18] J. R. Tesmer, M. Nastasi, Handbook of Modern Ion Beam Materials Analysis, Material research society, Pittsburgh Pennsylvania, 1995.

- [19] U. von Toussaint, T. Schwarz-Selinger, M. Mayer, S. Gori, Nucl. Instrum. Meth. B 268 (2010) 2115.
- [20] W. H. Bragg, R. Kleeman, Philos. Mag. 10 (1905) 318.
- [21] D. S. Sivia, Data Analysis, A Bayesian Tutorial, Clarendon Press, Oxford, 1996.

Figure captions

Fig. 1

Qualitative comparison of the energy dependence of the ${}^3\text{He}(p, \alpha)\text{D}$ nuclear reaction cross section and a typical D depth profile. The depth axis matches the energy axis according to the stopping power of a 4.5MeV ${}^3\text{He}$ ion beam.

Fig. 2

Comparison of the true input depth profile (Typical D-profile) with the deconvolution result using perfect depth discretization and SVD.

Fig. 3

Comparison of the artificial IBA data generated for the true input depth profile from Fig. 2. Also shown is the best fit obtained by a MCMC optimization of depth discretization and composition.

Fig. 4

Comparison of the true input depth profile (Typical D-profile) with the deconvolution result using a "reasonable guess" depth discretization and SVD.

Fig. 5

Comparison of the true input depth profile (Typical D-profile) with the deconvolution result using MCMC optimized depth discretization and SVD.

Fig. 6

Comparison of the true input depth profile (Typical D-profile) with the deconvolution result using MCMC optimization to determine the depth discretization and the composition.

Fig. 7

Real experimental NRA data obtained from D implanted W. Also shown is a comparison with the forward calculation result corresponding to the best fit depth profile.

Fig. 8

Comparison of the D depth profile obtained both from manual fitting and from MCMC optimization of the experimental data in Fig. 7. The error bars on the manual fit have been derived by NRADC.

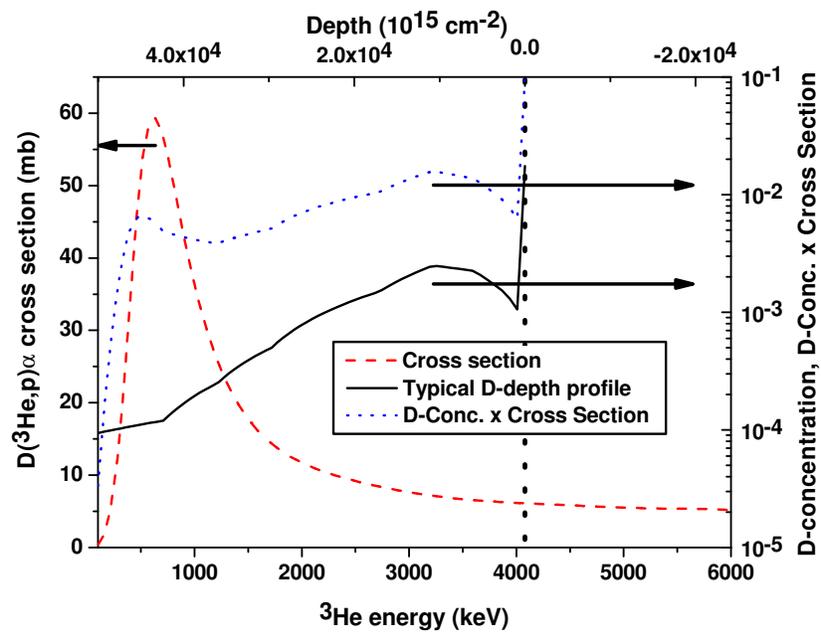


Figure 1:

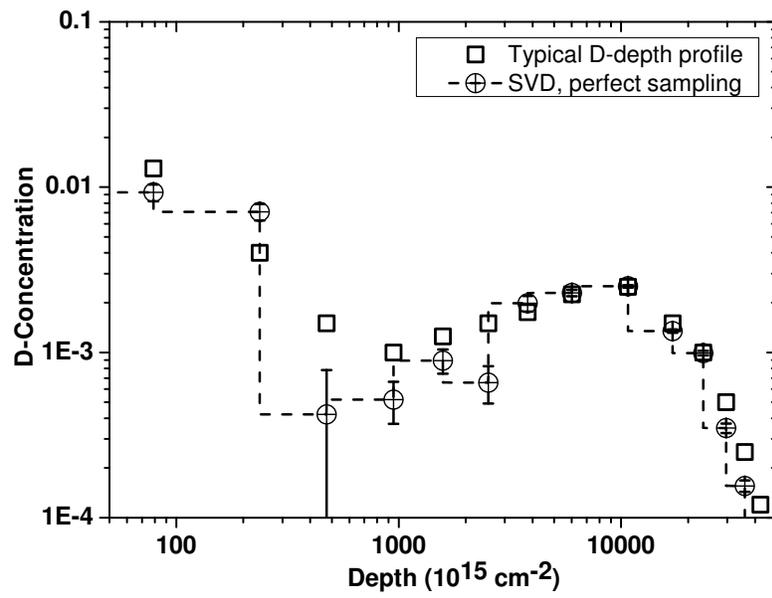


Figure 2:

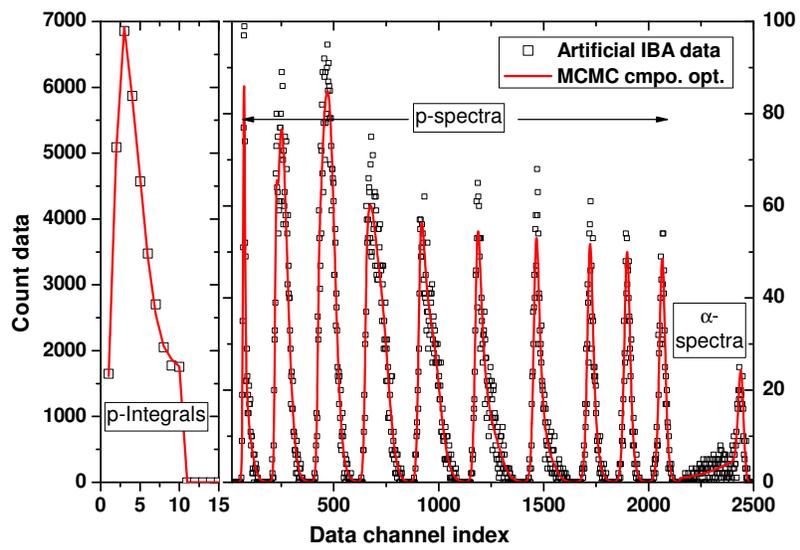


Figure 3:

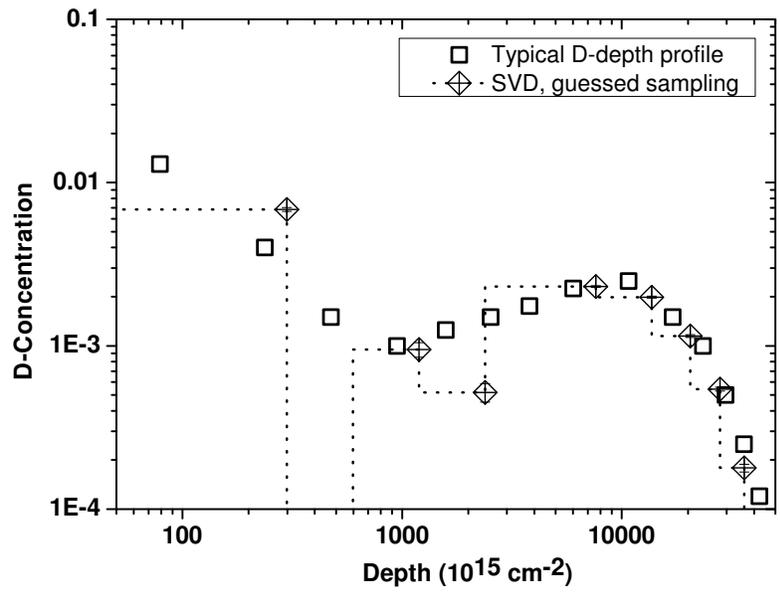


Figure 4:

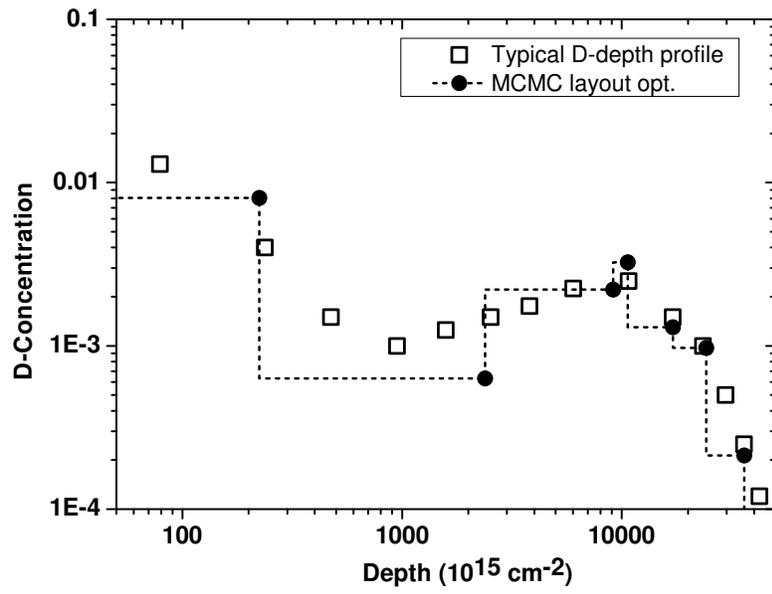


Figure 5:

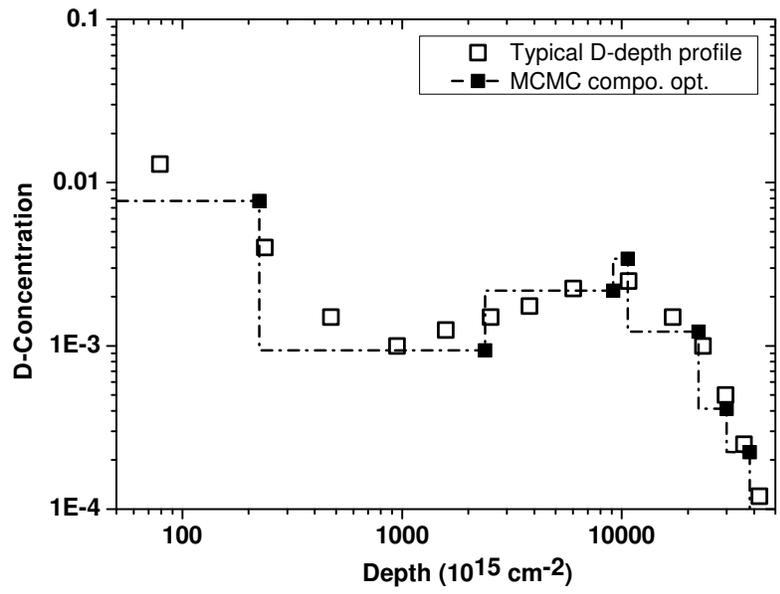


Figure 6:

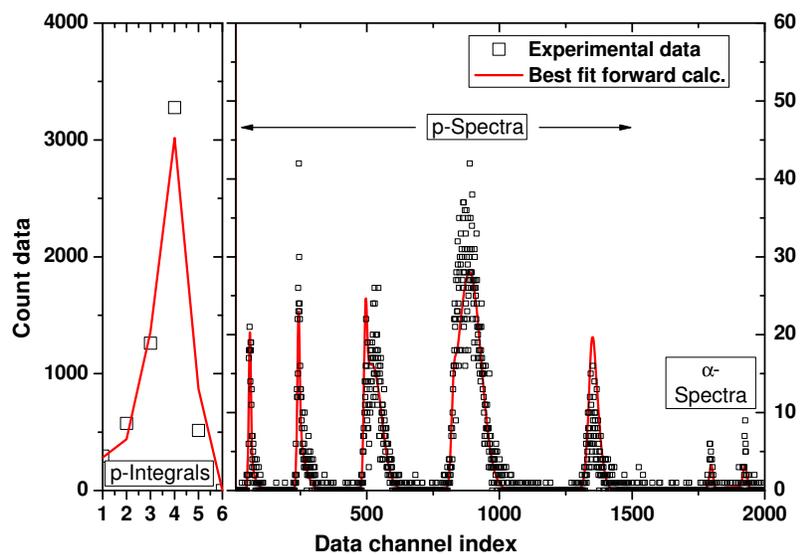


Figure 7:

